Recommendations to Promote Equity in Internal Medicine Clerkship Assessment and Grading

In this document, we define the terms “assessment” and “grades” in the following ways. Assessment can be in the form of formative feedback to foster students’ growth. Assessments also can be used in a summative manner to determine a student’s level of competence or achievement based on clerkship objectives and rubrics. Grades and summative narratives serve to communicate a student’s level of achievement or competence to internal and external stakeholders. Ideally, assessments and grades should be based on observation of students.

When adapted from the definition of health equity, educational equity broadly can be considered as the concept that all learners have the opportunity to attain their full learning potential and that no one is disadvantaged from achieving this potential because of structural or social barriers. Equity in assessment, as follows, is promoted when learners have “fair and impartial opportunities to learn, be evaluated, coached, graded, advanced, graduated, and selected for subsequent opportunities based on their demonstration of achievements that predict future success in the field of medicine and that neither learning experiences nor assessments are negatively influenced by structural or interpersonal bias related to personal or social characteristics of learners or assessors.” (Lucey)

These guidelines categorize best practices that internal medicine clerkship directors can consider when building an equitable assessment system for medical students on their clerkships.

Assessment Tools
Careful selection of assessment methods, ideally prioritizing those based on observations in the workplace, is critical, and must be paired with ongoing institutionally supported faculty and resident development on how to utilize these tools.

Utilize criterion-referenced and competency-based assessment forms with defined rubrics that include specific behaviors reflective of a certain level of achievement for a competency domain, e.g., a well-defined 4-point assessment rubric for a given competency domain (Appendix 1 and Appendix 2). Another option is to use standardized checklists in some situations, e.g., review of notes, directly observed patient encounters, oral presentations, or observed structured clinical encounters (OSCEs). This potentially reduces subjectivity in assessment which can introduce bias. Faculty/resident and other supervisors should be trained on how to apply these forms.

- The Association of American Medical Colleges guide for developing core entrustable professional activities and toolkits can be used as a guide for creating assessment rubrics. [https://www.aamc.org/what-we-do/mission-areas/medical-education/cbme/core-epas](https://www.aamc.org/what-we-do/mission-areas/medical-education/cbme/core-epas)

Increase the number of observations of students in clinical-based patient care. This might include workplace-based assessments (WPBA), e.g., observing a student communicating with a patient or with an interprofessional team member, review of notes or oral presentations, and OSCEs.
Having multiple observations and assessments can mitigate bias from any one individual supervisor. Assess “patient care” skills that are not traditionally assessed or are not weighted significantly in the determination of the final clerkship grade.

- These important activities and skills should be considered an essential clinical experience and should contribute to a student’s final grade, and can include patient advocacy (e.g., addressing patients’ social determinants of health), student initiative, team collaboration, and self-improvement. Formally assessing these skills, e.g., through evaluation of narrative written reflections, signal to students the competency domain’s value in patient care and contributes to a more holistic view of students’ contributions to patient care (Appendix 3).

Faculty/Resident Supervisor Development
Provide best practices for writing specific, behaviorally-based narrative assessments that include students’ strengths and areas for improvement. This should also include education on inequities in language use by gender and race. Examples of inequities in language:

- Change gender-influenced language- “She was a treat to have on the general medicine service” to behaviorally-based language: “Her written and oral presentations were thorough, she was responsive to patients, and her interactions with staff were always professional.”
- Change personality-focused language- “As is common in his culture, he was quiet and studious” to behaviorally-based language: “During rounds, when prompted his responses to direct questions about patient care reflected that he had done extensive reading about his patients.”
- Gender bias tool calculators for narratives are available on the internet (e.g., https://www.tomforth.co.uk/genderbias/)

Require that all faculty/resident supervisors participate in ongoing education, through the department or school, on bias reduction techniques, including education on implicit bias and influence on assessment of students, and identification of microaggressions and strategies on how to address with learners.

- Many medical schools provide local online and in-person workshops addressing these topics, and free resources are widely available, e.g., Implicit Association Test https://implicit.harvard.edu/implicit/education.html and AAIM website https://www.im.org/resources/diversity-inclusion/dei-resources.
- It is critical that these educational workshops are not simply a “one-time” occurrence. Ongoing, regular faculty/resident development, including “just in time” education prior to a stint working with learners is critical.

Clerkship Oversight of Grading Process
Recommend the use of departmental grading committees in all core clerkships to mitigate against potential effects of the individual bias of those responsible for assigning grades. Membership of the committee should be diverse (e.g., race/ethnicity, gender, years of experience).
Provide guidance on how to incorporate individual supervisors’ comments to generate a final detailed narrative summary that is free of bias. These summative narratives from the clerkship should include specificity regarding students’ level of achievement and should be reviewed to ensure avoidance of potentially biased language (e.g., gendered language).

- Use of a “Best Practices” document with examples can be helpful (e.g., from University of California San Francisco School of Medicine-
  https://meded.ucsf.edu/sites/meded.ucsf.edu/files/inline-files/Good%20assessment%20practice%20-%20evaluation%20examples.pdf)

**Standardized Exams**

If standardized exams are used in clerkship grading, limit the weight that standardized exam scores, including National Board of Medical Examiner’s subject or shelf exams, have in determining a student’s grade. Differences between population group outcomes in standardized examinations likely reflect unequal opportunities afforded to underrepresented in medicine (UIM) students. In addition, deemphasizing exam scores might allow students to shift their attention to other important patient skills that they need to develop.

Recommend against standardized test score cut-offs for Honors grades. Analysis of differences in clerkship performance based on UIM status suggests that attributes linked to performance on high-stakes multiple choice exams may be responsible for differences in clerkship performance assessment. Small magnitudes of difference in clerkship director ratings are amplified by institutional grading policies and institutionally-defined eligibility criteria to the Alpha Omega Alpha (AOA) honor society which leads to lower attainment of honors grades and likelihood of selection for AOA membership for UIM students when compared to non-UIM students. This amplification cascade can affect UIM student residency training and career options and choice.

**Criterion-Referenced Grading v Normative Grading**

Implement criterion-referenced grading. The clinical performance assessments that supervisors complete at end of the rotation, which often forms the majority of a student’s grade, are susceptible to individual supervisor’s bias. The use of a normative approach to grading students (e.g., using a predetermined number or percentage of students who can be assigned a certain grade) only serves to further magnify the challenges with bias in individual assessments. Criterion-referenced grading provides transparency about the level of performance required to achieve a grade.

**Clerkship Programmatic Evaluation (Educational Continuous Quality Improvement)**

Address policies, processes, and the environment to enhance a culture of respect and inclusion.

During the curriculum and equivalency of sites review, include a review of data to mitigate bias. This is aligned with the Liaison Committee of Medical Education (LCME) accreditation standard requiring schools to evaluate their program’s effectiveness. This might include the following:
Monitor clerkship variables that relate to equity, including gender and UIM status in distribution of clerkship grades and NBME shelf scores; mistreatment experiences reported by students on clerkship evaluations or school report; and student clerkship satisfaction with clerkship in areas related to race, ethnicity, and gender. This programmatic review should include outcomes of inquiry (e.g., clerkship responses to these reports).

If developing new standardized assessments, work with institutional leaders in student assessment and programmatic evaluation to decrease bias in the new tools.

References


**Acknowledgements**

These recommendations were developed by a work group of CDIM members: Cindy Lai, MD, Chavon M. Onumah, MD, MPH, FACP, AAHIVS, Nora Y. Osman, MD, Diane L. Levine, MD, FACP, Nadia J. Ismail, MD, MPH, MEd, Amber T. Pincavage, MD, FACP, and Amy Shaheen, MD, MSc, FACP.
### Appendix 1: Observed Patient Encounter (OPE) Scoring
(20 min. with pt; 25 min. debrief)

<table>
<thead>
<tr>
<th>Category</th>
<th>Sub Category</th>
<th>Answers (points)</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Gathering (EPA1)</td>
<td>History</td>
<td>Pertinent positives (3) Pertinent negatives (3) Complete &amp; accurate in organized fashion (3)</td>
<td>___out of 9</td>
</tr>
<tr>
<td>Observed with patient</td>
<td>Physical Exam</td>
<td>Pertinent components (2) Appropriate skill (2) Correct findings (2)</td>
<td>___out of 6</td>
</tr>
<tr>
<td>–20 min total *give 5 min warning</td>
<td>Doctor-patient communication</td>
<td>0– Lack of rapport, little empathy, failure to act on verbal or nonverbal cues 1- 2- 3- Good rapport with patient. Empathic. Recognizes and responds to verbal or nonverbal cues. 4- 5- 6- Good rapport with patient. Empathic. Recognizes and responds to verbal or nonverbal cues. Develops therapeutic alliance.</td>
<td>___out of 6</td>
</tr>
<tr>
<td>Differential Diagnosis (EPA 2)</td>
<td>Identified pivotal pts(1)</td>
<td>Most likely leading diagnoses (1)</td>
<td></td>
</tr>
<tr>
<td>Observed in debrief session</td>
<td>Appropriate can’t miss/alternate diagnosis (2)</td>
<td></td>
<td>___out of 4</td>
</tr>
<tr>
<td>Evaluation (EPA 3&amp;4)</td>
<td>Appropriate test for ruling in disease (1)</td>
<td></td>
<td>___out of 3</td>
</tr>
<tr>
<td>Observed in debrief session</td>
<td>Appropriate test to rule out disease (1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Management (EPA 4)</td>
<td>Basics of management (3)</td>
<td></td>
<td>___out of 3</td>
</tr>
<tr>
<td>Observed in debrief session</td>
<td>Appropriate rationale for decision to order a test (1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient Education (EPA 4)</td>
<td>Clear explanation in patient appropriate language</td>
<td></td>
<td>___out of 2</td>
</tr>
<tr>
<td>Observed with patient</td>
<td>Assessment of understanding Anticipatory guidance (2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall performance</td>
<td></td>
<td></td>
<td>___out of 3</td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td></td>
<td>___out of 36</td>
</tr>
</tbody>
</table>

Grade (Honors Manager 32-36\High Pass Interpreter 24-31 \Pass Reporter 18-23 \ Fail<18)

Appendix 2: Grading Rubric for a Comprehensive Note Write-up

Chief Complaint: 0, 1, 2 points
   0: none
   1: present
   2: includes patient’s main complaint, in patient’s words, and no additional
   information/patient information/other non-pertinent wording

Opening sentence: 0, 3, 5 points
   0: none
   3: present but lacks appropriate important information, or includes information that is not
   important to the differential
   5: includes appropriate history and not distractors

HPI: 0-15 points
   2: Organized
   2: Thorough
   4: Includes pertinent positive ROS
   4: Includes pertinent negative ROS
   3: Includes pertinent past history/family history/social history

Past Medical History: 0, 1, 2 points
   0: none
   1: disorganized, incomplete, paragraph format
   2: organized, thorough, bulleted format (includes surgical history, ob/gyn history if
   appropriate, vaccinations/developmental history if a child)

Medications: 0, 1, 2
   0: nothing written (if no medications, must state so)
   1: medications listed but uses abbreviations, trade names
   2: medications listed, no abbreviations, generic names

Allergies: 0, 1, 2 points
   0: nothing listed (if no allergies, must indicate such)
   1: allergies listed but not reactions
   2: allergies and reactions listed, or no allergies listed as “no known drug allergies"

Social History: 0, 1 point (point system does NOT reflect a lack of importance to this!!! Please
include alcohol, tobacco, drug use, living situation, social support)

Family History: 0, 1 point (point system does NOT reflect lack of importance)

ROS: 0, 1 point
   0: none or lists only a few, not organized, includes PE or other findings, repeats
   information already described in HPI
   1: thorough, excludes information written in HPI with “as in HPI” references, does not
   include any PE findings in ROS

Physical Exam: 0, 5, 10 points
   0: none
5: incomplete, unorganized R
10: includes vitals, organized in appropriate order, thorough, mentions pertinent findings and pertinent negatives findings I

Summary Statement: 0, 5, 10 points
0: none
5: present but unorganized, does not include pertinent information or includes information that is not pertinent or incorrect I
10: organized, includes pertinent HPI, PE and data leading to differential diagnosis M

Problem list, Assessment/Plan with differential: total of 50 points

Problem list: 0, 2, 5 points
0: none listed
2: present but incomplete I
5: organized, thorough, complete; includes cc; in order of acuity M

Differential diagnosis: 0, 10, 20 points
0: none R
10: less than 3 items on differential I
20: at least 3 items on the differential, includes the cc as a problem for clinical reasoning M

Clinical reasoning: 0, 5, 10, 15, 20 points
0: none
5: minimal reasoning, does not list most likely diagnosis or must not miss diagnosis R
10: more thorough, but not organized into “differential, work up, treatment”
15: thorough and organized, works through differential, describes why and why not diagnoses should be considered, includes most likely diagnosis (and describes this), includes must not miss diagnoses when appropriate; organized into “differential, work up, treatment plan” format I
20: differential and clinical reasoning “wows”; reasoning is advanced; M

Overall organization and prioritization: 0-4 points
Organized, extraneous information removed, edited information from auto-population

Reviewer: _________________________________________________________
Total points & Grade: ____________________________________________

### Appendix 3: Sample Assessment Item for Patient Advocacy

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Not observed</td>
</tr>
<tr>
<td>1</td>
<td>Does not identify sociocultural factors that impact patient care, including (but not limited to) race, religion, culture, gender identity, sexuality, primary language, immigration status, and disability (ability) as important to patient care.</td>
</tr>
<tr>
<td>2</td>
<td>Can identify sociocultural factors that impact patient care, including (but not limited to) race, religion, culture, gender identity, sexuality, primary language, immigration status, and disability (ability). AND is able to consider these factors in the context of an individual patient’s care.</td>
</tr>
<tr>
<td>3</td>
<td>Identifies sociocultural factors that impact patient care, including (but not limited to) race, religion, culture, gender identity, sexuality, primary language, immigration status, and disability (ability). AND incorporates these factors in the individual patient’s care AND implements solutions to overcome barriers.</td>
</tr>
</tbody>
</table>

From the University of California School of Medicine, San Francisco (UCSF) Internal Medicine Clerkship, 2021.