Appendix 1: Summary of articles included for assessing the Reliability and Validity of Objective Structured Clinical Examinations (OSCE) in Assessing Clinical Competency in Anaesthesia.

1	Afra A, Seneysel Bachari S, Ban M.	Design, Implementation and Evaluation of Anesthesia Students' Clinical Competency Based on the Virtual Objective Structured Clinical Examination	Anesthesia and Pain Medicine	2025	Cronbach's alpha for the exam is not explicitly reported, but the high correlation between stations suggests good internal consistency	Construct validity Pearson's r = 0.861, P < 0.001)
2	Rebel A, DiLorenzo A, Nguyen D, et al.	Should Objective Structured Clinical Examinations Assist the Clinical Competency Committee in Assigning Anesthesiology Milestones Competency?	Anesthesia and Analgesia	2019	No report on reliability indices	Pearson r = 0.43–0.70
3	Ruan X, Xu X, Pei L, et al.	Chinese Anesthesiology Milestones in Resident Evaluation: Reliability, Validity, and Correlation with Objective Examination Scores: A Cross-sectional Study	Anesthesia and Analgesia	2025	Cronbach's α: 0.916 to 0.978	Pearson correlation coefficients r = 0.51 -0.66
4	Greenwood JE, Ledvina M. AANA J.	Development and Validation of a Quantitative Grading Rubric for High-Fidelity Simulation Assessment	American Association of Nurse Anesthesiology Journal	2023	interrater reliability = 0.9092	no validity scoring
5	Baig, L.A., Violato, C.	Temporal stability of objective structured clinical exams: a longitudinal study employing item response theory	BMC Medical Education	2012	Generalizability coefficient (Ep^2) for the OSCE stations exceeded 0.70	Construct validity: Evidence of stable measurement properties supports validity of station reuse
6	Baig LA, Beran TN, Vallevand A, et al	Accuracy of portrayal by standardized patients:	BMC Med Educ.	2014	Cronbach's α ranged from 0.40 to 0.74	No validity metrics used

		results from four OSCE stations conducted for high stakes examinations				
7	Mitchell, John D. MD; Amir, Rabia MD; Montealegre-Gallegos, Mario MD	Summative objective structured clinical examination assessment at the end of anesthesia residency for perioperative ultrasound	Anesthesia & Analgesia	2018	Kuder–Richardson Formula 20 (KR–20) > 0.5	No validity metrics used
8	Dubois DG, Lingley AJ, Ghatalia J, McConnell MM.	Validity of entrustment scales within anesthesiology residency training	Canadian Journal of Anesthesiology	2021	G-coefficient = 0.73; Cronbach's α = 0.86	Validity (criterion-related): Correlations: OSCE (0.45), AKT-6 (0.51), CCC rankings (0.52)
9	Khalafi A, Abbasi A, Sarvi Sarmeydani N, et al.	Investigating the Effectiveness of Formative OSCE Combined with Visual Feedback in Improving Clinical Competence among Iranian Nurse Anesthesia Students: A Quasi-experimental study	J Advanced Med Education and Professionalism	2024	Inter-rater Reliability: ICC = 0.824	Content Validity Index (CVI): > 0.79
10	Everett TC, McKinnon RJ, Ng E, et J	Simulation-based assessment in anesthesia: an international multicentre validation study. Can	Canadian Journal of Anesthesiology	2019	ICC (global ratings): 0.85–0.94	Construct validity: Significant score differences between juniors, seniors, and staff (P = 0.04 to < 0.001)
11	Schwid HA, Rooke GA, Carline J, et al	Evaluation of anesthesia residents using mannequin-based simulation: a multiinstitutional study.	Anesthesiology	2002	Cronbach's α): 0.71–0.76	In-training exam scores: r = 0.44–0.49
12	P. J. Morgan, D. Cleave-Hogg, S. DeSousa, et al	High-fidelity patient simulation: validation of performance checklists	British Journal of Anaesthesia	2004	Cronbach's α for the 10 scenarios ranged from 0.16 to 0.93	No validity scores

13	Tanaka P, Park YS, Liu L, et al	Assessment Scores of a Mock Objective Structured Clinical Examination Administered to 99 Anesthesiology Residents at 8 Institutions.	Anesthesia and Analgesia	2020	G-coefficient: 0.56	Construct Validity illustrates evidence of significant variance from individual and program levels via G-theory
14	Ahmed A, Khan FA, Ismail S.	Reliability and validity of a tool to assess airway management skills in anesthesia trainees	Anaesthesiology and Clinical Pharmacology	2016	Kappa 0.91–0.99; test–retest Kappa ~0.39	Junior vs senior scores differ significantly
15	Weller, J.M. et al.	Evaluation of high-fidelity patient simulator in assessment of performance of anaesthetists	British Journal of Anaesthesia	2017	Overall ICC was 0.85	Face Validity: High (realism & participant feedback) Content Validity: Confirmed via linkage with recognized crisis management frameworks Construct Validity: Supported by literature showing novice vs experienced performance
16	McNeer RR, Dudaryk R, Nedeff NB, et al	Development and Testing of Screen-Based and Psychometric Instruments for Assessing Resident Performance in an Operating Room Simulator.	Anesthesiology Research and Practice	2016	ICC = 0.825	Convergent validity: Global Rating Scale (GRS) and Crisis Management Checklist (CMC) highly correlated (Spearman ρ = 0.948)