Accreditation Council for Graduate Medical Education

A GUIDEBOOK FOR IMPLEMENTATING AND CHANGING ASSESSMENT IN THE MILESTONE ERA

Eric J. Warm, MD

Laura Edgar, EdD

Matthew Kelleher, MD, MEd

Benjamin Kinnear, MD, MEd

Dana Sall, MD, MEd

Gina Luciano, MD

Eric S. Holmboe, MD

Michael J. Rosenblum, MD, FACP

Table of Contents

Background

In 2001, the ACGME launched the Outcome Project and formally introduced the six Core Competencies now familiar to the graduate medical education (GME) community. This competency framework moved GME beyond a predominant focus on medical knowledge, patient care, and procedural skills, to incorporate additional competencies critical for safe, effective, patient-centered care (Nasca et al. 2012). While the creation of the six Core Competencies had been informed by the Dreyfus stage model of development, only the general characteristics of the educational outcome had been described and defined (Nasca et al. 2012).

The Milestones were formally introduced into the GME system in 2013. Milestones simply describe the learning trajectory within the Core Competencies. Each Competency has been further subdivided into a small sets of Subcompetencies that allow for further specification of key ability areas. A subcompetency takes the resident or fellow from a beginner in the specialty or subspecialty, to ideally a proficient practitioner at the time of graduation. Residency and fellowship education are an intensely developmental process and experience, and the assessment program should reflect this by combining the right combination of assessments to guide development. The Milestones provide an opportunity for learners to demonstrate their attainment of competence to a level beyond the stage of proficiency, and just as importantly, allow for a mutual understanding of the expectations between the learner and the members of the faculty. The Milestones can provide a framework for GME programs that facilitate the desired outcome: *all* graduating residents and fellows are truly competent for unsupervised, self-guided practice. Assessment is an essential activity to ensure these outcomes for learners.

There is now substantial experience with the initial introduction of the Milestones system. For example, a continuous quality improvement mindset for an assessment program is important as programs learn what works and what does not, for whom and in what circumstances, using this information to make changes in their assessment and educational programs. As the GME community transitions to Milestones 2.0, an opportunity exists for programs to review their current assessment programs and how they view and use the Milestones to make necessary changes. Implementing change is hard, but a growing body of research can help guide programs in making meaningful and effective change.

Milestones as a Complex Educational Intervention
Implementing and assessing the Milestones should be viewed as a complex educational intervention (Holmboe 2015; Holmboe 2017; Holmboe, Call, and Ficalora 2016). Simply defined, a complex intervention is one that consists of several interacting parts and components (Holmboe, Call, and Ficalora 2016). Key dimensions of complexity and their associated correlates in medical education are highlighted in Table 1 (Holmboe, 2017).

Table 1: Key Dimensions of Complexity: Implications for GME

Dimensions of Complexity	Examples for GME Programs
Number of and interactions between components (e.g., people, tools, technology)	People (fellows, program directors, interprofessional faculty members, coordinators, etc.); assessment tools; learning management systems; electronic medical records
Number and difficulty of behaviors required by the those delivering and/or receiving the intervention	Assessment activities and behaviors, synthesis of assessment data; Milestones review and judgments, Clinical Competency Committee (CCC) group process
Number of individuals, groups, or organizations targeted by the intervention	Multiple, from individual learners to faculty members and other health professionals (e.g., nurses completing a multi-source feedback tool to inform the Milestones); CCCs, program and departmental leadership; ACGME; certification boards; specialty societies and program director organizations
Number and variability of outcomes	Number of subcompetencies within the specialty; other important outcomes not captured by the Milestones, such as procedures not captured within a specialty's Milestones
Degree of flexibility or tailoring of the intervention permitted	Important issues of local context in implementing Milestones; choice of assessment approaches and tools used; choices in CCC membership; etc.

*Adapted from Holmboe, Call, and Ficalora 2016.

The Milestones are also embedded within complex educational systems. The hallmark of complex systems is the interdependencies and interactions between all the "parts and components," with people (i.e., health professions faculty members, program directors, program coordinators, residents, and fellows) being the most important components of the system. When implementing a change, such as a new assessment, changes to an existing assessment, and finally using the new Milestones 2.0, it will be important for the GME community to attend to key aspects of implementation. Let's now examine approaches to implementation that can help you and your program get the most impact from assessment using the Milestones.
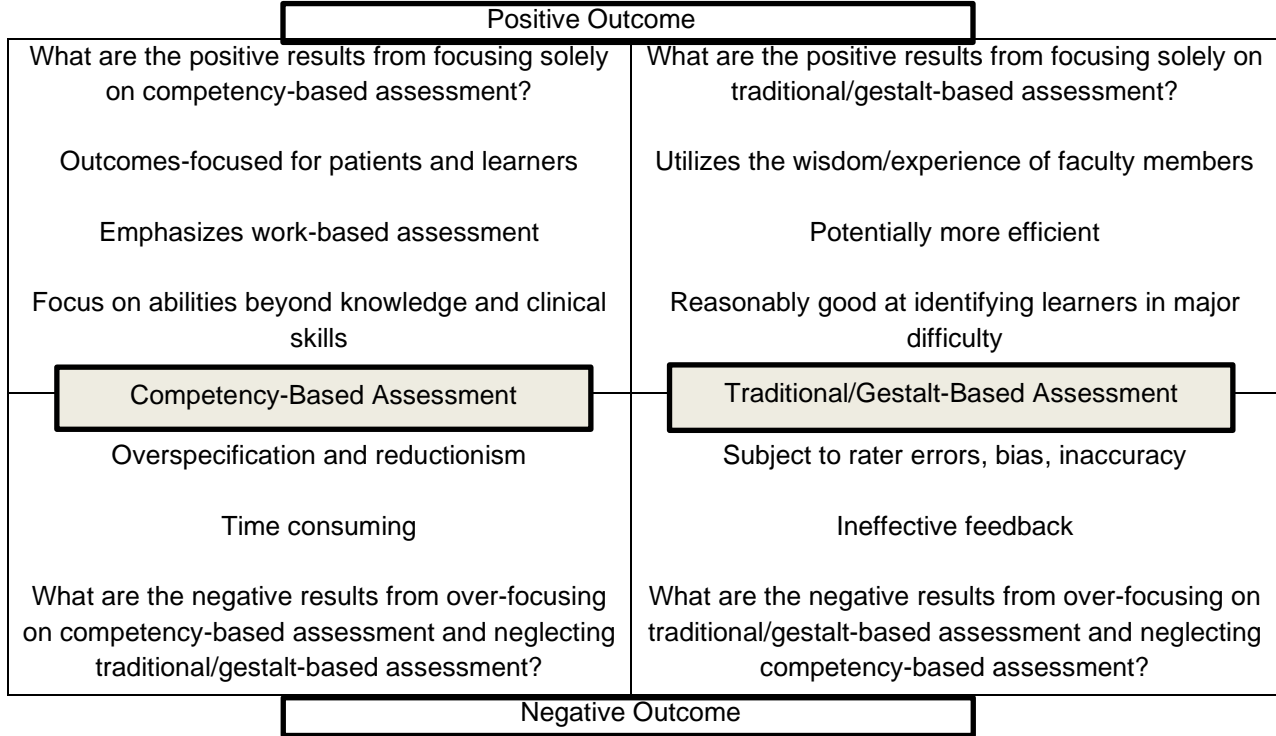
Milestones Implementation
Implementing and effectively using the Milestones within a program of assessment requires collaboration between institutions and programs and necessitates a significant transformation of educational culture (van der Vleuten et al. 2014). *John Kotter's 8 Steps for Leading Change* provide a helpful framework for planning, executing, and managing these efforts, and this Guidebook will discuss Milestones implementation through this lens,[6] but readers are encouraged to review this primary work (Kotter 1995). In addition, sections of this guidebook were adapted with permission from *The Toolkit Series: A Textbook for Internal Medicine Education Programs*, although the following principles apply to any type of ACGME-accredited program (Warm, Kinnear, and Sall 2017).

Managing Polarities in Milestones-Based Assessment
Before jumping into this work, readers should understand what kind of resistance they may face, especially if the Milestones have not yet been fully embraced at their institution. Milestones-based assessment is perceived by some as a solution to the pitfalls of time-based 'gestalt' or holistic assessment. Others see the Milestones as an onerous administrative requirement being forced upon them and prefer traditional methods of assessment. Typical criticisms of Milestones-based assessment include bureaucratic burden, lack of agreed upon definitions of competence, and lack of validity evidence for competency-based medical education (Boyd et al. 2018; Norman, Norcini, and Bordage 2014; Touchie and ten Cate 2016; Whitehead and Kuper 2017; Witteles and Verghese 2016). Rather than seeing either approach to assessment as a problem or solution, the two sides can be seen as polarities, or different values or points of views that are interdependent (Johnson 1992). In this polarity, the two poles are time-based 'gestalt' assessment versus competency- or Milestones-based assessment. (Figure 1)

'Traditionalists' argue that pre-Milestones assessment took less time and 'felt right,' and that faculty member and program director judgement was accurate and valid without having to measure numerous discrete elements. Milestones 'crusaders' on the other hand argue that traditional assessment lacks rigor, conflates formative and summative assessment, and leaves residents/fellows, programs, and patients at risk of suboptimal performance (Holmboe 2015; Holmboe 2017; Holmboe, Call, and Ficalora 2016; Nasca et al. 2012). Each pole has an upside value and a downside fear, and it is the fear of getting stuck with the downsides of one pole that keeps people from considering change. The goal of change should not be victory of one side over the other but maximizing the upsides of each pole while minimizing the downsides (Johnson 1992). Those at any stage of planning or already deeply embedded in implementation work should spend some time thinking about each quadrant in this polarity map (Figure 1). Change management requires a nuanced approach that acknowledges the downside fears held by traditionalists that are preventing them from becoming change agents. In this Guidebook, 'traditionalists' refers to people who resist adoption of Milestones-based assessment, and 'crusaders' refers to people who are change agents active in Milestones implementation. Certainly there is a spectrum, with many programs and institutions falling somewhere between these two poles.

Figure 1: Polarity Map of Competency-Based Assessment versus Traditional/Gestalt-Based Assessment (Touchie and ten Cate 2016)

Positive Outcome	
What are the positive results from focusing solely on competency-based assessment?	What are the positive results from focusing solely on traditional/gestalt-based assessment?
Outcomes-focused for patients and learners	Utilizes the wisdom/experience of faculty members
Emphasizes work-based assessment	Potentially more efficient
Focus on abilities beyond knowledge and clinical skills	Reasonably good at identifying learners in major difficulty
Competency-Based Assessment	**Traditional/Gestalt-Based Assessment**
Overspecification and reductionism	Subject to rater errors, bias, inaccuracy
Time consuming	Ineffective feedback
What are the negative results from over-focusing on competency-based assessment and neglecting traditional/gestalt-based assessment?	What are the negative results from over-focusing on traditional/gestalt-based assessment and neglecting competency-based assessment?
Negative Outcome	

Kotter's Eight Steps of Leading Change (Kotter, 1995)

Step 1: Establish a sense of urgency. Resistance to change in assessment may have multiple sources, including absence of a visible crisis, lack of external feedback on the system, low-candor/low-confrontation culture, busy or stressed-out faculty members, or lack of measurement for suboptimal processes. Developers of Milestones assessment systems should actively seek out reasons why change is unacceptable to some team members in order to deploy countermeasures for creating urgency. One strategy to engage traditionalists is to ask questions that promote cognitive dissonance. Such a narrative might sound like this: In addition to expediency, what is the measurable value in the current system? How do we currently define and measure "competence"? Is there validity evidence for what we are currently doing? Another strategy is to gather a group of faculty members and residents in a room, and ask them the following questions: How confident are you that we are accurately graduating competent physicians? What do you think patients would think of how we assess and graduate our learners? How often is the feedback you receive specific, timely, objective, and associated with a clear plan for improvement? How often do you hide your weaknesses from your supervisors for fear of being found out? How often are you hesitant to say, "I don't know"? How much of what you do is directly observed by your assessors? Ask faculty members: How often do you feel the current assessment process allows for an honest appraisal of resident performance? Are you directly observing your learners? Ask Clinical Competency Committee (CCC) members: How often is the aggregate feedback accurate, valid, and complete? How confident are you that you can use the data to find ways to help every resident improve? How often have you relied mainly on gut feeling to make remediation and promotion decisions? All of these questions explore the downside fears of the traditionalist pole and create a sense of urgency. However, also acknowledging the downside fears of the crusader pole (change itself is difficult, more faculty development needed, etc.) engages traditionalists with authenticity and avoids 'sugar-coating' the hard work of change. The change narrative should sound like this: How can we improve assessment of learners *while at the same time* minimizing bureaucratic burden and developing validity evidence for our work?

Building a sense of urgency can take many forms, such as asking the provocative questions above, sharing insightful anecdotes, or using data and evidence to support your position. For evidence, consult the Milestones bibliography that is updated approximately every six months and now contains over 350 articles, including articles on validity and effective practices for CCCs (link: https://www.acgme.org/What-We-Do/Accreditation/Milestones/Research). For example, to create cognitive dissonance in traditionalist assessors, it may be helpful to show the paucity of specific and accurate, and actionable assessment data for a resident who is known to be struggling. The disconnection between what assessors feel and what is documented may help serve as a nidus for change. Change agents must adapt arguments depending on the audience's need and core motivations. When people are inspired, they will begin to make the change argument for themselves: *maybe we should move from a summative, tangential, and incomplete assessment system to one that assures that every assessment provides meaningful feedback for learning and when aggregated can accurately track competency attainment over time.* Table 2 below lists several sources of complacency and strategies to address them.

Table 2: Sources of Complacency with regard to Milestones-Based Assessment and Potential Responses

Source of Complacency	How to Raise Urgency	Example
Absence of a visible crisis	Create a crisis	• Alert high-level stakeholders when data are lacking to support high-stakes summative decisions (e.g., remediation, termination)
Measurement of the wrong metrics	Find your weakness and measure it	Consider measuring: • Rate of on-time assessment completion • Amount of narrative feedback learners receive • Quality of assessments • Confidence CCC members have in your assessment data
Lack of feedback from external sources	Send the data to as many people as possible	Send your above measurements to: • Core education faculty • Other program directors at your institution • Your DIO • Your learners • Ask your patients what matters to them and what they would like the program to assess
Low-candor, low-confrontation culture	Bring in outsiders to start the conversation	• Host a visiting speaker who is a CBME/Milestones expert • Attend regional or national CBME conferences with key stakeholders • Engage patient ombudsman and advocacy groups for input about their expectations around assessment and feedback
Human nature – people are busy and stressed	Create opportunities for people who are involved	Faculty members (particularly junior faculty members) may be incentivized if you: • Develop a research agenda from your work that will allow for scholarly product • Present your work at regional and national conferences • Appoint faculty members to relevant committees

Step 2: Create a Guiding Coalition. Medical institutions contain silos of power and influence. A guiding coalition is a team that represent various levels and areas of the department who will help to lead the change. Building a guiding coalition under these circumstances is difficult but

critically important (Kotter 1995). It is tempting to launch into work that we are passionate about but developing any assessment program without the voices of key stakeholders will result in limited buy-in, as well as major omissions within the vision itself. Deliberate efforts should be made to include members with position power (authority and resources), expertise in Milestones assessment (or educational assessment methods), technology expertise, leadership skills, and credibility in the institution (Kotter 1995). There is also value in seeking out diversity (of culture, gender, expertise, experience) within a coalition to avoid group think (Kerr and Tindale 2004; Lorenz et al. 2011). Prepare members of the guiding coalition to represent and share the team's future work within their sphere of influence. Key stakeholders could include:

- □ Associate program directors
- □ Core faculty
- □ Key clinical faculty members (consider at least one from every division)
- □ Subspecialty education coordinators
- □ CCC members
- □ Residents/fellows
- □ Nurses
- □ Allied health professionals (social workers, pharmacists, case managers, etc.)
- □ Office staff members (program coordinators, administrative assistants, etc.)
- □ GME staff members (e.g., designated institutional official)
- □ Patients
- □ Others (who would be important on *your* team?)

While building a guiding coalition, use the methods from Step 1 to persuade those who may not be crusaders yet. One example is linking change initiatives to tangible outcomes, such as scholarship (e.g., abstracts, posters, peer-reviewed manuscripts) or credit toward academic promotion standards that may help with guiding coalition formation. When doing this, engage scholarship experts early to help plan the path forward, and avoid making promises that cannot be kept (e.g., "This will result in a published paper") while still acknowledging the effort that will be given toward such efforts (e.g., "I think it would be great to see if we could submit an abstract about our work"). Ask whether the guiding coalition only includes crusaders. Consider balancing the coalition by including other stakeholders and traditionalists who can offer their perspectives and influence. Consider engaging key traditionalists in a polarity mapping exercise. This exercise can help all members of the coalition understand each other's downside fears and create a more unified, shared mental model of the coalition's goals.

Step 3: Develop a Vision and Strategy. Once the guiding coalition is formed, create a strategic vision to develop master plans for assessment and strategies for implementation. Consider holding a retreat or series of meetings at times when everyone in the guiding coalition can be present. The group should consider creating a set of mission/vision/value statements for the work, and as noted above, consider using a polarity mapping exercise as a foundation. An example may look like the following:

- Mission (*overall purpose of the work*): Train physicians that deliver high-quality patient care using Milestones-based assessment methods.
- Vision (*description of the what the group will provide and accomplish*): Residents will receive numerous low stakes assessments of directly observed behaviors from multiple assessors and be given meaningful feedback for improvement and learning; residency

programs will use this information in education, remediation, promotion, and graduation decisions; education leaders and medical education researchers can collaborate to create validity evidence for this work.

- Value (*core priority in the group's culture*): The principle purpose of assessment is to improve learning and performance in each individual resident.

Next, the group should objectively identify how the current system is working to accomplish the vision and to meet the goals described above. It is important to identify and prioritize problems before focusing on plans for assessment and strategies for implementation. Methods for prioritizing are outside the scope of this Guidebook, but consider processes such as Q-sort, modified Delphi Technique, or Open Space Technique (Brown, 1996; Custer, Scarcella, and Stewart 1999; Stadler n.d.).

As the shared vision forms and problems are prioritized, the team can begin to craft master plans for assessment and strategies for implementation. Although the exact path will depend on local resources and challenges, questions that may help guide the process include:

- What would high quality assessment and feedback look like in our institution?
- How will diverse groups of assessors contribute to improving assessment?
- What techniques of assessment will we use and when/where will we use them?
- What tools do we have, and what tools do we need to build or acquire?
- What assessment and feedback expertise do we have or need to develop?
- Which rating scales should be used and how can we align these with clinician-assessor priorities?
- What specifically will we assess in a given environment/rotation/competency level?
- Who will complete these assessments and how often?
- How will we map front-line assessments to the subcompetencies/Milestones reporting?
- Where should the first efforts begin?
- What will be the roll-out plan after that?
- What process will be used for tracking this information?
- Will these new processes interface or replace systems already in place?
- Can other learners use these processes? (Interprofessional conversations may reveal that the issues facing residents are also faced by other learners).
- How will the process be periodically reviewed and improved?
- How will the assessments be shared with learners?
- Other (depending on your particular sets of resources and challenges)

Step 4: Communicate the Vision. Effective communication of the coalition's vision requires that it be clear, unambiguous, and understandable to any stakeholder. Efforts should be made to share the vision in various formats (meetings, one-on-one conversations, emails, signs, etc.) and on multiple occasions. Virtual town halls using various platforms, such as WebEx, Zoom, Skype, and others, can also help and maintain momentum when in-person large meetings are not possible or difficult to schedule. Virtual platforms can also be used asynchronously to provide background and updates to the institution. While it can be counterproductive to "brow-

beat" learners and faculty members with a message, communicating the vision only once is unlikely to have a significant or lasting impact. Consistently link everyday conversations with faculty members and learners back to the vision. Use specific and relatable stories to illustrate how the vision will be beneficial. For example, ask faculty members about a time when they had a struggling learner, but were unsure how to help. When discussing programmatic or curricular issues, connect how the envisioned assessment system will drive the improvements. It is critical to provide opportunities for others to give feedback on the vision and for the guiding coalition to actively listen. Addressing concerns or traditionalist views at this step can make the change process easier in later steps. Transformation requires as much support as possible. Members of the guiding coalition should act as representatives for their respective area of clinical or educational focus. Ask each person to take the questions in Step 3 back to their groups to discuss.

Step 5: Empower Others for Broad-Based Action. One way to empower faculty members and create buy-in is to invite the potential assessors to help build the system. This will create familiarity with the assessment tools as they are formed and promote intrinsic motivation as they are implemented. The Milestones were not intended to be used as direct assessment tools, but rather as a framework to track learner progression. Therefore, programs should develop observational assessment tools that align with their assessors, learners, and curriculum. These tools can later be connected to the ACGME Milestones. For example, ask a member of the Cardiology Division to make a list of abilities a learner should competently perform, dependent on the stage of training and current ability, on the rotations the Cardiology Division offers. This should include having them connect these abilities to the ultimate outcome – what a general internist must be able to do to care for patients with cardiovascular disease. Guide the faculty member to write learning objectives that require direct observation and how to think about doing this. Examples for a given cardiology rotation may include (University of Cincinnati Internal Medicine 2017):
1. Demonstrate basic EKG reading skills
2. Refer patients for appropriate cardiac imaging
3. Counsel patient on lifestyle modifications to reduce risk factors
4. Differentiate cardiac versus non-cardiac chest discomfort
5. Diagnose acute coronary syndrome (unstable angina, NSTEMI, STEMI)
6. Manage heart failure (acute, chronic, systolic, and diastolic)
7. Begin an initial management plan for common arrhythmias

The type and content of the direct observation will vary by rotation, specialty, program, and assessor. Procedure-based specialties, such as surgery or gastroenterology will have different expectations and focuses than specialties like psychiatry or general internal medicine, leading to unique types of assessments for Competencies such as medical knowledge and patient care. There may be significant inter-professional overlap with the Competencies of professionalism and interpersonal and communication skills, and where it is possible, programs within an institution may wish to share best practices. In addition, the Milestones 2.0 Work Groups are all using a "harmonized" Set of Milestones for interpersonal skills and communication, practice-based learning and improvement, professionalism, and systems-based practice. The intent is

that all specialties use a common language, altering as needed for a given context, making it easier for specialties within institutions to share assessment methods and tools. Patient care and medical knowledge milestones will remain specialty-specific. The key is to make each set of direct observations relevant for the work at hand.

Once an initial list of assessment items and abilities to be assessed have been created, present this at faculty and resident/fellow meetings for review. Use the same process with nurses, pharmacists, case managers, social workers, other allied health professionals and program staff members, and anyone else who observes resident work. Be sure to engage residents/fellows and chief residents, too, as they are often best positioned for peer assessment. Finally, consider ways to include patient-level data, such as satisfaction scores or patient outcomes. The final list of objectives will then belong to everyone, because they chose what to assess. For teams that lack the bandwidth to do this work, consider reviewing and adapting curricula already in existence (University of Cincinnati Internal Medicine 2017).

The next step is to map the final list of assessment elements, especially work-based assessments, to the Milestones and the Subcompetencies. As above, the team should engage a broad group of educators and assessors and begin with a clear vision before conducting the mapping process. An example rubric for mapping may be the following: *When assessing a resident on a given assessment element, would the assessor feel as if the mapped subcompetency were also being assessed?* (Kelleher et al. 2020; Warm et al. 2014) Teams can use techniques such as consensus building or modified Delphi technique to determine the final mapping choices (Custer, Scarcella, and Stewart 1999). Teams should be prepared to analyze the results of their mapping choices and adjust and revise over time. Over-mapping (assigning too many subcompetencies to each workplace-based assessment element) leads to lack of discrimination and increased noise, while under-mapping (assigning too few subcompetencies to each workplace-based assessment element) results in paucity of valuable information. Example mapping choices are shown below in Table 3 (harmonized milestones for all specialties indicated by an asterisk).

Table 3: Example Mapping Strategy for Workplace-Based Assessment (Internal Medicine)

Workplace-Based Assessment Element	Mapped ACGME Subcompetencies
Initiate basal bolus insulin therapy and manage blood glucose over time	PC-2 - Develops and achieves comprehensive management plan for each patient
	PC-3 - Manages patients with progressive responsibility and independence
	MK-1 - Clinical knowledge
	MK-2 - Knowledge of diagnostic testing and procedures
Demonstrate accurate medication reconciliation	PC-1 - Gathers and synthesizes essential and accurate information to define each patient's clinical problem
	MK-1 - Clinical knowledge
	SBP-2* - System Navigation for Patient-Centered Care

	ICS-3* - Communication Within Healthcare Systems
Minimize unfamiliar terms (medical jargon) during patient encounters	PROF-1* - Professional Behavior and Ethical Principles
	ICS-1* - Patient- and Family-Centered Communication
Perform central lines	PC-4 - Skill in performing procedures
Perform a PDSA (Plan-Do-Study-Act)	SBP-1* - Patient Safety and Quality Improvement
	SBP-3* - The Physician's Role in Healthcare Systems

PC = Patient Care; MK = Medical Knowledge; SBP = Systems-based Practice; ICS = Interpersonal and Communication Skills; PROF = Professionalism

To successfully implement the assessment system, faculty members and other assessors require formal education and training. Adapting or removing ineffective processes in real time can aid in avoiding creating unnecessary resistance. Pilot the system with those "champions" who first designed it. Work out the bugs using quality improvement techniques, such as Plan-Do-Study-Act cycles, Lean, or the Knowledge to Action Cycle (See Figure 2) (Courtlandt, Noonan, and Feld 2009; Randolph et al. 2009; Weinstock 2008). No matter how much testing is done, though, real-world use will uncover unanticipated issues. During the roll-out phase (and beyond), let users know that there will be issues, and they should identify and report these immediately. The educational design team must then quickly address these problems and communicate resulting solutions. Otherwise, a negative narrative about the assessment system may hinder future faculty member/assessor development opportunities.

Faculty member/assessor development should align core motivations with the aims of the system. Resorting to extrinsic motivators such as rewards, mandates, or punishments may work in the short term, but often has long-term consequences, including negative reactions to system designers. Behavior guided by intrinsic motivation is generally associated with better outcomes (Ryan and Deci 2000). Self-determination theory outlines three innate psychological needs for development of intrinsic motivation: 1) autonomy (control of one's own behavior); 2) competence (feeling of mastery for a specific action); and 3) sense of relatedness (feeling connected to others) (Ryan and Deci 2000). Fostering a sense of *autonomy* begins by engaging all assessors in development and maintenance of the system as described above. Increasing assessor *competence* can take the form of traditional didactic or online training courses, but teams may find greater success in developing systems of personalized feedback for each assessor (Warm et al. 2018). Building a sense of *relatedness* can take the form of regularly showing assessors how their assessment data fits into the framework helping to guide residents'/fellows' progress toward competence. Residents and fellows must be included in every step of this work as they too require a sense of autonomy, competence, and relatedness. Much as the members of the faculty need to learn and be assessed on use of the system, residents and fellows should be instructed and monitored on how to use it to optimize their performance.

In addition to faculty members, learners should also be included in designing, creating, and managing assessment systems. Ideal co-production of assessment for learning requires integrating the lived experiences and expertise of all users with a continuous feedback loop centered around optimizing performance (Englander et al. 2019). Learners should contribute at every stage of development, from determination of the most effective assessment items, to creation and use of assessment tools, to the use of information in visual displays, to including data in formative learning experiences.

Step 6: Generate short term wins. To sustain acceleration for successful implementation, the guiding coalition should not just hope for short-term wins, but plan for them. Identify likely improvements that will occur in the near-term and share the progress with others while linking it directly with the change efforts and your shared vision. When assessment and feedback begin to improve, share these stories broadly and frequently. If possible, consider a research agenda to study outcomes and develop validity evidence for this work (Cook et al. 2015). Revisit faculty meetings, attend resident councils, and let the medical center know of the success of the project. Lead with evidence but share anecdotes as well. Be positive about the successes and realistic about the failures, and always remind groups about the mission/vision/values of the project.

In addition to improving front-line feedback, short-term wins will occur at the CCC level. Let the guiding coalition, stakeholders, and assessors at large know how the information they created and collected is helping residents/fellows become competent physicians.

Step 7: Consolidate gains and produce more change. Work to this point may take months to years to accomplish. It is highly unlikely the perfect system will be created in the first iteration. Some traditionalists will have become crusaders, but resistance to change never fully disappears. Monitor how the assessment interface feels to the user through focus groups or unstructured conversations with diverse types of assessors. Make this part of the program evaluation or research agenda if possible. Poorly designed systems create negative energy that becomes a barrier to good assessment work. What barriers still exist? What is missing from the system? What should be augmented or reduced? What bias can be detected in the data? What actually occurs on the wards, and in operating rooms and clinics? Are the workplace-based assessments the correct objectives? Are the right assessors in a position to observe the things they are asked to assess? Are the assessment construct and thought processes of assessors the same?[29] What are the unintended consequences and balancing measures of the system? Have the mission/vision/values been achieved? Do not be surprised if the answer is not *yes* to all these questions. This will be an opportunity to reconvene the guiding coalition and perhaps the research team to further improve the work.

Step 8: Anchor new approaches in the culture. Assessment systems that maximize the upside values of all parties while minimizing the downside fears of each will be more successful than the opposite. The narrative of a successful Milestone-based assessment system implementation might look something like this: *Our assessment system collects granular* and *holistic information about trainees to track progression of competence over time, using work-*

place based assessment strategies embedded directly into our daily work, gathered by an intuitive and user-friendly assessment tool, for which we are developing or using existing validity evidence. It may take turnover of people for changes to become embedded in the culture. Consider using the natural matriculation of learners to propagate change. For new learners, changes that you have implemented over months to years simply becomes "the way we've always done it". Empower learners to spread changes you have made, bringing traditionalists into the fold along the way. Table 4 summarizes key points in this work.

Table 4: Key Points in Milestones Implementation (adapted from *The Toolkit Series: A Textbook for Internal Medicine Education Programs (Holmboe, Call, and Ficalora 2016)*

Kotter (1995) Stage of Leading Change	Strategies and Examples
Establish a Sense of Urgency	Generate a series of thoughtful questions that will inspire people to make the change argument for themselves (*see text for examples*). Use narrative storytelling to highlight the inadequacies of the current assessment system. Gather data and evidence to support why you are building a sense of urgency and share it with others. Use 'imagine if' statements: *Imagine if our milestones assessment system collected granular and global information about trainees to track progression of competence over time, used work-place based assessment strategies embedded directly into our daily work, and was gathered by an intuitive, user-friendly assessment tool, for which we have developing validity evidence.*
Create a Guiding Coalition	Assemble an interprofessional group from across the medical center. Example team members include: • Associate program directors • Key Clinical Faculty members • Clinical Competency Committee members • Residents/fellows • Nurses • Allied health professionals (social workers, pharmacists, case managers, etc.) • Office staff members (program coordinators, administrative assistants, etc.) • Graduate Medical Education Office staff members (e.g., designated institutional official) • Patients • Others (who would be important on your team?)

Develop a Vision and Strategy	Create a clear set of Mission/Vision/Value statements. For example: *Mission:* Train physicians that deliver high-quality patient care utilizing Milestones-based assessment methods. *Vision:* Residents will receive numerous low stakes assessments of directly observed behaviors from multiple assessors and be given meaningful feedback for learning; residency programs will use this information in remediation, promotion, and graduation decisions. *Value:* The principle purpose of assessment is to improve learning and performance. Generate a series of questions, given your particular set of resources and challenges, that will guide your work (*see text for examples*). Use quality improvement techniques to prioritize and organize the work.
Communicate the Vision	Share the vision with learners and faculty members using varied formats over multiple occasions. Link everyday discussions about programmatic problems or improvement efforts back to your vision. Actively seek out feedback on the vision and adapt as needed.

Empower Others for Broad-Based Action	Identify a champion for each group of assessors (e.g., physicians, nurses, social workers, office staff members, residents/fellows, etc.) to create specific workplace-based assessment items that make sense to them in their own environment (what would they expect of a learner in their unit?).
	Empower these lead creators to vet the work with their native groups; attend these meetings to clarify issues and answer questions.
	The core education team is responsible for collating these assessment elements into a cohesive whole. A subset of the team will choose a strategy to map workplace-based assessment elements to the Milestones and subcompetencies and monitor the effects of these choices.
	If possible, enlist a research team to develop validity evidence for this work.
	For teams with insufficient reserves for this work, identify resources already in existence, and use the techniques above to adapt for your local environment.
	Work with IT vendors to make the electronic user interface as simple to use as possible.
	Test out the system to 'work out the bugs' before implementation.
	After roll-out, ask users to report issues with the system in real time, and fix these immediately; set the expectation that all users can help in optimizing system performance.
	Foster internal motivation of assessors by focusing on the autonomy, competence, and relatedness of those who use the system; this may take the form of didactic presentations, one-on-one feedback on performance, or any other method that improves use and communicates the mission/vision/values of the project.
	Use external motivation (rewards, punishments, etc.) sparingly.
	Co-produce with learners and other users to determine the most effective assessment items, tools, strategies, and visual displays.

Generate Short-Term Wins	Identify short-term wins early; these may include: • improved frontline feedback from assessors to residents • superior data for the Clinical Competency Committee decision making • high quality information to develop personalized learning plans for residents/fellows Share these success stories broadly person to person, in faculty meetings, hospital staff meetings, resident councils, etc.
Consolidate Gains and Produce More Change	After the first six to 12 months, evaluate your assessment system and encourage participants to give feedback. Ask: (*additional questions in text*) • What is missing? • Are the items being assessed the correct items? • Are the right assessors in a position to observe the things they are asked to assess? • What are the unintended consequences and balancing measures of the system? • Have you achieved your mission/vision/values? Be honest about the successes and failures; fix processes and items that do not work, and share these fixes as another example of success. Be prepared that dissenters will never fully go away and continue to engage them for feedback.
Institute Change	Work towards culture change. Use the natural turnover of learners to embed changes at the time of matriculation so changes become "the way we do things here." Regularly meet with your guiding coalition and key stakeholders. Maximize the upside values of all parties while acknowledging and minimizing the downside fears.

Post-Implementation of Milestones 2.0

Yes, the implementation is the hard part of changing to Milestones 2.0. However, it is important to complete an evaluation of the assessment program after implementation, what might be called a formal post-implementation review (PIR). A PIR is an opportunity to compare how the newly implemented Milestones are being used and evaluated against conformance with the original plan. Think of it as the "S" and "A" in the Plan-Do-Study-Act (PDSA) cycle. The PIR has several steps:
1. Start with a gap analysis
2. Measure the satisfaction of stakeholders
3. Identify the benefits and costs of the changes
4. Identify areas that need further improvement
5. Identify lessons learned
6. Report the findings to your stakeholders

This review will help to enhance the work that has been done, improve areas that need it, and get a better understanding of how to make other changes in the program. [30-34] The best time to complete this review is after one or two cycles of Milestones evaluations. If possible, use reviewers who were not involved in the implementation process. To get the best results: be open to all comments and critiques; be objective when performing the various steps; document successes and failures; note those things that were not expected but affected the work; and finally, consider how to use this knowledge to better prepare for future projects.

Step 1: Gap Analysis
To start the gap analysis, be sure to review all the documents used in the implementation process. The shared vision forms that were created to guide the changes (see page 7) will be a great resource to start with. Going through each question on the form, consider whether the goals were met with the level of quality intended.

If there are gaps, determine whether they are intended (something was purposefully done differently) and if it is still fit for that purpose. In other words, examine whether the implementation met the original plan in terms of quality and usability despite these differences. For the goals that were not fully met, use the rest of the review process to determine which goals should be worked on first.

Step 2: Measure Satisfaction with Stakeholders
Determining the usability and value of the changes to all stakeholders is essential. The system may be the best ever designed, but if the CCC, faculty members, and residents/fellows will not use it, the project is **not** a success. Satisfaction can be determined through surveys, small focus groups, or interviews with various stakeholders. If there is dissatisfaction, small focus groups and interviews are the best way to get more specific information about the areas that might need improvement.

Step 3: Identify Benefits and Costs of the Changes

Identifying the costs and benefits can be difficult on projects for which the outcomes are not readily and regularly measured. Beyond monetary costs, be sure to include the amount of time required to develop and implement the project and to train the assessors. The benefits can be measured in various ways, including through surveys before and after the changes, time required to complete the original and new assessment, or even the usability of the results.

Step 4: Identify Areas that Need Further Improvement

When considering the results of the first three steps, first list the areas that need improvement and determine which can be improved. Some needed improvements may be too costly to implement, so take time to prioritize the improvements to the assessment system. Once the improvements are identified, begin to plan the methods necessary to make these changes (e.g., if more training is needed begin to schedule those events). If it is a systems issue, work within the system to identify areas where other departments/experts may be able to help. If all improvement cannot be made at once, develop a long-term plan with identified outcomes.

Step 5: Identify Lessons Learned

Be sure to document each of these steps so that what worked and what did not work can be easily identified. For those changes that did not work, consider the steps taken, and determine what went wrong and how the same problem can be avoided in your next steps. For the changes that were successful, think about what can be learned from the process.

Step 6: Report the Findings to the Stakeholders

This last step is critical – stakeholders deserve a final report. They have worked on the effort in various ways, completed training, and are using the product put in place. Share how the level of satisfaction has changed, benefits and costs, areas for improvement, and lessons learned. If additional work is being done to make further improvements, share the plan, and how whether or not it has met their needs will be measured.

Rarely does a project end with implementation; the same is true when implementing changes to the many facets of assessment and Milestones evaluation. Be sure to offer training annually for new faculty members and residents/fellows. Monitor the level of satisfaction and usability of the new system regularly. Through regular engagement with the CCC, faculty members, and residents/fellows, it will be possible to continue making small improvements to fit their ongoing needs and ensure a successful system.

References

Boyd, Victoria A., Cynthia R. Whitehead, Patricia Thille, Shiphra Ginsburg, Ryan Brydges, and Ayelet Kuper. 2018. "Competency-Based Medical Education: The Discourse of Infallibility." *Medical Education* 52, no. 1: 45-57.

Brown, Steven R. 1996. "Q Methodology and Qualitative Research." *Qualitative Health Research* 6, no. 4: 561-567.

Centers for Medicare and Medicaid Services. n.d. "PSDA Cycle Template." Accessed August 13, 2018. https://www.cms.gov/Medicare/Provider-Enrollment-and-Certification/QAPI/downloads/PDSACycledebedits.pdf.

Cook, David A., and Thomas J. Beckman. 2006. "Current Concepts in Validity and Reliability for Psychometric Instruments: Theory and Application." *Am J Med* 119, no. 2: 166.e7-16.

Cook, David A., Ryan Brydges, Shiphra Ginsburg, and Rose Hatala. 2015. "A Contemporary Approach to Validity Arguments: A Practical Guide to Kane's Framework." *Medical Education* 49, no. 6: 560-575.

Courtlandt, Cheryl D., Laura Noonan, and Leonard G. Feld. 2009. "Model for Improvement – Part 1: A Framework for Health Care Quality." *Pediatr Clin North Am* 56, no. 4: 757-778.

Custer, Rodney L., Joseph A. Scarcella, and Bob R. Stewart. 1999. "The Modified Delphi Technique – A Rotational Modification." *Journal of Career and Technical Education* 15, no. 2.

Englander, Robert, Eric Holmboe, Paul Batalden, Rosemary Caron, Carol Durham, Tina Foster, Greg Ogrinc, Nacide Ercan-Fang, and Maren Batalden. 2019. "Coproducing Health Professions Education: A Prerequisite to Coproducing Health Care Services?" *Academic Medicine.* Preprint, submitted December 24, 2019. https://doi.org/10.1097/ACM.0000000000003137

Google. n.d. "Post Implementation Review." Accessed August 13, 2018. https://docs.google.com/document/d/1FWf_qb1wrfmgrFa7nUqA5-Ad48sSoEA5baYQDZjqOqE/preview.

Holmboe, Eric S. 2015. "Realizing the Promise of Competency-Based Medical Education." *Academic Medicine* 90, no. 4: 411-413.

Holmboe, Eric S. 2017. "The Journey to Competency-based Medical Education - Implementing Milestones." *Marshall Journal of Medicine* 3, no. 1: 3-9.

Holmboe, Eric S., Stephanie Call, and Robert D. Ficalora. 2016. "Milestones and Competency-Based Medical Education." *JAMA Intern Med* 176, no. 11: 1601-1602.

Institute for Healthcare Improvement. n.d. "Science of Improvement: Testing Changes." Accessed August 13, 2018. http://www.ihi.org/resources/Pages/HowtoImprove/ScienceofImprovementTestingChanges.aspx.

Johnson, Barry. 1992. *Polarity Management: Identifying and Managing Unsolvable Problems.* Amherst MA: HRD Press Inc.

Kelleher, Matthew, Benjamin Kinnear, Sue E. Poynter Wong, Jennifer O'Toole, and Eric Warm. 2020. "Linking Workplace-Based Assessment to ACGME Milestones: A Comparison of Mapping Strategies in Two Specialties." *Teaching and Learning in Medicine* 32, no. 2: 194-203.

Kerr, Norbert L., and Scott Tindale. 2004. "Group Performance and Decision Making." *Annu Rev Psychol* 55: 623-655.

Kotter, John P. 1995. "Leading Change: Why Transformation Efforts Fail." *Harvard Business Review* 73, no. 2: 59-67.

Lorenz, Jan, Heiko Rauhut, Frank Schweitzer, and Dirk Helbing. 2011. "How Social Influence Can Undermine the Wisdom of Crowd Effect." *Proc Natl Acad Sci U S A* 108, no. 22: 9020-9025.

Mind Tools. n.d. "Post-Implementation Reviews." Accessed August 13, 2018. https://www.mindtools.com/pages/article/newPPM_74.htm.

Nasca, Thomas J., Ingrid Philibert, Timothy Brigham, and Timothy C. Flynn. 2012. "The Next GME Accreditation System – Rationale and Benefits." *New England Journal of Medicine* 366, no. 11: 1051-1056.

Norman, Geoff, John Norcini, and Georges Bordage. 2014. "Competency-Based Education: Milestones or Milestones?" *Journal of Graduate Medical Education* 6, no. 1: 1-6.

Randolph, Greg, Megan Esporas, Lloyd Provost, Sara Massie, and David G. Bundy. 2009. "Model for Improvement – Part Two: Measurement and Feedback for Quality Improvement Efforts." *Pediatr Clin North Am* 56, no. 4: 779-798.

Ryan, Richard M., and Edward L. Deci. 2000. "Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being." *Am Psychol* 55, no. 1: 68-78.

Stadler, Anne. n.d. "Doing an Open Space: A Two Page Primer." Accessed May 31, 2018. https://www.openspaceworld.org/files/tmnfiles/2pageos.htm.

Touchie, Claire, and Olle ten Cate. 2016. "The Promise, Perils, Problems and Progress of Competency-Based Medical Education." *Medical Education* 50, no. 1: 93-100.

University of Cincinnati Internal Medicine. 2017. "Resident Education Curriculum." Accessed May 14, 2018. http://med.uc.edu/docs/default-source/default-document-library/curriculum-2017-18-final.pdf?sfvrsn=0.

van der Vleuten, Cees P. M., Erik W. Driessen, Marjan J. B. Govaerts, and Sylvia Heeneman. 2014. "12 Tips for Programmatic Assessment." *Medical Teacher*: 1-6.

Warm, Eric J., Matthew Kelleher, Benjamin Kinnear, and Dana Sall. 2018. "Feedback on Feedback as a Faculty Development Tool." *Journal of Graduate Medical Education* 10, no. 3: 354-355.

Warm, Eric J., Benjamin Kinnear, and Dana Sall. 2017. "Section 5.3: Developing Evaluation Systems." In *A Textbook for Internal Medicine Education Programs*, edited by David A. Wininger and F. K. Williams. Alexandria VA: Alliance for Academic Internal Medicine.

Warm, Eric J., Bradley R. Mathis, Justin D. Held, Savita Pai, Jonathan Tolentino, Lauren Ashbrook, Cheryl K. Lee, David Lee, Sharice Wood, Carl J. Fichtenbaum, et al. 2014. "Entrustment and Mapping of Observable Practice Activities for Resident Assessment." *Journal of General Internal Medicine* 29, no. 8: 1177-1182.

Weinstock, Donna. 2008. "Lean Healthcare." *J Med Pract Manage* 23, no. 6: 339-341.

Westland, Jason. 2018. "What is Post-Implementation Review in Project Management?" Accessed August 13, 2018. https://www.projectmanager.com/blog/post-implementation-review.

Whitehead, Cynthia R., and Ayelet Kuper. 2017. "Faith-Based Medical Education." *Adv Health Sci Educ Theory* 22, no. 1: 1-3.

Witteles, Ronald M., and Abraham Verghese. 2016. "Accreditation Council for Graduate Medical Education (ACGME) Milestones – Time for a Revolt?" *JAMA Intern Med* 176, no. 11: 1599-1600.

Figure 2: Adapted Knowledge to Action Process Model for Milestones 2.0 Implementation
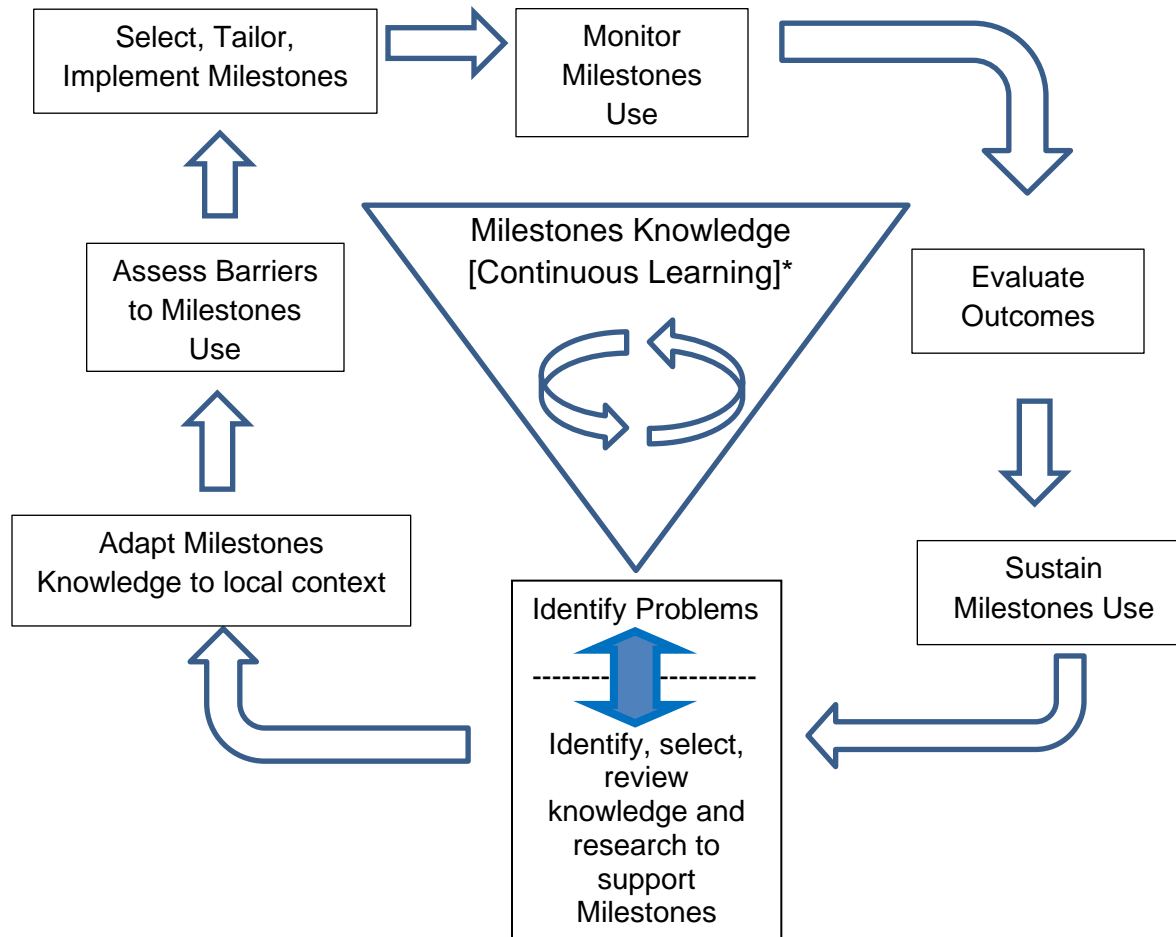
*Adapted from Stadler n.d. and University of Cincinnati Internal Medicine 2017. Implementing Milestones 2.0 requires using research and lessons learned in the first seven years of the Milestones system, while continuously generating new knowledge from the implementation of Milestones 2.0 and feeding that new knowledge back into the GME system.