

# Democratic evaluation of clinical clerkship in a medical school. A case study

*Piersante Sestini, Rossella Angotti, Claudia Commisso, Mario Messina*

*Department of Medicine, Surgery and Neurosciences, University of Siena. E-mail: [piersante.sestini@unisi.it](mailto:piersante.sestini@unisi.it)*

---

## ABSTRACT

We report a case study of democratic evaluation of a clinical clerkship for medical students. We used a mixed model combining interviews and questionnaires to students, teachers and nurses involved in 9 weekly clerkship rotations in 7 hospital wards. The main outcome variable in quantitative analysis was the score of perceived usefulness for professional development of 42 Observable Practice Activities (OPAs) that each student should perform during clerkship. Scores were higher when the OPA was performed more consistently, was performed in small groups, and when the teaching and the evaluation methods were perceived as valid. The overall satisfaction for the program of teacher was high and did not correlate with students' perceptions. The burden for patients was perceived as slightly higher in nurses, and did not correlate with teachers' perceptions. When returned to students, teachers and nurses, these data could contribute to improve the program.

**Key words:** Democratic evaluation, medical students, clinical clerkship

---

## RIASSUNTO

Contrariamente ai tipi di valutazione basati su standard normativi pre-determinati dalla istituzione stessa (valutazione burocratica) o da un corpo professionale esterno (autocratica), la valutazione democratica è un tentativo aperto, indipendente, non normativo (in quanto non vincolante per gli organizzatori né i partecipanti) di dare voce a tutti gli attori coinvolti nel programma valutato. Per questo è più probabile che fornisca una descrizione del programma più dal punto di vista dei partecipanti che da quello delle aspettative degli organizzatori, fornendo così ad entrambi i gruppi suggerimenti utili per svilupparlo e migliorarlo. Abbiamo utilizzato questo approccio per valutare l'impatto della introduzione in un tirocinio clinico del CdL in Medicina di una lista di attività cliniche che gli studenti dovevano effettuare, documentandole su un apposito libretto. Si è costituito un gruppo di valutazione indipendente autorizzato dal Comitato Didattico, costituito da un docente, una specializzanda, ed un gruppo di studenti. Abbiamo utilizzato una metodologia mista, basata su questionari, domande aperte e interviste. Quale principale variabile quantitativa abbiamo scelto la percezione dell'utilità di ognuna delle attività svolte per il proprio sviluppo professionale. Alcune attività non venivano svolte regolarmente. Sono risultate effettuate il 75%, con un massimo di 100% in due reparti ad un minimo del 44%. Limitando l'analisi a queste, la maggior parte delle attività è stata valutata positivamente. Fattori che sono risultati influenzare positivamente i punteggi delle singole attività sono la divisione in piccoli gruppi (6 o meno), la validità delle modalità di presentazione e di valutazione (se previsto), la consistenza di effettuazione fra le varie rotazioni. Questi dati, restituiti ai docenti e agli studenti, possono essere utili per migliorare il programma.

**Parole chiave:** Valutazione democratica, studenti di medicina, tirocinio clinico

## INTRODUCTION

Evaluation has a central role in the design and development of medical education [1]. Of the three main types of evaluation originally described by McDonalds [2, 3], however, (bureaucratic, autocratic and democratic), the latter, despite being largely acknowledged in the world of general education, is seldom used [4] or even mentioned [5] in the medical education literature. Unlike the other types of evaluation, which are based on pre-determined normative standards, stated either by the educational institution itself (bureaucratic) or by an external professional body (autocratic), democratic evaluation is a non-normative, independent, open attempt to give voice to all the actors affected by the educational program. As such, it is more likely to provide a picture of the program as is seen from the participants, than as is expected to be from organizers, and could provide to both groups useful hints for development and improvement [2, 3]. Here we report our experience of democratic evaluation of a change in the organization of the clinical clerkship for 4<sup>th</sup> year medical students at the University of Siena.

## SETTING

Unlike other western countries, medical education in Italy is still heavily based on a strong tradition of lectures and unstructured oral exams, while teaching of practical medical skills is generally unrewarded by academic authorities and marginalized in curricula.

In the School of Medicine of the University of Siena, the first three years are currently almost completely dedicated to biomedical and preclinical matters, so that the 4<sup>th</sup> year clinical clerkship constitutes the first main occasion of practical clinical learning for medical students. It consists in ten weekly rotations of groups of 18-20 students through 7 different wards: internal medicine (2 weeks), general surgery (2 weeks), cardiovascular diseases, respiratory medicine, occupational medicine, haematology and endocrinology. In 2013, the Education Committee of the School of Medicine at the University of Medicine, to promote a more effective practical learning, issued the requirement that during each rotation

each student should document the performance of 3-10 pre-determined clinical activities (Observable Practice Activities, 6) in each ward. These activities, as well as the teaching methods, were freely selected by the faculties of each discipline, and should have been classified as “Having seen”, “Having done” or “Having learnt to do”. Each activity had to be recorded by the students and signed by the teacher on a clerkship booklet.

In an attempt to evaluate the impact of this change and to identify methods to promote more effective practical and patient-oriented teaching, an independent study group was formed, composed by a teacher, a postgraduate student and a group of students, all affiliated to the local branch of organizations interested in the study and promotion of medical education (the Italian Society for Medical Pedagogy, SIPeM, and the Italian Secretariat of Medical Student, SISM). The group was acknowledged by the Education Committee (the local organism ruling the course) and was allowed to operate independently. The aim of the group was to evaluate the factors associated with better perceived usefulness of clinical activities taught to 4<sup>th</sup> year medical students during their first structured clinical clerkship, giving voice to all the actors of the process: students, teachers and nurses, the latter used as proxy for patients involved in clinical teaching.

## METHODS

As a conceptual guide for evaluation, the group selected the principles of democratic evaluation [7] as described in an authoritative textbook [8], substituting however the existentialist philosophical underpinning that characterize that author with a critical rationalistic background valuing Popperian concepts like truth as correspondence of propositions with facts, democratic education as a means to promote personal development rather than selection of excellence, piecemeal engineering, unintended effects of rational actions, and identification of failure as a tool to promote knowledge and improvement [9]. Accordingly, a mixed method approach was devised combining open comments, semi-structured interviews, and closed answer, quantitative questions [10].

At the end of the clerkship, in December 2013, the students were asked to complete a questionnaire asking, for each activity, whether they felt to have effectively accomplished it, whether they considered the method by which it was taught or presented to be valid, how they felt it was useful for their professional development (on a 10 point scale), and, for activities classified as “having learnt to do”, whether they considered valid the method by which they had been assessed. Open comments were also invited. A digital form of the questionnaire and an open forum were also offered on dedicated website built with the open source learning platform Moodle ([www.moodle.org](http://www.moodle.org)) for students who could not complete the paper version.

In addition, a group of students was charged to contact each teacher for a semi-structured interview about the activities in which they had participated, their experience and feelings about the clerkship, and the impact on patients and clinical activities. They were also asked to rate, on a 10 point scale, their overall satisfaction, their satisfaction about the students’ participation, and their perception of the burden of the clerkship program on the ward routine and on patients.

Furthermore, at least one nurse was interviewed in each ward or outpatient clinic involved in the clerkship, asking for comments on the impact of the activities on the care process and on patients. They were also asked whether nurses had been directly involved in the educational activities and to rate, on a 10 point scale, their assessment about the students’ behaviour and their perception of the burden of the clerkship program on the ward routine and on patients.

As many students reported a poor correspondence between the original classification of many of the learning activities (watch, do, learn to do) and their actual content, based on teachers’ interviews and on a focus group with students from several different clerkship groups, the activities were re-classified in 5 mutually exclusive categories: lessons, ward rounds, watching clinical procedures, performing clinical procedures on patients, discussing clinical cases and/or interpretation of reports, performing clinical procedures on peers. Furthermore, for each activity was recorded whether was performed

in subgroups, the method of assessment (if any), and whether it was performed consistently across the different groups/weeks.

There were 42 activities (see appendix): 12 were proposed as “having seen”, 2 as “having done”, and 28 as “having learnt to do”. We re-classified them 4 as lessons, 5 as ward rounds, 8 watching a procedure, 9 case discussion, 11 procedures on patients and 5 on peers. There was considerable overlapping between the types of activity (e.g. case discussions or procedures on patients where sometimes performed during ward rounds), and considerable variability from week to week (e.g. lesson or case discussions in place of ward rounds etc.) was reported for 13 activities. Twenty-eight were performed in groups of 6 students or less. Most of the activities were assessed only informally with questions to the group, only 5 were formally assessed using an OSCE.

A total of 146 students and 33 teachers participated in the clerkship. Questionnaires were filled by 92 students (63%), and interviews/questionnaires were obtained by 28 teachers (85%) and 10 nurses, representative of all the clinical settings involved. Questionnaire data were transcribed in digital form using the software Epi Info 7 (CDC, Atlanta, GA), and analysed using Stata 12 for Windows (Stata corp, College Station, TX).

Activities with a score of 1 or less and all other scores of zero or missing were considered as not having been performed.

Statistical analysis was performed using generalized linear models using Gaussian, binary or logit distributions as appropriate, including the student group as a random variable. The primary outcome variable was the perceived usefulness for professional development of each activity. A score for each discipline was computed using the average of all the activities performed during the rotations in that ward, or of all the teachers or nurses belonging to each ward. Except where specified, data from activities considered as not performed were excluded from analysis. Unless stated otherwise, quantitative data is presented as mean and 95% Confidence Intervals. [11]. All *P* values are two tailed.

Open comments and interview reports were subjected to content analysis and similar concepts

were grouped by family resemblance [12]. As most comments were rather terse and repetitive, this process was relatively straightforward.

Respecting the requirement of democratic evaluation for confidentiality, all the activities and wards are presented in anonymous form.

## RESULTS

Many students reported that, during their rotation, some of the activities were actually not performed. The rate of non-performance varied between activities and between wards, with some wards consistently performing virtually all activities during each rotation and others failing to do so in different measure, with one missing to perform more than 50% of the activities (Table 1).

Within the same ward, some activities were missed more often than others.

Procedures between peers were the ones performed most consistently (virtually 100% of the

times) and procedures on patients were the ones missed more often (performed 64% of the times), with all the others ranging between 76 and 78% of the times.

When the activities were performed, mean student scores varied greatly, from 4 to over 8, with 64% of them reaching a mean score of 6 or more (Figure 1). A strong correlation was observed between the percentage of performance and the mean score when they were performed, with less performed activities usually, but not always, scoring lower than the ones performed more consistently (Figure 2).

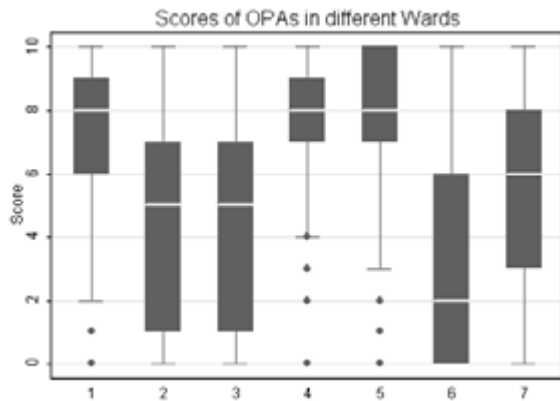
In multivariable analysis, the percentage of performance, considering to have effectively reached the aim of the activity, the appropriateness of the teaching method and the fact that was performed in small groups (6 or less) were all significantly associated to better scores, with the appropriateness of the method having the greater effect (a difference of 2.9 points, 2.4-3.4), followed by effectiveness (+1.7, 1.2-2.2) and small groups (+0.3 points, 0.1-

Wards	OPAs expected	% Performed
Ward #1	368	96%
Ward #2	368	68%
Ward #3	368	71%
Ward #4	644	99%
Ward #5	276	100%
Ward #6	920	44%
Ward #7	920	78%
Total	3,864	75%

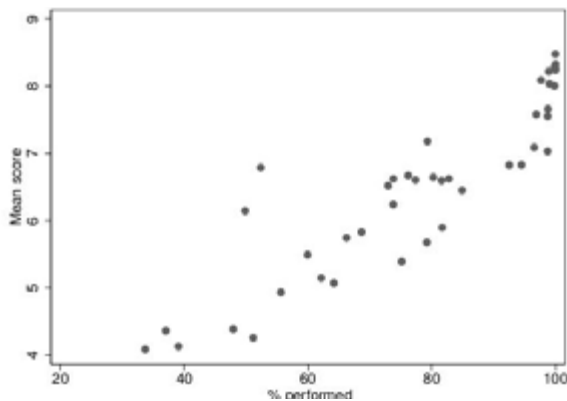
**Table 1.** Number of OPAs expected and percentage effectively performed during students' rotations in 7 wards

Factor	Mean difference in score
Performed in small groups (6 or less)	+0.3 (0.1 - 0.5)
Performed consistently among rotations	+1.9 (2.1 - 2.6)
I have effectively seen/performed/learnt	+1.7 (1.2 - 2.3)
The teaching method was valid	+2.9 (2.4 - 3.4)
The evaluation method was valid	<b>+1.9 (1.5 - 2.3)</b>

**Table 2.** Factors significantly associated with an increased score of perceived value for professional development of OPAs. Mean differences with 95% CI in parenthesis



**Figure 1.** Scores of individual OPAs performed in different wards (same ward numbers as in Table 1)



**Figure 2.** Correlation between the mean score of each OPA and the percent of times it was performed during rotations

0.5). After adjusting for these factors, ward rounds, lessons, and performing procedures on patients all scored similarly, watching procedures and case discussions scored slightly better (+0.3 to +0.4 points more, on average,  $p < 0.01$ ), and procedures on peers scored the most (+0.5 points, 0.3-0.7). These differences, however significant, were small.

In activities classified as “learning to do”, in which the perception of validity of the assessment method was asked, an affirmative answer was associated to a mean increase in score of +1.9 points (1.5-2.3). Assessment by OSCE further significantly increased the mean score by +0.4 (0.1-0.7) points.

Open comments by students were not particularly frequent. Most of them showed appreciation for the organization, the respect of schedule, the attention received by the teachers or the use of smaller groups in the wards that did so more consistently. Also, peer activities were often praised. A few complained of specific teachers or experiences. Two long comments argued against the use of peer activities at this stage of training.

In teacher questionnaires, mean score for satisfaction about the clerkship rotations were on average higher than student’s (7.4, 6.8-7.9) and were not significantly correlated with student scores for the same ward rotation. Satisfaction for the quality of student participation was also high (7.9, 6.8-7.5).

The score for added burden to clinical activities was variable (4.6, 3.6-5.7). Forty per cent mentioned increased time, a few complained about the morning timetable, when ward activities are busiest. The score for perceived burden on patients was low (2.7, 1.7-3.6). Five teachers reported that patients were happy to participate. Three complained about space restriction when students were present in the ward or in outpatient clinics. One mentioned uneasiness about privacy.

All the nurses reported that nurses were not directly involved in the teaching activity. They were generally satisfied of students’ behaviour (8.9, 7.6-10). One complained about some students chowing or using mobile phones in the ward. Except in one case, perception about burden on care activities was lower than teachers (3.9, 1.8-5.9), while perception about burden on patients was slightly higher (3.5, 1.6-5.4), mostly due to the presence of too many unknown people in the patients’ rooms, and was uncorrelated with the score given by teachers (doctors) of the same ward.

## DISCUSSION

According to MacDonald, Democratic evaluation is an independent information service to the whole community about the characteristics of an educational programme [2, 3]. Its main activity is the collection of definitions of, and reactions to, the programme. The key concepts of democratic evaluation are ‘confidentiality’, ‘negotiation’, and ‘accessibility’.

The justificatory concept is 'the right to know'. We believe that we fulfilled these strict requirements: although the evaluator team was not completely independent, as one teacher and several students were actually participating in the program, none of us was part of the Education Committee or was involved in the design, in the management and control of the activities. Furthermore, the group was fully autonomous in the design and development of the inquiry. Our only aim was to contribute to the students and to the Education Committee informations on the actual performance and on the effects of the program. Accordingly, we concentrated on the perceptions of students, teachers and health workers affected by the program rather than on technical achievements, which were outside of the scope of our inquiry. We guaranteed complete confidentiality and anonymity to all the parties involved, both during data gathering and in our report. We negotiated with the students the contents, methods and timing of the investigation, ensured a wide range of openness of themes and issues through the generous use of interviews and open comments in the questionnaires. Finally, we made our report fully accessible and open to comments on our website. Although the report contains some suggestions, as they have been contributed through the questionnaires or are implicit from the data, it does not present recommendations, leaving to the Education Committee, teachers, and students full freedom and responsibility for identifying and taking actions, if they want to.

Unlike the existentialist approach endorsed by the main experts and practitioners of democratic evaluation [8], we adopted a critic rationalistic (Popperian) approach, which was more familiar to us, to the faculty and to medical students. This led us naturally to the use of mixed methods of inquiry (10), combining quantitative methods (mostly Likert scales in questionnaires) with qualitative analysis of open comments. We found these to be rather concise and repetitive, so it was not difficult to analyse them without the use of specific software and techniques. Of course, a fully grounded theory or phenomenological study would have required different methods and expertise for data gathering and analysis, but that was not the aim of our work. We

selected the students perceived educational value for personal development of professionalism as the more relevant outcome measure of our inquiry. The use of quantitative methods, allowed us to examine the relations between students perceptions and different aspects of the program activities, providing useful information to the actors involved in the development and in the participation in the program. Most of these data are only of interest within the local environment, but some could also have external validity for other settings with a similar cultural and structural conditions.

First of all, we documented a difficulty by some of the clinical wards to cope with the demanding requirements of organization, reliability and reproducibility dictated by the sustained rate of students' rotations. Providing a reliable and reproducible set of clinical experiences to different groups of students for almost 12 consecutive weeks, during hours of maximal clinical activities, is a formidable task, so the lack of performance by some wards comes without surprise. Nevertheless, having documented this fact may help the Education Committees and/or the ward heads and staff to refine their performance by increasing organization and/or human resources dedicated at the program, or reducing the number of expected activities to these that can be reliably performed. In either case, students would take advantage of a better correspondence between propositions and facts (that is, between what they could expect from what is written in the program and what is actually performed).

Not surprisingly, activities that were not performed were rated very poorly by the students. When we excluded these activities, we still found a correlation between some characteristics of the clinical activities and their impact on the perception of educational usefulness experienced by the students. Of course, in some cases this was influenced by the intrinsic interest of the discipline (say, cardiology is probably more popular among students than occupational diseases, although both are equally important in their education). Clinical activities with peers ranked particularly high among students, but they carried the advantage of novelty (this was the first time that they were inserted in the curriculum) and were only carried by two wards with high intrinsic

interest and that also ranked high for reliability of organization. However, even when the types of activities were more evenly distributed among different rotations, some differences emerged: in particular, smaller groups, perceived effectiveness of the teaching methods and the perceived validity of the evaluation all were associated to higher scores of professional value by the students. Evaluation using OSCE rather better than the other methods (usually informal questions to the group, in a typical ward round style). However, OSCE was used by only one ward, that also had an overall high score, and this effect could not be adjusted for other factors in the analysis. These informations could be useful to the organizers to split some of the activities in smaller groups, when feasible, or to develop better methods for teaching and evaluation, and possibly for the Education Committee for organizing a program for faculty development focusing on these issues.

One of the characteristics of democratic evaluation is to attempt to give voice to all the subjects affected by the program, therefore we also interviewed the teachers and at least one nurse for each ward. Unfortunately, we had not the resources needed to interview the patients, so we used these two sources also as proxies for patients involvement. Teachers' satisfaction was usually high, both about the program and about the quality students participation, but was not correlated with students scores, suggesting a lack of communication and understanding between teachers and students. Difficulties about the impact of the presence of high number of students and on the workload imposed by the many rotations were frequently reported, but overall the impact of the program on clinical activities appeared to be limited and sustainable by most of them. None of the nurses was directly involved in the teaching activity (reflecting a low level of inter-professional education in our institution). Overall, the impact of the program on patients was considered to be limited. However, the lack of correlation between the perception of the impact of the program on patients between doctors and nurses of the same ward raises doubts on the reliability of these sources as proxies for patients. Nevertheless, the combined views of teachers, nurses and students could provide useful ground to promote a dialogue between these groups and to improve the program.

In conclusion, with this case study limited to a specific course activity, we provide an example of the feasibility and possible usefulness of democratic evaluation in medical education. This observation is in contrast with the scarcity of references to democratic evaluation in medical education [4, 5], suggesting that the barriers to the diffusion of this type of evaluation in the education of health professionals should be identified and its wider use should be promoted.

## REFERENCES

- [1] Frye A.W., Hemmer P.A. Program evaluation models and related theories: AMEE guide no. 67. *Medical Teacher*, 2012, 34: 288-299.
- [2] MacDonald B. Evaluation and the Control of Education. In: Murphy R., Torrance H. (eds.). *Evaluating Education: Issues and Methods: an Open University Reader*. New York: Harper & Row, 1987.
- [3] MacDonald B.A Political classification of evaluation studies in education. In: Hammersley M. (ed.). *Social Research: Philosophy, Politics and Practice*. New York: Sage Publications, 1993: 105-108.
- [4] Guilbert J.J. Comparison of the opinion of students and teachers concerning medical education programmes in Switzerland. *Med Educ*. 1998, 32: 65-69.
- [5] Laluna M.C., Ferraz C.A. The meanings of evaluation practice in nursing education. *Rev Lat Am Enfermagem*. 2009, 17: 21-27.
- [6] Warm E.J., Mathis B.R., Held J.D., Pai S. et al. Entrustment and mapping of observable practice activities for resident assessment. *J Gen Intern Med*. 2014, 29: 1177-1182.
- [7] MacDonald B., Walker, R. Case-study an the Social Philosophy of Educational Research. *Cambridge Journal of Education*, 1975, 5: 2-11.
- [8] Kushner S. *Personalizing Evaluation*. New York: Sage Publications, 2000.
- [9] Popper K.R. *The Open Society and Its Enemies*. London:, Routlege, 1945.
- [10] Maudsley G. Mixing it but not mixed-up: mixed methods research in medical education (a critical narrative review). *Med Teach*, 2011, 33: 92-104.
- [11] Cumming G. *Understanding The New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. London: Routlege, 2012.
- [12] Wittgenstein L. *Philosophical Investigations*. Oxford: Basil & Blackwell, 1953.

## APPENDIX. OBSERVABLE PRACTICE ACTIVITIES

Perform EKG on peers  
Assist to a cardiology visit  
Perform an echocardiogram  
Collect a cardiologic anamnesis  
Collect an endocrinologic anamnesis  
Interpretation of the main tests and imaging of hormonal function  
Interpretation of laboratory data for diabetic and obese patients  
Evaluation of calciotropic hormones  
Assist to the activities of the occupational respiratory function laboratory  
Collect an occupational anamnesis  
Know the legal implications of occupational diseases  
Participate in the ward round activities  
Assist to the measurement of PFT  
Assist to allergy skin tests  
Interpretation of the PFT  
Be able to use a pulse oximeter  
Perform spirometry on peers  
Perform walking test on peers  
Perform chest visit on peers  
Give supporting indications for red blood cell and platelet transfusion  
Read a normal blood smear  
Physical examination of the spleen  
Surgical suture  
Procedure for hand washing and preparation of the outpatient surgical field  
Watch rectal examination  
Watch insertion of bladder catheter  
Watch nasogastric intubation  
Care of surgical wound  
Physical examination of the abdomen  
Diagnosis of the main diseases of the abdomen  
Write the clinical record  
Approach to patients and relatives at the hospital admission  
Diagnosis of diseases presenting at the internal medicine ward  
Physical examination of the chest and evaluation of the respiratory patient  
Measure and interpretation of blood pressure  
Investigations for the patient with respiratory problems  
Investigations for the patient with cardiovascular problems  
Physical examination of the heart in patients with cardiovascular problems  
Symptoms and signs of heart failure  
Measure and interpretation of central and peripheral cardiac frequency  
Interpretation of chest pain  
Measure and interpretation of respiratory rate