EXECUTIVE SUMMARY

American Medical Association policy (H-295.862) from the 2014 Annual Meeting of the AMA House of Delegates supports the concept that assessment of physicians across the continuum should be based in the six competency domains of the Accreditation Council for Graduate Medical Education (ACGME): patient care, medical knowledge, interpersonal and communication skills, professionalism, practice-based learning and improvement, and systems-based practice. The current report uses this competency framework in the discussion of assessment methods.

The report describes the research on and methods used to assess knowledge and cognitive skills, clinical and communication skills, professionalism, and other competencies. While there has been less research on the reliability and validity of assessment methods for systems-based practice and practice-based learning and improvement when used for summative purposes, current methods appear useful for formative (educational) purposes.

A physician needs to understand his or her knowledge and skill gaps, so that they can be remedied through targeted education and practice. Self-assessment allows a physician to take responsibility for his or her learning and to build an ongoing educational program based on perceived needs. However, there is evidence that there are gaps in some physicians’ ability to independently assess their own knowledge, skills, or performance in a global content domain.

There have been attempts to utilize assessment methods to predict the performance of physicians at later stages of the continuum, as an aid in selection. For example, the results of the Medical College Admissions Test and the United States Medical Licensing Examination are used in selection for medical school and residency training, respectively. In summary, while performance on tests of knowledge tends to predict performance on later tests of knowledge, there is far less evidence for valid measures to predict performance at later stages of the continuum in other competency domains.

There is a need to create an organizing framework that would allow assessment along the medical education continuum related to the six competency domains. Workplace based assessment allows the results of various assessment methods to be aggregated so that a picture of composite performance can be developed. The results of workplace assessment would allow a cumulative judgment about the performance of an individual at a given stage of the medical education continuum and allow a determination if he or she is ready to progress to the next year of the program or phase of the continuum.

This report recommends that evaluation of physicians as they progress along the medical education continuum should include assessments of each of the six competency domains. Additional research is needed on competency-based progression within and across phases of the medical education continuum, on innovative methods of assessment related to the six competency domains of the ACGME/American Board of Medical Specialties, and on best practices for workplace-based assessment that allow performance data related to each of the six competency domains to be aggregated and to serve as feedback to physicians-in-training and in practice.
HOD ACTION: Council on Medical Education Report 10 adopted and the remainder of the report filed.

REPORT OF THE COUNCIL ON MEDICAL EDUCATION

CME Report 10-A-15

Subject: Aligning the Evaluation of Physicians Across the Medical Education Continuum

Presented by: William A. McDade, MD, Chair

Referred to: Reference Committee C
(Daniel B. Kimball, Jr., MD, Chair)

THE IDEAL CONTINUUM

In order to provide a framework, this report begins with a description of an ideal continuum that would allow determination of whether a medical student, resident or practicing physician has acquired and can demonstrate the competencies that characterize a physician. As the individual moves through medical school, residency training, and into practice, he or she should be able to demonstrate the knowledge, skills, attitudes and behaviors related to these competencies at levels of accomplishment that are appropriate to his or her stage of the medical education continuum. This requires that, for each of the competencies, there will be assessment methods, tools and metrics to test an individual’s achievement of expected outcomes. In the ideal continuum, the methods and tools used for assessment are able to determine, and in some cases predict, the individual’s level of accomplishment.

The ideal continuum for evaluation depends on the availability of the following:

- Agreed-upon outcome-based competencies;
- Performance benchmarks for each level of the continuum and for entry into and maintenance of practice; and
- A process and measurement tools to assess whether the learner can demonstrate achievement of the relevant competencies at an appropriate level.

PURPOSE AND SCOPE OF THIS REPORT

American Medical Association (AMA) Policy H-295.862, Alignment of Accreditation Across the Medical Education Continuum, adopted at the 2014 Annual Meeting of the House of Delegates (HOD), supports the concept that assessment of physicians across the continuum should be based on the six competency domains of the Accreditation Council for Graduate Medical Education (ACGME): patient care, medical knowledge, interpersonal and communication skills, professionalism, practice-based learning and improvement, and systems-based practice. The current report uses this competency framework in the discussion of assessment methods.

This report is the second in a series. Council on Medical Education Report 4-A-14, Alignment of Accreditation Across the Medical Education Continuum, discussed how accreditation could be aligned from medical school through residency. It concluded that there should be collaboration among interested stakeholder groups to identify guidelines for the general level of learners’ competencies as they move from one stage of the continuum to the next. The current report summarizes approaches to evaluating physicians across the continuum from entry to medical school into practice and describes the following:

© 2015 American Medical Association. All Rights Reserved.
• The methods and tools currently used to assess knowledge, skills, attitudes and behaviors relevant to competency domains during the stages of the medical education continuum.
• The status of efforts to use assessment to predict individuals’ success during training and in practice.
• The status of and potential approaches to using a competency framework for aligning assessment across the continuum.

THE DEFINITION OF ASSESSMENT AND RELATED CONCEPTS

The evaluation of physician learners across the continuum should include a variety of assessment methods to allow a judgment about an individual’s attainment of specific knowledge, skills, and behavioral and attitudinal outcomes. As described by Boulet and McKinley, assessments:

must be practical, yield sufficiently precise measures of ability, and allow one to make justifiable inferences concerning the qualities or abilities of those being evaluated.²

Through the use of relevant assessment methods, it can be accurately and consistently determined whether and at what level the expected competencies have been acquired by a given learner.

Validity and Reliability

The validity of an assessment tool (e.g., a multiple-choice test, a clinical skills examination) relates to whether it actually measures the “construct” (the characteristic) that it intends to measure (e.g., professionalism).³ The assessment tool needs to be tested (validated) to ensure that it is accurately measuring all the relevant components of the construct (e.g., medical knowledge). The reliability of an assessment tool relates to the consistency of scores when the tool is administered repeatedly within a short timeframe to the same learner.³ Both reliability and validity must be considered in determining if a given assessment tool is appropriate for the desired purpose.

Formative and Summative Assessment

Formative assessment is designed to provide feedback to individuals for purposes of their learning and improvement. In formative assessment, the results are used by the learner for his or her own purposes and the results do not contribute to an external decision about the learner, such as progress to the next level of training or grading. Summative assessment, in contrast, contributes to final judgments, such as pass-fail decisions. The same types of assessment methods or tools may be used for both formative and summative assessment, though the level of feedback to the learner will differ. In formative assessment, the learner receives detailed information about his or her performance, such as content areas where performance was strong or weak. In summative assessment, the learner likely will receive only a score/set of subscores or a decision, such as pass or fail.

OVERVIEW OF THE ASSESSMENT METHODS USED IN VARIOUS PHASES OF THE EDUCATIONAL CONTINUUM

Assessment of Knowledge and Cognitive Skills

Assessment of knowledge is often done through tests using multiple-choice questions (MCQs).⁴ The MCQ format came into prominence in the 1950s. For example, the National Board of Medical Examiners engaged in statistical studies related to the validity and reliability of tests using MCQs and found reliability and validity of the format appropriate for licensure examinations and superior
methods that had been used, such as essay questions.5 Today, tests using MCQs are used across the continuum from preadmission testing (the Medical College Admission Test) through the medical specialty board certification and re-certification examinations in the various specialties. The MCQ format allows a wide variety of objectives to be tested and the test to be easily scored.4 However, care must be taken that questions match the expected competency that the learner should demonstrate. For example, questions that simply expect the recall of previously learned information are not appropriate when the goal is to assess higher-level skills, such as medical reasoning or problem-solving.

Cognitive skills, such as problem-solving, also can be assessed through observation of the learner in a classroom or workplace setting. During medical school, formats such as problem-based learning allow an assessment of how well learners identify and utilize information related to clinical problems. As the medical student gains more experience, he/she applies this skill in the context of real patients during clerkships, where the skill is assessed through supervisor observation. In medical school and residency training, cognitive skills such as clinical judgment also can be assessed through more structured observational techniques, such as case-based discussion/chart stimulated recall.6 In these situations, the individual is observed demonstrating his/her thought processes related to the care of real patients. Observational assessment methods, even under controlled conditions, require appropriate training of evaluators.6

Assessment of Clinical and Communication Skills

Assessment of procedural skills may occur in isolation (i.e., the performance of a specific task, such as examining the abdomen or suturing) or along with assessment of cognitive skills (i.e., the performance of a physical examination with the results used by the examinee to develop a problem list or management plan). Similarly, communication skills may be assessed alone (e.g., the ability to ask open-ended questions or to put the patient at ease) or in the context of eliciting information that allows a specific diagnosis to be made.

There are a variety of assessment methods and tools that are used for the evaluation of clinical and communication skills within the clinical setting. All are based on observation of performance with real patients during a single clinical encounter or cumulative over time. For example, the mini-clinical evaluation exercise (mini-CEX), developed in the 1990s, is a workplace-based single encounter assessment that evaluates patient encounters in the clinical setting. It is useful for the assessment of a variety of competencies, including professionalism, interviewing/communication and physical examination and allows for immediate post-encounter feedback.8,9 The mini-CEX is used during medical school and residency training.10 In summary, many tools, such as checklists and rating scales, are used during medical school and residency training to assess students, residents and fellows in the clinical setting.10 Often the tools are developed and used within one medical school or residency program. According to a systematic review of the literature, few tools have been “thoroughly evaluated and tested” for their reliability and validity, the mini-CEX being one exception.10

Assessment of clinical skills also can occur in a simulated setting. The Objective Structured Clinical Examination (OSCE) first was described in 1975 as a way to enhance the reliability and validity of clinical skills assessment and to ensure that learners (medical students and residents) are systematically observed performing core clinical skills.11,12 OSCEs consist of a series of cases that require the individual to elicit information through history and/or physical examination and/or to use clinical information in follow-up, such as creating a differential diagnosis or management plan. OSCEs are widely used within individual medical schools for formative or summative purposes.13 The United States Medical Licensing Examination Step 2-Clinical Skills is an OSCE-based
examination. In the 2013-2014 academic year, 96 percent of MD-granting medical schools required students to take the examination and 67 percent required a passing score for advancement or graduation.\textsuperscript{13}

High-fidelity simulation has been noted to be useful in assessing both technical and non-technical skills. While there is evidence for the face validity of these measures, the evidence for their reliability and predictive validity is not as clear.\textsuperscript{6}

Methods that allow assessment of written communication skills include review of clinical documentation (e.g., chart review, patient write-ups). For example, there is widespread use of clinical documentation review during required clinical clerkships.\textsuperscript{7} Review of clinical records as an assessment methodology extends into residency training and, in some cases, into clinical practice. For example, in practice there could be assessment of the accuracy and adequacy of the clinical record and whether information has been shared with appropriate parties, such as patients and referring physicians. There is little information in the literature about the extent to which the review of physician records occurs in a systematic manner.

\textbf{Assessment of Professionalism}

Professionalism may be characterized in a variety of ways and each has implications for assessment. For purposes of this report, professionalism is considered to be a “characteristic or attribute that is identifiable within individuals”\textsuperscript{14} and is assessed through the observation of behavior in actual or simulated settings. The complexity of assessing professionalism arises from the different characteristics included in the definition (e.g., altruism, integrity) by different groups and the need to operationalize these characteristics into observable behaviors.\textsuperscript{14} For purposes of assessment, professionalism has been considered as a “global construct” (that is, a composite characteristic) or as a set of individual, though perhaps related, characteristics.\textsuperscript{15}

Regardless of the complexities, professionalism is widely evaluated during medical school and residency training and also is considered during the admission process.\textsuperscript{14} For example, medical schools use a variety of methods to assess professionalism (Table 1).

\begin{table}[h!]
\centering
\begin{tabular}{ll}
\hline
\textbf{METHOD} & \textbf{NUMBER AND % OF SCHOOLS} \\
\hline
Observation by clinical faculty during clerkships & 134 100\% \\
Observation during small group sessions in the preclinical years & 129 96\% \\
Observation by residents & 126 94\% \\
Observation during laboratory sessions & 118 88\% \\
OSCE with one or more professionalism stations & 99 74\% \\
Comments from other health professionals & 79 59\% \\
Comments from patients & 55 41\% \\
\hline
\end{tabular}
\caption{METHODS USED BY MEDICAL SCHOOLS IN THE ASSESSMENT OF PROFESSIONALISM (2012)}\label{tab:methods}
\end{table}

Assessment of professionalism can occur as a single point-in-time evaluation, such as the mini-CEX; composite performance over time, such as in an end-of-clerkship evaluation; or a critical incident, such as the reporting of an incident of unprofessional behavior.\textsuperscript{15} In residency training, the ACGME milestones for all specialties include an assessment of various aspects of professionalism.
over time. While the milestones for each specialty include an evaluation of professionalism, each
organizes the components of professionalism (and consequently the specific behaviors evaluated)
differently.

Other tools are being used to support formative and summative evaluation of professionalism.
Portfolios are being used in medical school and residency as a means to store information from a
variety of assessment methods, to allow the creation of a comprehensive view of the individual
over time. The move to electronic portfolios has increased their flexibility and utility, though
security of information remains an issue.

**Systems to Assess Multiple Competencies**

While the previous discussion focused on the tools and methods typically used to assess single
competency domains, there are systems and processes in place to address the physician’s
accomplishments across the six competency domains, including systems-based practice and
practice-based learning and improvement. The ACGME milestones project includes the ongoing
assessment of each of the six competency domains in each specialty. This system is designed to
monitor a resident’s ongoing progress in more than 30 areas per specialty so that the graduate’s
readiness for unsupervised practice can be documented. Similarly, the American Board of Medical
Specialties Maintenance of Certification (MOC) program addresses the six competency domains
using multiple methods for learning and assessment. Such comprehensive assessment systems can
provide information for physicians and others to use for tracking progress along the continuum.

This report has described many tools and methods that are available to assess medical knowledge,
patient care, interpersonal and communication skills, and professionalism. A variety of processes
exist, such as those used by the individual medical specialty boards for Part IV of the MOC
program, to assess systems-based practice and practice-based learning and improvement. In
general, however, less research has been conducted to determine their reliability and validity for
summative purposes. They are able to provide useful formative feedback to individual physicians
and their practices.

**THE BENEFITS AND LIMITATIONS OF SELF-ASSESSMENT**

Physicians need to understand their knowledge and skill gaps, so that they can be remedied through
targeted education and practice. Self-assessment allows physicians to take responsibility for their
learning and to build an ongoing educational program based on perceived needs. However, reviews
of the literature have cast doubts on physicians’ ability to independently assess their own
knowledge, skills, or performance in a global content domain as compared with an appropriate
external assessment measure. In summary, self-assessment is important but insufficient in itself
to allow physicians to identify areas in which they need to improve. To address this, researchers
have pointed to the importance of external assessments. In addition, the creation of objective
measurements or benchmarks of performance and the use of an external appraiser to facilitate self-
assessment could be useful.

**METHODS AND TOOLS TO PREDICT PERFORMANCE**

There has been a great deal of research on what tools/measures are useful to predict the
performance of a physician-in-training or a physician in a future phase of the continuum, including
in practice. Some of these measures, for example, the Medical College Admission Test (MCAT)
and the United States Medical Licensing Examination (USMLE), are widely used in admission
decisions to medical school and residency programs, respectively. In summary, though, reviews of
the literature indicate that clinical competence is complex and that no one measure is sufficient to predict overall performance after medical school graduation. This section summarizes research linking tools or measures with specific outcomes, such as future examination or clinical performance. In summary, while performance on tests of knowledge tends to predict performance on tests of knowledge, there is far less evidence for valid measures to predict performance at later stages of the continuum in other competence domains.

**Predicting Success in Medical School**

In their selection processes, medical schools typically rely to varying degrees on the MCAT, the college grade point average (GPA) and interviews. There has been much research done on how well the MCAT predicts performance during medical school. Statistical analyses reveal that the MCAT score has a significant relationship to USMLE Step 1 performance (predicts about 43 percent of the variance) and a much smaller relationship to Step 2 performance (predicts about 18 percent of the variance). In general, the total MCAT score has a medium predictive validity for basic science course performance (19 percent of the variance) and clinical (clerkship) performance (15 percent of the variance), and medical school grades were best predicted by a combination of MCAT scores and undergraduate GPA, though the percent of the variance explained was not high. These results indicate that there are other factors that influence performance in medical school.

Some form of interview is used as part of the admission process to, in part, assess nonacademic personal qualities and to predict nonacademic success. Concerns have been raised, however, about lack of consistency and objectivity in an unstructured interview. To address this issue, new formats have been created that exhibit more standardization. The multiple mini-interview (MMI) uses a number of brief encounters modeled after the OSCE. As utilized by the McMaster University MD program, candidates have a short period of time to respond to questions or situations alone or with other applicants. All applicants experience the same scenarios. In an early study, the MMI was independently predictive of performance on the Medical Council of Canada Qualifying Examination (MCCQE). The MCCQE is similar to the USMLE. The MMI is a type of situational judgment test. This type of assessment has been shown to be useful to select for a variety of nonacademic or professional attributes.

**Predicting Success in Residency Training**

USMLE Step 1 scores are commonly used by residency program directors to select applicants for interviews. However, USMLE performance can be influenced by a variety of factors, such as the curriculum of the medical school, the assessment methods used by the school, and the clinical experience of the student at the time the exam is taken.

A review of the literature did not show a statistically significant correlation between USMLE Step 1 and 2 scores and reliable measures of procedural and clinical skill acquisition among residents and fellows. There is, however, correlation between USMLE scores and the scores on MCQ-based medical specialty board examinations. There also was a significant correlation between USMLE Step 2 scores and the scores on the in-training examination in one specialty.

**Predicting Success in Practice**

A systematic review of the literature found few studies of the relationship between performance in the early stage of the continuum (i.e., medical school) and performance in practice. One substantive area of inquiry is related to the identification of individuals who would experience
future adverse actions. Studies have linked behaviors in medical school and residency training related to professionalism with the risk of disciplinary actions by state medical licensing boards. Behaviors in medical school that were statistically related to licensing board actions were defined by the authors as “severe irresponsibility” and “severely diminished capacity for self-improvement.” A national study of internal medicine residents found that low professionalism ratings on the Residents’ Annual Evaluation Summary predicted increased risk for disciplinary action by state medical licensing boards. The study also found that progressively increasing professionalism ratings and higher scores on the American Board of Internal Medicine certification examination were associated with less risk for subsequent disciplinary action.

Clinical performance at all levels of the continuum is complex, and little is known about the relationship between performance measures early in the continuum and longer-term practice outcomes. There is a need, therefore, for a more systematic approach to study of the predictive value of assessment methods and tools.

APPROACHES TO ALIGNING ASSESSMENT ACROSS THE CONTINUUM

What type of assessment system would allow the performance of an individual to be determined through valid and reliable means at various stages of the continuum? Based on research to date, external assessment of clinical knowledge using “written” tests can be both predictive from one stage of the continuum to the next and can have appropriate levels of reliability and validity. There would be a need, however, to ensure that the test blueprint (the number of questions per content area) samples appropriately from the discipline domain being tested and that the questions are at an appropriate level for the stage of the continuum. There are methods, such the OSCEs and mini-CEX, to assess distinct cognitive and procedural skills. These also, when properly developed and administered, have appropriate statistical properties. Performance benchmarks (e.g., passing scores) for all these could be set based on the stage of the continuum. There are many other domains of clinical competence that are assessed in various ways, but these assessments tend to occur in isolation and do not allow a composite picture of knowledge and skills at a given phase of an individual’s professional development.

Workplace-based Assessment as an Organizing Framework

There is a need to create an organizing framework that would allow assessment along the medical education continuum related to the six competency domains. Workplace-based assessment is defined as:

the assessment of working practices based on what doctors actually do in the clinical setting and predominantly carried out in the workplace itself.

Workplace-based assessment can be a format for collecting and aggregating performance data from quantitative and qualitative sources about a breadth of clinical skills. As such, it can be used to provide feedback about physicians’ development of these skills as they progress along the continuum. The tools that typically are used for workplace-based assessment can be categorized as:

- Documentation of work experience, such as patient encounter logs.
- Observation of individual clinical encounters, such as the mini-CEX.
- Discussion of individual clinical cases, such as chart stimulated recall.
- Feedback from peers and others on routine performance.
These techniques have been described in an earlier section of this report. Workplace-based assessment allows the results to be aggregated so that a picture of composite performance can be developed.

**Setting Benchmarks of Performance**

The results of workplace-based assessment would allow a cumulative judgment about the performance of an individual at a given stage of the medical education continuum and allow a determination of readiness for progression to the next year of the program or phase of the continuum. How then do we know if the level of performance that is achieved is appropriate? Benchmarks for individual measures, such as the passing score on a written test and an OSCE, are common. However, benchmarks for the aggregate performance of an individual are not. One example that has been implemented is the milestones component of the ACGME Next Accreditation System. Residency programs will evaluate residents in each of the competency domains at intervals and submit composite milestone data on residents to the ACGME every six months. The results of the milestone evaluations will place each resident along a performance continuum for each competency domain. While this information will be used as part of the accreditation process, it is not clear how it will be used in decisions within a residency program about progression for individual residents.

The issue of benchmarks for progression decisions is an important one, since competency-based curricula permit the advancement/promotion of an individual within medical school and from medical school to residency that is not time-based. That is, a medical student or a resident could complete the educational program in less than the standard time if he or she meets the requirements of the program’s competencies. While theoretically attractive, there is a need to set appropriate performance benchmarks to determine if the requirements have been met. This is not just a theoretical need. In the 2013-2014 academic year, 17 medical schools (12%) reported having a time flexible/competency-based curriculum for all students.

**SUMMARY AND RECOMMENDATIONS**

The goal of ensuring that physicians are knowledgeable and skilled depends on an assessment system that allows both formative feedback to improve performance and summative decisions based on valid and reliable measures. The system should be coordinated so that progression in knowledge and skill development can be monitored across the stages of the medical education continuum. While there has been progress in achieving this outcome, more work is needed in two areas. One is assessment of the competency areas of systems-based practice and practice-based learning and improvement. Another area is to move beyond the individual competency areas to ensure that physicians are prepared for the complexities of medical practice.

The Council on Medical Education recommends that the following recommendations be adopted and that the remainder of this report be filed.

1. That our American Medical Association (AMA) support the concept that evaluation of physicians as they progress along the medical education continuum should include the following:

   a. Assessments of each of the six competency domains of patient care, medical knowledge, interpersonal and communication skills, professionalism, practice-based learning and improvement, and systems-based practice; and
b. Use of assessment instruments and tools that are valid and reliable and appropriate for each competency domain and stage of the medical education continuum. (New HOD Policy)

2. That our AMA encourage study of competency-based progression within and between medical school and residency.
   a. Through its Accelerating Change in Medical Education initiative, our AMA should study models of competency-based progression within the medical school.
   b. Our AMA should work with the Accreditation Council for Graduate Medical Education (ACGME) to study how the Milestones of the Next Accreditation System support competency-based progression in residency. (Directive to Take Action)

3. That our AMA encourage research on innovative methods of assessment related to the six competency domains of the ACGME/American Board of Medical Specialties that would allow monitoring of performance across the stages of the educational continuum. (Directive to Take Action)

4. That our AMA encourage ongoing research to identify best practices for workplace-based assessment that allow performance data related to each of the six competency domains to be aggregated and to serve as feedback to physicians in training and in practice. (Directive to Take Action)

Fiscal Note: Less than $500.
REFERENCES

7. 2013-2014 LCME Part II Annual Medical School Questionnaire. Sent to the deans of the 140 MD-granting medical schools, with a 100% response rate.