

# TEACHER EVALUATION: BETWEEN POSSIBILITY AND REALITY

Roxana Maria GHIAȚĂU<sup>a\*</sup>

<sup>a</sup> “Alexandru Ioan Cuza” University of Iași, Romania

---

## Abstract

*More and more countries show an increased interest in establishing teacher evaluation systems as an integral part of educational policies. Beyond the realistic preoccupation of judging the merits of educators, there remain several questions about the effective achievement of this process. The objectives of the present study are theoretical, namely the investigation of the weaknesses and strengths of teacher evaluation models as presented in the Western literature. We will analyse the two fundamental perspectives on evaluation, namely traditional evaluation and recent evaluation based on models. Regarding recent approaches, we shall insist on the value-added assessment model, which has triggered the most controversy in the world. A series of conclusions and recommendations on evaluation implementation will be presented. The real reform of teacher assessment should consider the broader context of transforming the education system. We advocate the development of new teaching and learning systems that align the assessment of teachers and students with the goal of improving both.*

**Key words:** teacher evaluation, traditional evaluation, value-added models

---

## 1. Introduction

More and more countries show an increased interest in setting up teacher evaluation systems as an integral part of educational policies. We propose the following definition for the concept of teacher evaluation: a formal, systematic process of collecting information about the degree of achievement of the educational objectives and the interpretation of this information with the help of specific methodologies, to achieve a competent value judgment on the potential and

---

\* Corresponding author. Associate Professor PhD Roxana Maria Ghiațău  
E-mail address: roxanag@uaic.ro

performance of teachers in teaching primarily, as well as in other related (secondary) professional activities (Ghiațău, 2016). The objectives of the present study are theoretical, namely the investigation of the weaknesses and strengths of the teachers' evaluation models as presented in the Western literature. Teacher assessment is presented in a variety of forms, considering numerous criteria (the author making the assessment, the centralization / decentralization axis, the time of the evaluation, the position occupied within a program, process or activity). In this paper we are interested in evaluation from a diachronic perspective, namely traditional models and practices of evaluation and more recent evaluation models and practices.

From an epistemological point of view, shaping patterns in a field is particularly valuable. The model has been imposed as a universal and indispensable tool, the achievement of knowledge by models being suitable to all sciences (Bîrzea, 1995). By models and theories, the field makes the leap from empirical knowledge to scientific knowledge. Analysing the dynamics of the relationship between theory and model in Mario Bunge's conception, Mariana Momanu (2002) points out that the model is the meaning nucleus of the theory, whereas theory develops the nucleus it contains.

There is no universal recipe or model for teacher evaluation. For Victoria Gheorghe (2010), a model of professional evaluation is not only a specific way to proceed in a unique, invariable manner, but also a system of theses, principles, validated by experimental and practical research related to the professional competences, namely the knowledge, skills, abilities, attitudes, but also the personal qualities a teacher needs to have. The developed assessment model will constitute the basis for elaborating the evaluation methodology. The relation between the model and the methodology is the following (Gheorghe, 2010): the evaluation model represents the theoretical, explicative and comprehensive framework in relation to the diversity and complexity of pedagogical competence, while the methodology is administrative, comprising the evaluation stages in a well-established sequence.

## **2. Traditional evaluation**

Many researchers have spoken about the shortcomings of traditional assessment. Darling-Hammond, Wise & Pease (1983) argued that existing systems for evaluating educators were illogical, simplistic, unfair, counterproductive or simply unproductive. In their syntheses on teachers' assessment methods, Frase & Streshly (1994) also highlighted their formal, "purely ceremonial" and inaccurate role. There are trenchant opinions about this way of conducting assessment, for example that of Scriven: "Teacher evaluation is a disaster. The practices are

shoddy, and the principles are unclear. [...] Using classroom visits ...to evaluate teaching is not just incorrect, it is a disgrace" (1981, pp. 244-251). Despite the criticism, this evaluation system enjoys a strong tradition, has many followers, any attempt for renewal are extremely difficult.

A standard evaluation model has been used in schools for decades, being truly the "dominant model" for evaluating teachers (Haefele, 1993). This model continues to be the most widely used and accepted in public school systems. In most schools, teachers without tenure are observed two or three times a year through classroom visits. Evaluations are made by the school principal (in accordance with certain standard forms) and last between 20 and 30 minutes. Tenured teachers are observed yearly but less frequently (Haefele, 1993). Sometimes, the observation is followed by a meeting where the principal's responsibility is to describe the strengths and weaknesses of the teaching performance and offer suggestions for the teacher's improvement (Haefele, 1993). This process usually involves communication in one direction, from evaluator to teacher (Gitlin and Smyth, 1990, apud Mertler, 1997). All teachers are assessed using a common instrument and established procedures.

The advantages of this dominant model are:

- (a) school principals with qualifications and expertise can significantly help teachers to improve their teaching performance;
- (b) the principal can demonstrate a concern for quality training;
- (c) the evaluation steps (for renewing the work contract, for position appointment, or for salary-related matters) can be clearly determined through this model.

However, research has shown that even this deficient model is not applied in a consistent, deep manner. Huddle (1985, apud Mertler 1997) synthesizes the results of the *High School and Beyond Teacher and Administrator Survey* from the *National Institute of Education*. Teachers were asked how often they had been observed by a supervisor during the previous year. A quarter (26%) of the respondents indicated "no", another quarter (27%) answered "once", and almost a quarter (23%) answered "twice". As an assessment system, the dominant model is devoid of three essential components (Mertler,1997): (a) the evaluators lack the two essential sets of skills needed to evaluate teachers: skills in performance assessment and skills in communicating the results of the evaluation process to the assessed teachers; (b) there is insufficient time for evaluation and subsequent improvement activities; (c) the process of connecting staff development to teacher evaluation is unclear or does not exist.

For Charlotte Danielson and Thomas L. McGreal (2000, p. 3), the traditional evaluation system has six major issues: criteria for evaluation are outdated, limited; few values are shared about what "good teaching" means; lack of precision in performance evaluation (the satisfactory /

unsatisfactory dichotomous scale); communication is hierarchical, unidirectional. The school principal (or the inspector) conducts the observation, takes notes, drafts the report and then gives the maximum mark. In the absence of clear evaluation criteria, this feed-back is likely to be idiosyncratic; It does not consider the difference between experienced and novice practitioners; limited expertise of the principal. This combination of factors leads to a culture of passivity and protection. It is unlikely that a teacher will be honest about the difficulties he/ she may come across if he/ she is afraid that the described difficulties may be included in the final assessment report as deficiencies.

### **3. Recent models and practices in teacher evaluation**

Concerns in the field of teacher evaluation are not new, but in the last twenty, thirty years there has been a significant change, at least theoretical, in terms of evaluation stakes and methods. To change the existing system, obviously, the flaws of traditional perspectives had to be mitigated. Therefore, the attention of the creators of teacher evaluation systems has been directed in several directions: establishing the basic principles for the evaluation, rethinking the purposes of evaluation, integrating assessment into school life, and methodological decentralization (evaluation based not just on two, three methods, namely observation and class visits). These concerns are in line with two tendencies that seem to be developing (Figazzolo, 2013). The first trend is the use of so-called value-added approaches, and the second trend, sometimes in contradiction with the former, is the growing support of comprehensive assessment approaches.

#### ***3.1. A Model of Evaluation Based on Standards***

*The Duty-Based Teacher Evaluation Model (DBTE model), Michael Scriven*

The author places the concept of duty at the centre, building a list of core-duties, used to assess the teachers' competence. A statement of values in the teaching profession is very incomplete; moreover, it is left to the traditional concept of "implied duties." Vagueness is a feature, hence the idea that "anyone can teach". Professionalism refers to high competence in teaching and commitment to satisfactory performance at work. There are four general dimensions of competence and excellence in teaching, associated with the teacher's merit (Scriven, 1988, p. 320): *quality* in the taught and learned content, *quantity* in learning, *ethics* and *professionalism* in the conducted process. Thus, we are not at risk of confusing "good teaching" and "teaching by the rules". The fundamental duty of a teacher is to teach students valuable knowledge. These include

both constitutive tasks and secondary duties. Scriven (1988, p. 25) draws a distinction between merit and value. The value is extrinsic, situational, merit is intrinsic, professional.

The teacher's duties are described in Table 1 (Scriven, 1994).

Table 1. The teacher's duties (Scriven, 1994)

1. Knowledge of the subject to be taught	- in the field of tenure, for example, secondary mathematics; - in areas beyond the specific curriculum, for example, composition, spelling
2. Instructive competence	a) communication skills (using the appropriate vocabulary for the age of children, providing examples, using the inflections of voice and body language); b) management skills: ✓ class management (knowledge of reward strategies and disciplinary practices, etc.); ✓ management of the individual progress of the student; ✓ management of emergency situations (fires, tornadoes, earthquakes, floods, stroke, violent attack); c) building the course and course-improving skills: ✓ design of courses; ✓ selection and creation of teaching materials; ✓ use of special resources I. Local sites II. Media III. Specialists; ✓ evaluation of the course, teaching, materials and curriculum
3. Competence in evaluation	a) knowledge of the students' assessment options; b) knowledge of test construction and test management skills; c) knowledge of scoring, hierarchy, appraisal practices; ✓ knowledge of the scoring process (ensuring fairness, anonymity - where it is required); ✓ knowledge of outputs (the results meet the standards); d) recording and reporting the students' results; ✓ knowledge of options and obligations in reporting results; ✓ appropriate reporting process (for students, administrators, parents, others)
4. Professionalism	a) professional ethics; b) professional attitude; c) professional development; d) devotion to the profession: ✓ knowledge about the profession; ✓ help offered to beginners and colleagues; ✓ work for professional organizations; ✓ research on teaching; e) knowledge of non-standard but contractual duties, for example, supervising chapel services in a theological school; f) knowledge about school and community

The validity of the duty-based approach derives from a single fact, namely the employee's obligation to comply with these duties as far as possible, with existing resources.

### **3.2. Value-added assessment models**

To understand value-added assessment, we need to resort to certain notions from the evaluation theory. There are several ways of achieving evaluation and reporting the results: in the normative assessment the students' results are compared with those of their peers, so that half of the students are below average; in criterial evaluation the student's result is compared with an ideal standard of achievement, from the curriculum. Characteristic of the value-added assessment is the fact that it compares the current level of the student with his / her previous level. Therefore, *value-added assessment is defined as a method of analysing student test data to establish their progress in learning by comparing the current level with their previous level (School directors' Handbook, p. 3)*. The value-added methodology does not involve a different type of test, but a new way of looking at scores. Value-added (VA) is a new concept that can be used in many contexts: to evaluate initial teacher training, to evaluate students, principals, etc. The most innovative and controversial proposal for using VA is teacher evaluation for formative or summative purposes. We mention that the philosophy of applying professional standards is hostile to standardized and value-added tests.

Contestation of this trend has not been late: we cannot put the sign of equality between the evaluation of the performance of a forming, spiritual activity and the evaluation of the performance of a material activity. The productivity of teachers' work is special. Educators are not like factory workers, they do not produce things. In teaching the results are not counted like pieces built on the turntable. The "obligation of outcomes" (Perrenoud, 1996) is not practicable in education, as it turns didactic professionalism upside down. Appropriate for the educators' work is the "obligation of competence", the ability to deal with complex educational situations. We cannot claim X results in Z time, but we can demand observance of general principles. Some comparisons are possible, however between more homogeneous subassemblies.

Linda Darling-Hammond (2008) lists several issues generated by the application of the value-added methodology, such as: statistical manipulation and missing data threaten the validity of results, there is required data collected over long periods, certain types of tests are needed, it is impossible to separate the influences of other teachers on the students, teachers' results are unstable from year to year, etc. Santiago et al. (2009) estimate that only one-third of US countries have state-of-the-art computer systems to "bind" teachers and students in the way needed to apply value-added assessment models. These require major national investment and even years to implement. However, the VA model has many supporters.

Tennessee is the state identified with the highest rate of value-added assessment. The system dates to 1992, when value-added assessment was implemented as an integral part of educational

reform. Using a complex statistical method developed by William Sanders, a Tennessee University statistician, the *Tennessee Value-Added Assessment System (TVAAS)* aims to:

- provide the public with information on school performance and appropriate data for administrators on teacher performance;
- inform teachers, parents and the public on how schools support each child to make academic progress each year;
- inform school directors on the identification of weaknesses, even in the best schools.

TVAAS is a statistical methodology that starts with testing each student in each class in several school subjects. In 1997, Tennessee tested students in the second to the eighth grades in Reading, Mathematics, Language, Science and Social Studies. The system determines the effectiveness of schools and teachers, relying on the academic growth of the student over time. An integral part of TVAAS is a massive, longitudinally merged database linking students and their outcomes to the schools and systems in which they are enrolled, as well as to the teachers to whom the results are attributed (Sanders, Horn, 1998). Research conducted using information from the TVAAS databases showed that ethnic origin, socio-economic level, class size, and class heterogeneity are poor predictors of student growth. Rather, the effectiveness of the teacher is a major determinant of academic progress. The "effects" of teachers on student outcomes proved to be additional, with little evidence that effective teachers can later compensate for the impact of ineffective ones (Sanders, Horn, 1998). Studies have shown that approximately 69% of teachers cannot be accurately assessed in this way, namely: beginner teachers, teachers in areas not tested with standardized annual tests, gymnasium first-grade teachers (if previous tests results are not available), special education teachers.

Audrey Amrein-Beardsley (2008) presents a series of shortcomings of a famous *Education Value-Added Assessment System (EVAAS)*, the most sophisticated value-added assessment model:

- *Lack of studies on the validity of the model, namely content validity and criteria validity.*
- *Use of data in formative ways*

Of the schools that implemented the model, too few seem to have used the output data in a formative way. It was found that confusing data reports and the lack of training of teachers and principals in comprehension of data reports prevented schools from using data to improve student learning.

- *Developers have not made this method fully open to peer review.*

Specifically, they own property information of the computational algorithms needed to manage and solve large systems of linear equations. This makes peer-review by external statisticians impossible.

- *Missing data*

Many specialists report problems caused by incomplete data records. Value-added models require complete and longitudinal high-quality data that is currently lacking in many states. EVAAS developers claim that the model can work regardless of the amount of missing or fractured data found in large-scale databases. They argue that data can be merged at the extent of 90%, but their argument seems improbable, especially if the percentage is expected to rise over a period of 5 years.

- *Average regression*

A serious problem is that the model measures the effectiveness of teachers by deviating from the average, or teachers who teach classes with fewer students are drawn towards the average regardless of whether they are thoroughly or poorly prepared. An expert teacher, just because he / she instructs a larger number of students, may be evaluated above average, whereas an equally good teacher, but with significantly fewer students, may not be evaluated equally well because of the fewer student records. Conversely, an inefficient teacher teaching a large class could be penalized because of the below average evaluation, whereas an equally inefficient teacher teaching a small class may remain undetected.

- *Foreign variables*

Under the EVAAS model, student risk factors (race and poverty) are not controlled. This makes EVAAS – the only sophisticated analytical model to measure student outcomes – not to consider students' background factors, factors that have been proven in decades of research as affecting learning achievements. EVAAS developers say that the effects of race and socio-economic status on students' achievements are negligible (since, as they argue, student progress is analysed from one year to another and from one field to another, and the influence that these variables may have on learning is controlled).

Larsen (2005) identifies the general and specific effects of current evaluation policies to show how these reforms work to reshape the nature of teaching. Firstly, attempting to capture the complexity of teaching through tests has resulted in *the simplification of the art of teaching to*



*linear testing formats and performance checklists.* Secondly, another weak point of current approaches is the fact that there is cultivated a forced, artificial visibility of the performance instead of a genuine concern for the profession. The whole teacher evaluation based on the responsibility model relies on the notion of visibility (Larsen, 2005). It is true that standardized models are easier to administer and evaluate, but they do not allow the capitalization of pluralism, differences and disagreements. Instead of focusing the educational reform exclusively on the political purpose of delivering public accountability, it would be more useful to channel education to create excellence and acknowledge teaching as a creative, versatile, diverse and complex activity. This is the challenge faced by decision makers in educational policies in the early 21<sup>st</sup>-century years.

#### **4. Conclusions and recommendations**

Evaluation is just one component of the global teacher development system. It is not sufficient to improve the quality of teachers. Recruiting, training, employing resources for equipping schools are equally important. The TALIS report contains an important breakthrough, pointing out the distance between stated goals of teacher evaluation and its real effects. The real reform of teacher assessment should consider the broader context of transforming the education system. We advocate the development of new teaching and learning systems that align the assessment of teachers and students with the goal of improving both. Student learning standards must have a systemic connection with teacher training and assessment. Teacher evaluation cannot be achieved without a set of leadership principles. The principles guide the appreciative activity, maintaining a unitary framework of the whole approach. We have attempted to outline such principles, based on the suggestions of some experienced authors in the field (Peterson, 2000, Stronge & Tucker, 2003) or the illustrations offered by professional associations (NEA's Professional Standards and Practice Committee, 2011).

*Firstly, safe and open collaboration is required.* If teacher evaluation systems are introduced, these should be developed and agreed upon by and with teachers and organizations. When teachers' assessment of teacher practices is transparent, teachers can build professional communities to learn from each other. This process can only take place in non-threatening, formative assessment environments. Secondly, professional development should be a starting point for evaluation, but the purpose of *evaluation should be to strengthen self-esteem and teacher confidence.* Thirdly, *key decisions on evaluation systems need to be approached as*

*closely as possible to the local level and in partnership with teachers and their representatives.* This is the most important principle. Ideals and visions must be balanced, responding to the local context. There is no single solution at the national level. Evaluations should consider the context of teacher work: class size, evaluated discipline, students with special educational needs, available materials, electronic equipment, etc. Fourthly, *assessments must be carried out by highly qualified supervisory authorities* or other locally certified evaluators, whose work is regularly reviewed to ensure the validity and reliability of the evaluation results. Fifthly, *integrated systems should link evaluation procedures with curricular standards, professional development activities and decisions on human capital* (selection, tenure, career advancement). The reform of teacher evaluation must take place with the understanding that it is part of a system designed to increase efficiency. Finally, *validated assessment tools are essential.* They must be based on widely accepted standards of teaching, which attempt to describe a series of teaching behaviours using several evaluation methods.

## References

- Amrein-Beardsley, A. (2008). Methodological Concerns About the Education Value-Added Assessment System. *Educational Researcher*, 37- 65.
- Bârzea, C. (1995). *Arta și știința educației* [Art and science of education]. Bucharest: Didactic and Pedagogical Publishing House.
- Danielson, Ch., & McGreal, Th. (2000). *Teacher evaluation to enhance professional practice.* Educational Testing Service (ETS).
- Darling-Hammond, L. (2008). Teacher Quality Definition Debates: What is an Effective Teacher?. in Thomas L. Good (Ed.), *21<sup>st</sup>-Century Education: A Reference Handbook*. Sage Publications, II-12, II-23.
- Darling-Hammond, L., Wise, A., & Pease, S. (1983). Teacher Evaluation in the Organizational Context: A Review of the Literature. *Review of Educational Research*, 53, 3, 285-328.
- Figazzolo, L. (2013). The use and misuse of teacher appraisal, an overview of cases in the developed world. *Education International*. Retrieved from <http://download.eiie.org/Docs/WebDepot/TeacherAppraisal.pdf>.
- Frase, L. E., & Streshly, W. (1994). Lack of accuracy, feedback, and commitment in teacher evaluation. *Journal of Personnel Evaluation in Education*, 8(1), 47-57.

- Gheorghe, V. (2010). Rezumatul tezei de doctorat *Metodologia evaluării personalului didactic din învățământul preuniversitar. Competențe și instrumente*, [Summary of the Doctoral Thesis Methodology of Assessment of Teaching Staff in Pre-University Education. Competences and tools]. Bucharest: University of Bucharest.
- Ghițău, R. (2016). *Evaluarea profesorilor. Evoluții și provocări* [Teacher evaluation. Evolutions and challenges]. Iași: The European Institute.
- Haefele, D.L. (1993). Evaluating teachers: A call for change. *Journal of Personnel Evaluation in Education*, 7(1), 21-31.
- Larsen, M. (2005). A critical analysis of teacher evaluation policy trends. *Australian Journal of Education*, 49(3), 292-305.
- Mertler, C. A. (1997) *Students as stakeholders in teacher evaluation: Teacher perceptions of a formative feedback model*. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Chicago, Illinois. (ERIC Document Reproduction Service, no. 413 328).
- Momanu, M. (2002). *Teoria educației* [Education theory]. Iași: Polirom.
- NEA's Professional Standards and Practice Committee, Cartea Albă, 2011, Retrieved from <http://www.nea.org/assets/docs/Transformingteaching2012.pdf>.
- Perrenoud, P. (1996). L'évaluation des enseignants: entre une impossible obligation de résultats et une stérile obligation de procédure. *Éducateur*, 10, 24-30.
- Peterson, K. (2000). *Teacher evaluation: a comprehensive guide to new directions and practices*. Corwin Press, second edition.
- Sanders, W., & Horn, S. (1998). Research Findings from the Tennessee Value-Added Assessment System (TVAAS) Database: Implications for Educational Evaluation and Research. *Journal of Personnel Evaluation in Education*, 12(3), 247-256.
- Santiago, P., Roseveare D., van Amelsvoort, G., Manzi, J., & Matthews, P. (2009). *Teacher Evaluation in Portugal – OECD Review*, Retrieved from <http://www.oecd.org/dataoecd/17/32/43327186.pdf>.
- School directors' handbook*, Evergreen FreedomFoundation, Retrieved from <http://www.ewfa.org/pdfs/Value-Added.pdf>.
- Scriven, M. (1981). Summative Teacher Evaluation. In J. Millman & L. Darling-Hammond (eds), *The New Handbook of Teacher Evaluation: Assessing elementary and secondary school teachers* (pp. 244-271). Newbury Park, CA. Sage.
- Scriven, M. (1994). Duties of the Teacher. *Journal of Personnel Evaluation in Education*, 8(2), 151-184.
- Stronge, J. H., & Tucker, P. D. (2003). *Handbook on teacher evaluation: Assessing and improving performance*. Larchmont, NY, Eye on Education.

Weiss, E., Weiss M., & Gary, S. (1998). *New Directions in Teacher Evaluation*, ERIC Clearinghouse on Teaching and Teacher Education Washington DC. ERIC Digest, Retrieved from <http://www.ericdigests.org/1999-4/new.htm>.