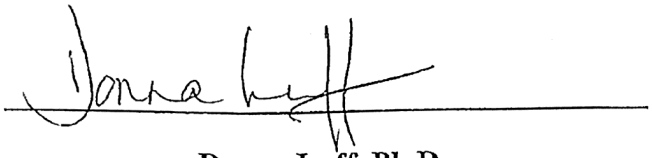
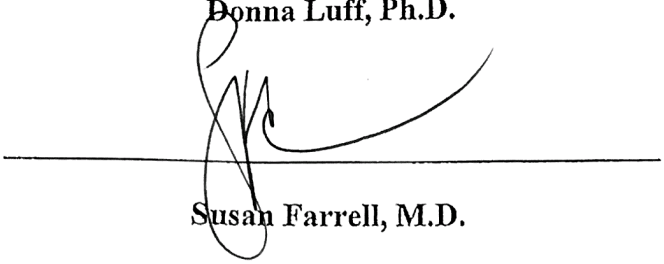
Implementing Milestones: A Qualitative Analysis of Impact

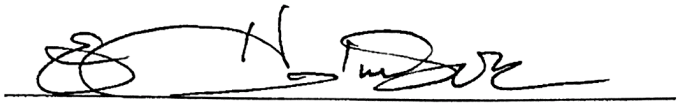
This Thesis, The Impact of the Accreditation Council for Graduate Medical Education Milestones on Programs of Assessment in Graduate Medical Education, presented by Kristina Dzara, and Submitted to the Faculty of The Harvard Medical School in Partial Fulfillment of the Requirements for the Master of Medical Sciences in Medical Education has been read and approved by:



Donna Luff, Ph.D.



Susan Farrell, M.D.



Eric Holmboe, M.D.

Date: April 12, 2017

IMPLEMENTING MILESTONES: A QUALITATIVE ANALYSIS OF IMPACT

KRISTINA DZARA, PH.D.

A Thesis Submitted to the Faculty of

The Harvard Medical School

in Partial Fulfillment of the Requirements

for the Degree of Master of Medical Sciences in Medical Education

Harvard University

Boston, Massachusetts.

May, 2017

Thesis Mentors: Dr. Daniel J. Schumacher, Dr. Jennifer Kesselheim Kristina Dzara, Ph.D.

Implementing Milestones: A Qualitative Analysis of Impact

Abstract

Background: Changes to programs of assessment following the shift to milestones-based assessment in the Next Accreditation System have not been fully characterized. Better understanding is critical to helping programs across all specialties capitalize on emerging best practices and overcome shared challenges to meaningful milestones implementation.

Methods: Semi structured interviews were conducted with 15 residency program directors or associate program directors in 6 medical specialties at 8 academic medical centers. Questions explored four areas: 1) reviews of pre-milestones assessment programs to prepare for milestones-based assessment and reporting; 2) what programs retained from previous assessment efforts; 3) what programs added or changed from previous assessment efforts; and 4) what was the impact of the shift to milestones-based assessment on the program. Thematic analysis was used to identify codes, sub-themes, and themes.

Results: Four themes were defined in the data: 1) challenges to effective implementation; 2) adaptability and making milestones work; 3) value and utility of assessment changes; and 4) accountability and transparency to society, program, and residents.

Conclusions: Three of the four themes identified elaborated the impact of the shift to milestones, such as identification of gaps in curricula or the ability to obtain previously lacking objective performance data. Program leaders desired more guidance and support in redesigning programs of assessment but are resourceful in adapting milestones to their context. This underscores the importance of supporting programs with practical strategies and resources to support success in milestones-based assessment.

Table of Contents

Tables

Acknowledgements

Kathleen Huth, M.D., Susan Farrell, M.D., Eric Holmboe, M.D., Donna Luff, Ph.D., Daniel J. Schumacher, M.D., M.Ed.

This work was conducted with support from Students in the Master of Medical Sciences in Medical Education program of Harvard Medical School. The content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard University and its affiliated academic health care centers.

Chapter 1: Background

1.1 Background

In 2009, the ACGME launched its Next Accreditation System, an effort to define milestones for the competencies they originally defined a decade earlier "to enhance the ability of the peer-review system to prepare physicians for practice in the 21st century, to accelerate the ACGME's movement toward accreditation on the basis of educational outcomes, and to reduce the burden associated with the current structure and process-based approach (p. 1051)."(1)

Since 2013, the first seven specialties began reporting milestones to the Accreditation Council for Graduate Medical Education (ACGME) and by 2015 all specialties were fully engaged.(1,2) Early guidance for how programs should transition their existing programs of assessment to encompass the milestones was ambiguous and variable for many specialties and continues to be an evolving process.(3-6) Most efforts have focused on foundational advice, such as utilizing milestones to track learner development over time to ensure that graduates have the skills and knowledge for unsupervised post-residency practice.(3, 6-7). Guidance to date also includes using clinical competency committees to assess learner progress over time by synthesizing information from a combination of assessment sources,(6-8) including direct observations, simulation, in-service training exams, multi-source feedback, audit and performance data, and global evaluations.(4, 6, 9). Finally, the shift to milestones has been advocated as an opportunity to intentionally design programs of assessment to provide learners with frequent, formative feedback to indicate areas where they are doing well and areas where improvement is needed.(3, 9)

While general guidance in the literature may help programs as they implement milestones-based assessment, it is likely of variable utility given program- and specialty-specific

differences and related practical challenges. This notion is supported by a recent description of the experiences with milestones implementation in the internal medicine community thus far.(10)

We sought to explore this area, explicitly responding to Holmboe's recent call for more research considering what works, for whom, in what circumstances, and why.(10,11) Grounded in a worldview of constructivism and relativism,(12) we explored changes to programs of assessment as a function of the shift to milestones-based assessment in the Next Accreditation System across a variety of specialties. Better understanding this area is critical to helping programs across all specialties capitalize on emerging best practices and overcome shared challenges to meaningful implementation of milestones-based assessment.

Chapter 2: Data and Methods

2.1 Setting and Participants

Between August and December 2016, we purposefully sampled residency program directors (PDs) or associate program directors (APDs) from a range of program sizes in six specialties at eight academic medical centers in a large, urban city (Boston, MA). While located in an urban, academic city, we intentionally included programs that were not located at tertiary or quaternary hospitals. In order to triangulate findings and compare potential differences between specialties, we chose medical (pediatrics and internal medicine), surgical (general surgery and orthopedics), and hospital-based (emergency medicine and anesthesiology) specialties.

2.2 Data Collection

Semi-structured interview questions were informed by realistic evaluation and developed by the research team in consultation with experts in competency-based medical education (CBME). Questions focused in four areas: 1) what, if any, reviews of their current system programs conducted to ensure they are prepared for milestones-based assessment and reporting; 2) what programs retained from previous assessment efforts; 3) what programs added or changed from previous assessment efforts; and 4) what was the impact of the shift to milestones-based assessment on the program.

2.3 Data Analysis

Interviews were audio recorded, transcribed and reviewed for quality and completeness by the first author (KD). We conducted a thematic analysis following the five key stages to

qualitative research framework developed by Ritchie and Spencer.[13] NVivo (11) was used to facilitate data management while coding. First, two authors (KD and KH) familiarized themselves with the data, each independently reading the first two interview transcripts to create a preliminary list of codes. Codes were compared and definitions refined as they were applied to the data. The iterative process allowed new questions to be asked in the interviews to explore patterns emerging in the analysis. These authors developed a codebook based on an immersive reading of the first five interviews, which represented a range of programs. After coding all subsequent interviews, they discussed recurring patterns in the data, and codes were combined into broad sub-themes. Relationships between sub-themes were explored to identify emerging themes. The coders independently reread the coded data within each theme to ensure coding consistency and to identify illustrative quotations.

As an educator (KD) and a physician (KH) with interests in postgraduate medical education, the coders practiced reflexivity by maintaining awareness of how their own views on milestones-based assessment may influence interpretation of the data. Throughout this process, they stayed close to the data and wrote interpretive memos to ensure the findings would represent the authentic views of the program leaders.

The project was deemed exempt by the institutional review board at Harvard Medical School.

Chapter 3: Results

3.1 Program and Participant Demographics

Fifteen PDs or associate PDs agreed to participate (Table 1). Interviews ranged between 20 and 58 minutes, the median interview being 37 minutes long.

Table 1: Program Description

Program Specialty	*Number of PDs Participating*
Internal Medicine	3
Pediatrics	3
Orthopedic Surgery	1
General Surgery	3
Anesthesiology	3
Emergency Medicine	2
Total	15
Interviewee Role	*Number of PDs by Role*
Program Director	13
Associate Program Director	2
Program Size	*Number of Residents*
Range	24 – 180
Mean	56
Median	67

3.2 Major Themes

Based on the data, 51 codes emerged which coalesced into 11 sub-themes and four major themes: 1) challenges to effective implementation; 2) adaptability and making milestones work; 3) value and utility of assessment changes; and 4) accountability and transparency to society, program, and residents. While we were open to the possibility of finding differences between specialties, we did not note major differences between specialties in participants' responses.

3.3 Theme #1: Challenges to Effective Implementation

Within the challenges to effective implementation theme, four sub-themes emerged: 1) desire for milestones guidance; 2) administrative burden; 3) implementation experience; and 4) faculty development opportunities and challenges.

Desire for Milestones Guidance

Nearly all PDs wanted more guidance for implementing milestones, and some felt stuck without knowing how to move forward. Indeed, early in the milestones process, participants expressed there was a general feeling of being mandated to change practice without funding or support.

> "Everyone had to do this individually, which was so incredibly frustrating about the whole process and I understand why…what works for [one specialty] is not going to work for [another] but the mandate was huge. It was huge…" [Interview #3]

This desire for implementation guidance led to considerable variability in how milestones were incorporated into programs. Those programs which were not as challenged with the shift to the milestones tended to be programs that either reported already having a strong program of assessment in place prior to the milestones or programs that didn't fully grasp the intensity they would need to review, consider, and potentially replace their program of assessment. Many participants expressed a strong desire for better tools to assess residents, especially measures with validity and reliability.

> "Can [the ACGME] just give us the form that [they] want, that [they] vetted, and we will do it, rather than us basically working slowly through multiple iterations?" [Interview #5]

Lacking engagement in the process and not knowing where the milestones fit into the ACGME's overarching plan made it difficult for program leaders to find the drive and desire to make large-scale changes to their programs of assessment.

> "I'm worried that in 5 years the next big thing will come along and as program directors, we don't have a lot of time to make innovative metrics roll out new evaluation systems every few years and our faculty would riot." [Interview #13]

Administrative Burden

Part of the emotional reaction to milestones implementation included the weight of having to not only review and potentially redesign programs of assessment but doing so while garnering the support of rotation directors and faculty. The administrative burden, including time and resource needs, helped encapsulate the overall weight participants felt as milestones were implemented in their programs.

> "We had to transition to a new online residency management plan, we had to create these forms which probably took us a year to do it, then we had to develop the faculty...we had to re-tool the CCC in that 2 out of the 12 meetings spend we spend purely doing this, and not talking about the residents per se…we've been doing it now for 2 years at least, maybe 3 years. So now it just seems like this is what we do. But it was a major

undertaking." [Interview #6]

Implementation Experience

Program directors noted a number of specific implementation challenges. Concerns about data quality and utility were clearly expressed, and linked to the desire for useful, vetted assessment tools. Additional concerns about faculty understanding the importance of assessment and desire to assess learners were expressed. These concerns were tethered to a lack of leverage to encourage faculty to complete assessments, either in total or in a timely manner.

> "I think what hasn't worked is being able to figure out how do you get these really stretched faculty to get me all the data. How do you get the data? Like, it's a lot of data that you need. And I think we haven't figured out how to get really a lot of data." [Interview #4]

However, other participants described ways that milestones pushed them to identify a mechanism to obtain more assessment information in their program:

> "We have more robust text coming in, and we hired a coordinator who has finally found the secret, which is nagging, to getting them done. So we went from about a 30% return to…nearing 90% return rate on evaluations. So we just have more volume now, and so that happens to be post milestones in part because we felt like, it was impossible to do the milestones without that." [Interview #11]

A small number of programs attempted to implement the milestones by incorporating them verbatim into direct assessment tools. Each of these programs reported having a failed first attempt at milestones implementation, which proved frustrating, as program directors then had to revert to their previous assessment tools, or identify or develop new tools.

> "When we heard the milestones were coming out, we actually changed all of our evaluations into milestone based evaluations. So we used the milestone wording, and that is what we had the evaluators scoring on…And it didn't work well. The faculty didn't understand the wording, and they didn't like using the evaluation forms used with the milestone wording. [Interview #4]

Faculty Development Opportunities and Challenges

Both barriers to and opportunities for faculty development were identified as key variables affecting milestones implementation. Here, there was a tension between having the time and resources to plan and organize training, determine faculty needs, and offering effective programming beyond a one-off, brief introduction to milestones or new assessment tools.

Some programs attempted to make assessment of residents easier by limiting the range of assessment tools. Others tried to shield or shelter faculty from the need for higher-level understanding of the purpose and function of the milestones, encouraging them to assess residents in the clinical setting but not offering a clear rationale for changes such as revised assessment tools or increased frequency of assessment.

> "We firewalled off our faculty that are division directors and are doing those evaluations from our day to day to faculty who [are not doing them]…The folks that are involved in our clinical competency committee or are doing rotational evaluations, or division directors got some additional coursework or training - both formal and more largely, informal - on how to wrap their heads around the milestones." [Interview #13]

Still, some program directors expressed hope that a combination of faculty development and overall changes to programs of assessment would gradually change faculty behavior to create a stronger culture of assessment.

> "There were some initial desires [that] the form itself could change behavior…[that] the evaluation form itself could become a form of faculty development. So we were kind of hoping that some of that would occur. I don't know whether that's happened or not. My guess is probably not…" [Interview #6]

3.4 Theme #2: Adaptability and Making Milestones Work

Three sub-themes emerged from the data for the second theme focused on adaptability and making milestones work: 1) integrating the milestones with prior program of assessment; 2) adaptability and changing approach to assessment; and 3) complete overhaul.

Integrating the Milestones with Prior Program of Assessment

As programs searched for ways of integrating the milestones into their current programs of assessment, having an existing clinical competency committee (CCC) in some form made the

transition easier.

> "We had a sort of evaluation committee that we basically renamed the CCC. So, we had it, we didn't call it the CCC, because that was new terminology." [Interview #8]

In large part, the format of assessment remained the same, such as collecting assessment data from end-of-shift or rotation evaluations, direct observation, or even informally.

> "The milestones have been additive in terms of how we assess residents. I would not use the word transformative…They have helped clarify when we're not sure about, for instance, why we might not be comfortable with a particular resident, but the process is still the same. We've incorporated the milestones into it. " [Interview #12]

A small number of programs, frustrated with or disinterested in the milestones, made no changes to their programs of assessment. Here, they assumed that their existing program of assessment would be sufficient, or they would rely on a mix of intuition and historical data.

> "I'll let you in on my secret, my dirty secret, is that I do not take the milestones seriously as a detailed exercise...I try to take everybody for whatever milestone level they were at before, and I spend almost no time at all on pondering what score they should have." [Interview #10]

Adaptability and Changing Approach to Assessment

Nearly all programs took the milestones as an opportunity to revise their assessment tools. In some cases, additional opportunities, such as having clinical faculty assess learners more than once per rotation, were identified. Programs also made changes to the clinical or didactic curriculum in an effort to ensure they were offering residents a range of educational experiences that would inform milestones reporting. Once CCCs were in place, programs had to adapt to their new function, in some cases trying out different organizational styles or processes before determining what worked best.

Complete overhaul

For many participants, milestones implementation prompted them to rethink their sources of assessment data and identify gaps in assessment within their programs. Some programs mapped sources of assessment to the milestones to ensure multiple data points were available for each milestone.

> "We took every rotation and every milestone sub-competency and mapped out basically which rotation is going to be assessing which of those sub competencies and made sure that there were multiple domains…so that was a whole overhaul actually of all of those rotations." [Interview #4]

For some, this process encouraged introspection, and the realization that additional opportunities for assessment were necessary.

"From the milestones grid [we asked] 'which of these are we currently observing, and where?' And in that process we discovered there are many things that we don't get to observe properly...So we had to think about what to do for those." [Interview #9]

3.5 Theme #3: Value and Utility of Assessment Changes

Two sub-themes emerged from the data regarding the value and utility of assessment changes: 1) increased focus on residents; and 2) gestalt to granular.

Increased Focus on Residents

Many program directors felt there was an increased focus on learners with the implementation of milestones. One way this was demonstrated was through renewed emphasis on direct observation. Some programs were able to achieve this:

"We have a lot more direct observations actually going on so what we've realized over time within the CCC is that we need more direct observations to really make sense of the milestones." [Interview #4]

A number of program directors knew that more direct observation would be good for the program, the faculty, and the residents, but lacked a clear path to achieving this given the additional resource needs, administrative burden, and lack of faculty leverage. This said, many expressed a sentiment that they were slowly moving in that direction:

"We're trying to…get more direct observation because I think that is going to be, direct

observation tools, and ways to implement that. That will probably be better for the system." [Interview #11]

An increased focus on learners was also demonstrated via the CCC through the development of a more structured and robust process for using assessment information for summative purposes. For example, participants' CCCs tended to spend more time on struggling residents or outliers, a group likely with the greatest need for receiving external feedback and guidance in their development. However, this did sometimes come at the cost of dedicating time to the review of typically performing residents.

"The good thing is we are talking about people who are lagging behind in much more depth and understanding them and discussing that…So this is what is gained. What is lost is that we don't talk about anyone who's average or above average in terms of where they can further improve." [Interview #9]

There were mixed perspectives on whether the milestones were useful for individual learner assessment. Milestones can result in an overall increase in assessment efforts, as noted previously. They can also offer faculty a framework for thinking about and offering feedback on where a resident is at in their development and where they should be going. Those who responded affirmatively noted specific domains – such as professionalism or communication – or felt that the milestones helped them move from viewing residents as globally struggling or succeeding but instead having both strengths and areas for improvement. This was reinforced by a sense that the milestones may be less effective at identifying "problem" residents and more

effective at pinpointing where, exactly, residents need additional focus or remediation.

> "I think the biggest value of the milestones for us by far has been able to describe where somebody is struggling better than we used to before." [Interview #11]

For some, the ability to characterize deficiencies was utilized as part of feedback to learners, showing their trajectory and movement towards expected performance at graduation.

> "I would say the main impact, well two main impacts, that most is that it has just helped us organize our approach to assessment in a way that's useful to learners, where they are moving along and they can see that. They are not just crossing our line and then they are done. But they can see the progress moving towards. That's helped us organize it that way." [Interview #15]

Gestalt to Granular

A number of program directors described the milestones as helping them move from a 'gestalt' system of assessment to a more 'granular' lens that they think increases the value and utility of their assessment efforts. For many program leaders, this was tempered by a desire to move to a more multidimensional system of assessment. Although some had successfully integrated assessments from sources such as peers, nurses, patients, simulations, procedure logs, examinations, OSCEs, or 360 degree evaluations, or had instituted daily or weekly faculty assessments, this was generally the exception rather than the rule. For the large part, decisions about learners were made based on faculty rotation assessments, with some program-specific

augmentation from the sources above. Still, some noted a lack of clarity as to whether the milestones offered value beyond the 'gestalt' consensus of experts.

> "It has certainly made us look at things like our quality and safety, and the way we manage that, much like the OSCE exams we are going to do for professionalism and communication. We needed something more discreet to nail down these metrics so it's kind of driving me, and our program, to come up with more discreet or more tools to drive assessments. Whether that's demonstrably better than the gestalt that we have, well had in the past, I don't know. But it will at least be somewhat more objective and that's probably a good thing." [Interview #13]

3.6 Theme #4: Accountability and Transparency to Society, Program, and Residents

One major impact of the milestones was an increased sense of accountability and transparency to society, programs, and learners. In this theme, three sub-themes emerged from the data: 1) pathway to unsupervised practice; 2) learner-centeredness and transparency; and 3) accountability and value.

Pathway to Unsupervised Practice

The milestones were useful in helping to clarify what was expected as part of the pathway to unsupervised practice. However, some programs noted it was not possible to obtain reliable data about residents for some milestones. This sometimes led to a 'benchmark' approach to assigning residents milestones levels, where a resident is assumed to be competent at a pre-established level based on postgraduate year, largely due to an absence of any – or reliable –

data.

Related to this, participants desired a more clear-cut description of where their residents should be at various time points during residency, as well as what it means to be a competent graduate.

> "What I want to know is where should someone be in terms of when should they be independent. When should they be hitting which milestone at which year across all our programs?" [Interview #4]

Learner-centeredness and Transparency

Participants clearly noted that milestones offer a higher degree of learner-centeredness and transparency for residents. Some programs required residents to read the milestones and bring questions before advising meetings; a small number required residents to use the milestones as self-assessment, filling out their own milestones so as to compare to the clinical competency committee's evaluation. Regardless of the process, residents benefitted by receiving a clearer, more defined portrait of areas where they are doing well and areas where additional developmental focus are needed. This transparency – a clearer view of the expectations for performance, the overall assessment process, and residents' progress – was a noted programmatic improvement.

> "I really like that [milestones-based assessment] allows us to externally tell residents 'If you look at your milestones, you can see that you are struggling in professionalism and you are struggling here and here.' And so, that is the other thing that is different is that

> we can now articulate someone who is globally struggling versus struggling in one area in a numerical graphical factual way for residents which I think it helpful." [Interview #2]

Accountability and Value

Despite the challenges raised, participants felt the milestones offer value to society.

"At the end of the day we need to turn to the public and say 'we think this person is ready for *independent practice in medicine* because they've achieved these behaviors and you can look at our forms, and you can see exactly what that means.' So there's an *accountability* there." [Interview #6]

Numerous program directors still struggle to determine the right processes for their programs, but felt that they were moving towards the right approach in a process akin to continuous quality improvement, balancing competing learner, faculty, program, and accreditation constraints. Although some program directors felt that the milestones impacted their programs minimally, there was still appreciation for their role in increasing the value of the residency process.

> "I think the philosophy here and the reasoning for all of this is sound. It makes sense. It's acknowledging that people progress at their own rate. It's a formative type of feedback." [Interview #6]

Chapter 4: Conclusion

4.1 Study Summary

In the shift to milestones-based assessment, this study sought to understand what programs from a representative range of specialties did to review their previous assessment efforts, what they kept from their previous efforts, what they added or changed, and what the impact of moving to milestones was for them. Participants offered the most rich and deep insights for the *impact* of the shift to milestones-based assessment.

Milestones have been described as a complex service intervention (CSI), with the underlying premise that learners' educational outcomes will improve if implemented effectively. This will, in turn, improve patient care outcomes.(11) However, implementation is a long, fragile, non-linear journey that changes based on local culture and needs and may not be implemented as originally intended.(11) This study uncovers ways that milestones have not been implemented as intended. However, it also describes implementation experiences that shed light on what works and what does not work to consider in efforts moving forward.(11)

4.2 What Was Reviewed, Kept, and Added in the Shift to Milestones-based Assessment

The second theme of adaptability and making milestones work gave rich insight into what programs reviewed, kept, and added as they shifted to milestones. Almost all programs reviewed and revised the assessment tools used in their programs. Sometimes this was a complete overhaul to develop tools more explicitly linked to the milestones. However, programs noted that this effort was not always successful for them, as would be expected of a CSI.(11) Milestones were not intended to be used verbatim for assessment, and this finding may lend support to that intention.(14, 15)

Most programs already had a committee that reviewed residents' performance, but participants in this study formalized this process with milestones implementation. Making summative decisions that are valid, reliable, and educationally meaningful is important.(16) While our study did not assess these variables, having a formal process for collecting and then making decisions on assessment data, which our study found was a positive change in the shift to milestones, is an important foundation for achieving these goals.

The shift to milestones also gave some programs insight into gaps in their residency curriculum. Failing to assess milestones was seen not just as a mismatch between milestones and curriculum, but rather a mismatch between educational experiences that residents need (as represented in the milestones) and the current curriculum. This is an encouraging finding, as the goal of CBME is to first define what patients need and then map backwards to defining the curriculum and assessment.(17) This model ensures learners achieve the educational outcomes necessary to meet the needs for patient care in each specialty – represented by the milestones. Thus, finding that programs identified gaps in their curriculum provides some evidence that, in the era of milestones-based assessment, programs are better preparing residents to achieve necessary learning outcomes to provide care for their current and future patients. Moreover, this suggests that some programs have utilized their implementation experiences formatively to continue improving curricula and assessment.(11)

4.3 Impact of the Shift to Milestones-based Assessment

Participants offered extensive insight into the positive and negative impacts that milestones have had on their programs of assessment.

Negative Impacts/Challenges

It is clear that participants desired more guidance for how to implement milestones in their programs. The ACGME has now published guidebooks for clinical competency committees as well as milestones. However, these resources only recently became available, and early efforts in the shift to milestones have been likened to 'building a plane in flight.'(14, 15) While much is learned through this approach, this study and others suggest that some additional guidance may benefit programs, especially as implementation of a CSI is a long and fragile journey.(10, 11, 18)

CBME and the milestones have not been accepted by medical educators without controversy, which may be in part attributable to a lack of a shared mental model by multiple stakeholders.(11, 18-21) While the ACGME and milestones working groups communicated the rationale and purpose of shifting to milestones as well as vision for implementation, the extent of messaging likely varied by specialty and the degree of uptake likely varied by individual.(1, 22-25) While our study did not uncover major differences between participants from different specialties, it identified differences between individual program leaders in this area.

Many of the challenges participants described are not unique to milestones, but rather to assessment in general. Our data highlight areas in which programs would benefit from additional support for their programs of assessment, especially in the era of CBME. These areas include garnering support of rotation directors and faculty to meaningfully participate in assessment efforts, ensuring opportunities for direct observation, and dedicating time and resources at the program level for assessment efforts. A number of program directors would have preferred to be given validated assessment tools. However, unless tools were accompanied by rater training, scores are likely to have high variation and thus limited validity.(26) Perhaps a better understanding of best practices for using assessment tools among program leaders, rather than

"better" tools, is most important. Indeed, faculty development in assessment – including rater training and cognition – has been described as the 'missing link' in CBME.(27, 28) Based on our findings, assessor training bundles for faculty development would likely be appreciated and utilized by many program leaders, and their development and dissemination should be considered.

Positive Impacts/Benefits

While programs did raise concerns about the negative impacts of the shift to milestones-based assessment, they also offered several potential positive impacts. Some program directors approached the milestones as an improvement opportunity, rather than a punitive mandate, and attributed successes in part to their change mindset. For example, some programs noted that the shift to milestones has improved their ability to obtain objective performance data about residents that was previously lacking. For them, milestones helped lead to an actual program of assessment, with data intentionally collected from a multitude of sources to be used later in time by the CCC.(29) This allowed them provide more guidance for residents and also provide a roadmap for where learners should be in their development, a benefit described in several specialties to date.(14, 30) Although the milestones lend themselves to a continuous quality improvement approach that was embraced by some, not all programs embraced this and tried to continually build upon successes and re-imagine areas of weakness.(11)

Participants also saw the potential that milestones offer for giving an increased sense of accountability and transparency to society, programs, and learners. They felt they had a better shared understanding and common language with learners about where they are on the path to unsupervised practice and specific areas that may require attention, further aligning with the CSI

framework. As participants described, this then provides assurances to all stakeholders, perhaps most importantly the public, that a trainee is ready for unsupervised practice. Despite the challenges described by participants, this potential offering of the milestones is reassuring. Indeed, the call for public accountability in the medical profession has increased in recent years,(1, 31) and assurances to assuage these calls will be important to cultivate in the coming years.

4.4 Limitations

While our data included a variety of specialties, we conducted our study in a single city. There may be geographic differences and differences in types of institutions (e.g., community hospital versus academic) regarding milestones implementation, and future studies should seek to further understand this. We also only conducted our study at one moment in time; reactions to milestones likely change and different themes or conclusions might be identified if data were collected longitudinally. Moreover, we only talked to program leaders – future research should also include resident's experiences and perspectives, as they are key stakeholders. Interprofessional team members can also offer important insights about whether resident assessment is on target. Patients and families should also be engaged, as they are the ultimate GME stakeholders.

4.5 Future Directions

While program leaders in this study desired more guidance with milestones implementation from the ACGME, they also noted that what works for one program or specialty may not work for another. With this in mind, national guidance from specialty communities, who

partnered with the ACGME in developing milestones for each specialty, may be more successful. As an example, the pediatrics milestone working group defined a path forward for milestones before implementation(24) and has followed this up with multiple national studies supported by the Association of Pediatric Program Directors and the American Board of Pediatrics.(32, 33) There is a need for an open and ongoing dialogue between the specialties and the ACGME as more information about what works for programs accumulates.

It is also clear that there is a need to build capacity for program leaders to implement milestones, through standard tools or training based on emerging best practices. Programs seem to be more effective with tasks such as curriculum development than with developing or refining programs of assessment. This may be because curricular development may be more straightforward and more commonly undertaken in administering a residency program.(34) In contrast, designing a robust program of assessment is more nuanced based on specialty, institution, and program. Rater training techniques may or may not work for certain groups of assessors. Different types of assessment (e.g., direct observations, simulation, multisource feedback, OSCEs, chart review) have varying levels of fit for purpose between specialties and programs. Future efforts should expand opportunities for program leaders to develop and expand skills in building successful programs of assessment. The ACGME should continue to play a meaningful role in these efforts, but specialty communities such as boards and program director organizations may be best situated to engage their communities in this way.

4.6 Conclusion

This study offers rich insights into the *impact* of the shift to milestones-based assessment. Program leaders described experiences and challenges navigating the paradigm shift in

assessment prompted by CBME, highlighting the importance of supporting programs with strategies and resources for success in milestones-based assessment.

Bibliography

1. Nasca TJ, Philibert I, Brigham T, Flynn TC. The next GME accreditation system--rationale and benefits. N Engl J Med. 2012; 366(11):1051-1056.

2. Swing SR, Beeson MS, Carraccio C, et al. Educational milestone development in the first 7 specialties to enter the next accreditation system. J Grad Med Educ. 2013; 5(1):98-106.

3. Holmboe ES. Realizing the Promise of Competency-Based Medical Education. Acad Med. 2015; 90(4):411-413.

4. Holmboe ES, Batalden P. Achieving the Desired Transformation: Thoughts on Next Steps for Outcomes-Based Medical Education. Acad Med. 2015.

5. Williams RG, Dunnington GL, Mellinger JD, Klamen DL. Placing constraints on the use of the ACGME milestones: a commentary on the limitations of global performance ratings. Acad Med. 2015; 90(4):404-407.

6. Colbert CY, Dannefer EF, French JC. Clinical Competency Committees and Assessment: Changing the Conversation in Graduate Medical Education. J Grad Med Educ. 2015; 7(2):162-165.

7. Philibert I, Brigham T, Edgar L, Swing S. Organization of the educational milestones for use in the assessment of educational outcomes. J Grad Med Educ. 2014; 6(1):177-182.

8. Accrediation Council for Graduate Medical Education. Clinical Competency Committees: A Guidebook for Programs. 2015.

9. Holmboe ES, Yamazaki K, Edgar L, et al. Reflections on the First 2 Years of Milestone Implementation. J Grad Med Educ. 2015; 7(3):506-511.

10. Holmboe ES, Call S, Ficalora RD. Milestones and competency-based medical education in internal medicine. JAMA Intern Med. 2016.

11. Holmboe ES. The Journey to Competency-based Medical Education-Implementing Milestones. Marshall Journal of Medicine. 2017; 3(1):3.

12. Bergman E, de Feijter J, Frambach J, et al. AM last page: A guide to research paradigms relevant to medical education. Acad Med. 2012; 87(4):545.

13. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. In: Bryman A, Burgess B, eds. Analyzing qualitative data: Routledge; 1994.

14. Iobst WF. Building the Plane As We Fly It. J Grad Med Educ. 2015; 7(2):259-261.

15. Carter WA, Jr. Milestone myths and misperceptions. J Grad Med Educ. 2014; 6(1):18-20.

16. Van Der Vleuten CP. The assessment of professional competence: Developments, research and practical implications. Adv Health Sci Educ Theory Pract. 1996; 1(1):41-67.

17. Frenk J, Chen L, Bhutta ZA, et al. Health Professionals for a New Century: Transforming Education to Strengthen Health Systems in an Interdependent World. Lancet. 2010; 376(9756):1923-1958.

18. Witteles RM, Verghese A. Accrediation Council for Graduate Medical Education (ACGME) Milestones-Time for a Revolt? JAMA Intern Med. 2016.

19. Norman G, Norcini J, Bordage G. Competency-based education: milestones or millstones? J Grad Med Educ. 2014; 6(1):1-6.

20. Hawkins RE, Welcher CM, Holmboe ES, et al. Implementation of competency-based medical education: are we addressing the concerns and challenges? Med Educ. 2015; 49(11):1086-1102.

21. Whitehead CR, Kuper A. Faith-based medical education. Adv Health Sci Educ Theory Pract. 2017; 22(1):1-3.

22. Nasca TJ. The next step in the outcomes-based accreditation project. ACGME Bull. May. 2008:2-4.

23. Hicks PJ, Schumacher DJ, Benson BJ, et al. The pediatrics milestones: conceptual framework, guiding principles, and approach to development. J Grad Med Educ. 2010; 2(3):410-418.

24. Hicks PJ, Englander R, Schumacher DJ, et al. Pediatrics milestone project: next steps toward meaningful outcomes assessment. J Grad Med Educ. 2010; 2(4):577-584.

25. Green ML, Aagaard EM, Caverzagie KJ, et al. Charting the road to competence: developmental milestones for internal medicine residency training. J Grad Med Educ. 2009; 1(1):5-20.

26. van der Vleuten CP, Schuwirth LW, Scheele F, Driessen EW, Hodges B. The assessment of professional competence: building blocks for theory development. Best Pract Res Clin Obstet Gynaecol. 2010; 24(6):703-719.

27. Kogan JR, Conforti LN, Iobst WF, Holmboe ES. Reconceptualizing variable rater assessments as both an educational and clinical care problem. Acad Med. 2014; 89(5):721-727.

28. Holmboe ES, Ward DS, Reznick RK, et al. Faculty development in assessment: the missing link in competency-based medical education. Acad Med. 2011; 86(4):460-467.

29. van der Vleuten CP, Schuwirth LW, Driessen EW, Govaerts MJ, Heeneman S. 12 Tips for programmatic assessment. Med Teach. 2014:1-6.

30. Schumacher DJ, Lewis KO, Burke AE, et al. The pediatrics milestones: initial evidence for their use as learning road maps for residents. Acad Pediatr. 2013; 13(1):40-47.

31. Weinstein DF. The Elusive Goal of Accountability in Graduate Medical Education. Acad Med. 2015.

32. Hicks PJ, Margolis M, Poynter SE, et al. The Pediatrics Milestones Assessment Pilot: Development of Workplace-Based Assessment Content, Instruments, and Processes. Academic Medicine. 2016; 91(5):701-709.

33. Li S-TT, Tancredi DJ, Schwartz A, et al. Competent for Unsupervised Practice: Use of Pediatric Residency Training Milestones to Assess Readiness. Academic Medicine. 2017; 92(3):385-393.

34. Kern DE. Curriculum development for medical education: a six-step approach. JHU Press; 1998.