Students, Supervisors and Principals as Evaluators: A Western Perspective

Amanuel Gebru*

ABSTRACT: *Evaluations are performed by three parties: students, supervisors and principals (who often act supervisors) with varying degrees of accountability, professionalism and evaluative weight accorded. In the Ethiopian school system the evaluation of teachers by students was recently, but controversially introduced. The literature, which aims to provide an international background to evaluation practice in the Ethiopian school system indicates that supervisory evaluations are the most accountable and professional implying the differential weight that should be attached. Though students are the principal "consumers" of education, their evaluative judiciousness is shrouded in some uncertainty. The literature also shows that although teachers generally welcome evaluations, they often have serious concerns that they may not be objectively, fairly and reliably evaluated, especially when the evaluation is clearly summative. The qualities desired of professional and ethical evaluators and evaluation systems that can ease such teacher fears are discussed as they relate mainly to primary and secondary school teaching staff.*

Introduction

Societal representations that schools as service givers have failed to satisfy their customers, the society at large and parents and their school children in particular have led to the need for stringent evaluations. A number of reasons have been advanced for teacher evaluation at all levels. The following by Marland (1987) deserve particular mention:

> *Self-knowledge*: it is necessary that teachers know the difference between how well they think they are performing and what their students and supervisors believe they are actually doing
>
> *Curriculum Planning*: evaluation must be included in curriculum planning and development in so far as it supplies a reason for planning
>
> *Comprehensive planning of school*: it is essential that an in depth-appraisal is conducted by schools to support in policy decision
>
> *Professional development*: evaluation provides opportunity for teacher development through diagnosis of professionalism.
>
> - *Accountability*: the consumers of education and their sponsors require

* Lecturer, Department of Foreign Languages and Literature, Addis Ababa University

Students, Supervisors and Principals as Evaluators: A Western Perspective

Amanuel Gebru*

ABSTRACT: *Evaluations are performed by three parties: students, supervisors and principals (who often act supervisors) with varying degrees of accountability, professionalism and evaluative weight accorded. In the Ethiopian school system the evaluation of teachers by students was recently, but controversially introduced. The literature, which aims to provide an international background to evaluation practice in the Ethiopian school system indicates that supervisory evaluations are the most accountable and professional implying the differential weight that should be attached. Though students are the principal "consumers" of education, their evaluative judiciousness is shrouded in some uncertainty. The literature also shows that although teachers generally welcome evaluations, they often have serious concerns that they may not be objectively, fairly and reliably evaluated, especially when the evaluation is clearly summative. The qualities desired of professional and ethical evaluators and evaluation systems that can ease such teacher fears are discussed as they relate mainly to primary and secondary school teaching staff.*

Introduction

Societal representations that schools as service givers have failed to satisfy their customers, the society at large and parents and their school children in particular have led to the need for stringent evaluations. A number of reasons have been advanced for teacher evaluation at all levels. The following by Marland (1987) deserve particular mention:

> *Self-knowledge*: it is necessary that teachers know the difference between how well they think they are performing and what their students and supervisors believe they are actually doing
>
> *Curriculum Planning*: evaluation must be included in curriculum planning and development in so far as it supplies a reason for planning
>
> *Comprehensive planning of school*: it is essential that an in depth-appraisal is conducted by schools to support in policy decision
>
> *Professional development*: evaluation provides opportunity for teacher development through diagnosis of professionalism.
>
> - *Accountability*: the consumers of education and their sponsors require

---

* Lecturer, Department of Foreign Languages and Literature, Addis Ababa University

assurance that schools and practitioners are honouring their contractual and moral commitment , which should lead to their retention. The concepts of *merit pay*, *career ladder* and *master teacher* are linked to accountable performance and evaluation (Schott, 1989).

As Murgatoyd and Morgan (1992:5) write, *the other driving factor (behind the need for evaluations) is the growing recognition of the importance of educational investment and performance to the competitive strategy of nations in the world.* In the shrinking global village, international competitiveness is associated with national educational standards, and the allocation of vast resources to this vital sector apparently needs a quality audit, a constant improvement of the performance of schools whose demand for human and financial resources fiercely competes with demands from other sectors.

The two authors also argue that the evaluation process itself requires substantial resources whether it is done by the consumers of education or by school administrators. Yet questions are being asked: are the evaluation systems useful or effective at all, what are their shortfalls, should schools rely on their students' testimonials or their administrators' appraisals or on a cocktail method ? A comparison of the evaluation modes is presented and discussed .

Supervisory Evaluation

Supervisory evaluation has been introduced in the school system for quite a long time. Yet it seems to have been at times, *a meaningless ritual* as teacher incompetence has often been decried partly because such evaluation has not led (if at all it was properly done) to the rewarding of superior teachers, the encouragement of mediocre ones, and the removal of incompetent ones. Inefficient application has been recorded, casting a shadow of doubt about its merits. Though ideally evaluation should have a healthy purpose of instructional improvement, Styles et al (1979:17) regret that ... the teacher evaluation process has been poorly handled by administrators. They view each evaluation in terms of job orientation rather than career development. Seldom is the objective to improve the quality of education and to further professional development realised.

Such misuse would definitely give rise to mistrust, fear and feelings of insecurity which would apparently be pedagogically counterproductive. The undesirable consequences of errant administration of evaluation have led to many concerns of teachers given voice by Kult (in Styles et al, 1979) as follows:

- The majority of persons observing and evaluating teachers are lacking in proportional academic and pedagogical competence.
- Forms surveyed by evaluators including survey sheets which, at best, are highly subjective and in many cases, entirely useless for teacher improvement since they are devoid of any substance criteria or content. Their value thus is limited to what has been called *watchdog techniques.*
- The absence of any similar evaluation for the evaluators as either evaluators or administrators, which presupposes that teachers comprise the only educational segment that needs improvement.
- The lack of any active and defensible forms of formal evaluation for all professional members of the educational complex which inhibits professional growth at all levels.

There is no question that the evaluation equally needs to be evaluated in terms of its objectivity, reliability and validity. Since there is not a single universally acceptable definition of *effective teachers* and *effective teaching*, the task of setting evaluation items becomes even more difficult. In fact it is for this reason that we have a wide variety of rating forms in different but same level secondary education settings.

Such differences owing in part to differences in the definition of goals and purposes have led to wide mistrust of the evaluation processes as a whole. Evaluation procedures fail to be distinguished, as if one procedure is as good as or as bad as another. There is an overall reluctance to examine what effective evaluation means. Moxley (1978:62) and Styles et al (1979:35) present as an instance a commonly used teacher rating form which appears porous in terms of objectivity and validity of its evaluative items.

The format contains columns. The first column lists 3 categories of teacher characteristics: personal, professional and teaching performance factors which must be rated against a 3-point scale: outstanding, satisfactory or unsatisfactory. Items in each category are:

- *Personal factors*- appearance, co-operation, sense of humour, tactfulness, health, attendance and personality.
- *Professional factors*- flexibility, loyalty to school system, judgement, professional ethics, rapport with staff, rapport with students, rapport with parents, teaching performance.

_ *Teaching performance and classroom management*--classroom organisation and performance, mastery of subject, techniques, command of English, reports and records.

The evaluation is supposed to be based on a classroom observation and follow-up conference. A comment may be included with the average point in the evaluation.

Thus such pitfalls have led to the observation that there is now a widespread agreement that present evaluation methods are most commonly ineffectual at best. Because many evaluation systems currently in use can barely be taken seriously, they usually do not lead to the removal of incompetent teachers who remain in the school system, giving rise to *horror stories* about teachers (Andrews, 1995:1).

Also Wiessel et al, (1984) in Andrews (1995) concluded that *most secondary school teacher evaluation systems in use were not suitable for rating teachers for the merit pay or master teacher programme* .. They also regret that most school districts studied did not have a sufficiently developed evaluation system. Thus evaluation was not followed by administrative action.

Needless to say, a system of evaluation is as good as the evaluators it commands. In this connection, Kult's assertion (in Styles, 1979) that most classroom observers and evaluators are academically and pedagogically ill-prepared relative to classroom teachers seems to have an element of truth. Their professionalism and accountability is often called into question. Andrews quotes a teacher who complains:

> *My last evaluation was very complementary, but it was based on nothing more than casual observation on how I interact with the kinds. What if I was not an effective teacher? My principal does not make in-class evaluations and is not familiar with the classroom. He does not hold conference* *writes every one a complementary evaluation. For the last two year* *missed the district deadline and predated his evaluations* (Andrews 1

Sufficient training for administrators has often been felt to be wanting. N (1990) in Andrews (1995) makes a research based recommenc administrators need to be trained in (1) observation (2) recording classroo (3) conducting pre-and post-evaluation conference with teachers. There not much that can be done about biasing factors including purpose of obs preferred teaching method but it may be important to consider some biases to be aware of in an observation and rating session as articulately presented by Latham et al (in Fiddler, k1992:200):

- *Contrast effect*: observers should be cautious about the tendency to grade performance relative to others rather than on absolute performance in the details of the components of the job.
- *Hallow effect:* one aspect of instruction which is either very good or bad may lead to over-generalisation about performance and hence contaminate evaluation.
- *Similar to the error:* this is a tendency of overrating similar to self-teaching styles and downgrade differing ones.

The removal of the few incompetent teachers has not been a comfortable decision in the least. There are many legal hurdles standing in the way they harm keeping poor teachers which is a damage to the image of the teaching profession. Understandably, the granting of tenure needs a considerable administrative preparation and care since as Wheeler et al (1993) in Andrews (195) write it provides substantial protection against the teacher due to the process, procedures and other protection that may not be available to the non-tenured teacher.

It is important then that the evaluation system and those who implement it must have professional and ethical readiness and power. Hunter (1988a) in Andrews (1995:34) describes that summative evaluation is the common method in schools. He states:

> The purpose of the district's evaluation is summative, based on a years' professional performance, the summative evaluation certifies a professional as belonging to category from outstanding to unacceptable. Summative evaluation is viewed as extremely important but is only a small part of the total time devoted to staff development. It is the final assessment of the district's and the teacher's efforts.

> The summative evaluation is fair and just because it has the following three qualities. First, it is based on many performance samples (not on one observation or on hearsay); second, it is conducted only by an adequately trained evaluator; last, it is based on stipulated criteria with meanings common to teacher and evaluator.

> The evaluators are competent and demonstrate expertise in two key areas. They possess knowledge of the research based, cause-effect relationships between teaching and learning, and they demonstrate competence in observation and conferencing skills.

Also Bridges (1990) outlines the requisite qualities an efficient professional evaluator needs to possess.

- the ability to describe and analyse what is happening in a teacher's classroom
- the ability to provide an unbiased rating of a teacher's performance
- the ability to prescribe remediation that is appropriate to the teachers classroom deficiencies
- the ability to conduct conferences with teachers regarding their instructional performance
- the ability to document matters related to the above 5 points
- knowledge of the legal bases for evaluating and dismissing incompetent teachers.

Wise et al (1985) studied and found as exemplary different but effective systems of evaluation in four American schools involving co-operation and commitment of the staff and administration. They indicate the commonalties leading to the evaluation success as follows (Wise et al. 1985: 101).

- top-level leadership and institutional resources were applied to the evaluation process
- evaluators charged with the task of implementing the evaluation system had the necessary expertise to perform their task administrator-teacher
- collaboration enabled a common understanding of evaluation goals and process and, the evaluation process was comparable with the districts' overall goals and organisational context.

It was also observed that the rating forms in these schools reflected the philosophy educational goals and community values of the schools in the study. Contributory t' even greater evaluative success could be the inclusion of clarity of a framework c criteria for teacher professionalism (Washington, 1970).

Menatt's (Menatt, Palmer, and Hidlebough, 1970) *Management by Objectiv* *Evaluation,* two commonly discussed models have been in use in many Americ schools. In common, these models have (a) set educational goals, (b) the participation of teachers in the appraisal process (c) core teaching effectiveness standards and criteria. Menatt's model is a broad system of evaluation encompassing teaching staff, administrators, and the programme of education itself. The four steps of the model include:

- determination by the school system of the minimum acceptable standards
  of teaching performance. According to Gudrige (1980) these may include
  instructional productivity, interpersonal skills, classroom management,
  scholarly provocativeness, and outside class behaviour
- pre-observation conference with teacher, self-evaluation, classroom
  observations, and post-observation conferences
- setting discrete and reliably scaleable job-targets (3-5) for instructional
  improvement by teacher
- re-evaluation of teacher and setting of new instructional targets.

Principally, the models have the intention of helping the professional development of
the classroom teacher and harmonising teacher performance goals with school
policies.

Student Evaluations

Since as early as the 1920s, students have participated in the rating of their teachers.
However, despite massive research, student evaluations remain intriguingly
controversial. Whether students can confidently and reliably assess their teachers
whom they consider certified as knowledgeable and who are employed by people who
know better is replete with unabated controversies.

For one thing, students as evaluators seem to have several advantages. One merit
clearly is that the average student has the opportunity to know his/her teacher who has
a considerably lengthier contact than a principal or supervisor, who may help to
accumulate sufficient data about the considerably lengthier contact than a principal or
supervisor, who may help to accumulate sufficient data about the teacher for end-of-
semester evaluations by the students. But can this really sufficiently help students to
judge instructional excellence reliably and objectively, especially when effective
teaching is difficult to accurately define and teacher rating forms are often inexpertly
written up and their validity questionable? The answers do not come forth so easily.

A major problem has been the alleged link between marks and evaluations. There are
doubts that lenient graders may receive generous evaluations, and vice versa, hence
the contamination effects marks have, reducing the validity and reliability of student
evaluations. The research on the subject often does not offer complete comfort. One
of the strongest criticisms of student ratings, Hocutt (in Andrews 1995) argues that
student ratings do *not yield an objective measure of the teachers' performance but a*

subjective index of student satisfaction. He adds that studies on the subject often report contradictory findings. He further says that *the greatest grading inflation in educational history* has come about with teachers concerned about their employment generously awarding undeserved marks apparently in return for expected lenient ratings. This survival technique seems to be certain to contaminate student testimonials of teaching effectiveness.

Rotem and Glasman (1979) doubt students' evaluative competence because they are arguably inexperienced, incompetent, short of perspective and influenced by matters not directly relevant to instructional quality. Such suspicious extraneous factors have been sex, teacher rank, class size, class level and grade expectations which some studies have found to influence student ratings. For instance, Villano (in Elmore and Pohimann, 1978) found that full professors and associate professors received higher ratings than instructors and assistant professors, while Cashin (1990) found that sex, age, teaching experience, and research record have insignificant relationship with student evaluations.

In a study of faculty attitude toward student evaluations, Marsh (1982) reported that although 80% felt that students ratings are important, only 38% believed that such ratings give a correct picture of instructional performance. Course difficulty (72%), grading behaviour (68%), teacher popularity, student interest in subject (62%), class size (55%) were mentioned as influencing student evaluation of teachers. In a different study, Marsh (1984) found that the students evaluate small classes more favourably. Intriguingly, while 35-100 student classes received least favourable evaluations, classes of 10-15 students received highest ratings. But other studies found no significant causal link between classes size and student ratings (Cashin, 1990).

Though there is literature supporting the occurrence of evaluation bias, a number of studies seem to suggest the reliability of student ratings. Lazovik (1972), Centra (1979) and Costin (1968) have found student ratings are to a large extent consistent and reliable. Also Hamond et al (1983) have reported a reliability of .8 to .9 in their review of the subject.

Nevertheless, given the possibility that students may often be lacking perspective and qualification as evaluators as some studies have shown, it becomes easier to understand teachers occasional fears about and lack of confidence in student evaluations. Hence, there is a need for the participation of teachers in the development of evaluation items understandable to students and agreeable to teachers. Such an arrangement would ally teacher fears and insecurity and restore confidence in the

evaluation system. Added to this, a greater involvement in the endeavours of professional classroom observers would contribute even more to the restoration of teacher confidence and removal of fears of summative effects.

Conclusion

The literature survey indicates that the subject of evaluation is probably the most researched sub-field in education. Over two thousand researches, notwithstanding, the area continue to offer academic and practical challenges. Research has often come up with contradictory results, resulting in the combative past many evaluation systems have had to experience.

The practical implications have been discussed in the highest of places--in parliamentary circles. Teacher concern has been high in the teaching community. Indeed often their concern that there should be guarantees of a due process should be appreciated. Evaluator competence, both ethical and professional, may often be lacking as the literature seems to indicate; hence the need for the professional competence of evaluators so that teaching staffs have confidence in the system.

The evaluation programme should also build up a healthy relationship between the administration and teaching staff members, encouraging the involvement of the latter in the design and implementation of appraisals. Teacher fears of summative consequences should be allied, and developmental needs should override. Possibly evaluation will continue to have problems, but much like exams, it will remain a necessary evil.

References

Aleamoni, L.M.(1981). *Student Ratings of Instruction* In J. Millman (ed.), **Handbook of Teacher Evaluation**. Beverley Hills: Saga Publications.

Andrews, H.(1995). **Teachers Can Be Fired: The Quest for Quality**. Chicago: Catfeet Press.

Atsede Beyene (1991). An Investigation of the Relationship Between Students' Evaluation of Teachers and Evaluation of Teachers By Principals in Ethiopian Teacher Training Institutes. MA Thesis (unpublished).

Bridges, E.M (1990). *valuation for Tenure and Dismissal.* In J. Millman and L. Darling Hamond, (eds), **The New Handbook of Teacher Evaluation: Assessing Primary and Secondary School Teachers**. Des Moines, IA: Iowa Association of School Boards.

Cashin, W.E. (1990) *Student Ratings of Teaching: a Summary of the Research*: **Management Newsletter**, 4, 1, 2-7.

Centra, J. A. (1979). **Determining Faculty Effectiveness**. San Francisc: Jossey Bass.

Costin, F. et al (1971). S*tudent Ratings of College Teaching: Reliability, Validity and Usefulness.* **Review of Educational Research.** 41 (5) 511-35.

Elmore P. and J. Pohlmann (1978). *Effect of Teacher, Student and Class Characteristics on the Evaluation of College Instructors.* **Journal of Educational Psychology.** 70(2) 187-192.

Fiddler, B and R. Cooper (1992). **Staff Appraisal and Staff Management in schools and Colleges: A Guide to Implementation**. Longman.

Gudridge, B. M. (1980). **Teacher Competency: Problems and Solutions:** Arlington, va: American Association of School Administrators.

Hammond etal, (1983). *Teacher Evaluation in the Organizational Context: A Review of the Literature.* **Review of Educational Research.** 53(3) 285-328.

Lazovik, G. (1989). *Evaluation of College Teaching-Guidelines for Summative and Formative Procedures.* Association of American Colleges, Occasional paper.

Marland, M. (1987). *Appraisal and Evaluation Chimera, Fantasy or Practicality* in Bunnell, S. (ed.). **Teacher Appraisal in Practice.** London, Heinemann Educational.

Marsh, C. (1992). **Key Concepts for Understanding Curriculum**. The Falmer Press.

Mennat, R. P., K. L.Palmer, and E. Hidlebough (1976). **Evaluating Teacher Performance with Improved Rating Scales.** NASSP Bulletin, 60(401), 21-23.

Moxley, R. (1978). T*eacher Evaluation. Images and Analysis.* **Journal of Teacher Education**. 29 (6) 61-66.

Murgatroyd, S. E C. Morgan, (1992). **Total Quality Management and the School**. Open University Press.

Redfern, G. B. (1980). **Evaluating Teachers and Administrators: A Performance Objectives Approach.** Boulder, Colo.: West Press.

Rotem A. and N. Glasman (1979). *On the Effectiveness of Students Evaluative Feedback to University Instructors.* **Review of Educational Research.** 49(3) 497-511.

Schott, C.J. (1989). **Improving Training and Evaluation of Teachers at the Secondary School Level.** The Edwin Mellen Press. Lampter. Wales.

Stiggins, R.J. and Duke, D. (1988). **The Case for Commitment to Teacher Growth. Research on Teacher Education.** Albony, State University of New York.

Strike, K. A. (1990). *The Ethics of Educational Evaluation.* In J. Millinan and L. Darling-Hammond (eds.). **The New Handbook of Teacher Evaluations: Assessing Elementary and Secondary School Teachers.** Newbury Park. CA: Saga Publications.

Styles, G. (1979). *The Supervision of Teacher Evaluation. A Brief Overview.* **Journal of Teacher Education.** 30(2) 17-24.

Wise et al. (1985). *Teacher Evaluation: A study of Effective Practices.* **Elementary School Journal.** (8)6 61-119.