HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

IMPROVING COMPETENCY-BASED ASSESSMENT IN HUMAN ANATOMY EDUCATION: INTEGRATING OSPE, DIGITAL TESTING, AND PRACTICAL SKILLS EVALUATION

Ashuraliyeva Gulhumor
Department of Anatomy and clinic anatomy, Andijan State medical institute
doi:https://doi.org/10.5281/zenodo.23575451

Abstract: Competency-based assessment in human anatomy should establish whether learners can recognize anatomical structures, explain three-dimensional relationships, interpret clinically relevant images, and transfer foundational knowledge to patient-centered problems. This narrative methodological review examines the complementary roles of the objective structured practical examination (OSPE), well-designed digital assessments, and observed practical performance. Drawing on assessment theory and health professions education literature, it proposes a blueprint-driven, multimodal assessment framework linking learning outcomes, tasks, scoring rubrics, feedback, and quality assurance. The framework emphasizes validity evidence, standard setting, accessibility, examiner calibration, academic integrity, and longitudinal monitoring. Because no original cohort data were collected, claims about effectiveness are presented as pedagogical propositions rather than measured outcomes.

Keywords: human anatomy; competency-based education; OSPE; digital assessment; practical skills; assessment validity; medical education.

Introduction

Human anatomy remains an essential foundation of medical education, yet assessment can become disproportionately focused on memorizing names and locations. Recognition of a tagged specimen is important, but clinical practice also requires the ability to reason from structure to function, explain anatomical variation, and locate relevant structures in radiological images. A comprehensive assessment system must therefore sample several kinds of evidence rather than equate success on a single examination with competence [1,2].

Competency-based education organizes instruction and evaluation around explicit capabilities that learners must demonstrate. For anatomy, those capabilities include anatomical identification, spatial orientation, application to clinical scenarios, accurate use of terminology, safe specimen handling, and communication of relevant anatomical information. The shift does not make factual knowledge less important. Instead, factual knowledge is assessed together with interpretation and application in progressively authentic contexts [2,3].

The objective structured practical examination offers standardized stations and explicit criteria, while digital testing provides efficient coverage of a broad curriculum and can generate rapid feedback. Practical assessment adds evidence of what students actually do when handling models, imaging

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

resources, or anatomical material. Each method has limitations; an integrated system is intended to offset those limitations and make consequential decisions more defensible [4,5].

Aim and Review Approach

The aim of this article is to develop a coherent pedagogical approach to competency-based assessment in human anatomy through the coordinated use of OSPE stations, digital tests, and practical-skills evaluation. Specific objectives are to define assessable competencies, align assessment formats with outcomes, identify feasible scoring and feedback procedures, and describe a quality-assurance strategy suitable for medical universities.

This paper is a narrative methodological review rather than a systematic review or empirical trial. It draws on established literature in medical assessment, learning theory, and anatomy education, with attention to widely used concepts such as assessment blueprints, constructive alignment, formative feedback, validity evidence, and programmatic assessment [1–10]. The proposed model should be locally piloted and evaluated before any claims of improved examination performance or clinical competence are made.

Defining the Anatomy Competency Framework

A competency framework should begin with observable performances. The first domain is foundational anatomical knowledge: correct identification and description of bones, muscles, vessels, nerves, organs, and their relationships. The second domain is spatial understanding: mentally transforming two-dimensional diagrams into three-dimensional relationships and correctly interpreting anatomical planes. The third domain is clinical application: using anatomy to explain symptoms, procedural risks, physical examination findings, and imaging appearances.

A fourth domain involves practical and professional behavior, including appropriate handling of anatomical specimens, respect for donated bodies, compliance with laboratory safety requirements, and accurate communication. A fifth domain is self-regulated learning: recognizing knowledge gaps, acting on feedback, and documenting progress. Not every domain requires the same assessment format. A written question can assess conceptual reasoning, but it cannot directly establish whether a student can orient a specimen safely or explain a structure to a colleague in real time [2,3].

Competencies should be written with action verbs and defined performance conditions. For example, instead of the broad outcome “understands brachial plexus anatomy,” an assessable outcome could state: “Given a labeled diagram and a short clinical vignette, identify the likely injured nerve, justify the answer using its anatomical course, and distinguish it from two plausible alternatives.” Such wording supports both teaching design and transparent scoring.

Constructive Alignment and Assessment Blueprinting

Constructive alignment requires the stated learning outcomes, teaching activities, and assessments to reinforce one another. When the curriculum promises clinically applicable spatial understanding but examinations reward only isolated recall, students receive contradictory signals. A

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

blueprint makes the alignment visible by mapping body regions, cognitive demands, competencies, assessment formats, and relative weighting [6].

A workable blueprint can distribute evidence across the musculoskeletal, nervous, cardiovascular, respiratory, gastrointestinal, and urogenital systems while avoiding overrepresentation of easily photographed structures. Within each region, tasks should sample recall, interpretation, spatial reasoning, and clinical application. A digital test may efficiently cover a wide range of content, whereas OSPE stations can probe targeted performances with controlled prompts and standardized time limits.

Blueprinting also protects fairness. Two cohorts should not receive assessments of markedly different difficulty merely because different instructors prepared their stations. A review panel can compare content coverage, task complexity, ambiguity, and the cognitive work expected from students. Documenting this process provides part of the validity argument for score interpretation [4,7].

OSPE as a Structured Practical Assessment

The objective structured practical examination adapts structured station-based testing to practical disciplines. In anatomy, stations may use cadaveric prosections, plastinated material, skeletal specimens, high-quality photographs, 3D models, and radiological images. A station presents a defined task, a specified response format, a time limit, and a scoring scheme. Standardization can reduce irrelevant differences in questioning and marking, although it cannot eliminate all measurement error [5].

A well-designed OSPE station should assess a meaningful unit of competence rather than reward guessing a label. For example, a station displaying the femoral triangle may ask the student to identify the femoral nerve, artery, and vein; explain their medial-to-lateral relationships; and identify a safety implication for vascular access. Scoring can distinguish accurate identification, spatial explanation, and clinical relevance. This produces more interpretable feedback than a single all-or-nothing mark.

Station design should also consider authenticity and feasibility. Some competencies can be assessed through images or models, while others require direct interaction with specimens. Station time must permit the intended reasoning. Excessively short stations may measure reading speed and anxiety rather than anatomy. Conversely, very long stations reduce the breadth of content sampled. Piloting with representative learners can reveal unclear instructions, unrealistic timing, and unreliable prompts.

Examiners need training to use rubrics consistently. Where a task permits multiple valid explanations, the rubric should credit equivalent scientifically correct responses. For higher-stakes examinations, double marking of selected stations, post-exam review of unusual scores, and structured examiner debriefing can improve defensibility. The term “objective” should not be mistaken for proof that the assessment is automatically valid or reliable [4,5].

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

Digital Testing and Image-Based Assessment

Digital platforms support multiple-choice questions, extended matching, image hotspots, short-answer responses, and interactive anatomical labeling. Their principal advantage is scalable sampling: a large number of learners can encounter diverse regions and question types, with automated marking for appropriately structured items. Digital tests can also provide timely diagnostic feedback and item-level analytics [7,8].

However, digital assessment quality depends on item design. Questions that ask for an isolated label may be suitable for foundational recognition but insufficient for higher-order reasoning. Case-based items should require students to infer an anatomical explanation from a clinical presentation. Imaging tasks can test identification across axial, sagittal, and coronal planes, but image resolution, display size, accessibility, and platform compatibility must be controlled.

Question banks should be tagged by competency, region, difficulty, and cognitive demand. Items require periodic review for ambiguity, technical errors, cueing, and differential performance across student groups. Item difficulty and discrimination can inform revision, but these statistics should be interpreted with content expertise rather than used mechanically. Digital convenience does not substitute for a clear validity argument [4,7].

Security is a further consideration. Randomization, controlled access, versioned question banks, and appropriate invigilation may reduce some forms of misconduct. Nevertheless, excessive surveillance can introduce privacy and equity concerns. Low-stakes formative quizzes may benefit more from open-book reasoning, immediate explanation, and repeated practice than from strict examination-style controls.

Assessing Practical Skills and Professional Behaviors

Practical skills evaluation addresses capabilities that are difficult to infer from written answers alone. In anatomy laboratories, students may be asked to orient a specimen, trace the course of a nerve, locate a clinically relevant landmark, explain anatomical relationships using a model, or demonstrate safe and respectful conduct. Observed performance makes the assessment more authentic but also introduces potential variability between examiners and settings [3,5].

Behaviorally anchored rubrics can improve clarity. A criterion such as “explains the relationship between the recurrent laryngeal nerve and the thyroid gland” should distinguish incomplete naming, correct anatomical explanation, and clinically relevant interpretation. A separate professionalism criterion can cover respectful specimen handling and adherence to laboratory rules. Safety-critical behavior should not be averaged away by strong performance on unrelated knowledge items.

Assessment should avoid conflating confidence with competence. A fluent speaker may provide an anatomically incorrect explanation, while a hesitant student may possess accurate knowledge. Examiners should focus on observable correctness, reasoning, and safe actions. Where communication is an explicit competency, the rubric should define the communication behavior being evaluated rather than reward personality or accent.

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

A Proposed Integrated Assessment Model

The proposed model contains four connected phases. In the diagnostic phase, brief digital pre-assessments identify baseline knowledge and common misconceptions. During the formative phase, low-stakes quizzes, peer explanation, and instructor observation provide feedback while students work with specimens and digital resources. The structured performance phase combines OSPE stations with practical demonstrations. Finally, a synthesis phase asks students to apply anatomical knowledge to short clinical or imaging cases.

No single percentage weighting is universally appropriate. Departments should determine the contribution of each component according to curricular priorities, resources, and the stakes of decisions. Importantly, a high score on factual recall should not automatically compensate for failure to meet a defined safety-critical standard. Programs can use minimum performance thresholds for selected essential competencies while reporting other domain scores separately [3,9].

Evidence should accumulate across several observations rather than depend exclusively on one testing day. A student portfolio may include feedback summaries, completed practical tasks, and reflections on recurring errors. Such records are not substitutes for standardized examinations, but they help educators identify patterns of development and offer targeted remediation. This approach is consistent with the broader concept of programmatic assessment, in which multiple data points support consequential judgments [9].

Feedback, Remediation, and Learner Development

Assessment becomes educationally valuable when it changes subsequent learning. Feedback should identify what the learner did well, which anatomical concept was inaccurate, and what action would improve performance. A message such as “revise neuroanatomy” is less useful than “review the distinction between upper and lower motor neuron lesions and use the corticospinal tract diagram to justify each clinical sign” [8,10].

Digital systems can deliver rapid explanations after formative questions. OSPE feedback can summarize station-specific weaknesses, such as poor identification of sectional anatomy or difficulty translating relationships into procedural implications. Practical instructors can provide brief observed-action feedback. Coordinating these messages helps prevent students from receiving fragmented or contradictory advice.

Remediation should be competency-specific. Learners who struggle with spatial orientation may benefit from rotating 3D models and interpreting serial imaging sections; learners who struggle with clinical application may require guided cases and think-aloud explanations. A follow-up assessment should verify improvement in the targeted competency rather than merely repeat the original item. Feedback literacy should be taught explicitly so that students can interpret evidence, set priorities, and monitor progress.

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

Validity, Reliability, Fairness, and Standard Setting

Assessment validity concerns the evidence supporting particular interpretations and uses of scores, not an intrinsic property conferred by the name of an examination. Relevant evidence includes content alignment, response processes, internal structure, relationships with other measures, and consequences of decisions. For anatomy, educators should ask whether the tasks represent intended competencies, whether students understand instructions as intended, and whether passing standards correspond to defensible expectations [4].

Reliability is affected by sampling, scoring, examiner variation, and testing conditions. Increasing the number of well-designed independent stations may improve the stability of OSPE scores, but the optimal design depends on resources and the competencies assessed. Digital item banks likewise need enough high-quality questions to avoid unstable conclusions based on a narrow sample. Periodic review of station performance, item behavior, and borderline decisions is essential [5,7].

Standard setting should be planned before high-stakes use. Criterion-referenced approaches, such as structured expert judgment about minimally competent performance, are often more aligned with competency-based education than arbitrary fixed pass marks. Departments should document panel composition, training, judgments, and subsequent review. The method selected must fit the assessment format and institutional regulations.

Fairness requires attention to disability accommodations, visual accessibility, language clarity, digital literacy, and consistent access to practice materials. Anatomical images should be sufficiently clear and appropriately labeled for the task. Students should not be disadvantaged by unrelated technological barriers. At the same time, accommodations must preserve the competence that the assessment is designed to measure.

Implementation in Resource-Constrained Medical Universities

Implementation does not require immediate investment in expensive simulation infrastructure. A phased model can begin with a small blueprinted OSPE, a curated bank of high-quality digital questions, and common practical-skills rubrics. Existing anatomical specimens, printed sectional images, freely available educational models, and institutional learning platforms may support the first cycle. Faculty development is more important than the mere purchase of software.

A practical implementation team should include anatomists, clinical teachers, assessment specialists, and technical staff. The team can identify priority competencies, write and peer-review tasks, train examiners, pilot stations, and review performance data. Students should receive clear information about expected competencies, sample station formats, feedback procedures, and appeals mechanisms before a high-stakes assessment.

Continuous improvement should proceed through documented cycles. After each assessment, educators can examine item statistics, examiner observations, student feedback, equipment failures, and differences between intended and actual timing. Changes should be recorded so that the next

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

cohort benefits from lessons learned. The ultimate objective is an assessment culture that supports learning while maintaining credible standards.

Limitations and Future Research

The model presented here is conceptual and has not been experimentally validated in a specific institution. The cited assessment literature supports the underlying principles but does not establish that this particular combination of OSPE, digital tests, and practical observations will improve learning in every context. Outcomes may depend on staff training, student preparedness, assessment stakes, available specimens, technology quality, and curricular alignment.

Future studies should compare clearly defined assessment packages using appropriate designs, report baseline differences, and examine more than immediate test scores. Relevant outcomes include retention of anatomical knowledge, performance on unfamiliar clinical cases, reliability of station scores, student perceptions, workload, cost, and equity. Mixed-methods evaluation could clarify why a model works well for some competencies but not others. Where feasible, longitudinal follow-up may help determine whether improvements persist into clinical training.

Conclusion

Competency-based anatomy assessment is strongest when evidence from several complementary formats is aligned with explicit educational outcomes. OSPE stations can assess structured practical and interpretive performances, digital tests can sample a broad curriculum efficiently, and direct observation can evaluate authentic laboratory behaviors. Their integration should be guided by blueprinting, transparent rubrics, formative feedback, validity evidence, standard setting, and fairness safeguards.

For medical universities, the proposed framework offers a feasible pathway from predominantly recall-based examinations toward a more balanced assessment of anatomical knowledge, spatial reasoning, clinical application, and professional practice. Its effectiveness should be tested locally through careful implementation and evaluation rather than assumed from the use of new assessment technologies.

References

1. Turney BW. Anatomy in a modern medical curriculum. Ann R Coll Surg Engl. 2007;89(2):104–107. doi:10.1308/003588407X168244.
2. Harden RM. Outcome-based education: the future is today. Med Teach. 2007;29(7):625–629. doi:10.1080/01421590701729930.
3. Frank JR, Snell LS, Cate OT, et al. Competency-based medical education: theory to practice. Med Teach. 2010;32(8):638–645. doi:10.3109/0142159X.2010.501190.
4. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. Med Educ. 2015;49(6):560–575. doi:10.1111/medu.12678.

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

5. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. Br Med J. 1975;1(5955):447–451. doi:10.1136/bmj.1.5955.447.
6. Biggs J. Enhancing teaching through constructive alignment. High Educ. 1996;32:347–364. doi:10.1007/BF00138871.
7. Downing SM. Twelve steps for effective test development. In: Downing SM, Yudkowsky R, eds. Assessment in Health Professions Education. Routledge; 2009.
8. Hattie J, Timperley H. The power of feedback. Rev Educ Res. 2007;77(1):81–112. doi:10.3102/003465430298487.
9. van der Vleuten CPM, Schuwirth LWT, Driessen EW, et al. A model for programmatic assessment fit for purpose. Med Teach. 2012;34(3):205–214. doi:10.3109/0142159X.2012.652239.
10. Nicol DJ, Macfarlane-Dick D. Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. Stud High Educ. 2006;31(2):199–218. doi:10.1080/03075070600572090.