Setting a common standard in clinical skills assessment:

The experience of the California Consortium for the Assessment of Clinical Competence

Candace Pau, MD (KPSOM)
Stacy Charat, MD (UCSD)
Daniel Kahn, MD (UCLA)
Alan Liu, MD (USC)
Theresa Murdock-Vlautin, MD (UCD)
Andrew Nevins, MD (Stanford)

Kerensa Peterson (UCR)
Abigail Phillips, MD (UCSF)
Tamara Shankel, MD (LLU)
Alisa Wray, MD (UCI)
Christy Boscardin, PhD (UCSF)

ALOHA AAMC Western Group Collaborative Spring Conference

Disclosures

The presenters have no financial interest or affiliation concerning material discussed in this presentation.

The California Consortium for the Assessment of Clinical Competence

- Established in 1991 to develop and share resources around standardized-patient based examinations for the assessment of clinical skills
- Currently includes 10 allopathic California medical schools
- Annual administration of the Clinical Performance Examination (CPX)

Clinical Performance Examination (CPX)

- 8-case OSCE (6 common core cases, 2 institutional elective cases)
- Encounter checklists scored by standardized patients
- Interstation exercises scored by trained clinical raters
- Typical timing is end of M3 / beginning of M4 year
- Passing CPX is required for graduation at all schools

Case Content

- Bank of approximately 40 bread and butter cases
- Range of acuity levels
- Includes pediatrics cases (telehealth visits with patient's parent; adolescent health)
- Case presentations cover multiple organ systems
- Learner tasks: pertinent history and physical examination, discuss diagnosis and plan with the patient
- Interstation exercises: traditionally modeled after Step 2 CS note → now being adapted to assess other skills

Approach to Standard Setting: Before

- Most schools require passing each domain to pass the exam:
  - History
  - Physical Examination
  - Patient Education and Counseling
  - Patient-Provider Interaction
- One school requires passing 6 out of 8 cases
- Most schools used a norm-based approach to determine passing thresholds for each domain
- Cut score generally 1.5-2 standard deviations below the mean
- One school used a modified Angoff approach (criterion-based) to determine the passing score for each domain

Toward a Common Standard

- Elimination of Step 2 CS has left a gap for clinical skills benchmarking in UME
- Validity evidence for results from clinical skills exams is enhanced by a robust standard setting process
- Criterion-based standards are preferable—align with CBME and are not affected by variability among institutions and cohorts
- Many criterion-based methods are time- and resource-intensive

Dual Approach

Modified Angoff method

- 3 content experts x 2 schools per case
- Review case checklist and provide ratings for each item
- Raters given opportunity to adjust ratings based on existing cross-institutional cohort data

Borderline regression method

- Standardized patients in each case asked to rate students as “pass”, “borderline”, or “not pass”
- Item appended to existing checklist
- Training provided to SPs on how to score
- Ratings regressed on student performance to identify standard

Results: Modified Angoff vs Borderline Regression

Case	Modified Angoff Cut Score	Borderline Regression Cut Score
Case 1 (Peds)	53-89%	46-87%
Case 2 (Neuro)	55-77%	42-94%
Case 3 (Pulmonary)	47-55%	39-89%
Case 4 (Acute Cardiac)	82-91%	52-91%
Case 5 (Chronic Cardiac)	49-80%	42-88%
Case 6 (ENT)	70-82%	42-90%
Overall (6 cases)	**70.0%**	**70.2%**

Results: Comparison to normative approach

	School 1	School 2	School 3	School 4	School 5	School 6	School 7	School 8	Total
Mean score	79%	76%	77%	81%	76%	78%	75%	78%	77%
2 SD below	69%	66%	67%	73%	66%	68%	65%	68%	67%
Effect of switching to common cut score on pass rate	↔	↓	↓	↑	↓	↔	↓	↔	↓

Learnings

Modified Angoff:

- Definition of “minimally competent”
- How to interpret and use cohort data
- Meaning of the percentage rating

Borderline regression:

- SP concerns with rating student competence
- Holistic impressions vs critical checklist items
- Confusion with pre-existing global ratings

Rater selection and training is key!

Borderline Rating Language

Based on your impression of the student's overall performance across all domains (history-taking, physical exam, patient education and counseling, and communication) in this encounter, please select the most appropriate category for their performance in the case.

Pass: The student performed at or above the level expected for a third-year medical student. Overall, the student did reasonably well on this case. Performance in this category may range anywhere from "just barely passing to excellent."

Borderline: The student is at the border between meeting and not meeting the expected level of performance for a third-year medical student. The student doesn't clearly fall into Pass or Not Pass. It is not clear that they are minimally competent as a third-year medical student. Behaviors observed might include some level of disorganization or not sure they understood what was going on, unprofessional or did not inspire trust.

Not Pass: The student has not demonstrated the level of performance expected for a third-year medical student. There were significant issues and concerns with the student's performance and interaction. Behaviors observed might include significant disorganization, lack of professionalism, errors or omissions that impacted my experience.

Conclusions & Future Directions

- Modified Angoff (clinician) and borderline regression (SP) ratings yielded similar results
- Variability across cases suggests case specificity
- Challenges with SP training/perception

- Comparing SP vs clinician borderline regression ratings
- Considering domain-specific borderline ratings
- Institutions must determine how to use/adapt the common standard according to their own curricular context

Thank You!

Abigail Phillips, MD
CCACC Chair

UCSF
abigail.phillips@ucsf.edu

Christy Boscardin, PhD

UCSF
christy.boscardin@ucsf.edu

Candace Pau, MD

KPSOM
candace.y1.pau@kp.org

Alisa Wray, MD

UC Irvine
awray@hs.uci.edu

References

Dwyer T, Wright S, Kulasegaram KM, Theodoropoulos J, Chahal J, Wasserstein D, Ringsted C, Hodges B, Ogilvie-Harris D. How to set the bar in competency-based medical education: standard setting after an Objective Structured Clinical Examination (OSCE). BMC Med Educ. 2016 Jan 4;16:1. doi: 10.1186/s12909-015-0506-z. PMID: 26727954; PMCID: PMC4700734.

John JT, Gowda D, Schlair S, Hojsak J, Milan F, Auerbach L. After the Discontinuation of Step 2 CS: A Collaborative Statement from the Directors of Clinical Skills Education (DOCS). Teach Learn Med. 2022 Mar 14:1-6. doi: 10.1080/10401334.2022.2039154. Epub ahead of print. PMID: 35287502.

Milan FB, Grochowalski JH. A resource efficient and reliable standard setting method for OSCEs: Borderline regression method using standardized patients as sole raters in clinical case encounters with medical students. Med Teach. 2022 Aug;44(8):878-885. doi: 10.1080/0142159X.2022.2041586. Epub 2022 Mar 2. PMID: 35234562.

Nevins AB, Boscardin CK, Kahn D, May W, Murdock-Vlautin T, Pau CY, Phillips A, Racataian-Gavan R, Shankel T, Wilkerson L, Wray A, Charat S. A Call to Action From the California Consortium for the Assessment of Clinical Competence: Making the Case for Regional Collaboration. Acad Med. 2022 Sep 1;97(9):1289-1294. doi: 10.1097/ACM.0000000000004663. Epub 2022 Mar 8. PMID: 35263299.