



USMLE-Related Publications 2017–2009

2017

Bennett R, von Davier M. Advancing human assessment: a synthesis over seven decades. In: Bennett R, von Davier M, eds. *Advancing Human Assessment*. New York, NY: Springer International Publishing; 2017:635–87.

Braun H, von Davier M. The use of test scores from large-scale assessment surveys: psychometric and statistical considerations. *Large-scale Assessments in Education*. 2017;5:17. doi:10.1186/s40536-017-0050-x.

Carlson J, von Davier M. Item response theory. In: Bennett R, von Davier M, eds. *Advancing Human Assessment*. New York, NY: Springer International Publishing; 2017:133–178.

Clauser A, Raymond M. Specifying the content of credentialing tests. In: Davis-Becker S, Buckendahl C, eds. *Testing in the Professions: Credentialing Policies and Practice*. New York, NY: Rutledge; 2017:21–40.

Clauser B, Margolis M, Swanson D. Issues of validity and reliability for assessments in medical education. In: Holmboe E, Durning S, Hawkins R, eds. *Practical Guide to the Evaluation of Clinical Competence*. Amsterdam, the Netherlands: Elsevier; 2017:22–36.

Clauser BE, Baldwin P, Margolis MJ, Mee J, Winward M. An experimental study of the internal consistency of judgments made in bookmark standard setting. *Journal of Educational Measurement*. 2017;54(4):481–497.

Clauser BE, Margolis MJ, Clauser JC. Validity issues for technology-enhanced innovative assessments. In: Jiao H, Lissitz RW, eds. *Technology Enhanced Innovative Assessment: Development, Modeling, and Scoring from an Interdisciplinary Perspective*. Charlotte, NC: Information Age Publishing; 2017:139–161.

Cuddy MM, Young A, Gelman A, Swanson DB, Johnson DA, Dillon GF, Clauser BE. Exploring the relationships between USMLE performance and disciplinary action in practice: a validity study of score inferences from a licensure examination. *Academic Medicine*. 2017;92(12):1780–1785.

Dong T, Zahn C, Saguil A, Swygert K, Yoon M, Servey J, Durning S. The associations between clerkship objective structured clinical examination (OSCE) grades and subsequent performance. *Teaching and Learning in Medicine*. 2017;29(3):280–285.

Feinberg R, Jurich D. Decision visualization for incomplete test administrations. *Educational Measurement: Issues and Practice*. 2017;36(2):Cover.

Feinberg R, Jurich D. Guidelines for interpreting and reporting subscores. *Educational Measurement: Issues and Practice*. 2017;36(1):5–13.

Grabovsky I, Wainer H. A guide for setting the cut-scores to minimize weighted classification errors in test batteries. *Journal of Educational and Behavioral Statistics*. 2017;42(3):264–281.

Grabovsky I, Wainer H. The cut-score operating function: a new tool to aid in standard setting. *Journal of Educational and Behavioral Statistics*. 2017;42(3):251–263.

Haist S, Butler A, Paniagua M. Testing and evaluation: the present and future of the assessment of medical professionals. *Advances in Physiology Education*. 2017;41(1):149–153.

He Q, von Davier M, Greiff S, Steinhauer EW, Borysewicz PB. Collaborative problem solving measures in the programme for international student assessment (PISA). In: von Davier AA, Kyllonen PC, Zhu M, eds. *Innovative Assessment of Collaboration*. Dordrecht, the Netherlands: Springer; 2017:95–111.

Indik J, Duhigg L, McDonald F, Lipner R, Rubright J, Haist S, Botkin N, Kuvin J. Performance on the cardiovascular in-training examination in relation to the ABIM cardiovascular disease certification examination. *Journal of the American College of Cardiology*. 2017;69(23):2862–2868.

Jang H, Pak S. Perfectionism and high school adjustment: self-directed learning strategies as a mediator. *Journal of Asia Pacific Counseling*. 2017;7(1):1–16.

Kane M, Clauser B, Kane J. A validation framework for credentialing tests. In: Davis-Becker S, Buckendahl C, eds. *Testing in the Professions: Credentialing Policies and Practice*. New York, NY: Rutledge; 2017:21–40.

Kirsch I, Lennon ML, Yamamoto K, von Davier M. Large- scale assessments of adult literacy. In: Bennett R, von Davier M, eds. *Advancing Human Assessment*. New York, NY: Springer International Publishing; 2017:285–310.

Knight C, Windish D, Haist S, et al. The SGIM TEACH program: a curriculum for teachers of clinical medicine. *Journal of General Internal Medicine*. 2017;32(8):948–952.

Miller E, Heitz C, Ross LP, Beeson MS. Emergency medicine student end-of-rotation examinations: where are we now? *Western Journal of Emergency Medicine*. <https://escholarship.org/uc/item/1jz0p8jm>. Published December 2017. Accessed January 1, 2018.

Paniagua M. 100 days of rain: a reflection on the limits of physician resilience. National Academy of Medicine. <https://nam.edu/100-days-of-rain-a-reflection-on-the-limits-of-physician-resilience>. Published January 20, 2017. Accessed November 29, 2017.

Stankov L, Lee J, von Davier M. A note on construct validity of the anchoring method in PISA 2012. *Journal of Psychoeducational Assessment*. Published April 4, 2017. doi:10.1177/0734282917702270.

Stites SD, Karlawish J, Harkins K, Rubright, JD, Wolk D. Awareness of mild cognitive impairment and mild Alzheimer's disease dementia diagnoses associated with lower self-ratings of quality of life in older adults. *Journal of Gerontology: Psychological Sciences*. 2017;72(6):974–985.

von Davier M, Shin HJ, Khorramdel L, Stankov L. The effects of vignette scoring on reliability and validity of self reports. *Applied Psychological Measurement*. Published September 27, 2017. doi:10.1177/0146621617730389.

von Davier, M. New results on an improved parallel EM algorithm for estimating generalized latent variable models. In: van der Ark LA, Wiberg M, Culpepper SA, Douglas JA, Wang WC, eds. *Quantitative Psychology: Proceedings of the 81st Annual Meeting of the Psychometric Society*. Asheville, NC: Springer; 2017:1–8.

Walsh K, Harik P, Mazor K, et al. Measuring harm in health care: optimizing adverse event review. *Medical Care*. 2017;55(4):436–441.

2016

Clauser J, Hambleton R, Baldwin P. The effect of rating unfamiliar items on Angoff passing scores. *Educational and Psychological Measurement*. 2016; Oct 10. [Epub ahead of print]

Margolis MJ, Mee JM, Clauser BE, Winward M, Clauser JC. Effect of content knowledge on Angoff-style standard setting judgments. *Educational Measurement: Issues and Practice*. 2016;35(1):29-37.

Cuddy MM, Winward ML, Johnston MM, Lipner RS, Clauser BE. Evaluating validity evidence for USMLE Step 2 Clinical Skills data gathering and data interpretation scores: Does performance predict history-taking and physical examination ratings for first-year internal medicine residents? *Academic Medicine*. 2016;91:133-139.

Prober CG, Kolars JC, First LR, Melnick DE. A plea to reassess the role of United States Medical Licensing Examination Step 1 scores in residency selection. *Academic Medicine*. 2016;91:12-15.

Raymond MR, Ling Y, Grabovsky I. Investigating the performance of second language medical students on lengthy clinical vignettes. *Evaluation & the Health Professions*. 2016; Oct 19. [Epub ahead of print]

Katsufakis PJ, Uhler TA, Jones LD. The residency application process: pursuing improved outcomes through better understanding of the issues. *Academic Medicine*. 2016;91:1483-1487.

2015

Cuddy MM, Winward ML, Johnston MM, Lipner RS, Clauser BE. Evaluating validity evidence for USMLE Step 2 Clinical Skills data gathering and data interpretation scores: does performance predict history-taking and physical examination ratings for first-year internal medicine residents? *Academic Medicine*. 2015. Sept 21 [Epub ahead of print]

Dong T, LaRochelle JS, Durning SJ, Saguil A, Swygert KA, Artino AR. Longitudinal effects of medical students' communication skills on future performance. *Military Medicine*. 2015;180(4 Suppl):24-30.

Feinberg RA, Raymond MR, Haist SA. Repeat testing effects on credentialing exams: are repeaters misinformed or uninformed? *Educational Measurement: Issues and Practice*. 2015;34(1):34-39.

Furman G. Proving our worth: foundational literature supporting the standardized patient educational methodology. *ASPE eNews*. 2015; December 8. <http://multibriefs.com/briefs/aspeorg/ASPEORG120815.php>. Accessed 01/13/2016

Kahraman N, Brown CB. Using multigroup confirmatory factor analysis to test measurement invariance in raters: a clinical skills examination application. *Applied Measurement in Education*. 2015;28:350-366.

Lane S, Raymond MR, Haladyna TM, Downing SM. Test development process. In: Lane S, Raymond MR, Haladyna TM, eds. *Handbook of test development*. 2nd ed. New York: Routledge; 2015:3-18.

Ouyang W, Cuddy MM, Swanson DB. US medical student performance on the NBME Subject Examination in Internal Medicine: do clerkship sequence and clerkship length matter? *Journal of General Internal Medicine*. 2015;30:1307-1312.

Peitzman SJ, Cuddy MM. Performance in physical examination on the USMLE Step 2 Clinical Skills examination. *Academic Medicine*. 2015;90:209-213.

Prober CG, Kolars JC, First LR, Melnick DE. A plea to reassess the role of United States Medical Licensing Examination Step 1 scores in residency selection. *Academic Medicine*. 2015. Aug 3. [Epub ahead of print]

2014

Clauser JC, Clauser BE, Hambleton RK. Increasing the validity of Angoff standards through analysis of judge-level internal consistency. *Applied Measurement in Education*. 2014;27:19-30.

Clauser JC, Margolis MJ, Clauser BE. An examination of the replicability of Angoff Standard setting results Within a generalizability theory framework. *Journal of Educational Measurement*. 2014;51:127-140.

Dillon GF, Johnson DA. Implementing strategic changes to the USMLE. *Journal of Medical Regulation*. 2014;100(3):19-23.

Dong T, Swygert KA, Durning SJ, Saguil A, Gilliland WR, Cruess DF, DeZee KJ, LaRochelle J, Artino AR. Validity evidence for medical school OSCEs: associations with USMLE step assessments. *Teaching and Learning in Medicine*. 2014;26:379-386.

Dong T, Swygert KA, Durning SJ, Saguil A, Zahn CM, Dezee KJ, Gilliland WR, Cruess DF, Balog EK, Servey JT, Welling DR, Ritter M, Goldenberg MN, Ramsay LB, Artino AR. Is poor performance on NBME clinical subject examinations associated with a failing score on the USMLE Step 3 Examination? *Academic Medicine*. 2014;89:762-766.

Grabovsky I, Hess BJ, Haist SA, Lipner RS, Hawley JL, Woodward S, Engleberg NC. The relationship between performance on the infectious diseases in-training and certification examinations. *Clinical Infectious Diseases*. 2014; 2014 Nov 18. [Epub ahead of print]

Haist SA, Katsufakis PJ, Dillon GF. Basic science content in the USMLE Step 1--reply. *JAMA*. 2014;311:1359-1360.

Holtzman KZ, Swanson DB, Ouyang W, Dillon GF, Boulet JR. International variation in performance by clinical discipline and task on the United States Medical Licensing Examination Step 2 Clinical Knowledge component. *Academic Medicine*. 2014;89:1558-1562.

Morrison CA, Ross LP, Sample L, Butler A. Relationship between performance on the NBME Comprehensive Clinical Science Self-Assessment and USMLE Step 2 Clinical Knowledge for USMGs and IMGs. *Teaching and Learning in Medicine*. 2014;26:373-378.

Norcini JJ, Boulet JR, Opalek A, Dauphinee WD. The Relationship Between Licensing Examination Performance and the Outcomes of Care by International Medical School Graduates. *Academic Medicine*. 2014;89(8):1157-1162.

Peitzman SJ, Cuddy MM. Performance in physical examination on the USMLE Step 2 Clinical Skills Examination. *Academic Medicine*. 2014;Nov 18. [Epub ahead of print].

Raymond MR, Mee J, Haist SA, Young A, Dillon GF, Katsufakis PJ, McEllhenney SM, Johnson DA. Expectations for physician licensure: a national survey of practice. *Journal of Medical Regulation*. 2014;100(1):15-23.

Stoffel H, Raymond MR, Bucak SD, Haist SA. Editorial changes and item performance: implications for calibration and pretesting. *Practical Assessment, Research & Evaluation*. 2014;19(14):available online: <http://pareonline.net/genpare.asp?wh=4&abt=stoffel>

Wainer H. The route to USMLE; the shibboleth of modern medical licensure. *Journal of Medical Regulation*. 2014;100(4):21-28.

2013

Brown CB, Kahraman N. Exploring psychometric models to enhance standardized patient quality assurance: evaluating standardized patient performance over time. *Academic Medicine*. 2013;88:866-871.

Chavez AK, Swygert KA, Peitzman SJ, Raymond MR. Within-session score gains for repeat examinees on a standardized patient examination. *Academic Medicine*. 2013;88:688-692.

Clauser BE, Mee J, Margolis MJ. The effect of data format on integration of performance data into Angoff judgments. *International Journal of Testing*. 2013;13:65-85.

Cuddy MM, Swanson DB, Drake RL, Pawlina W. Changes in anatomy instruction and USMLE performance: empirical evidence on the absence of a relationship. *Anatomical Sciences Education*. 2013;6:3-10.

Dillon GD, Swanson DB, McClintock JC, Gravlee GP. The Relationship Between the American Board of Anesthesiology Part 1 Certification Examination and the United States Medical Licensing Examination. *Journal of Graduate Medical Education*. 2013;6:276-283.

Haist SA, Katsufakis PJ, Dillon GF. The evolution of the United States Medical Licensing Examination (USMLE): enhancing assessment of practice-related competencies. *JAMA*. 2013;310:2245-2246.

Hoppe RB, King AM, Mazor KM, Furman GE, Wick-Garcia P, Corcoran-Ponisciak H, Katsufakis PJ. Enhancement of the assessment of physician-patient communication skills in the United States Medical Licensing Examination. *Academic Medicine*. 2013;88:1670-1675.

Kahraman N, Cuddy MM, Clauser BE. Modeling Pacing Behavior and Test Speededness Using Latent Growth Curve Models. *Applied Psychological Measurement*. 2013;37:343-360.

King AM, Hoppe RB. "Best practice" for patient-centered communication: a narrative review. *Journal of Graduate Medical Education*. 2013;5:385-393.

Mee J, Clauser BE, Margolis MJ. The impact of process instructions on judges' use of examinee performance data in Angoff standard setting exercises. *Educational Measurement: Issues and Practice*. 2013;32(3):27-35.

Swygert KA, Haist SA. A response to "Bridging the gender gap in communication skills" by Wu and McLaughlin (2012). *Advances in Health Sciences Education: Theory and Practice*. 2013;18:133-134.

Winward ML, Lipner RS, Johnston MM, Cuddy MM, Clauser BE. The relationship between communication scores from the USMLE Step 2 Clinical Skills examination and communication ratings for first-year internal medicine residents. *Academic Medicine*. 2013;88: 693-698.

2012

Feinberg RA, Swygert KA, Haist SA, Dillon GF, Murray CT. The impact of postgraduate training on USMLE Step 3 and its computer-based case simulation component. *Journal of General Internal Medicine*. 2012;27:65-70.

Kahraman N, De Champlain AF, Raymond MR. Modeling the psychometric properties of complex performance assessment tasks using confirmatory factor analysis: a multistage model for calibrating tasks. *Applied Measurement in Education*. 2012;25:79-95.

Raymond MR, Swygert KA, Kahraman N. Measurement precision for repeat examinees on a standardized patient examination. *Advances in Health Sciences Education: Theory and Practice*. 2012;17:325-337.

Raymond MR, Swygert KA, Kahraman N. Psychometric equivalences of ratings for repeat examinees on a performance assessment for physician licensure. *Journal of Educational Measurement*. 2012;49:339-361.

Swygert KA, Cuddy MM, van Zanten M, Haist SA, Jobe AC. Gender differences in examinee performance on the Step 2 Clinical Skills data gathering (DG) and patient note (PN) components. *Advances in Health Sciences Education Theory and Practice*. 2012; 17:557-571.

van Zanten M, McKinley DW, Durante Montiel I, Pijano CV. Medical education accreditation in Mexico and the Philippines: impact on student outcomes. *Medical Education*. 2012;46(6):568-592.

2011

Baldwin P, Baldwin SG, Haist SA. F-type testlets and the effects of feedback and case-specificity. *Academic Medicine*. 2011;86(10 Suppl):S55-S58.

Cuddy MM, Swygert KA, Swanson DB, Jobe AC. A multilevel analysis of examinee gender, standardized patient gender, and United State Medical Licensing Examination Step 2 Clinical Skills communication and interpersonal skills scores. *Academic Medicine*. -2011;86(10 Suppl):S17-S20.

Dillon GF, Clauser BE, Melnick DE. The role of USMLE scores in selecting residents [letter]. *Academic Medicine*. 2011;86:793-794.

Katsufakis PJ, Scoles PV, Melnick DE. Correcting a misperception [letter]. *Academic Medicine*. 2011;86:1333.

Raymond MR, Harik P, Clauser BE. The impact of statistically adjusting for rater effects on conditional standard errors of performance ratings. *Applied Psychological Measurement*. 2011;35:235-246.

Raymond MR, Kahraman N, Swygert KA, Balog KP. Evaluating construct equivalence and criterion-related validity for repeat examinees on a standardized patient examination. *Academic Medicine*. 2011;86:1253-1259.

Raymond MR, Mee JM, King AM, Haist SA, Winward ML. What new residents do during their initial months of training. *Academic Medicine*. 2011;86(10 Suppl):S59-S62.

2010

Furman GE, Smee S, Wilson C. Quality assurance best practices for simulation-based examinations. *Simulation in Healthcare: Journal of the Society for Simulation in Healthcare*. 2010;5:226-231.

Margolis MJ, Clauser BE, Winward M, Dillon GF. Validity evidence for USMLE examination cut scores: results of a large-scale survey. *Academic Medicine*. 2010;85(10 Suppl):S93-S97.

Morrison C, Ross LP, Fogle T, Butler A, Miller JG, Dillon GF. Relationship between performance on the NBME Comprehensive Basic Sciences Self-Assessment and USMLE Step 1 for U.S. and Canadian medical school students. *Academic Medicine*. 2010;85(10 Suppl):S98-S101.

Raymond MR, Clauser BE, Furman GE. The impact of statistical adjustment on conditional standard errors of measurement in the assessment of physician communication skills. *Advances in Health Sciences Education: Theory and Practice*. 2010;15:587-600.

Swygert KA, Balog KP, Jobe AC. The impact of repeat performance on examinee performance for a large-scale standardized-patient examination. *Academic Medicine*. 2010;85:1506-1510.

Swygert KA, Muller ES, Scott CL, Swanson DB. The relationship between USMLE Step 2 CS patient note ratings and time spent on the note: do examinees who spend more time write better notes? *Academic Medicine*. 2010;85(10 Suppl):S89-S92.

2009

Baldwin SG, Harik P, Keller LA, Clauser BE, Baldwin P, Rebbecchi TA. Assessing the impact of modifications to the documentation component's scoring rubric and rater training on USMLE Integrated Clinical Encounter scores. *Academic Medicine*. 2009;84(10 Suppl):S97-S100.

Boulet JR, Smee SM, Dillon GF, Gimpel JR. The use of standardized patient assessments for certification and licensure decisions. *Simulation in healthcare: Journal of the Society for Simulation in Healthcare*. 2009;4:35-42.

Clauser BE, Balog KP, Harik P, Kahraman N. A multivariate generalizability analysis of history-taking and physical examination scores from the USMLE Step 2 Clinical Skills examination. *Academic Medicine*. 2009;84(10 Suppl):S86-S89.

Dillon GF, Clauser BE. Computer-delivered patient simulations in the United States Medical Licensing Examination (USMLE). *Simulation in healthcare: Journal of the Society for Simulation in Healthcare*. 2009;4:30-34.

Harik P, Clauser BE, Grabovsky I, Nungester RJ, Swanson DB, Nandakumar R. An examination of rater drift within a generalizability theory framework. *Journal of Educational Measurement*. 2009;46:43-58.

Harik P, Cuddy MM, O'Donovan S, Murray CT, Swanson DB, Clauser BE. Assessing potentially dangerous medical actions with the computer-based case simulation portion of the USMLE Step 3 examination. *Academic Medicine*. 2009;84(10 Suppl):S79-S82.

Holtzman KZ, Swanson DB, Ouyang W, Hussie K, Albee K. Use of multimedia on the Step 1 and Step 2 Clinical Knowledge components of USMLE: a controlled trial of the impact on item characteristics. *Academic Medicine*. 2009;84(10 Suppl):S90-S93.

Johnson DA. An Assessment of USMLE examinees found to have engaged in Irregular Behavior. *Journal of Medical Regulation*. 2009/2010;95(4):26-35.

Melnick DE. Licensing examinations in North America: is external audit valuable? *Medical Teacher*. 2009;31:212-214.

Raymond MR, Clauser BE, Swygert KA, van Zanten M. Measurement precision of spoken English proficiency scores on the USMLE Step 2 Clinical Skills examination. *Academic Medicine*. 2009;84(10 Suppl):S83-S85.

Swanson DB, Holtzman KZ, Johnson DA. Developing test content for the United States Medical Licensing Examination. *Journal of Medical Licensure and Discipline*. 2009;95(2):22-29.

Swanson DB, Sawhill AJ, Holtzman KZ, Bucak SD, Morrison C, Hurwitz S, DeRosa GP. Relationship between performance on Part I of the American Board of Orthopaedic Surgery Certifying Examination and scores on USMLE Steps 1 and 2. *Academic Medicine*. 2009;84(10 Suppl):S21-S24.

Swygert KA, Muller ES, Swanson DB, Scott CL. The relationship between USMLE Step 2 CS Communication and Interpersonal Skills (CIS) ratings and the time spent by examinees interacting with standardized patients. *Academic Medicine*. 2009;84 (10 Suppl):S1-S4.

van Zanten M, Boulet JR. Medical education in the Caribbean: variability in Educational Commission for Foreign Medical Graduate certification rates and United States Medical Licensing Examination attempts. *Academic Medicine*. 2009;84(10 Suppl):S13-S16.