

The Need to Develop Peer Rating Procedure for Optimal Achievement in Schools

Sunday Okegbile Akanmu

Federal College of Education (Technical) Bichi, Nigeria

Abstract

Improving students' level of school achievement optimally has been the onerous task for educators and researchers. One of the major activity or component of the instructional system is the educational assessment and its procedure. Being a determinant of the level of educational achievement, students' assessment needs to be comprehensive and exhaustive. This study observes the need to actively involve learners in the process of assessment through learners peer rating procedure. It is observed that the concept of peer rating is not limited to non-cognitive domain and that peer rating procedure (PRP) has not been well adopted in all areas of students' assessment despite its applicability, fairness, validity, reliability, veracity, quick impact and cost effectiveness as established by researchers. The paper highlights relevant research findings on the use of PRP and also reports findings of studies carried out by the writer which established significant correlations ranging from 0.929 to 0.691 between teachers' and students' rating of course mates in two courses. It observes that basic hindrances and limitations to the use of PRP in addition to other evaluation techniques are linked to availability of an adaptable procedure for peer rating. This paper proposes adaptable PRP procedures or model to include identification and teaching of subject matter; instrumentation; schedule of ratings; testing or demonstration; pre rating discussion; peer rating; post rating discussion and reports. The paper emphasizes utilization of opportunities presented during PRP procedures and concludes that PRP be recommended for use by educators as an additional skill that will enhance teachers' efficiency and students' achievement.

1. Introduction

Teachers and researchers have labored severally in their attempt to improve learners' achievement by applying various techniques and approaches. Since evaluation or assessment of learning objective is a major component and determinant of the instructional system, it is very necessary to develop activities and skills associated with the component. Assessment is placing a value on man's ability in behaving to some specified expectation as a result of

instruction. It has more often than not been misconceived as an activity carried out by teachers or instructors only. Falchikov [7] posited that alternative to traditional, summative forms of assessment include self, peer and collaborative assessment.

Pope et al. [18], while emphasizing student centered and collaborative approach to assessment in the form of peer nomination, peer rating and peer ranking stated that of these mode of peer assessment techniques, peer rating (PR) is the most useful for provision of feedback. Hence, Nicol [12] assertion that 'producing feedback is cognitively more demanding than receiving it, as it involves higher levels of reflection' gives vital hint on the cognitive and non-cognitive values of peer rating. That is, PR could be linked with improved learning.

One of the skills that is worth emphasizing therefore vis-à-vis comprehensive evaluation is peer rating. The skill has been erroneously thought to be applicable only to the non-cognitive domain more especially the affective domain. Peer Rating procedure (PRP) is seemingly applicable to all the domains though for instance when used in a sociogram, it is referred to as the socio-metric technique.

2. Peer Rating

Peer rating is a technique involving a group of people working independently or separately as the case may be. It is not synonymous with peer appraisal in the sense that it is a technique in peer appraisal whereby pupils rate their peers (fellow students) on the same rating device allowable by the teacher.

The common or traditional peer appraisal techniques include guess who, sociometric and nominating techniques but the broad procedure of students PRP is open to teachers' initiative and versatility. If the teachers could rate students activity in class, their projects, assignments and solutions to sums, and in like manner, why would the students not be requested to rate themselves ahead of teachers' rating for comparison and when necessary for discussion/conference? If assessment of the

cognitive and psychomotor domains is being done by the teacher alone, then it is a one-way affair.

The doubt expressed by researchers such as Thomas et al. [25] about the efficacy of peer rating especially in the cognitive domain could have arisen from concerns for the following inquiry:

- Can students rate their peers objectively and accurately?
- Can students rate a superior colleague?
- How will the relationship between peer rating and teachers rating be?
- What is the standard procedure for peer rating?
- Of what use is PR?

These are some of the many challenges to be continually addressed by research in education.

3. The Gains of Using Peer Rating Procedure

From Rowntree's contribution [21], the awareness of a student as a result of his observation after rating others can encourage learning that is motivated (extrinsically) which he called "the ignoble satisfaction of feeling that one is better than someone else".

Sax and Newton [23] observed that the use of PR allows group members who know each other well to nominate each other for different activities. With the introduction of the instructional process, the complex nature of students' evaluation is reduced. Data gotten through this means serve as formative process too and each student involved develops rating skills with every rating and learns how mistakes are committed and penalized.

From the implications of various studies ([12], [17], [19], [21], [22], [23]); peer rating may be an additional approach at enhancing students' sense of judgment, objectivity and optimum performance. If eventually it develop students' autonomous thinking, it can reduce emphasis on teachers' verdict only, raise students level of confidence and understanding of the subject matter involved, develop students' assessment skills for future challenges, emphasize teachers' strength and weaknesses and reduce teachers' stress of marking or errors of educational measurement. It aids the understanding of marking scheme and standards, changes teachers' role from assessor to facilitator and improves learners' understanding of presentation skills.

Furthermore, the use of Peer rating in teaching will reduce teachers workload, make students learn from others' mistakes or success, engage in dialogue about their decisions, understand the value of feedback, construct meanings of their own, help develop clearer assessment criteria, make learners take up the role of appraisers and over time acquire

the skills they need to make judgements on their own work (self-assessment).

4. Review of Relevant Studies

Nash [11] discovered that pupils as young as eight years were able to say which children in the class were better than them in reading, writing, and numbers and their self-perceived class rank correlated highly with the ranking made by the teacher. On a different note, Stubblebine [24] in a survey of 200 employed graduate and undergraduate business students investigated differences in individuals' acceptance and perceptions about peer evaluation vs. supervisor evaluation. Results showed that, compared to peer evaluations, acceptance of supervisor evaluations is higher. Supervisor evaluations are perceived as more accurate and more trusted, supervisors are seen as having more qualifications to evaluate performance and more opportunities

Bood and Loblins work cited in Beard and Hartley [1] used a mid-term examination where engineering students were randomly allocated an unnamed paper from one of the other students. It was reported that the students liked the procedure. Ramsey et al. [20] confirmed that in medical education, peer rating is feasible to obtain assessment from professional associates of practicing physicians in areas such as clinical skills, humanistic qualities and communication skills. The conclusion noted that, peer rating provides a practical method of assessing clinical performance in humanistic qualities and communication skills which are difficult to assess with others. Nicol [12] however found out that peer rating tool was suitable for post-graduate because of maturity, work experience and group size.

Ramsey et al. [20] noted that peer rating provided practice - based information on cognitive and interpersonal areas as well as professionalism in the training of medical students. He reported that 7 - 12 observations were needed to provide reliable rating of overall clinical grade. He also found out that peer rating are not biased substantially by the method of selection of the peers or the relationship between the rater and the subject.

In a 2007 study, faculty members, having used a teaching perspective inventory developed by Daniel D. Pratt, rated video recording of a lecture twice and inter-rater reliability improved significantly [9]. Okegbile [13] observed in a study of teachers classroom performance that students seem capable of rating the teaching ability of trainee teachers or teaching. This calls for the need to further investigate the objectivity of rating ability of students for peer or self.

Epkins [6] studied peer rating of depression, anxiety, and aggression in impatient and elementary

school children and found that children's self-reports on each of the three traits were significantly related to their peers' ratings of the same trait.

A military based research by Chuan et al. [4] studied validity of peer ratings under developmental versus evaluative contexts and found that peer ratings are more reliable and valid under the developmental context. That is students awareness of high administrative values or implications might lead to low validity.

Long et al. [14] wrote on correlates of satisfaction with a peer evaluation system by examining the effects of performance level and individual differences. Subjects were reported to indicate their satisfaction with and reaction to the peer evaluation system. Ohland and Layton [13] persevered on the need to develop and use peer rating technique. They improved the reliability of peer rating instrument by listing explicitly characteristics and criteria of rating and describing what each form of rating mean in terms of individual performance and admonition that ratings should reflect individual level of performance but not perceived academic ability. The instruments were used by students to evaluate their teammates' contributions to the team deliverable in oral and written presentations of their solutions to a technical design problem. Each student was rated by his or her teammate, producing multiple ratings.

The peer procedure used included:

- Students learn about the peer evaluation.
- Students reflect on the evaluation.
- Students get feedback on how the group assessed each member's work to date.
- The instructor was alerted if groups were not functioning.

Findings revealed high reliability of the procedure but noted that good team work can improve peer evaluation.

A Google based procedure by Grader [8] emphasizes four main practical approaches useful in peer rating procedure. That is

- Teacher creates an assignment.
- Students submit their solutions.
- Students grade each other's work.
- Teacher assigns final grades.

In a study similar to Ohland and Layton [13], the writer taught two undergraduate courses while on leave in a newly established university in Nigeria.

The following questions were raised.

- Can undergraduates rate their peers objectively?
- Is there any relationship between students' and teachers' rating.

The two corresponding null hypotheses were: there is no significant difference between students' and

teachers' ratings in theoretical course. There is no significant relationship between students' rating of colleagues in statistics course and teachers' rating at undergraduate level.

5. Methodology

The study was correlational involving all the eight foundation students offering the two courses – one theoretical and the other statistical. The theoretical course is Human Learning tagged HLM while the statistical is Statistics for Behavioural Scientists tagged (SBS).

The procedure involved two lecturers, one teaching the subject matter while the other was supervising the peer rating task. Seven stages involved in the data collection procedure include teaching instrument development, time tabling, testing, pre-rating, post rating discussion, teachers' ratings and report. No treatment (pre and post) was introduced because of the kind and stage of the study.

Each student rated all other students and the multiple ratings and that of the lecturer were correlated for the two subjects. Since the number of raters was below 10, Kendall's Ta was used [3]. Using the SPSS package, students' ratings were tagged STR, Teacher Rating was tagged TER as shown in the table below.

6. Results

From the non-parametric correlations (see Table 1) between the students' ratings, all coefficients are significant and nine are significant at 1.00 percent level. Correlating teachers' rating with those of students showed a positive and significant relationship in all cases. The highest is 0.929 and the lowest is 0.714 which is significant at 5% level. For the SBS course, there was, therefore, no significant difference between the ratings of students and that of the teacher.

Table 1. Correlation Matrix of BSS Ratings

	STR 1	STR 2	STR 3	STR 4	STR 5	STR 6	STR 7	STR 8	TER 1
STR 1	1.00								
STR 2	.643*	1.00							
STR 3	.667*	.964**	1.00						
STR 4	.837**	.837**	.793*	1.00					
STR 5	.764**	.837**	.793**	.963**	1.00				
STR 6	.764**	.909**	.944**	.778**	.741*	1.00			
STR 7	.643*	.714*	.741*	.764**	.691*	.691*	1.00		
STR 8	.837**	.837**	.869**	.852**	.815**	.926**	.618*	1.00	
TER 1	.929**	.714*	.741*	.909**	.837**	.837**	.714*	.909**	1.00

*: Significant at 0.05 level (2-tailed)

** : Significant at 0.01 level (2-tailed)

Table 2. Correlation Matrix of HLM Ratings

	STR 1	STR 2	STR 3	STR 4	STR 5	STR 6	STR 7	STR 8	TER 1
STR 1	1.00								
STR 2	.909*	1.00							
STR 3	.643*	.764**	1.00						
STR 4	.764**	.815**	.546	1.00					
STR 5	.837**	.889**	.691*	.889**	1.00				
STR 6	.741*	.793**	.519	.868**	.717*	1.00			
STR 7	.857**	.909**	.643*	.909**	.837**	.889**	1.00		
STR 8	.929**	.982**	.714*	.837**	.909**	.815**	.929**	1.00	
TER 1	.982**	.926**	.691*	.741*	.815**	.717*	.837**	.909**	1.00

*: Significant at 0.05 level (2-tailed)

** : Significant at 0.01 level (2-tailed)

Table 2 also revealed that positive significant relationship existed between each student's rating and the rating by the teacher. In two cases between the students' rating, there were insignificant relationship but with the teacher's rating all were significant varying from 0.691 to 0.982. This also confirmed that significant relationship exists between teachers' rating and students' ratings for the course HLM.

Observation of students in the course of rating procedures revealed that the idea of rating colleagues seems novel, interesting, and educative. Results so far obtained may not be at variance with assertions of other researchers except in procedures, assumptions or scope. Doyle and Green [5] used mathematical technique of data envelopment analysis for their work in self and peer appraisal in higher education, with specific focus on its rationale, the prerequisite for use, and insights that it yields. The study by Ramsey et al. [20] which involved eleven peer physicians established the feasibility of peer rating procedure in the training of physicians. It could be inferred that students can be objective raters where they are given the opportunity or adequately trained while at the same time ensuring objectivity and promoting learning.

7. Adaptable Peer Rating Procedure

To successfully carry out the use of PRP for the assessment of students' achievement in learning of concepts, skills and the like, there is the need for modeling after workable steps. Such steps may require adaptation since there are various forms in which peers could be used in PRP. The following conditions necessitate adoption of any procedure.

The Nature of the test

Type of skills

The relevance and appropriateness of groups and grouping

The nature of the subject matter

The following procedure explains a model.

1. Teaching of the subject matter (with the objectives and fitting assessment technique as main focus.)

2. Developing the peer rating instrument. This may be constructed by the teacher or in conjunction with a parallel peer group. It could be modified by the peer group at a later stage before the peer rating. This stage runs concurrently with the first stage.

3. Scheduling of the PRP period. This depends on the nature of the subject matter.

4. Testing, examining or demonstration stage. All conditions and rules of the test administration should be adhered to.

5. Pre rating discussion. This may be done before some forms of observations but usually done after testing in cognitive based tests or paper and pencil test. The stage is characterized by methods, objectives, instrument and other details of the instrumentation to be unveiled including the various dimensions of possible responses. It is usually noted at the stage that care should be taken in attaching stronger summative importance to their rating because it could lead to low reliability. This depends therefore on the power of the discussion.

6. Peer rating stage: The peers and the teacher rate the students. In case of paper and pencil, the teacher rates after that of the peers.

7. Post rating discussion. This gives room for students' comments on students' difficulties in learning the concept/skill, common errors committed by colleagues, other dimensions of responses not considered, how to improve teaching/learning of the concept or skill. The level of attainment of objectives of teaching/learning is optimal at this stage and teacher has to explore such opportunities.

8. Teachers' rating: This could come up before post rating discussion or after depending on the nature of the observation required. The teacher is to apply the instrument strictly to grade the students for comparison sake.

9. Reports: Reports of teachers' rating and students' peer rating are compared and analyzed statistically (descriptive/inferential) and interpreted appropriately. The results help the teacher and students to be confident in the entire process of teaching and learning.

8. Recommendations

PRP is recommended for explicit inclusion in policies of school assessment as a technique that could be used for optimal students' school achievement.

PRP should be used more in Faculties and Institutes to assess students' presentations at all levels for proposals/theses defense, practical, awards promotions, cognitive and non-cognitive based tests etc. Its use at elementary levels requires more investigation. Validation of PRP procedure is required through research for further establishment of its applicability usability worldwide

9. References

- [1] Beard, R.M., Hartley, J., (1986). Teaching and learning in Higher Education. London: Harper & Row.
- [2] Bernstein, J.D., Johnson, J. & Smith, K., (2000) "An Examination of the Implementation of Peer Review of Teaching," in *New Directions for Teaching and Learning* 83, 73-86
- [3] Borg, W.R., and Gall, M.D., (1983). Educational research: An introduction. New York: Longman.
- [4] Chuan, T.P., Ramaya, R., Can, K.Y., and Ang, Y.M., (2005). A paper presented at the 47th International Military Testing Association
- [5] Doyle, J.R. & Green, R.H., (1994). Self and peer appraisal in higher education. *Higher Education Journal*, 28(2), 241 – 264.
- [6] Epkins, C.C., (1994). Peer ratings of depression, anxiety, and aggression in inpatient and elementary school children: rating biases and influence of rater's self-reported depression, anxiety, and aggression. *Journal of Abnormal Psychology*, 22, <http://www.questia.com/google Scholar.qst?docId=5000252460>
- [7] Falchikov, N., (1986). "Product Comparison and Process Benefits of Collaborative Peer Group Assessment", *Assessment and Evaluation in Higher Education*, Vol. 11, No. 2, pp. 146-166.
- [8] Crowd Grader (2015) Peer grading for your classroom. <http://www.crowdgrader.org/>
- [9] CELT, Iowa state University, (2009). Summative Peer Evaluation of Teaching: Literature Review and Best Practices. http://www.celt.iastate.edu/teaching-resources/document-your-teaching/peer-evaluation-teaching/#_ftn32
- [10] Long, S.W., Long, J.E & Dobbins, H.G., (1998). Correlates of satisfaction with a peer evaluation system: investigation of performance levels and individual differences. *Journal of Business and Psychology*, 12(3).
- [11] Nash, R., (1976). Teacher expectations and pupils learning. London: Routledge and Kegan Paul.
- [12] Nicol, D., (2010). Enhancement themes: Developing students ability to construct Feedback. The Quality Assurance Agency for Higher Education.
- [13] Ohland, W. M. & Layton, A.R., (2001). Comparing the Reliability of Two Peer Evaluation Instruments. Proceedings of the 2001 American Society for Engineering Education Annual Conference & Exposition.
- [14] Ohland, W.M., Loughry, L.M., Carter, L.R., Bullard, G.L., Felder, M.R., Finelli, J. C., Schmucker, G.D., (2005). Developing a Peer Evaluation Instrument that is Simple, Reliable, and Valid. Proceedings of the 2005 American Society for Engineering Education Annual Conference & Exposition.
- [15] Okegbile, A.S., (2001). The Applicability & Meaningfulness of Peer Rating Technique in the New Millennium. *Knowledge Review: A Multidiscipline Journal Published by National Association for the Advancement of Knowledge NAFK*, 3(1).
- [16] Okegbile, A.S., (2001). Students and supervisors evaluation of trainee teachers' classroom performance. *The Nigeria Academic Forum*, 1(1) Pg 83-87.
- [17] Okegbile, A.S., (2008). The applicability of peer rating technique in the higher institutions. *Nigerian journal of educational research and evaluation*, 8(2), 1-7.
- [18] Pope, N., Forrest, E., and Gatfield, T. (n.d.), An examination of the use of peer rating for formative assessment in the context of interactive education. [an_examination_of_the_use_of_pee.htm](http://www.cbpp.uaa.alaska.edu/afef/), <http://www.cbpp.uaa.alaska.edu/afef/>. (Access date: 6 December 2015).
- [19] Race, P., (2001). The Lecturers Tool kit. London Kogan Page, pp. 94 – 95.
- [20] Ramsey, P.G., Wenrich, M.D., Carline, J.D., Inui, T.S., Larson, E.B., and Logerto, J.P., (1993). Use of peer rating to evaluate physician performance. *The Journal of American Medical Association (JAMA)*, 269(13), 1655-1660.
- [21] Rowntree, D. (1987). *Assessing students: How shall we know them?* London: Kegan Paul.
- [22] Sadler, D. R., (2010). Beyond feedback. Developing students capability appraisal, Assessment and evaluation in Higher Education.
- [23] Sax, G., and Newton, J.W., (1997). Principles of educational and psychological measurement and evaluation. London: Wadsworth.
- [24] Stubblebine, C.P., (2001). Perception and acceptance of evaluations by supervisors and peers. *Current Psychology*, 20, (1), 85-94.
- [25] Thomas, A.P., Gebo, A.K., and Hellmann, B.D., (1999). A Pilot Study of Peer Review in Residency Training. *Journal of General Internal Medicine*, 14(9), 551-554. doi: 10.1046/j.1525-1497.1999.10148.x