ORIGINAL REPORTS

Program Director Perceptions of the General Surgery Milestones Project

Brian C. Drolet, MD,*,† Jayson S. Marwaha, BS,‡ Abdul Wasey, BS,‡ and Adam Pallant, MD, PhD§

*Department of Plastic Surgery, Vanderbilt University Medical Center, Nashville, Tennessee; †Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee; ‡The Warren Alpert Medical School of Brown University, Rhode Island Hospital, Providence, Rhode Island; and §Department of Pediatrics, Rhode Island Hospital, Providence, Rhode Island

OBJECTIVE: As a result of the Milestones Project, all Accreditation Council for Graduate Medical Education accredited training programs now use an evaluation framework based on outcomes in 6 core competencies. Despite their widespread use, the Milestones have not been broadly evaluated. This study sought to examine program director (PD) perceptions of the Milestones Project.

DESIGN, SETTING, AND PARTICIPANTS: A national survey of general surgery PDs distributed between January and March of 2016.

RESULTS: A total of 132 surgical PDs responded to the survey (60% response rate). Positive perceptions included value for education (55%) and evaluation of resident performance (58%), as well as ability of Milestones to provide unbiased feedback (55%) and to identify areas of resident deficiency (58%). Meanwhile, time input and the ability of Milestones to discriminate underperforming programs were less likely to be rated positively (25% and 21%, respectively). Half of PDs felt that the Milestones were an improvement over their previous evaluation system (55%).

CONCLUSIONS: Using the Milestones as competency-based, developmental outcomes measures, surgical PDs reported perceived benefits for education and objectivity in the evaluation of resident performance. The overall response to the Milestones was generally favorable, and most PDs would not return to their previous evaluation systems. To improve future iterations of the Milestones, many PDs expressed a desire for customization of the Milestones' content and structure to allow for programmatic differences. (J Surg Ed ▮:▮▮▮-▮▮▮. Published by Elsevier Inc on behalf of the Association of Program Directors in Surgery)

KEY WORDS: milestones, assessment, evaluation, surgery, graduate medical education

COMPETENCIES: Practice-Based Learning and Improvement, Patient Care and Procedural Skills, Systems-Based Practice, Medical Knowledge, Interpersonal and Communication Skills, Professionalism

INTRODUCTION

Since the implementation of the Accreditation Council for Graduate Medical Education (ACGME) Outcomes Project in 1999, outcomes-based education and assessment has become a cornerstone of graduate medical education.[1-3] Nearly a decade later, the Next Accreditation System was implemented to further promote this outcomes focus.[4] A central feature of Next Accreditation System is the specialty-specific Milestones, which are competency-based developmental outcomes that form the basis for evaluative metrics within the framework of the core competencies.[5,6]

Although the Milestones are now used for resident and fellow evaluations at all ACGME-accredited training programs, their use in practice has not been broadly studied and some concerns have been raised. An earlier study of the 1999 Outcome Project demonstrated significant barriers to successful utilization, specifically including lack of time, funding, and faculty support as well as resistance to the ACGME mandate.[7] Authors of another study, which examined similar competency-based evaluations outside of medicine (K-12 education and the department of defense), found several concerning features of the Milestones that may lead to failure, including differences in learner styles as well as evaluators' assessment constructs and the time needed for direct observation to perform these evaluations.[8]

In this study, we sought to evaluate program directors' (PD) experience with and perceptions of the Milestones in general surgery.

Correspondence: Inquiries to Brian C. Drolet, MD, Department of Plastic Surgery, Vanderbilt University Medical Center, Medical Center N, D-4219, Nashville, TN 37232; fax: (615)936-0167; e-mail: brian.c.drolet@gmail.com

TABLE 1. Summary of Main Survey Results

	Survey Response		
	Negative	**Neutral**	**Positive**
	% Respondents (95% CI)		
How would you rate Milestones in terms of:			
Value for education	9.8 (6.3-13.4)	34.8 (29.2-40.5)	**55.3 (49.4-61.2)**
Frequency of evaluation	7.6 (4.4-10.7)	35.6 (29.9-41.3)	**56.8 (50.9-62.7)**
Evaluation of resident performance	12.1 (8.2-16)	29.5 (24.1-35)	**58.3 (52.4-64.2)**
Ability to provide unbiased evaluations	12.9 (8.9-16.9)	31.8 (26.3-37.4)	**55.3 (49.4-61.2)**
Time input for completion	36.4 (30.6-42.1)	38.6 (32.8-44.5)	25 (19.8-30.2)
Comparison to your previous evaluation system	15.9 (11.5-20.3)	28.8 (23.4-34.2)	**55.3 (49.4-61.2)**
	Negatively	No change	Positively
How have Milestones changed your practice as PD with regard to:			
Teaching of residents	5.3 (2.7-8)	**74 (68.8-79.3)**	20.6 (15.8-25.4)
Identifying resident deficiencies	6.1 (3.2-9)	35.9 (30.1-41.6)	**58 (52.1-63.9)**
Decreasing bias in evaluations	3.1 (1-5.1)	48.9 (42.9-54.8)	48.1 (42.1-54.1)
Moving your evaluative process toward competency	6.1 (3.2-9)	37.4 (31.6-43.2)	**56.5 (50.6-62.4)**
Your overall feelings about being a PD	13.7 (9.6-17.9)	**61.8 (56-67.6)**	24.4 (19.3-29.6)
	Disagree	Neutral	Agree
Indicate your level of agreement:			
Milestones enhances trainee competence	34.1 (28.4-39.8)	34.1 (28.4-39.8)	31.8 (26.3-37.4)
I have adequate support (e.g., protected time and trained coordinator) for Milestones requirements	28.2 (22.9-33.6)	26.7 (21.4-32)	**45 (39.1-51)**
I wish Milestones were customizable for aspects unique to my program	26.5 (21.2-31.8)	34.1 (28.4-39.8)	39.4 (33.6-45.2)
Milestones enhances my effectiveness as a PD	31.8 (26.3-37.4)	32.6 (27-38.2)	35.6 (29.9-41.3)
Milestones have positively influenced my career plans	40.2 (34.3-46)	**54.5 (48.6-60.5)**	5.3 (2.6-8)
Milestones data would be able to discriminate between successful and underperforming training programs	**50 (44-56)**	28.8 (23.4-34.2)	21.2 (16.3-26.1)
I was given enough instruction/training on how to complete the Milestones	23.5 (18.4-28.5)	37.1 (31.4-42.9)	39.4 (33.6-45.2)

Bold indicates statistically significant plurality or majority response ($P < 0.05$).

MATERIALS AND METHODS

Data Collection

We began by identifying the developmental themes and goals of outcomes-based evaluation in Graduate Medical Education through a literature review.[1-6] From this article, we developed a survey instrument to evaluate PD experience with the Milestones. We used a 5-point Likert scale for all survey responses with the exception of one question ("Overall, how do you feel about the Milestones?"), which offered 5 possible responses (Table 1). We pretested the survey with faculty and then piloted the survey with PDs on the graduate medical education committee at one institution. Throughout this process, we iteratively revised the survey construct for content and clarity. Cronbach's alpha (0.83) demonstrated excellent internal consistency of the questionnaire. The project was granted exempt status by the institutional review board.

The reference set for this national survey included all general surgery PDs ($N = 249$) listed by the ACGME in the Accreditation Data System. We obtained functional email addresses for 219 PDs (88%) from multiple sources including the ACGME Accreditation Data System, American Medical Association Residency and Fellowship Database (FREIDA) and a broad internet search. The anonymous, electronic survey was then distributed in 3 rounds to all contacts by individualized email between January and March 2016.

Statistical Analysis

For our analysis, we studied 5-point Likert scale responses as 3 groups (negative/disagree, neutral/no change, and positive/agree) by combining the proportions of similar responses. For example, we combined the proportions of "Strongly agree" and "Agree" responses into 1 group. Two-sided confidence intervals were calculated for the proportion of respondents in each group using the standard error of proportions with a fixed population correction and an $\alpha = 0.05$. We considered there to be a significant difference between groups when the confidence intervals did not overlap. All analysis was performed in SPSS, version 22.0 (IBM Corp).