

# Staff Publications 1923-Present



NBME® staff publications contribute to the body of scholarship on assessment, explore emerging test constructs, and demonstrate the continuous reliability and validity of existing examinations. The following list of NBME publications highlights the body of work produced by staff. We hope you enjoy exploring these scholarly contributions.

Please contact the Office of Research Strategy at [ors@nbme.org](mailto:ors@nbme.org) if you would like more information about a publication.

## Contents

2020 TO PRESENT .....	4
2020 PUBLICATIONS .....	4
2010 TO 2019.....	7
2019 PUBLICATIONS .....	7
2018 PUBLICATIONS .....	9
2017 PUBLICATIONS .....	11
2016 PUBLICATIONS .....	13
2015 PUBLICATIONS .....	16
2014 PUBLICATIONS .....	19
2013 PUBLICATIONS .....	21
2012 PUBLICATIONS .....	23
2011 PUBLICATIONS .....	26
2010 PUBLICATIONS .....	28
2000 TO 2009.....	32
2009 PUBLICATIONS .....	32
2008 PUBLICATIONS .....	34
2007 PUBLICATIONS .....	36
2006 PUBLICATIONS .....	39
2005 PUBLICATIONS .....	41
2004 PUBLICATIONS .....	44
2003 PUBLICATIONS .....	46
2002 PUBLICATIONS .....	48
2001 PUBLICATIONS .....	51
2000 PUBLICATIONS .....	53
1990 TO 1999.....	58
1999 PUBLICATIONS .....	58

1998 PUBLICATIONS .....	59
1997 PUBLICATIONS .....	61
1996 PUBLICATIONS .....	66
1995 PUBLICATIONS .....	68
1994 PUBLICATIONS .....	71
1993 PUBLICATIONS .....	73
1992 PUBLICATIONS .....	75
1991 PUBLICATIONS .....	77
1990 PUBLICATIONS .....	79
1980 TO 1989.....	81
1989 PUBLICATIONS .....	81
1988 PUBLICATIONS .....	81
1987 PUBLICATIONS .....	82
1986 PUBLICATIONS .....	82
1985 PUBLICATIONS .....	83
1984 PUBLICATIONS .....	83
1983 PUBLICATIONS .....	84
1982 PUBLICATIONS .....	85
1981 PUBLICATIONS .....	85
1980 PUBLICATIONS .....	86
1970 TO 1979.....	88
1979 PUBLICATIONS .....	88
1978 PUBLICATIONS .....	89
1977 PUBLICATIONS .....	89
1976 PUBLICATIONS .....	90
1975 PUBLICATIONS .....	92
1974 PUBLICATIONS .....	93
1973 PUBLICATIONS .....	94
1972 PUBLICATIONS .....	95
1971 PUBLICATIONS .....	95
1970 PUBLICATIONS .....	96
1960 TO 1969.....	97
1969 PUBLICATIONS .....	97
1968 PUBLICATIONS .....	97
1967 PUBLICATIONS .....	97
1966 PUBLICATIONS .....	98
1965 PUBLICATIONS .....	98
1964 PUBLICATIONS .....	98
1963 PUBLICATIONS .....	99
1962 PUBLICATIONS .....	99

<b>1961 PUBLICATIONS</b> .....	99
<b>1960 PUBLICATIONS</b> .....	99
<b>1923 TO 1959</b> .....	101
<b>1950s PUBLICATIONS</b> .....	101
<b>1940s PUBLICATIONS</b> .....	102
<b>1930s PUBLICATIONS</b> .....	102
<b>1920s PUBLICATIONS</b> .....	102

---

## 2020 TO PRESENT

### Simulations, Psychometrics, and Standard Setting: 2010s

Articles on the assessment of professionalism, as well as the use of multisource feedback–type assessments, appeared into the 2010s, while more recent efforts focused on setting performance standards.

## 2020 PUBLICATIONS

Paniagua M, DelVescovo M, Dyrbye LN. Re-examining exams: National Board of Medical Examiners' Efforts on Wellness (RENEW). In: Byyny RL, Christensen S, Fish JD, eds. *Medical Professionalism Best Practices: Addressing Burnout and Resilience in Our Profession*. Alpha Omega Alpha Honor Medical Society; 2020:73-80.

Feinberg RA, von Davier M. Conditional subscore reporting using iterated discrete convolutions. *Journal of Educational and Behavioral Statistics*. 2020;45(5):515-533.

Harik P, Feinberg RA, Clauser BE. How examinees use time. In: Margolis MJ & RA Feinberg, eds. *Integrating Timing Considerations to Improve Testing Practices*. Routledge; 2020:90-103.

Hu K, Hicks PJ, Margolis MJ, et al. Reported pediatrics milestones (mostly) measure program, not learner performance. *Academic Medicine*. 2020;95(11S):S89-S94.

Jurich DP, Daniel M, Hauer KE, et al. Does delaying the United States Medical Licensing Examination Step 1 to after clerkships affect student performance on clerkship subject examinations? *Teaching and Learning in Medicine*. 2020;33(4):366-381.

Jurich DP, Santen SA, Paniagua M, et al. Effects of moving the United States medical licensing examination step 1 after core clerkships on step 2 clinical knowledge performance. *Academic Medicine*. 2020;95(1):111.

Jurich DP. A history of test speededness: tracing the evolution of theory and practice. In: Margolis MJ & RA Feinberg, eds. *Integrating Timing Considerations to Improve Testing Practices*. Routledge; 2020:90-103.

Leventhal BC, Grabovsky I. Adding objectivity to standard setting: evaluating consequence using the conscious and subconscious weight methods. *Educational Measurement: Issues and Practice*. 2020;39(1):30-36.

Margolis MJ, Feinberg RA. *Integrating Timing Considerations to Improve Testing Practices*. Routledge; 2020.

- McDonald FS, Jurich DP, Duhigg LM, et al. Correlations between the USMLE Step examinations, American College of Physicians in-training examination, and ABIM internal medicine certification examination. *Academic Medicine*. 2020;95(9):1388-1395.
- Park YS, Morales A, Ross L, Paniagua M. Reporting subscore profiles using diagnostic classification models in health professions education. *Evaluation & the health professions*. 2020;43(3):149-158.
- Stites S, Cao H, Gill J, Harkins K, Rubright JD, Flatt J. The CoGenT3 study: examining gender's impact on education and cognition trends in three American generations. *Innovation in Aging*. 2020;4(Suppl 1):696.
- Tolsgaard MG, Boscardin CK, Park YS, Cuddy MM, Sebok-Syer SS. The role of data science and machine learning in health professions education: practical applications, theoretical contributions, and epistemic beliefs. *Advances in Health Sciences Education*. 2020;25:1057–1086.
- Allem J-P, Attonito J, Avol E, et al. Acknowledgment of Reviewers for 2019. *Evaluation & the Health Professions*. 2020;43(1):71.
- Baldwin P, Margolis MJ, Clauser BE, Mee J, Winward M. The choice of response probability in bookmark standard setting: an experimental study. *Educational Measurement: Issues and Practice*. 2020;39(1):37-44.
- Ha L, Yaneva V, Harik P, Pandian R, Morales A, Clauser BE. Automated prediction of examinee proficiency from short-answer questions. *Proceedings of the 28th International Conference on Computational Linguistics*. 2020: 893–903.
- Hammoud MM, Foster LM, Cuddy MM, Swanson DB, Wallach PM. Medical student experiences with accessing and entering patient information in electronic health records during the obstetrics-gynecology clerkship. *American Journal of Obstetrics and Gynecology*. 2020;223(3):435-e6.
- Jodoin MG, Rubright JD. When examinees cannot test: the pandemic's assault on certification and licensure. *Educational Measurement: Issues and Practice*. Published online July 23, 2020. doi:10.1111/emip.12361
- Margolis MJ, Clauser BE. Automated scoring in medical licensing. In: Margolis MJ & Clauser BE, eds. *Handbook of Automated Scoring*. Chapman and Hall/CRC; 2020:445-468.
- Runyon CR. Using multisite instrumental variables to estimate treatment effects and treatment effect heterogeneity. *UT Electronic Theses and Dissertations*. 2020. <http://dx.doi.org/10.26153/tsw/13096>
- Baldwin P. A problem with the bookmark procedure's correction for guessing. *Educational Measurement: Issues and Practice*. Published online November 24, 2020. <https://doi.org/10.1111/emip.12400>

Clauser BE, Kane M, Clauser JC. Examining the precision of cut scores within a generalizability theory framework: A closer look at the item effect. *Journal of Educational Measurement*. 2020;57(2):216-229.

Eraslan S, Yesilada Y, Yaneva V, Harper S. Eye-tracking scanpath trend analysis for autism detection. *ACM SIGACCESS Accessibility and Computing*. 2020(128):1-8.

Margolis MJ, von Davier M, Clauser BE. Timing considerations for performance assessments. In: Margolis MJ & RA Feinberg, eds. *Integrating timing considerations to improve testing practices*. Routledge; 2020:90-103.

Peterson LE, Boulet JR, Clauser BE. Associations between medical education assessments and American Board of Family Medicine Certification Examination score and failure to obtain certification. *Academic Medicine*. 2020;95(9):1396-1403.

Yaneva V, Eraslan S, Yesilada Y, Mitkov R. Detecting high-functioning autism in adults using eye tracking and machine learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2020;28(6):1254-1261.

---

## 2010 TO 2019

### Simulations, Psychometrics, and Standard Setting: 2010s

Articles on the assessment of professionalism, as well as the use of multisource feedback–type assessments, appeared into the 2010s, while more recent efforts focused on setting performance standards.

## 2019 PUBLICATIONS

Baldwin P. Some problems with the analytical argument in support of RP67 in the context of the bookmark standard setting method. *Applied Psychological Measurement*. October 2019;43(6):481–492.

Baldwin P, Margolis MJ, Clauser BE, Mee J, Winward M. The choice of response probability in bookmark standard setting: an experimental study. *Educational Measurement: Issues and Practices*. 2019; 39(1):37–44.

Daniel M, Jurich D, Santen SA. In reply to Le et al. *Academic Medicine*. July 2019;94(7):925–926.

Dekhtyar M, Ross LP, D'Angelo J, et al. Validity of the Health Systems Science examination: relationship between examinee performance and time of training. *American Journal of Medical Quality*. 2020;35(1):63–69.

Ha LA, Yaneva V. Automatic question answering for medical MCQs: Can it go further than information retrieval? *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*. 2019:418–422.

Ha LA, Yaneva V, Baldwin P, Mee J. Predicting the difficulty of multiple-choice questions in a high-stakes medical exam. *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*. 2019:11–20.

Jurich D, Santen SA, Paniagua MA, et al. Effects of moving the United States Medical Licensing Examination Step 1 after core clerkships on Step 2 Clinical Knowledge performance. *Academic Medicine*. March 2019;94(3):371–377.

Khorramdel L, von Davier M, Pokropek A. Combining mixture distribution and multidimensional IRTree models for the measurement of extreme response styles. *British Journal of Mathematical and Statistical Psychology*. 2019;72(3):538–559.

Khorramdel L, Jeon M, Wang LL. Advances in modelling response styles and related phenomena. *British Journal of Mathematical and Statistical Psychology*. 2019;72(3):392–400.

Liu C. Comparison of two item preknowledge detection approaches using response time. *83rd Annual Meeting of the Psychometric Society*. May 2019:355–365.

Morrison C, Ross LP, Baker G, Maranki M. Implementing a new score scale for the clinical science subject examinations: validity and practical considerations. *Medical Science Educator*. June 2019;29(3):841–847.

Ouyang W, Harik P, Clauser BE, Paniagua MA. Investigation of answer changes on the USMLE Step 2 Clinical Knowledge examination. *BMC Medical Education*. October 2019;19(1):389.

Raymond MR, Stevens C, Bucak SD. The optimal number of options for multiple-choice questions on high-stakes tests: application of a revised index for detecting nonfunctional distractors. *Advances in Health Sciences Education*. 2019;24(1):141–150.

Salt J, Harik P, Barone MA. In reply to Spadafore and Monrad. *Academic Medicine*. July 2019;94(7):926–927.

Santos KCP, Torre JDL, von Davier M. Adjusting person fit index for skewness in cognitive diagnosis modeling. *Journal of Classification*. 2019.

Sarker A, Klein AZ, Mee J, Harik P, Gonzalez-Hernandez G. An interpretable natural language processing system for written medical examination assessment. *Journal of Biomedical Informatics*. 2019;98:103268.

Shin HJ, von Davier M, Yamamoto K. Investigating rater effects in international large-scale assessments. In: Veldkamp BP, Sluijter C, eds. *Theoretical and Practical Advances in Computer-based Educational Measurement*. New York: Springer; 2019:249–268.

Feinberg RA, Jurich DP. Visualizing hierarchical score inferences. *On the Cover, Educational Measurement: Issues and Practice*. 2019;38(2).

Jurich D, Daniel M, Paniagua M, Fleming A, Harnik V, Pock A, Swan-Sein A, Barone MA, Santen SA. Moving the United States Medical Licensing Examination Step 1 after core clerkships: an outcomes analysis. *Acad Med*. 2019;94(3):371-377.

Knetka E, Runyon C, Eddy S. One size doesn't fit all: using factor analysis to gather validity evidence when using surveys in your research. *CBE-Life Sciences Education*. 2019;18(1):rm1.

Raymond MR, Stevens C, Bucak SD. The optimal number of options for multiple-choice questions on high-stakes tests: application of a revised index for detecting nonfunctional distractors. *Adv Health Sci Educ Theory Pract*. 2019;24(1):141-150.

Salt J, Harik P, Barone MA. Leveraging natural language processing: toward computer-assisted scoring of patient notes in the USMLE Step 2 Clinical Skills Exam. *Acad Med*. 2019;94(3):314-316.



Salt J, Harik P, Barone MA. In reply to Spadafore and Monrad. *Acad Med*. 2019;94(7):926-927.

von Davier M, Cho Y, Pan T. Effects of discontinue rules on psychometric properties of test scores. *Psychometrika*. 2019;84(1):147-163.

von Davier M, Lee YS. *Handbook of Diagnostic Classification Models*. Springer International Publishing; 2019.

Wallach PM, Foster LM, Cuddy MM, Hammoud MM, Holtzman KZ, Swanson DB. Electronic health record use in internal medicine clerkships and sub-internships for medical students graduating from 2012 to 2016. *J Gen Intern Med*. 2019;34(5):705-711.

---

## 2018 PUBLICATIONS

Carey EC, Paniagua M, Morrison LJ, Levine SK, Klick JC, Buckholz GT, Rotella J, Bruno J, Liao S, Arnold RM. Palliative care competencies and readiness for independent practice: a report on the American Academy Of Hospice and Palliative Medicine review of the U.S. Medical Licensing Step Examinations. *Journal of Pain and Symptom Management*. 2018;56(3):371-378.

Cui Z, Liu C, He Y, Chen H. Evaluation of a new method for providing full review opportunities in Computerized Adaptive Testing — Computerized Adaptive Testing with salt. *Journal of Educational Measurement*. 2018;55(4):582-594.

Edwards MC, Slagle A, Rubright JD, Wirth RJ. Fit for purpose and modern validity theory in clinical outcomes assessment. *Quality of Life Research*. 2018;27(7):1711-1720.

Feinberg RA, Jurich D, Lord J, Case H, Hawley J. Examining the validity of the North American Veterinary Licensing Examination (NAVLE) time constraints. *Journal of Veterinary Medical Education*. 2018;45(3):381-387.

Feinberg RA, Jurich DP. Providing utility, not scores: visualizations to support subscore inferences. *On the Cover, Educational Measurement: Issues and Practice*. 2018;37(3).

Feinberg RA, Jurich DP, Foster LM. Effects and unforeseen consequences of accessing references on a Maintenance of Certification examination. *Academic Medicine*. 2018;93(4):636-641.

Felgoise SH, Feinberg RA, Stephens HB, Barkhaus P, Boylan K, Caress J, Simmons Z. ALS specific quality of life short form (ALSSQOL-SF): A brief, reliable and valid version of the ALSSQOL-R. *Muscle and Nerve*. 2018;58(5):646-654.

Foster L, Cuddy M, Swanson DB, Holtzman KZ, Hammoud M, Wallach P. Medical student use of electronic and paper health records during inpatient clinical clerkships: results of a national longitudinal study. Association of American Medical Colleges Learn Serve Lead: Proceedings of the 57th Annual Research in Medical Education Sessions. *Academic Medicine*. 2018;93(suppl):14-20.

Franzen D, Cuddy MM, Ilgen JS. Trusting your test results: building and revising multiple-choice

examinations. *Journal of Graduate Medical Education*. 2018;10(3):337-338.

Harik P, Clauser BE, Grabovsky I, Baldwin P, Margolis M, Bucak D, Jodoin M, Walsh W, Haist S. A comparison of experimental and observational approaches to assessing the effects of time constraints in a medical licensing examination. *Journal of Educational Measurement*. 2018;55(2):308-327.

Jiang Z, Raymond MR. The use of multivariate generalizability theory to evaluate the quality of subscores. *Applied Psychological Measurement*. 2018;42:595-612.

Jurich D, Duhigg LM, Plumb TJ, Haist SA, Hawley JL, Lipner RS, Smith L, Norby SM. Performance on the Nephrology In-Training Examination and ABIM Nephrology Certification Examination Outcomes. *Journal of the American Society of Nephrology*. 2018;13(5):710-717.

Kirsch I, Thorn W, von Davier M. Guest editorial. *Quality Assurance in Education*. 2018;26(2):150-152.

Liu C, Kolen MJ. A comparison of strategies for smoothing parameter selection for mixed-format tests under the random groups design. *Journal of Educational Measurement*. 2018;55(4):564-581.

Michalec B, Cuddy MM, Hafferty P, Hanson MD, Kanter SL, Littleton D, Martimianakis MAT, Michaels R, Hafferty FW. It's happening sooner than you think: spotlighting the pre-medical realm. *Medical Education*. 2018;52(4):359-361.

Miller ES, Heitz C, Ross L, Beeson MS. Emergency Medicine student end-of-rotation examinations: where are we now? *West J Emerg Med*. 2018;19(1):134-136.

Paniagua M, Katsufakis P. The National Board of Medical Examiners®: Testing and Evaluation in the United States and Internationally. *Imagia Comunicacion*. 2018;8(29):5-12.

Paniagua M, Salt J, Swygert K, Barone M. Perceived utility of the USMLE Step 2 Clinical Skills Examination from a GME Perspective. *Journal of Medical Regulation*. 2018;104(2):51-57.

Park YS, Hicks PJ, Carraccio C, Margolis M, Schwartz A, PMAC Module 2 Study Group. Does incorporating a measure of clinical workload improve workplace-based assessment scores? Insights for measurement precision and longitudinal score growth from ten pediatrics residency programs. Association of American Medical Colleges Learn Serve Lead: Proceedings of the 57th Annual Research in Medical Education Sessions. *Academic Medicine*. 2018;93(suppl):21-29.

Pohl S, von Davier M. Commentary: on the importance of the speed-ability trade-off when dealing with not reached items. *Frontiers in Psychology*. 2018;30(9):1988.

Raymond MR. Integrating competency modeling with traditional job analysis. *CLEAR Exam Review*. 2018;27(2):21-27.

Rubright JD. Impact of both local item dependencies and cut-point locations on examinee classifications. *Educational Measurement: Issues and Practice*. 2018;37(3):40-45.

Short K, Bucak SD, Rosenthal F, Raymond MR. When listening is better than reading: performance gains on cardiac auscultation test questions. *Academic Medicine*. 2018;93(5):781-785.

Stites SD, Harkins K, Rubright JD, Karlawish J. Relationships between cognitive complaints and quality of

life in older adults with mild cognitive impairment, mild Alzheimer disease dementia, and normal cognition. *Alzheimer Disease & Associated Disorders*. 2018;32(4):276-283.

Stites SD, Rubright JD, Karlawish J. What features of stigma do the public most commonly attribute to Alzheimer's disease dementia? Results of a survey of the U.S. general public. *Alzheimer's & Dementia*. 2018;14(4):925-932.

Tackett S, Raymond M, Desai R, Haist SA, Morales A, Gaglani S, Clyman SG. Crowdsourcing for assessment items to support adaptive learning. *Medical Teacher*. 2018;40(8):838-841.

von Davier M. Automated item generation with recurrent neural networks. *Psychometrika*. 2018;83(4):847-857.

von Davier M. Diagnosing diagnostic models: from Von Neumann's elephant to model equivalencies and network psychometrics. *Measurement: Interdisciplinary Research and Perspectives*. Routledge; 2018;16(1):59-70.

von Davier M. Detecting and treating errors in tests and surveys. *Quality Assurance in Education*. 2018;26(2):243-262.

von Davier M, Carstensen C, Langeheine R, Eid M. In memoriam Jurgen Rost (1952-2017). *Psychometrika*. 2018;83(3):782-784.

von Davier M, Shin HJ, Khorramdel L, Stankov L. The effects of vignette scoring on reliability and validity of self-reports. *Applied Psychological Measurement*. 2018;42(4):291-306.

---

## 2017 PUBLICATIONS

Baker A J, Raymond M R, Haist S A, Boulet J R. Using national health care databases and problem-based practice analysis to inform integrated curriculum development. *Academic Medicine*. 2017;92(4):448-454.

Bennet R, von Davier M. *Advancing Human Assessment*. New York, NY: Springer International Publishing; 2017.

Bennett R, von Davier M. Advancing human assessment; a synthesis over seven decades. In: Bennett R, von Davier M, ed. *Advancing Human Assessment*. New York, NY: Springer; 2017:635-687.

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).

Carlson J, von Davier M. Item Response Theory. In: Bennett R, von Davier M, ed. *Shaping the Landscape of Educational Measurement and Evaluation*. New York, NY: Springer; 2017.

Clauser A, Raymond M. Specifying the content of credentialing tests. In: Davis-Becker S, Buckendahl C, ed. *Testing in the Professions: Credentialing Policies and Practice*. New York, NY: Routledge; 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;54:481-497.

Clauser BE, Margolis MJ, Clauser JC. Validity issues for technology-enhanced innovative assessments. In: Jiao H, Lissitz RW, ed. *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 A, 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. Guidelines for interpreting and reporting subscores. *Educational Measurement: Issues and Practice*. 2017;36:5-13.

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

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

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.

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

Haist S A, 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, Netherlands: Springer; 2017:95-111.

Indik J, Duhigg L, McDonald F, Lipner R S, Rubright J D, Haist S A, 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, Caluser B, Kane J. A validation framework for credentialing tests. In: Davis-Becker S, Buckendahl C, ed. *Testing in the Professions: Credentialing Policies and Practice*. New York: Routledge; 2017:21-40.

- Kirsch I, Lennon M L, Yamamoto K, von Davier M. Large-scale assessments of adult literacy. In: Bennett R, von Davier M, ed. *Advancing Human Assessment*. New York, NY: Springer; 2017:285-310.
- Knight C, Windish D, Haist S, Karani R, Chheda S, Rosenblum M, Basaviah P, Spencer A, Aagaard E. 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*. 2017(December).
- Paniagua M. 100 days of rain: a reflection on the limits of physician resilience. National Academy of Medicine; 2017;2017(November 29).
- Raymond M, Ling L, Grabovsky I. Investigating the performance of second language medical students on lengthy clinical vignettes. *Evaluation & the Health Professions*. 2017;40(2):151-158.
- Stankov L, Lee J, von Davier M. A Note on Construct Validity of the Anchoring Method in PISA 2012. *Journal of Psychoeducational Assessment*. 2017(April 4).
- 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. *The Journals of Gerontology: Series B*. 2017;72(6):974-985.
- 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, North Carolina, 2016*. New York, NY: Springer; 2017:1-8.
- von Davier M, Shin HJ, Khorramdel L, Stankov L. The effects of vignette scoring on reliability and validity of self reports. *Applied Psychological Measurement*. 2017(September 27).
- Walsh K, Harik P, Mazor K, Perfetto D, Anatchkova M, Biggins C, Wagner J. Measuring harm in health care: optimizing adverse event review. *Medical Care*. 2017;55(4):436-441.

---

## 2016 PUBLICATIONS

- Anderson M. A peer-reviewed collection of short reports from around the world on innovative approaches to medical education. *Medical Education*. 2016;50:560-561.
- Anderson M. Patient and public involvement in medical education: Is a new pedagogy necessary? *Medical Education*. 2016;50:8-10.
- Anderson M, Tolsgaard M G, Wolvaart J E, Duvivier R. Introduction. *Medical Education*. 2016;50:562-563.
- Brennan L, Siderowf A, Rubright J D, Rick J, Dahodwala N, Duda J E, Hurtig H, Stern M. The Penn Parkinson's Daily Activities Questionnaire-15: psychometric properties of a brief assessment of cognitive



- instrumental activities of daily living in Parkinson's disease. *Parkinsonism Related Disorders*. 2016;25:21-26.
- Brennan L, Siderowf A, Rubright J D, Rick J, Dahodwala N, Duda J E, Hurtig H, Stern M, Xie S X, Rennert L, Karlawish J, Shea J A, Trojanowski J Q, Weintraub D. Development and initial testing of The Penn Parkinson's Daily Activities Questionnaire. *Movement Disorders*. 2016;31:126-134.
- Clauser A L, Wainer H. A tale of two tests (and of two examinees). *Educational Measurement: Issues and Practice*. 2016;35(2):19-28.
- Clauser B E, Clauser J. Generalizability theory. In: Wells C, Faulkner-Bond M, ed. *Educational Measurement*. New York, NY: Guilford Publications; 2016:89-104.
- Clauser BE, Margolis M J, Clauser J C. Issues in simulation-based assessment. In: Dragow F, ed. *Technology and Testing*. New York, NY: Routledge; 2016:49-78.
- Clauser J, Hambleton R, Baldwin P. The effect of rating unfamiliar items on Angoff passing scores. *Educational and Psychological Measurement*. 2016;77(6):901-916.
- Collichio FA, Hess B J, Muchmore E A, Duhigg L, Lipner R S, Haist S A, Hawley J L, Morrison C, Clayton C P, Raymond M J, Kayoumi K M, Gitlin S D. Medical knowledge assessment by hematology and medical oncology in-training examinations are better than program director assessments at predicting subspecialty certification examination performance. *Journal of Cancer Education*. 2016;32(3):647-654.
- Cuddy M M, Winward M L, Johnston M M, Lipner R S, Clauser B E. 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(1):133-139.
- Feinberg R A, Clauser A L. Can item keyword feedback help remediate knowledge gaps? *Journal of Graduate Medical Education*. 2016;8:541-545.
- Feinberg R A, Rubright J D. Conducting simulation studies in psychometrics. *Educational Measurement: Issues and Practice*. 2016;35(2):36-49.
- Gomella L G, Haist S A. *Clinician's Pocket Desk Reference*. 14th ed. New York: McGraw-Hill; 2016.
- Hicks P J, Margolis M J, Poynter S E, Chaffinch C, Tenney-Soeiro R, Turner T L, Waggoner-Fountain L, Lockridge R, Clyman S G, Schwartz A. APPD LEARN-NBME Pediatrics Milestones Assessment Group. The Pediatrics Milestones Assessment Pilot: development of workplace-based assessment content, instruments, and processes. *Academic Medicine*. 2016;91:701-709.

Katsufraakis P J, Uhler T A, Jones L D. The residency application process: pursuing improved outcomes through better understanding of the issues. *Academic Medicine*. 2016;91:1483-1487.

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

Michalec B, Smith L, Ross L, Butler A, Smith C. Learning through self-assessment: investigating the relationship between performance on the NBME® clinical science mastery series self-assessments and clinical science subject examinations. *Medical Science Educator*. 2016;26(4):665–672.

Mills C, Briethaupt K. Current issues in computer-based testing. In: Wells C, Faulkner-Bond M, ed. *Educational Measurement: From Foundations to Future*. New York, NY: Guilford Press; 2016:208-220.

Natesan P, Nandakumar R, Minka T, Rubright J D. Bayesian prior choice in IRT estimation using MCMC and Variational Bayes. *Frontiers in Psychology*. 2016;7:1422.

Paniagua M, Swygert K A, Haist S A, Merrill J, Hussie K, Deruchie K, Billings M, Tyson J. Constructing Written Test Questions for the Basic and Clinical Sciences. 4th ed. Philadelphia, PA: NBME; 2016.

Prober C G, Kolars J C, First L R, Melnick D E. A plea to reassess the role of United States Medical Licensing Examination Step 1 scores in residency selection. *Academic Medicine*. 2016;91:12-15.

Rubright J D, Nandakumar R, Karlawish J. Identifying an appropriate measurement modeling approach for the Mini-Mental State Examination. *Psychological Assessment*. 2016;28(2):125-133.

Schwartz A, Margolis M J, Multerer S, Haftel H M, Schumacher D. APPD LEARN-NBME Pediatrics Milestones Assessment Group. A multi-source feedback tool for measuring a subset of Pediatrics Milestones. *Medical Teacher*. 2016;38:995-1002.

Small C, Land A M, Haist S A, Estrada C A, Snyder E D. Managing cognitive load to uncover an unusual cause of syncope: exercises in clinical reasoning. *Journal of General Internal Medicine*. 2016;31(2):247-251.

Steinberg L, Wainer H. VAMS and baseball. *Chance*. 2016;29(2):64.

Wainer H. Discussion of David Thissen's bad questions: an essay involving item response theory. *Journal of Educational and Behavioral Statistics*. 2016;41:100-103.

Wainer H. Defeating deception: escaping the shackles of truthiness by learning to think like a data scientist. *Chance*. 2016;29(1):61-64.

Wainer H. Don't try this at home. *Significance*. 2016;13(1):22-23.

Wainer H. Extracting sunbeams from cucumbers: how to design a better table. *Significance*. 2016.

Wainer H. The staunchions of statistics. *Science*. 2016;352:779.

Wainer H. Truth or truthiness: distinguishing fact from fiction by learning to think like a data scientist. New York: Cambridge University Press; 2016.

Wainer H, Spence I. William Playfair and the invention of statistical graphs. In: Black A, Luna P, Lund O, Walker S, eds. *Information Design: Research and Practice*. New York: Routledge; 2016:43-60.

Wenghofer H, Henzel T, Miller S, Norcross W, Boal P. The value of general medical knowledge examinations in performance assessment of practicing physicians with potential competence and performance deficiencies. *Journal of Continuing Education in the Health Professions*. 2016;36(2):113-118.

---

## 2015 PUBLICATIONS

Anderson MB. Really good stuff: lessons learned through innovation in medical education: a peer-reviewed collection of short reports from around the world on innovative approaches to medical education. *Medical Education*. 2015;49:1137-1138.

Anderson MB, Varpio L, Finn G, Youssry I. Really good stuff: lessons learned through innovation in medical education: introduction. *Medical Education*. 2015;49:509-512.

Baldwin P. Weighting components of a composite score using naïve expert judgments about their relative importance. *Applied Psychological Measurement*. 2015;39:539-550.

Baldwin P, Wainer H. Item response theory: a statistical theory of measurement based on fungible items. In: Henly SJ, ed. *Routledge International Handbook of Advanced Quantitative Methods in Nursing Research*. New York, NY: Routledge; 2015:58-80.

Bogle D, Thulasiram S. Multi-period accounting in Accounts Payable and how to integrate it to projects : OAUG Insight. *Oracle Applications Users Group*. 2015(Winter):26-30.

Clauser BE, Margolis MJ, Clauser JC. Issues in simulation-based assessment. In: Drasgow F, ed. *Technology and Testing: Improving Educational and Psychological Measurement*. New York, NY: Routledge; 2015:49-78.



Cook R, Wainer H. Joseph Fletcher, thematic maps, slavery and the worst places to live in the UK and the US. In: Kostelnick C, Kimball M, ed. *Visible Numbers: Essays on the History of Statistical Graphics*. New York, NY: Routledge; 2015:83-106.

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.

Dong T, Durning SJ, Gilliland WR, Swygert KA, Artino AR. Development and initial validation of a program director's evaluation form for medical school graduates. *Military Medicine*. 2015;180(4 Suppl):97-103.

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.

Feinberg RA, Wainer H. How much is enough? A reply to Sinharay, Haberman, and Boughton. *Educational Measurement: Issues and Practice*. 2015;34(3):9.

Feinberg RA, Wainer H. For want of a nail: why unnecessarily long tests may be impeding the progress of Western civilisation. *Significance*. 2015;12(1):16-21.

Furman G. Proving our worth: foundational literature supporting the standardized patient educational methodology. *ASPE eNews*. 2015;December 8.

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*. 2015;60(677-683).

Guernsey J. Perspectives on testing, responses to your questions: response. *CLEAR Exam Review*. 2015;25(1):30-32.

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. *Handbook of Test Development*. 2nd ed. New York, NY: Routledge; 2015.

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, NY: Routledge; 2015:3-18.

Lohr KM, Clauser A, Hess BJ, Gelber AC, Valeriano-Marcet J, Lipner RS, Haist SA, Hawley JL, Zirkle S, Bolster MB, American College of Rheumatology Committee on Rheumatology Training and Workforce Issues. Performance on the Adult Rheumatology In-Training Examination and relationship to outcomes on the Rheumatology Certification Examination. *Arthritis and Rheumatism*. 2015;67:3082-3090.

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;30:1307-12.

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

Peterson LN, Rusticus SA, Ross LP. Comparability of the national board of medical examiners comprehensive clinical science examination and a set of five clinical science subject examinations. *Academic Medicine*. 2015;90:684-690.

Raymond MR. Job analysis, practice analysis and the content of credentialing examinations. In: Lane S, Raymond MR, Haladyna TM, eds. *Handbook of Test Development*. 2nd ed. New York, NY: Routledge; 2015:144-164.

Saguil A, Dong T, Gingerich RJ, Swygert KA, LaRoche JS, Artino AR, Cruess DF, Durning SJ. Does the MCAT predict medical school and PGY-1 performance? *Military Medicine*. 2015;180(suppl 4):4-11.

Skorupski WP, Wainer H. The Bayesian flip: correcting the prosecutor's fallacy. *Significance*. 2015;August:16-20.

Small C, Land AM, Haist SA, Estrada CA, Snyder ED. Managing cognitive load to uncover an unusual cause of syncope: exercises in clinical reasoning. *Journal of General Internal Medicine*. Nov 5. [Epub ahead of print] ed. 2015.

Swygert KA, Williamson DM. Using performance tasks in credentialing tests. In: Lane S, Raymond MR, Haladyna TM, eds. *Handbook of Test Development*. 2nd ed. New York: Routledge; 2015:294-312.

Thompson BM, Haidet P, Borges NJ, Carchedi LR, Roman BJ, Townsend MH, Butler AP, Swanson DB, Anderson MP, Levine RE. Team cohesiveness, team size and team performance in team-based learning teams. *Medical Education*. 2015;49:379-385.

Wainer H. *Truth or Truthiness: Distinguishing Fact from Fiction by Learning to Think Like a Data Scientist*. New York: Cambridge University Press; 2015.

Wainer H. A review of D. Yan, A. A. Von Davier, & C. Lewis (2014). *Computerized Multistage Testing: Theory and Applications*. *Psychometrika*. 2015;80(1):259-261.

Wainer H, Friendly M, Millan-Martinez P. Graphs R Us: a discussion of Antony Unwin's *Graphical Data Analysis With R*. *Journal of Educational and Behavioral Statistics*. 2015;40:665-670.

Wainer H, Rubin DB. Causal inference and death. *Chance*. 2015;28(2):59-64.

---

## 2014 PUBLICATIONS

Anderson MB. Introduction. *Medical Education*. 2014;48:1103.

Anderson MB. Introduction. *Medical Education*. 2014;48:520-521.

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.

Feinberg RA, Wainer H. When can we improve subscores by making them shorter? The case against subscores with overlapping items. *Educational Measurement: Issues and Practice*. 2014;33(3):47-54.

Feinberg RA, Wainer H. A simple equation to predict a subscore's value. *Educational Measurement: Issues and Practice*. 2014;33(3):55-56.

Haist SA, Katsufakis PJ, Dillon GF. Basic science content in the USMLE Step 1--reply. *Journal of the American Medical Association*. 2014;311:1359-60.

Hicks PJ, Schwartz A, Clyman SG, Nichols DG. The Pediatrics Milestones: pursuit of a national system of workplace-based assessment through key stakeholder collaboration. *Academic Pediatrics*. 2014;14(suppl 2):10-12.

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.

LeBlanc KE, Muncie HL, LeBlanc LL. Hip fracture: diagnosis, treatment, and secondary prevention. *American Family Physician*. 2014;89:945-951.

Margolis MJ, Clauser BE. The impact of examinee performance information on judges' cut scores in modified Angoff standard-setting exercises. *Educational Measurement: Issues and Practice*. 2014;33(1):15-22.

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.

Raymond MR, Mee J, Haist SA, Young A, Dillon GF, Katsufakis PJ, McElhenney SM, Johnson D. 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).

Wainer H. Life follows art: gaming the missing data algorithm. *Chance*. 2014;27(2):56-57.

Wainer H. Musing about changes in the SAT: is the College Board getting rid of the bulldog? *Chance*. 2014;27(3):59-63.

Wainer H. Happiness and causal inference. *Chance*. 2014;27(4):61-64.

Wainer H. William Playfair (version 3). *StatProb: The Encyclopedia Sponsored by Statistics and Probability Societies*. <http://statprob.com/encyclopedia/WilliamPLAYFAIR.html>. 2014.

Wainer H. Cheating: Some ways to detect it badly. In: Kingston N, Clark AK, ed. *Test Fraud: Statistical Detection and Methodology*. New York: Routledge; 2014:8-20.

Wainer H. Book review: Licensed to Practice: the Supreme Court Defines the American Medical Profession. *Journal of Medical Regulation*. 2014;100(1):24.

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

Wainer H. On the crucial role of empathy in the design of communications: genetic testing as an example. *Chance*. 2014;27(1):45-50.

Wainer H. Medical illuminations: using evidence, visualization and statistical thinking to improve healthcare. Oxford UK: Oxford University Press; 2014.

---

## 2013 PUBLICATIONS

Anderson MB. Really good stuff: lessons learned through innovation in medical education. Introduction. *Medical Education*. 2013;47:513.

Anderson MB. Really good stuff: lessons learned through innovation in medical education. *Medical Education*. 2013;47:1117-1118.

Baldwin P. On mean-sigma estimators and bias. *British Journal of Mathematical and Statistical Psychology*. 2013;66:277-289.

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.

Cook R, Wainer H. Plotting evidence to affect social policy: guns, murders, life, death, and ignorance in contemporary America. *Chance*. 2013;26(2):38-44.

- 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 GF, 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;5:276-283.
- Galbraith RM. Got feedback? *Medical Education*. 2013;47:224-225.
- Gomella LG, Haist SA, Adams AG. Clinician's Pocket Drug Reference. 11th ed. New York, NY: McGraw-Hill; 2013.
- Gomella LG, Haist SA, Adams AG. Clinician's Pocket Drug Reference. 12th ed. New York, NY: McGraw-Hill; 2013.
- Gomella PT, McMullan M, Oettinger G, Schoenvetter D, Zielewicz J, Haist S. EMS Pocket Drug Guide. 2nd ed. New York, NY: McGraw-Hill; 2013.
- Haist SA, Katsufakis PJ, Dillon GF. The evolution of the United States Medical Licensing Examination (USMLE): enhancing assessment of practice-related competencies. *Journal of the American Medical Association*. 2013;310:2245-2246.
- Harik P, Baldwin P, Clauser BE. Comparison of automated scoring methods for a computerized performance assessment of clinical judgment. *Applied Psychological Measurement*. 2013;37:587-597.
- 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.

- Melnick DE. International assessment of medical students: should it matter anymore where the school is located? *Innovations in Global Medical & Health Education*. 2013;5.
- Raymond MR, Luecht RL. Licensure and certification testing. In: Geisinger K, ed. *Handbook of Testing in Assessment and Psychology*. Washington, DC: American Psychological Association; 2013:391-414.
- Schuwirth L, Colliver J, Gruppen L, Kreiter C, Mennin S, Onishi H, Pangaro L, Ringsted C, Swanson DB, van der Vleuten C, Wagner M. Research on assessment practices. In: McGaghie W, ed. *International Best Practices for Evaluation in the Health Professions*. London, UK: Radcliffe Publishing; 2013:59-75.
- Schuwirth L, Swanson DB. Standardised versus individualised assessment: related problems divided by a common language. *Medical Education*. 2013;47:627-631.
- Swanson DB, Marsh JL, Hurwitz S, DeRosa GP, Holtzman KZ, Bucak SD, Baker A, Morrison C. Utility of AAOS OITE scores in predicting ABOS Part I outcomes: AAOS exhibit selection. *Journal of Bone and Joint Surgery American*. 2013;95(12):e84.
- Swanson DB, van der Vleuten CP. Assessment of clinical skills with standardized patients: state of the art revisited. *Teaching and Learning in Medicine*. 2013;25(suppl 1):17-25.
- 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.
- Wainer H. How the rule of 72 can provide guidance to advance your wealth, weight, career, and gas mileage. *Chance*. 2013;26(3):47-48.
- Wainer H. Taking a chance: an interview with William F. Eddy and Stephen E. Fienberg. *Chance*. 2013;26(4):30-34.
- Wainer H, Clauser BE. Reflections on a too extreme idea. *Educational Psychology Review*. 2013;25:325-330.
- Wainer H, Harik P, Neter J. Stigler's law of eponymy and Marey's train schedule: did Serjev do it before Ibray, and what about Jules Petiet? *Chance*. 2013;26(1):53-56.
- 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 PUBLICATIONS



Anderson MB. Introduction; a peer-reviewed collection of reports on innovative approaches to medical education. *Medical Education*. 2012;46:1101.

Anderson MB. Introduction; a peer-reviewed collection of reports on innovative approaches to medical education. *Medical Education*. 2012;46:503.

Babcock B, Albano A, Raymond MR. Nominal weights mean equating: a method for very small samples. *Educational and Psychological Measurement*. 2012;72:608-628.

Barberia JA, Gomelle LG, Adams AG, Haist SA. Nurse's Pocket Drug Guide 2012. 8th ed. New York, NY: McGraw-Hill-Medical; 2012.

Cook R, Wainer H. A century and a half of moral statistics in the United Kingdom: variations on Joseph Fletcher's thematic maps. *Significance*. 2012;9(3):31-36.

Dauphinee WD, Anderson MB. Maturation (and déjà vu) comes to the research in medical education program at age 51. *Academic Medicine*. 2012;87:1307-1309.

Dillon GF. The importance of testing medical students' knowledge of what is least likely [letter]. *Academic Medicine*. 2012;87:1454.

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.

Hammer D, Anderson MB, Brunson WD, Grus C, Heun L, Holtman M, Mashima T, McGuinn K, Nunez L, Register S, Ross L, Ruffin A, Frost JG. Defining and measuring construct of interprofessional professionalism. *Journal of Allied Health*. 2012;41:e49-53.

Hubert L, Wainer H. *A Statistical Guide for the Ethically Perplexed*. Boca Raton, FL: Chapman and Hall/CRC; 2012.

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 equivalence of ratings for repeat examinees on a performance assessment for physician licensure. *Journal of Educational Measurement*. 2012;49:339-361.



Sales D, Sturrock A, Swanson DB. Machine markable knowledge testing. *Excellence in Medical Education*. 2012;12(3):23-29.

Scoles PV. The significance of significance: commentary on an article by Robert Grunfeld, MD, et al: "An assessment of musculoskeletal knowledge in graduating medical and physician assistant students and implications for musculoskeletal care providers". *The Journal of Bone and Joint surgery. American volume*. 2012;94:e28.

Sondheimer HM, Anderson MB. Introduction. In: *A Snapshot of the New and Developing Medical Schools in the United States and Canada*. Washington DC: Association of American Medical Colleges; 2012:3-6.

Swygert KA, Cuddy MM, Van Zanten M, Haist SA. 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.

Wainer H. Cheating: some ways to detect it badly. *Chance*. 2012;25(3):54-57.

Wainer H. How statistics rescued a damsel in distress. *NJEA Review*. 2012;85(5):16-19.

Wainer H. Moral statistics and the thematic maps of Joseph Fletcher. *Chance*. 2012;25(1):43-46.

Wainer H. More statistics: a contribution to one hundred great ideas for higher education. *Academic Questions*. 2012;25(4):69.

Wainer H. Piano virtuosos and the four-minute mile. *Significance*. 2012;9(2):28-29.

Wainer H. Review of Erich Lehmann's *Fisher, Neyman and the Creation of Classical Statistics*. *Journal of Educational Measurement*. 2012;49:335-338.

Wainer H. The survival of the fittists. *The American Scientist*. 2012;100:358-361.

Wainer H. Waiting for Achilles. *Chance*. 2012;25(4):50-51.

Wainer H. When nothing is not zero: a true saga of missing data, adequate yearly progress, and a Memphis charter school. *Chance*. 2012;25(2):49-51.

Wainer H, Savage S, McGrayne, Sharon Bertsch (2011). *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines and Emerged Triumphant from Two Centuries of Controversy*. New Haven, CT: Yale University Press. Book review. *Journal of Educational Measurement*. 2012;49:214-219.

---

## 2011 PUBLICATIONS

Anderson MB. A peer-reviewed collection of reports on innovative approaches to medical education. *Medical Education*. 2011;45:1131-1132.

Anderson MB. Introduction. *Medical Education*. 2011;45:1133.

Baldwin P. A strategy for developing a common metric in item response theory when parameter posterior distributions are known. *Journal of Educational Measurement*. 2011;48:1-11.

Baldwin P. Book review: Bayesian Item Response Modeling: Theory and Applications. *Journal of Educational Measurement*. 2011;48:357-359.

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

Barberio JA, Gomella LG, Adams AG, Haist SA. Nurse's Pocket Drug Guide 2011. 7th ed. New York NY: McGraw-Hill; 2011.

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

De Champlain AF, Grabovsky I, Scoles PV, Pannizzo L, Winward ML, Dermine A, Himpens B. Collecting evidence of content validity for the International Foundations of Medicine examination: an expert-based judgmental approach. *Teaching and Learning in Medicine*. 2011;23:144-147.

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

Feinberg RA, Wainer H. Extracting sunbeams from cucumbers. *Journal of Computational and Graphical Statistics*. 2011;20:793-810.

Gomella LG, Haist SA, Adams AG. Clinician's Pocket Drug Reference. 9th ed. New York NY: McGraw-Hill; 2011.

Holmboe ES, Ward DS, Reznick RK, Katsufakis PJ, Leslie KM, Patel VL, Ray DD, Nelson EA. Faculty development in assessment: the missing link in competency-based medical education. *Academic Medicine*. 2011;86:460-467.

Kahraman N, Thompson T. Relating unidimensional IRT parameters to a multidimensional response space: a review of two alternative projection IRT models for scoring subscales. *Journal of Educational Measurement*. 2011;48:146-164.

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

Mazor K, Holtman MC, Shchukin Y, Mee JM, Katsufrakis PJ. The relationship between direct observation, knowledge, and feedback: results of a national survey. *Academic Medicine*. 2011;86(suppl 10):63-67.

Melnick DE. Commentary: balancing responsibility to patients and responsibility to aspiring physicians with disabilities. *Academic Medicine*. 2011;86:674-676.

Norcini JJ, Anderson MB, Bollela V, Burch V, Costa MJ, Duvivier R, Galbraith RM, Hays R, Kent A, Perrott V, Roberts T. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. *Medical Teacher*. 2011;33:206-214.

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(suppl 10):59-62.

Richmond M, Canavan C, Holtman MC, Katsufrakis PJ. Feasibility of implementing a standardized multisource feedback program in the graduate medical education environment. *Journal of Graduate Medical Education*. 2011;3:511-516.

Schuwirth L, Colliver J, Gruppen L, Kreiter C, Mennin S, Onishi H, Pangaro L, Ringsted C, Swanson DB, van Der Vleuten C, Wagner-Menghin M. Research in assessment: Consensus statement and recommendations from the Ottawa 2010 Conference. *Medical Teacher*. 2011;33:224-233.

Sinharary S, Haberman S, Wainer H. Do adjusted subscores lack validity? Don't blame the messenger. *Educational and Psychological Measurement*. 2011;71:789-797.

Wainer H. Value-added models to evaluate teachers: a cry for help. *Chance*. 2011;24(1):11-13.

Wainer H. The first step toward wisdom. *Chance*. 2011;24(2):60-61.

Wainer H. How much is tenure worth? *Chance*. 2011;24(3):54-57.

Wainer H. *Uneducated Guesses : Using Evidence to Uncover Misguided Education Policies*. Princeton NJ: Princeton University Press; 2011.

Wainer H. A remarkable horse: an inquiry into the accuracy of medical predictions. *Chance*. 2011;24(4):55-57.

Wainer H. The Pleasures of Statistics: The Autobiography of Frederick Mosteller. Book review. *Psychometrika*. 2011;76:155-157.

Wainer H. Some reflections on data display and evidence. *Journal of Computational and Graphical Statistics*. 2011;20:8-15.

Wainer H. How should we screen for breast cancer: using evidence to make medical decisions. *Significance*. 2011;8:28-30.

Wainer H. A profile of Karl G. Joreskog. *Journal of Educational and Behavioral Statistics*. 2011;36:403-412.

Wainer H. Waiting for Achilles. *Newark Star Ledger*. Newark, NJ; 2011;2012(January 18):Op Ed Essay.

Wainer H. Assessing teachers from student scores: on the viability and fairness of value-added models for STEM Teachers; Op Ed. *US News & World Report*. 2011(January 18).

Wainer H, Hubert L. A statistical guide for the ethically perplexed. In: Panter AT, Sterba S, ed. *Handbook of Ethics in Quantitative Methodology*. New York: Routledge; 2011:61-124.

Wainer H, Hubert L. Assessing long-term risk with short-term data. *Significance*. 2011;8:170-171.

---

## 2010 PUBLICATIONS

Barberio JA, Gomella LG, Adams AG, Haist SA. *Nurse's Pocket Drug Guide 2010*. 6th ed. New York, NY: McGraw-Hill; 2010.

Canavan C, Holtman MC, Richmond M, Katsufakis PJ. The quality of written comments on professional behaviors in a developmental multisource feedback program. *Academic Medicine*. 2010;85(Suppl 10):S106-S109.

- Clauser BE, Margolis MJ, Holtman MC, Katsuftrakis PJ, Hawkins RE. Validity considerations in the assessment of professionalism. *Advances in Health Sciences Education: Theory and Practice*. 2010;17:165-181.
- Cromley JG, Snyder-Hogan LE, Luciw-Dubas UA. Cognitive activities in complex science text and diagrams. *Contemporary Educational Psychology*. 2010;35:59-74.
- De Champlain AF, Cuddy MM, Scoles PV, Brown M, Swanson DB, Holtzman KZ, Butler A. Progress testing in clinical science education: results of a pilot project between the National Board of Medical Examiners and a U.S. medical school. *Medical Teacher*. 2010;32:503-508.
- 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.
- Hawkins RE, Margolis MJ, Durning SJ, Norcini JJ. Constructing a validity argument for the mini-clinical evaluation exercise: a review of the research. *Academic Medicine*. 2010;85:1453-1461.
- Karnieli-Miller O, Vu TR, Holtman MC, Clyman SG, Inui TS. Medical students' professionalism narratives: a window on the informal and hidden curriculum. *Academic Medicine*. 2010;85:124-133.
- Katsuftrakis PJ, Nussbaum MRH. Adolescent sexuality. In: South-Paul J, Matheny S, Lewis E, eds. *Current Diagnosis & Treatment in Family Medicine*. 3rd ed. New York, NY: Lange Medical Books/McGraw-Hill; 2010.
- Katsuftrakis PJ, White TD. Caring for lesbian, gay, bisexual, and transgender patients. In: South-Paul J, Matheny S, Lewis E, eds. *Current Diagnosis & Treatment in Family Medicine*. 3rd ed. New York, NY: Lange Medical Books/McGraw-Hill; 2010.
- Katsuftrakis PJ, Workowski KG. Sexually transmitted diseases. In: South-Paul J, Matheny S, Lewis E, eds. *Current Diagnosis & Treatment in Family Medicine*. 3rd ed. New York, NY: Lange Medical Books/McGraw-Hill; 2010.
- Keller LA, Clauser BE, Swanson DB. Using multivariate generalizability theory to assess the effect of content stratification on the reliability of a performance assessment. *Advances in Health Sciences Education: Theory and Practice*. 2010;15:717-733.
- Langer MM, Swanson DB. Practical considerations in equating progress tests. *Medical Teacher*. 2010;32:509-512.

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(suppl 10):93-97.

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(suppl 10):98-101.

Ramsay JO, Wainer H. Inside-out plots. *Chance*. 2010;23(3):57-62.

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.

Raymond MR, Luciw-Dubas UA. The second time around: accounting for retest effects on oral examinations. *Evaluation & the Health Professions*. 2010;33:386-403.

Raymond MR, Nagy P. Developing and verifying the psychometric integrity of the certification examination for imaging informatics professionals. *Journal of Digital Imaging*. 2010;23:241-245.

Rosner MH, Berns JS, Parker M, Tolwani A, Bailey J, DiGiovanni S, Lederer E, Norby S, Plumb TJ, Qian Q, Yeun J, Hawley JL, Owens S, , ASN In-Training Examination Committee. Development, implementation, and results of the ASN in-training examination for fellows. *Clinical Journal of the American Society of Nephrology*. 2010:328-334.

Schmidt W. From Wireframes to Code, Part 1. *UX Matters*. 2010(December 20).

Subhiyah RG, Boyce JR. North American Veterinary Licensing Examination pacing study. *Journal of Veterinary Medical Education*. 2010;37:377-382.

Swanson DB, Holtzman KZ, Butler A, Langer MM, Nelson MV, Chow JWM, Fuller R, Patterson JA, Boohan M. Collaboration across the pond: The multi-school progress testing project. *Medical Teacher*. 2010;32:480-485.

Swanson DB, Holtzman KZ, Butler A, The Case Western Reserve University School of Medicine Cumulative Achievement Testing Study Group. Cumulative achievement testing: Progress testing in reverse. *Medical Teacher*. 2010;32:516-520.

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(suppl 10):89-92.

Tarassenko YN, Wackerbarth SB, Love MM, Joyce JM, Haist SA. Colorectal Cancer Screening: Patients' and Physicians' Perspectives on Decision-Making Factors. *Journal of Cancer Education*. 2010;27:65-70.

Wainer H. Preface. In: *Semiology of Graphics: Diagrams, Networks, Maps*. Redlands CA: Esri Press; 2010:xi-xii.

Wainer H. Pies, spies, roses, lines and symmetries. *Chance*. 2010;23(4):58-61.

Wainer H. 14 conversations about 3 things. *Journal of Educational and Behavioral Statistics*. 2010;35:5-25.

Wainer H. Exams and disabilities. *Princeton Alumni Weekly*. 2010;110(7):11-12.

Wainer H. Schroedinger's cat and the conception of probability in item response theory. *Chance*. 2010;23(1):53-56.

Wainer H. Commentary on the graphic displays in the 2008 National Healthcare Quality Report and state snapshots. *Chance*. 2010;23(2):47-53.

Wainer H, Bradlow E, Wang X. Detecting DIF: many paths to salvation. *Journal of Educational and Behavioral Statistics*. 2010;35:489-493.

Wang X, Baldwin SG, Bradlow E, Wainer H, Reeve B, Smith A, Bellizzi K, Baumgartner K. Using testlet response theory to analyze data from a survey of attitude change among breast cancer survivors. *Statistics in Medicine*. 2010;29:2028-204.

---



## 2000 TO 2009

### **Simulations, Psychometrics, and Standard Setting: 2000s**

In the first decade of the 2000s, researchers published on Objective Structured Clinical Examinations (OSCEs), on refinements to computerized case simulations, and on new statistical methods to improve exam scoring.

## 2009 PUBLICATIONS

Baldwin P, Bernstein J, Wainer H. Hip psychometrics. *Statistics in Medicine*. 2009;28:2277-92.

Baldwin P, Wainer H. A little ignorance: how statistics rescued a damsel in distress. *Chance*. 2009;22(3):51-55.

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*. 2009;4:35-42.

Clauser BE, Balog K, 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.

Clauser BE, Harik P, Margolis MJ, McManus IC, Mollon A, Chis L, Williams S. An empirical examination of the impact of group discussion and examinee performance information on judgments made in the Angoff standard-setting procedure. *Applied Measurement in Education*. 2009;22:1-21.

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

Griffith CH, Wilson JF, Haist SA, Albritton TA, Bognar BA, Cohen SJ, Hoesley CJ, Fagan MJ, Ferenchick GS, Pryor OW, Friedman E, Harrell HE, Hemmer PA, Houghton BL, Kovach R, Lambert DR, Loftus TH, Painter TD, Udden MM, Watkins RS, Wong RY. Internal medicine clerkship characteristics associated with enhanced student examination performance. *Academic Medicine*. 2009;84:895-901.

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(suppl 10):79-82.

Hauer KE, Ciccone AL, Henzel TR, Katsufakis PJ, Miller SH, Norcross WA, Papadakis MA, Irby DM. Remediation of the deficiencies of physicians across the continuum from medical school to practice: a thematic review of the literature. *Academic Medicine*. 2009;84:1822-1832.

Hawkins RE, Katsufakis PJ, Holtzman MC, Clauser BE. Assessment of medical professionalism: who, what, when, where, how, and ... why? *Medical Teacher*. 2009;31:348-361.

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(suppl 10):90-93.

Kahraman N, De Boeck P, Janssen R. Modeling DIF in complex response data using test design strategies. *International Journal of Testing*. 2009;9:151-166.

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(suppl 10):83-85.

Raymond MR, Neustel S, Anderson D. Same-form retest effects on credentialing examinations. *Educational Measurement: Issues and Practice*. 2009;28(2):19-27.

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(suppl 10):21-24.

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(suppl 10):1-4.

Wainer H. A centenary celebration for Will Burtin: a pioneer of scientific visualization. *Chance*. 2009;22(1):51-55.

Wainer H, Larsen M. Pictures at an exhibition. *Chance*. 2009;33(2):46-47.

Wainer H, Robinson DH. Profiles in courage: Linda S. Gottfredson. *Journal of Educational and Behavioral Statistics*. 2009;34:395-427.

Wells CS, Baldwin S, Hambleton RK, Sireci SG, Karatonis A, Jirka S. Evaluating score equity assessment for state NAEP. *Applied Measurement in Education*. 2009;22:394-408.

Winward ML, De Champlain AF, Grabovsky I, Scoles PV, Swanson DB, Holtzman KZ, Pannizzo L, Sousa N, Costa ML. Gathering evidence of external validity for the Foundations of Medicine Examination: a collaboration between the National Board of Medical Examiners and the University of Minho. *Academic Medicine*. 2009;84(suppl 10):116-119.

---

## 2008 PUBLICATIONS

Berg K, Winward M, Clauser BE, Veloski JA, Berg D, Dillon GF, Veloski JJ. The relationship between performance on a medical school's clinical skills assessment and USMLE Step 2 CS. *Academic Medicine*. 2008;83(10 Suppl):S37-S40.

Boulet JR, Van Zanten M, De Champlain AF, Hawkins RE, Peitzman SJ. Checklist content on a standardized patient assessment: an ex post facto review. *Advances in Health Sciences Education*. 2008;13:59-69.

Clauser BE. A Review of the EDUG Software for Generalizability Analysis [book review]. *International Journal of Testing*. 2008;8:296-301.

Clauser BE. War, enmity, and statistical tables. *Chance*. 2008;21(4):6-11.

Clauser BE, Harik P, Margolis MJ, Mee JM, Swygert KA, Rebbecchi T. The generalizability of documentation scores from the USMLE Step 2 Clinical Skills Examination. *Academic Medicine*. 2008;83(10 Suppl):S41-S44.

Clauser BE, Margolis MJ, Swanson DB. Issues of validity and reliability for assessments in medical education. In: Holmboe ES, Hawkins RE, ed. *A Practical Guide to the Evaluation of Clinical Competence*. Philadelphia, PA: Mosby; 2008:10-23.

Cuddy MM, Swanson DB, Clauser BE. A multilevel analysis of examinee gender and USMLE Step 1 Performance. *Academic Medicine*. 2008;83(10):S58-S62.

- Furman GE. The role of standardized patient and trainer training in quality assurance for a high-stakes clinical skills examination. *Kaohsiung Journal of Medical Science*. 2008;24:651-5.
- Galbraith RM, Hawkins RE, Holmboe ES. Making self-assessment more effective. *Journal of Continuing Education in the Health Professions*. 2008;28:20-4.
- Gilliland WR, La Rochelle J, Hawkins RE, Dillon GF, Mechaber AJ, Dyrbye L, Papp KK, Durning SJ. Changes in clinical skills education resulting from the introduction of the USMLE Step 2 Clinical Skills (CS) examination. *Medical Teacher*. 2008;30:325-7.
- Haist SA, Lineberry MJ, Griffith CH, Hoellein AR, Talente GM, Wilson JF. Sexual history inquiry and HIV counseling: improving clinical skills and medical knowledge through an interactive workshop utilizing standardized patients. *Advances in Health Sciences Education Theory and Practice*. 2008;13:427-434.
- Hawkins RE, Boulet JR. Direct observation: standardized patients. In: Holmboe ES, Hawkins RE, ed. *Practical Guide to the Evaluation of Clinical Competence*. Philadelphia, PA: Mosby; 2008:102-118.
- Hawkins RE, Holmboe ES. Constructing an evaluation system for an educational program. In: Holmboe ES, Hawkins RE, ed. *Practical Guide to the Evaluation of Clinical Competence*. Philadelphia, PA: Mosby; 2008:216-237.
- Hawkins RE, Swanson DB. Using written examinations to assess medical knowledge and its application. In: Holmboe ES, Hawkins RE, ed. *Practical Guide to the Evaluation of Clinical Competence*. Philadelphia, PA: Mosby; 2008.
- Holmboe ES, Hawkins RE. *Practical Guide to the Evaluation of Clinical Competence*. Philadelphia, PA: Mosby; 2008.
- Holtman MC. A theoretical sketch of medical professionalism as a normative complex. *Advances in Health Sciences Education: Theory and Practice*. 2008;13:233-245.
- Kahraman N, Clauser BE, Margolis MJ. A comparison of alternative item weighting strategies on the data gathering component of a clinical skills performance assessment. *Academic Medicine*. 2008;83(suppl 10):72-75.
- Lee G, Velleman P, Wainer H. Giving the finger to dating services. *Chance*. 2008;21(3):59-61.
- Ling Y, Swanson DB, Holtzman KZ, Bucak SD. Retention of basic science information by senior medical students. *Academic Medicine*. 2008;83(suppl 10):82-85.

- Lockyer JM, Clyman SG. Multisource feedback (360-degree evaluation). In: Holmboe ES, Hawkins RE, ed. *Practical Guide to the Evaluation of Clinical Competence*. Philadelphia, PA: Mosby; 2008:75-84.
- Mazmanian PE, Galbraith RM, Miller SH, Schyve PM, Kopelow M, Thompson JN, Aparicio A, Davis DA, Kahn NB. Accreditation, certification, and licensure: How six general competencies are influencing medical education and patient care. *Journal of Medical Licensure and Discipline*. 2008;94(1):8-14.
- Mazor KM, Canavan CT, Farrell M, Margolis MJ, Clauser BE. Collecting validity evidence for an assessment of professionalism: findings from think-aloud interviews. *Academic Medicine*. 2008;83(suppl 10):9-12.
- Norcini JJ, Holmboe ES, Hawkins RE. Evaluation challenges in the era of outcomes-based education. In: Holmboe ES, Hawkins RE, ed. *A Practical Guide to the Evaluation of Clinical Competence*. Philadelphia, PA: Mosby; 2008:1-9.
- Ramineni C, Clauser BE, Harik P, Swanson DB. Contrast effects in the USMLE Step 2 Clinical Skills Examination. *Academic Medicine*. 2008;83(suppl 10):45-48.
- Savage S, Wainer H. Until proven guilty: false positives and the war on terror. *Chance*. 2008;21(1):55-58.
- Scoles PV. Comprehensive review of the USMLE. *Advances in Physiology Education*. 2008;32(2):109-10.
- Swanson DB, Holtzman KZ, Albee K. Measurement characteristics of content-parallel single-best-answer and extended-matching questions in relation to number and source of options. *Academic Medicine*. 2008;83(suppl 10):21-24.
- Wackerbarth SB, Peters JC, Haist SA. Modeling the decision to undergo colorectal cancer screening: insights on patient preventive decision making. *Medical Care*. 2008;46(9 suppl 1):17-22.
- Wainer H. Why is a raven like a writing desk? *American Scientist*. 2008;96:446-449.
- Wainer H. Improving graphic displays by controlling creativity. *Chance*. 2008;21(2):46-52.
- Wang X, Bradlow E, Wainer H, Muller E. A Bayesian method for studying DIF: A cautionary tale filled with surprises and delights. *Journal of Educational and Behavioral Statistics*. 2008;33:363-84.

---

## 2007 PUBLICATIONS

- Baldwin SG. Book review of Wainer H, et al. Testlet response theory and its applications. *Journal of Educational and Behavioral Statistics*. 2007;32:333-6.

Braun H, Wainer H. Value-added modeling. In: Rao CR, Sinharay S, ed. *Handbook of Statistics* 26. Amsterdam, The Netherlands: Elsevier; 2007:867-892.

Clauser BE. The life and labors of Francis Galton: a review of four recent books about the father of behavioral statistics. *Journal of Educational and Behavioral Statistics*. 2007;32:440-444.

Cuddy MM, Swanson DB, Clauser BE. A multilevel analysis of the relationships between examinee gender and United States Medical Licensing Exam (USMLE) Step 2 CK content area performance. *Academic Medicine*. 2007;82(10 Suppl):S89-S93.

De Champlain AF, Cuddy MM, LaDuca A. Examining contextual effects in a practice analysis: an application of dual scaling. *Educational Measurement: Issues and Practice*. 2007;26(3):3-10.

Hess B, Subhiyah RG, Giordano C. Convergence between cluster analysis and the Angoff method for setting minimum passing scores on credentialing examinations. *Evaluation in the Health Professions*. 2007;30:362-375.

Hoadley D, Wang S, Wang N. Construct equivalence of a national certification examination that uses dual languages and audio assistance. *International Journal of Testing*. 2007;7:255-268.

Holtman MC. Disciplinary careers of drug-impaired physicians. *Social Sciences and Medicine*. 2007;64:543-553.

Katsufraakis PJ. Caring for gay, lesbian, bisexual & transgender patients. In: South-Paul JE, Matheny SC, Lewis EL, eds. *Current Diagnosis and Treatment in Family Medicine*. 2nd ed. New York, NY: McGraw-Hill; 2007:664-673.

Katsufraakis PJ, Nusbaum MRH. Adolescent sexuality. In: South-Paul JE, Matheny SC, Lewis EL, eds. *Current Diagnosis and Treatment in Family Medicine*. 2nd ed. New York, NY: McGraw-Hill; 2007:124-132.

Katsufraakis PJ, Workowski KA. Sexually transmitted diseases. In: South-Paul JE, Matheny SC, Lewis EL, eds. *Current Diagnosis and Treatment in Family Medicine*. 2nd ed. New York, NY: McGraw-Hill; 2007:146-164.

Mazor K, Clauser BE, Holtman MC, Margolis MJ. Evaluation of missing data in an assessment of professional behaviors. *Academic Medicine*. 2007;82(suppl 10):44-47.

McGaha AL, Garrett E, Jobe AC, Nalin P, Newton WP, Pugno PA, Kahn NB. Responses to medical students' frequently asked questions about family medicine. *American Family Physician*. 2007;76:99-106.

Ramineni C, Harik P, Margolis MJ, Clauser BE, Swanson DB, Dillon GF. Sequence effects in the United States Medical Licensing Examination (USMLE) Step 2 Clinical Skills (CS) Examination. *Academic Medicine*. 2007;82(suppl 10):101-104.

van Zanten M, Boulet JR, McKinley DW, De Champlain AF, Jobe AC. Assessing the communication and interpersonal skills of graduates of international medical schools as part of the United States Medical Licensing Exam (USMLE) Step 2 Clinical Skills (CS) Exam. *Academic Medicine*. 2007;82(suppl 10):65-68.

Wainer H. A psychometric cicada: Educational Measurement returns. Book review. *Educational Researcher*. 2007;36:485-6.

Wainer H. Taking a chance: an interview with William F. Eddy and Stephen E. Fienberg. *Chance*. 2007;20(4):33-9.

Wainer H. Science and the SAT (letter). *Princeton Alumni Weekly*. 2007;8(4).

Wainer H. L'equazione piu pericolosa. *Le Scienze*. 2007(470):80-87.

Wainer H. Improving data displays: ours and the media's. *Chance*. 2007;20(3):8-15.

Wainer H. The most dangerous equation. *American Scientist*. 2007;95:249-256.

Wainer H. Insignificant is not zero: rescoring the SAT as an example. *Chance*. 2007;20(1):55-58.

Wainer H. Galton's normal is too platykurtic. *Chance*. 2007;20(2):57-58.

Wainer H, Bradlow ET, Wang X. Testlet Response Theory and Its Applications. New York: Cambridge University Press; 2007.

Wainer H, Gelman A. A catch-22 in assigning primary delegates. *Chance*. 2007;20(4):6-7.

Wainer H, Robinson DH. Profiles in research: Fumiko Samejima. *Journal of Educational and Behavioral Statistics*. 2007;32:206-222.

Wainer H, Robinson DH. Profiles in Research: Roderick P. McDonald. Interview by Howard Wainer and Daniel H. Robinson. *Journal of Educational and Behavioral Statistics*. 2007;32:315-32.

Wainer H, Robinson DH. Profiles in Research: Susan E. Embretson. Interview by Howard Wainer and Daniel H. Robinson. *Journal of Educational and Behavioral Statistics*. 2007;32:431-439.



## 2006 PUBLICATIONS

Boulet JR, Swanson DB, Cooper RA, Norcini JJ, McKinley D. A Comparison of the characteristics and examination performances of US and non-US citizen international medical graduates who sought ECFMG certification: 1995-2004. *Academic Medicine*. 2006;81(10 Suppl):S116-S119.

Clauser BE, Harik P, Margolis MJ. A multivariate generalizability analysis of data from a performance assessment of physicians' clinical skills. *Journal of Educational Measurement*. 2006;43:173-91.

Clauser BE, Margolis MJ. Item Generation for Test Development [book review]. *International Journal of Testing*. 2006;6:310-4.

Clauser BE, Margolis MJ, Case SM. Testing for licensure and certification in the professions. In: Brennan RL, ed. *Educational Measurement*. 4th ed. Westport, CT: American Council on Education/Praeger; 2006:701-731.

Cuddy MM, Swanson DB, Dillon GF, Holtman MC, Clauser BE. A multi-level analysis of selected examinee characteristics and USMLE Step 2 Clinical Knowledge performance: revisiting old findings and asking new questions. *Academic Medicine*. 2006;81(10):S103-S107.

De Champlain AF, Sample L, Dillon GF, Boulet JR. Modeling longitudinal performances on the United States Medical Licensing Examination and the impact of sociodemographic covariates: an application of survival data analysis. *Academic Medicine*. 2006;81(10 Suppl):S108-S111.

De Champlain AF, Swygert KA, Swanson DB, Boulet JR. Assessing the underlying structure of the United States Medical Licensing Examination Step 2 test of clinical skills using confirmatory factor analysis. *Academic Medicine*. 2006(10 Suppl):S17-S20.

Galbraith RM, Holtman MC, Clyman SG. The use of assessment to reinforce competency in patient safety. *Quality and Safety in Healthcare*. 2006;15 suppl 1:i30-i33.

Gilliland WR, Pangaro LN, Downing S, Hawkins RE, Omori DM, Marks ES, Adamo G, Bordage G. Applied research: standardized versus real hospitalized patients to teach history-taking and physical examination skills. *Teaching and Learning in Medicine*. 2006;18:188-195.

Hallock JA, Melnick DE, Thompson JN. The Step 2 Clinical Skills Examination. *Journal of the American Medical Association*. 2006;295:1123-1124.

Hannon L, Cuddy MM. Neighborhood ecology and drug dependence mortality: an analysis of New York City census tracts. *The American Journal of Drug and Alcohol Abuse*. 2006;32:453-463.

Harik P, Clauser BE, Grabovsky I, Margolis MJ, Dillon GF, Boulet JR. Relationship among subcomponents of the USMLE Step 2 Clinical Skills Examination, the Step 1, and the Step 2 Clinical Knowledge examinations. *Academic Medicine*. 2006;81(suppl 10):21-24.

Henzel TR, Ciccone AL, Cain F, Clothier CA, Hawkins RE. Implementing assessment of practicing physicians: the development and benefits of a collaborative model. *Journal of Medical Licensure and Discipline*. 2006;92(4):31-39.

LaDuca A. Commentary: a closer look at task analysis: reactions to Wang, Schnipke, and Witt. *Educational Measurement: Issues and Practices*. 2006;25(2):31-33.

Margolis MJ, Clauser BE. A regression-based procedure for automated scoring of a complex medical performance assessment. In: Williamson DM, ed. *Automated Scoring of Complex Tasks in Computer-Based Testing*. Mahwah, NJ: Lawrence Erlbaum Associates; 2006:123-167.

Margolis MJ, Clauser BE, Cuddy MM, Ciccone AL, Mee JM, Harik P, Hawkins RE. Use of the Mini-CEX to rate examinee performance on a multiple-station clinical skills examination: a validity study. *Academic Medicine*. 2006;81(suppl 10):56-60.

McKinley DW, Boulet JR, Swanson DB, Swygert KA, Scott CL. Effects of case characteristics on encounter time in a high-stakes standardized patient examination. *Academic Medicine*. 2006;81(suppl 10):61-64.

Melnick DE. From defending the walls to improving global medical education: fifty years of collaboration between the ECFMG and the NBME. *Academic Medicine*. 2006;81(suppl 12):30-35.

Melnick DE. An examination of clinical skills in the United States Licensing Examination™ (USMLE™). *AAMC Reporter*. 2006;15(7).

Melnick DE, Clauser BE. Computer-based testing for professional licensing and certification of health professionals. In: Bartram D, Hambleton RJ, ed. *Computer-based Testing and the Internet: Issues and Advances*. London, UK: John Wiley & Sons; 2006:163-186.

Swanson DB, Holtzman KZ, Albee K, Clauser BE. Psychometric characteristics and response times for content-parallel extended-matching and one-best-answer items in relation to number of options. *Academic Medicine*. 2006;81(10 Suppl):S52-S55.

Wainer H. Book review: L Wilkinson (2005). The grammar of graphics, 2d ed. *Psychometrika*. 2006;71:603.



Wainer H. Book review: Richard P. Phelps, ed. *Defending Standardized Testing*. *Journal of Educational Measurement*. 2006;43:77-84.

Wainer H. Chance Conversation with Judith Tanur. *Chance*. 2006;19(4):52-57.

Wainer H. Using graphs to make the complex simple: the Medicare drug plan as an example. *Chance*. 2006;19(2):55-56.

Wainer H. On model-based inferences: A fitting tribute to a giant. In: Hantula D, ed. *Advances in Social and Organizational Psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates; 2006:61-73.

Wainer H, Brown L. Three statistical paradoxes in the interpretation of group differences: illustrated with medical admission and licensing data. In: Roao CR, Sinharay S, ed. *Handbook of Statistics* 26. Amsterdam: Elsevier; 2006:893-918.

Wainer H, Brown LM, Bradlow ET, Wang WP, Skorupski WP. An application of testlet response theory in the scoring of a complex certification exam. In: Williamson DM, ed. *Automated Scoring of Complex Tasks in Computer-Based Testing*. Mahwah, NJ: Lawrence Erlbaum Associates; 2006:169-199.

Wainer H, Gessaroli ME, Verdi M. Finding what is not there through the unfortunate binning of results: The Mendel effect. *Chance*. 2006;19(2):49-52.

Wainer H, Robinson D. Profiles in research: Julian Cecil Stanley. *Journal of Educational and Behavioral Statistics*. 2006;31:231-240.

Wainer H, Velleman PF. Statistical graphics: A guidepost for scientific discovery. In: Green JL, Camilli G, Elmore PB, eds. *Complementary methods for research in education*. 3rd ed. Washington, D.C: American Educational Research Association; 2006:605-621.

Wainer H, Zwerling HL. Evidence that smaller schools do not improve student achievement. *Phi Delta Kappan*. 2006;88:300-303.

Wallach PM, Crespo LM, Holtzman KZ, Galbraith RM, Swanson DB. Use of a committee review process to improve the quality of course examinations. *Advances in Health Sciences Education*. 2006;11:61-68.

---

## 2005 PUBLICATIONS

Babcox E. Commentary [an excerpt from Nicholas Nickleby]. *Academic Medicine*. 2005;80:456-457.

Clauser BE, Margolis MJ. Free response data scoring. In: Everitt BS, Howell DC, ed. *Encyclopedia of Statistics in Behavioral Science*. Chichester, UK: John Wiley & Sons; 2005:668-673.

De Champlain AF, Scoles PV, Holtzman KZ, Angelucci K, Flores MC, Mendoza E, Martin M, De Calvo OL. Assessing the reliability and validity of a residency selection process examination: a preliminary study between the National Board of Medical Examiners and the University of Panama Faculty of Medicine. *Teaching and Learning in Medicine*. 2005;17:14-20.

Dillon GF, Scoles PV. An examination of clinical skills in the United States Medical Licensing Examination (USMLE). *ACGME Bulletin*. 2005(December):16.

Fletcher EA. The National Board of Medical Examiners subject examination update. *ADMSEP Association of Directors of Medical Student Education in Psychiatry Newsletter*. 2005;17(1):5.

Galbraith RM, Clyman SG. Emerging trends in the U.S. physician workforce: implications for licensure and professional standards. *Journal of Medical Licensure and Discipline*. 2005;91(1):14-20.

Gessaroli ME, DeChamplain AF. Assessment of test dimensionality. In: Everitt BS, Howell DC, ed. *Encyclopedia of Statistics in Behavioral Science*. Chichester, UK: John Wiley & Sons; 2005:2014-2021.

Hammoud MM, Cox SM, Goff B, Goepfert A, Butler A, Swanson DB, Holtzman KZ, Allbee K, Katz NT, Erickson SS. The essential elements of undergraduate medical education in obstetrics and gynecology: a comparison of the Association of Professors of Gynecology and Obstetrics Medical Student Educational Objectives and the National Board of Medical Examiners Subject Examination. *American Journal of Gynecology and Obstetrics*. 2005;193:1773-1779.

Hawkins RE, Swanson DB, Dillon GF, Clauser BE, King AM, Scoles PV, Whelan GP, Burdick WP, Boulet JR, Homan AG. The introduction of clinical skills assessment into the United States Medical Licensing Examination (USMLE): A description of USMLE Step 2 Clinical Skills (CS). *Journal of Medical Licensure and Discipline*. 2005;91(3):21-5.

Playfair W, Wainer H, Spence I. *The Commercial and Political Atlas and Statistical Breviary*. New York, NY: Cambridge University Press; 2005.

Scoles PV. USMLE Update. *ADMSEP Association of Directors of Medical Student Education in Psychiatry Newsletter*. 2005;17(1):4-5.

Stern DT, Ben-David MF, De Champlain AF, Hodges B, Wojtczak A, Schwarz MR. Ensuring global standards for medical graduates: a pilot study of international standard-setting. *Medical Teacher*. 2005;27:207-13.

Swanson DB, Holtzman KZ, Clauser BE, Sawhill AJ. Psychometric characteristics and response times for one-best-answer questions in relation to number and source of options. *Academic Medicine*. 2005;80(suppl 10):93-96.

Swanson DB, Lazarus CJ, Dillon GF, Melnick DE. Coverage of the behavioral and social sciences on the United States Medical Licensing Examination (USMLE). *Annals of Behavioral Science and Medical Education*. 2005;11:30-36.

Swygert KA. Book review: Automated Essay Scoring: A Cross Disciplinary Perspective. *Journal of Educational Measurement*. 2005;42:215-218.

Wainer H. Chance Conversations: Former director of the U.S. Census Bureau gets personal. *Chance*. 2005;18(4):48-51.

Wainer H. Reflections: shopping for colleges when what we know ain't. *Journal of Consumer Research*. 2005;32:337-42.

Wainer H. *Graphic Discovery: A Trout in the Milk and Other Visual Adventures*. Princeton, NJ: Princeton University Press; 2005.

Wainer H. Graphical presentation of longitudinal data. In: Everitt BS, Howell DC, ed. *Encyclopedia of Statistics in Behavioral Science*. Chichester, UK: John Wiley & Sons; 2005:762-772.

Wainer H. Nonrandom samples. In: Everitt BS, Howell DC, ed. *Encyclopedia of Statistics in Behavioral Science*. Chichester, UK: John Wiley & Sons; 2005:1430-1433.

Wainer H. Visual Revelations: Old Mother Hubbard and the United Nations: an adventure in exploratory data analysis. *Chance*. 2005;18(3):38-45.

Wainer H. Visual Revelations: stumbling on the path toward the visual communication of complexity. *Chance*. 2005;18(2):53-4.

Wainer H, Clauser BE. Truth is slower than fiction: Francis Galton as an illustration. *Chance*. 2005;18(4):52-54.

Wainer H, Skorupski WP. Was it ethnic and social-class bias or statistical artifact? Logical and empirical evidence against Freedle's method for reestimating SAT scores. *Chance*. 2005;18(2):17-24.

Wainer H, Spence I. William Playfair and his graphical inventions. *The American Statistician*. 2005;59:224-229.

Wainer H, Wang XA, Skorupski WP, Bradlow ET. A Bayesian method for evaluating passing scores: the PPop curve. *Journal of Educational Measurement*. 2005;42:271-81.

---

## 2004 PUBLICATIONS

Arbet S, Morrison C, Griffin R. Proctored and secure examinations administered over the Internet. *CLEAR Exam Review*. 2004;XV(2):19-21.

Boulet JR, Swanson DB. Psychometric challenges of using simulations for high-stakes assessment. In: Dunn D, ed. *Simulators in Critical Care Education and Beyond*. Philadelphia, Pa: Lippincott, Williams and Wilkins; 2004:119-130.

Braun H, Wainer H. Numbers and the remembrance of things past. *Chance*. 2004;17(1):44-48.

Chapman DM, Hayden S, Sanders AB, Binder LS, Chinnis A, Corrigan K, LaDuca A, Dyne P, Perina DG, Smith-Coggins R, Sulton L, Swing S. Integrating the Accreditation Council for Graduate Medical Education core competencies into the model of the clinical practice of emergency medicine. *Academic Emergency Medicine*. 2004;11:674-685.

Cuddy MM, Dillon GF, Clauser BE, Holtzman KZ, Margolis MJ, McEllhenney SM, Swanson DB. Assessing the validity of the USMLE Step 2 clinical knowledge examination through an evaluation of its clinical relevance. *Academic Medicine*. 2004;79(10 Suppl):S43-S45.

De Champlain AF. Ensuring that the competent are truly competent: an overview of common methods and procedures used to set standards on high-stakes examinations. *Journal of Veterinary Medical Education*. 2004;31:61-65.

De Champlain AF, Schoeneberger J, Boulet JR. Assessing the impact of examinee and standardized patient ethnicity on test scores in a large-scale clinical skills examination: gathering evidence for the consequential aspect of validity. *Academic Medicine*. 2004;79(10 Suppl):S12-S14.

De Champlain AF, Winward M, Dillon GF, De Champlain JE. Modeling passing rates on a computer-based medical licensing examination: an application of survival data analysis. *Educational Measurement: Issues and Practice*. 2004;23(3):15-22.

Dillon GF, Boulet JR, Hawkins RE, Swanson DB. Simulations in the United States Medical Licensing Examination (USMLE). *Quality & Safety in Health Care*. 2004;13 Suppl 1:i41-i45.

Featherman CM, Nelson MV, Landau E, Sims A, Butler A. The NBME medical school resource site: a multi-purpose application for communicating with medical schools. *CLEAR Exam Review*. 2004;XV(1):17-20.

Friendly M, Wainer H. Nobody's perfect. *Chance*. 2004;17(2):48-51.

Hawkins RE, MacKrell-Gaglione M, LaDuca A, Leung C, Sample L, Gliva-McConvey G, Liston W, De Champlain AF, Ciccone AL. Assessment of patient management skills and clinical skills of practicing physicians using computer case simulations and standardized patients. *Medical Education*. 2004;38:958-968.

Holmboe ES, Hawkins RE, Huot SJ. Effects of training in direct observation of medical residents' clinical competence: a randomized trial. *Annals of Internal Medicine*. 2004;140:874-881.

Margolis MJ, Clauser BE, Harik P. Scoring the computer-based case simulation component of USMLE Step 3: a comparison of preoperational and operational data. *Academic Medicine*. 2004;79(suppl 10):62-64.

Melnick DE. Physician performance and assessment and their effect on continuing medical education and continuing professional development. *Journal of Continuing Education in the Health Professions*. 2004;24(suppl 1):38-49.

Sawhill AJ, Butler A, Ripkey DR, Swanson DB, Subhiyah R, Thelman J, Walsh W, Holtzman KZ, Angelucci K. Using the NBME self-assessments to project performance on USMLE Step 1 and Step 2: impact of test administration conditions. *Academic Medicine*. 2004;79(suppl 10):55-57.

Swygert KA, Muller ES, Clauser BE, Dillon GF, Swanson DB. The impact of timing changes on examinee pacing on the USMLE Step 2 exam. *Academic Medicine*. 2004;79(suppl 10):52-54.

Wainer H. Curbstoning IQ and the 2000 presidential election. *Chance*. 2004;17(4):43-6.

Wainer H. An editor's gratitude: reviewer acknowledgement. *Journal of Educational and Behavioral Statistics*. 2004;29:489-490.

Wainer H. The promises and pitfalls of making national educational assessments adaptive: America's assessment as an example. *Metodologia de las Ciencias del Comportamiento*. 2004;5:209-222.

Wainer H. Introduction to a special issue of the Journal of Educational and Behavioral Statistics on value-added assessment. *Journal of Educational and Behavioral Statistics*. 2004;29(1):1-3.

Wainer H, Bridgeman B, Najarian M, Trapani C. How much does extra time on the SAT help? *Chance*. 2004;17(2):19-24.

Wainer H, Brown LM. Two statistical paradoxes in the interpretation of group differences: illustrated with medical school admission and licensing data. *The American Statistician*. 2004;58:117-123.

Wainer H, Mee J. On assessing the quality of physicians' clinical judgment. *Evaluation & the Health Professions*. 2004;27:369-82.

Wang X, Wainer H, Bradlow ET. User's Guide for SCORIGHT (Version 3.0): A Computer Program for Scoring Tests Built of Testlets Including a Module for Covariate Analysis. *ETS Technical Report RR-04-49*. Princeton, NJ: Educational Testing Service; 2004.

---

## 2003 PUBLICATIONS

Boulet JR, De Champlain AF, McKinley DW. Setting defensible performance standards on OSCEs and standardized patient examinations. *Medical Teacher*. 2003;25:245-9.

De Champlain AF, Melnick DE, Scoles PV, Subhiyah R, Holtzman KZ, Swanson DB, Angelucci K, McGrenra C, Fournier JP, Benchimol D, Rampal P, Staccini P, Braun M, Kohler C, Guidet B, Claudepierre P, Prevel M, Goldberg J. Assessing medical students' clinical sciences knowledge in France: a collaboration between the NBME and a consortium of French medical schools. *Academic Medicine*. 2003;78:509-17.

Fournier JP, De Champlain AF, Benchimol D, Staccini P, Subhiyah R, Braun M, Kohler C, Guidet B, Claudepierre P, Prevel M, Scoles PV, Holtzman KZ, Swanson DB, Angelucci K, McGrenra C, Goldberg J, Rampal P, Melnick DE. [Transposition of an American-designed comprehensive medical student examination within the framework of the forthcoming French nationwide comprehensive examination. A preliminary study]. *Annales de Medecine Interne (Paris)*. 2003;154:148-56.

Hockberger RS, LaDuca A, Orr NA, Reinhart MA, Sklar DP. Creating the model of a clinical practice: the case of emergency medicine. *Academic Emergency Medicine*. 2003;10:161-8.

Holmboe ES, Huot S, Chung J, Norcini J, Hawkins RE. Construct validity of the MiniClinical Evaluation Exercise (MiniCEX). *Academic Medicine*. 2003;78:826-830.

Margolis MJ, Clauser BE, Swanson DB, Boulet JR. Analysis of the relationship between score components on a standardized patient clinical skills examination. *Academic Medicine*. 2003;78(suppl 10):68-71.



Muller ES, Harik P, Margolis MJ, Clauser BE, McKinley DW, Boulet JR. An examination of the relationship between clinical skills examination performance and performance on USMLE Step 2. *Academic Medicine*. 2003;78(suppl 10):27-29.

Pasquina PF, Kelly S, Hawkins RE. Assessing clinical competence in physical medicine & rehabilitation residency programs. *American Journal of Physical Medicine and Rehabilitation*. 2003;82:473-478.

Sawhill AJ, Dillon GF, Ripkey DR, Hawkins RE, Swanson DB. The impact of postgraduate training and timing on USMLE Step 3 performance. *Academic Medicine*. 2003;78(suppl 10):10-12.

Scoles PV, Blakemore LC. Congenital and pediatric disorders of the cervical spine. In: Emery SE, Boden SD, ed. *Surgery of the Cervical Spine*. Philadelphia PA: W.B. Saunders; 2003.

Scoles PV, Hawkins RE, LaDuca A. Assessment of clinical skills in medical practice. *Journal of Continuing Education in the Health Professions*. 2003;23:182-190.

Swanson DB, Jacovino SK, Holtzman KZ, Ripkey DR, Arbet S, Subhiyah R. CBT for high-stakes licensure and certification examinations: impact of examinee volume on test design and program operation. *CLEAR Exam Review*. 2003;XXIV(1):17-23.

Swygert KA. The relationship of item-level response times with test-taker and item variables in an operational CAT environment. *LSAC Computerized Testing Report 98-10*. Newtown, PA: Law School Admission Council; 2003.

Swygert KA, Margolis MJ, King AM, Siftar T, Clyman SG, Hawkins RE, Clauser BE. Evaluation of an automated procedure for scoring patient notes as part of a clinical skills examination. *Academic Medicine*. 2003;78(suppl 10):75-77.

Wainer H. La diffusion de quelques idées: a master's voice. *Chance*. 2003;16(3):58-61.

Wainer H. How long is short? *Chance*. 2003;16(2):55-7.

Wainer H. A graphical legacy of Charles Joseph Minard: two jewels from the past. *Chance*. 2003;16(1):56-60.

Wainer H. Editor's Forward to: "Comparing harm done by mobility and class absence: missing students and missing data" by Michelle C. Dunn, Joseph B. Kadane and John R. Garrow. *Journal of Educational and Behavioral Statistics*. 2003;29:267-8.

Wainer H. John Wilder Tukey: statistical inventor, discoverer and revolutionary. *Statistical Science*. 2003;18(3):1-2.



Wainer H. One cheer for null hypothesis significance testing. In: Kazdin AE, ed. *Methodological Issues & Strategies in Clinical Research*. 3rd ed. Washington, DC: American Psychological Association; 2003:461-464.

Wainer H, Koretz D. A political statistic. *Chance*. 2003;16(4):45-7.

Wainer H, Robinson DH. Shaping up the practice of null hypothesis significance testing. *Educational Researcher*. 2003;32(7):22-30.

---

## 2002 PUBLICATIONS

Aronson S, Butler A, Subhiyah R, Buckingham RE, Cahalan MK, Konstandt S, Mark J, Ramsay J, Savage R, Savino J, Shanewise JS, Smith J, Thys D. Development and analysis of a new certifying examination in perioperative transesophageal echocardiography. *Anesthesia and Analgesia*. 2002;95:1476-82.

Clauser BE. Advances in computerized scoring of complex item formats. *Applied Measurement in Education*. 2002;15:335-6.

Clauser BE, Kane MT, Swanson DB. Validity issues for performance-based tests scored with computer-automated scoring systems. *Applied Measurement in Education*. 2002;15:413-32.

Clauser BE, Margolis MJ, Swanson DB. An examination of the contribution of computer-based case simulations to the USMLE Step 3 examination. *Academic Medicine*. 2002;77(10 Suppl):S80-S82.

Clauser BE, Schuwirth Lambert WT. The use of computers in assessment. In: Norman GR, ed. *International Handbook of Research in Medical Education*. Dordrecht, The Netherlands: Kluwer; 2002;2:757-792.

Clauser BE, Swanson DB, Harik P. A multivariate generalizability analysis of the impact of training and feedback on judgments made in an Angoff-style standard-setting procedure. *Journal of Educational Measurement*. 2002;39:269-290.

Clyman SG, Galbraith RM, Melnick DE. Trends affecting the future of medical licensure assessment. *Journal of Medical Licensure and Discipline*. 2002;88(1):28-39.

Dillon GF, Clyman S G, Clauser B E, Margolis M J. The introduction of computer-based case simulations into the United States Medical Licensing Examination. *Academic Medicine*. 2002;77(10 Suppl):S94-S96.

- Farmer EA, Beard J D, Dauphinee W D, LaDuca A, Mann K V. Assessing the performance of doctors in teams and systems. *Medical Education*. 2002;36:942-8.
- Floreck LM, Guernsey MJ, Clyman SG, Clauser BE. Examinee performance on computer-based case simulations as part of the USMLE Step 3 examination: are examinees ordering dangerous actions? *Academic Medicine*. 2002;77(10 Suppl):S77-S79.
- Garibaldi RA, Subhiyah R, Moore ME, Waxman H. The In-Training Examination in Internal Medicine: an analysis of resident performance over time. *Annals of Internal Medicine*. 2002;137:505-510.
- Gessaroli ME, Folske JC. Generalizing the reliability of tests comprised of testlets. *International Journal of Testing*. 2002;2:277-95.
- Holtzman K, Case SM, Ripkey DR. Developing high-quality items quickly, cheaply, consistently - pick two. *CLEAR Exam Review*. 2002;13(1):16-19.
- Jones LS, Paulman LE, Thadani R, Terracio L. Medical student dissection of cadavers improves performance on practical exams but not on the NBME Anatomy Subject Exam. *The Meducator*. 2002;2(1):10-16.
- Jozefowicz RF, Koeppen BM, Case SM, Galbraith RM, Swanson DB, Glew RH. The quality of in-house medical school examinations. *Academic Medicine*. 2002;77:156-161.
- Luecht RM, Clauser BE. Test models for complex computer-based testing. In: Mills CN, Potenza MT, Fremer JJ, Ward CW, eds. *Computer-based testing: Building the foundation for future assessments*. Mahwah, NJ: Lawrence Earlbaum; 2002:67-88.
- Margolis MJ, Clauser B E, Harik P, Guernsey M J. Examining subgroup differences on the computer-based case simulation component of USMLE Step 3. *Academic Medicine*. 2002;77(suppl 10):83-85.
- Mazor KM, Clauser BE, Field T, Yood RA, Gurwitz JH. A demonstration of the impact of response bias on the results of patient satisfaction surveys. *Health Services Research*. 2002;37:1403-18.
- Melnick DE, Asch DA, Blackmore DE, Klass DJ, Norcini JJ. Conceptual challenges in tailoring physician performance assessment to individual practice. *Medical Education*. 2002;36:931-935.
- Melnick DE, Dillon GF, Swanson DB. Medical licensing examinations in the United States. *Journal of Dental Education*. 2002;66:595-9; discussion 610-611.

Rethans JJ, Norcini JJ, Baron-Maldonado M, Blackmore D, Jolly BC, LaDuca A, Lew S, Page GG, Southgate LH. The relationship between competence and performance: implications for assessing practice performance. *Medical Education*. 2002;36:901-9.

Robinson DH, Wainer H. On the past and future of null hypothesis significance testing. *Journal of Wildlife Management*. 2002;66:263-271.

Rosenfeld M, Keiser S, Goldsmith S. Issues of special concern in licensing and certification. In: Ekstrom RB, Smith DK, ed. *Assessing Individuals With Disabilities in Educational, Employment, and Counseling Settings*. Washington, DC: American Psychological Association; 2002:235-248.

Scoles PV. An evaluation of clinical skills in the United States Medical Licensing Examination: a report from the National Board of Medical Examiners. *Journal of Medical Licensure and Discipline*. 2002;88:66-69.

Swanson DB, Clauser B E, Case SM, Nungester Ronald J, Morrison Carol. Analysis of differential item functioning (DIF) using hierarchical logistic regression models. *Journal of Educational & Behavioral Statistics*. 2002;27:53-75.

Wainer H. Clear thinking made visible: redesigning score reports for students. *Chance*. 2002;15(1):56-8.

Wainer H. On the automatic generation of test items: some whens, whys and hows. In: Irvine S, Kyllonen P, ed. *Item Generation for Test Development*. Hillsdale, N.J: Lawrence Erlbaum Associates; 2002:287-305.

Wainer H. Remembering Sam Messick. In: Irvine S, Kyllonen P, ed. *Item Generation for Test Development*. Mahwah, NJ: Lawrence Erlbaum Associates; 2002:xxxi.

Wainer H. The BK-Plot: Making Simpson's Paradox clear to the masses. *Chance*. 2002;15(3):60-62.

Wainer H. Reporting test results to institutions and nations. *Chance*. 2002;15(2):1-4.

Wainer H. ..and still champion: Review of E.R. Tufte, *The Visual Display of Quantitative Information*. *Psychometrika*. 2002;67:173-178.

Wainer H, Zabell S. A small hurrah for the Black Death. *Chance*. 2002;15(4):58-60.

Wang X, Bradlow ET, Wainer H. A General Bayesian model for testlets: Theory and applications. *Applied Measurement in Education*. 2002;26:109-128.

## 2001 PUBLICATIONS

- Blakemore LS, Scoles PV, Poe-Kochert C, Thompson GH. Submuscular Isola rod with or without limited apical fusion in the management of severe spinal deformities in young children: preliminary report. *Spine*. 2001;26:2044-2048.
- Buckley G, LaDuca A. A dialogue on teaching: resolving a dilemma. *Medical Education*. 2001;35:178-179.
- Carson JD. Legal issues in standard setting for licensure and certification. In: Cizek CJ, ed. *Setting Performance Standards: Concepts, Methods and Perspectives*. Mahwah NJ: Lawrence Erlbaum; 2001:427-444.
- Case SM, Holtzman KZ, Ripkey DR. Developing an item pool for CBT: a practical comparison of three models of item writing. *Academic Medicine*. 2001;76(76 (10 Suppl)):S111-S113.
- Chang HH, Qian J, Ying Z. Stratified multistage computerized adaptive testing with b blocking. *Applied Psychological Measurement*. 2001;25:333-341.
- Clauser BE, Nungester RJ. Classification accuracy for tests that allow retakes. *Academic Medicine*. 2001;76 (10 Suppl):S108-S110.
- De Champlain AF, Margolis MJ, Macmillan MK, Klass DJ. Predicting mastery on a large-scale standardized patient test: a comparison of case and instrument score-based models using discriminant function analysis. *Advances in Health Sciences Education Theory and Practice*. 2001;6:151-158.
- Floreck LM, De Champlain AF. Assessing sources of score variability in a multisite medical performance assessment: an application of hierarchical linear modeling. *Academic Medicine*. 2001;76 (10 Suppl):S93-S95.
- Holtman MC, Swanson DB, Ripkey DR, Case SM. Using basic science subject tests to identify students at risk for failing Step 1. *Academic Medicine*. 2001;suppl 10:48-51.
- LaDuca A. Competence and the laying of blame. *Medical Education*. 2001;35:1170-1171.
- Sample L, LaDuca A, Leung C, Hawkins RE, Gaglione M, Liston W, De Champlain AF, Guernsey MJ, Ciccone AL, Illige M, Korinek E. Comparing patient-management skills of referred physicians and non-referred physicians on a computer-based-simulation examination. *Academic Medicine*. 2001;suppl 10:24-26.

Sireci SG, Clauser BE. Issues to be considered in setting standard on computerized-adaptive tests. In: Cizek CJ, ed. *Setting Performance Standards: Concepts, Methods and Perspectives*. Mahwah, NJ: Lawrence Erlbaum; 2001:355-369.

Swanson DB, Case SM, Ripkey DR, Clauser BE, Holtman MC. Relationships among item characteristics, examinee characteristics, and response times on USMLE Step 1. *Academic Medicine*. 2001;76(suppl 10):114-116.

Thissen D, Wainer H. *Test Scoring*. Mahwah, NJ: Lawrence Erlbaum; 2001.

Wainer H. Graphical details: a review of Leland Wilkinson's *The Grammar of Graphics*. *Psychometrika*. 2001;66:307-310.

Wainer H. Review of *Presenting Your Findings: A Practical Guide for Creating Tables* by Adelheid A. Nichol. *Teachers College Review*. 2001;103:93-98.

Wainer H. Order in the court. *Chance*. 2001;14:43-46.

Wainer H. New tools for exploration data analysis: Ill smoothing & nearness engines. *Chance*. 2001;14:43-46.

Wainer H. On the alienation of content and evidence from commercial design. *Chance*. 2001;14:37-39.

Wainer H. Sex, smoking and life insurance. *Chance*. 2001;14:42-45.

Wainer H. Winds across Europe: Francis Galton and the graphic discovery of weather patterns. *Chance*. 2001;14:44-47.

Wainer H, Spence I. William Playfair (1759-1823): an inventor and ardent advocate of statistical graphics. In: Heyde CC, Seneta S, ed. *Statisticians of the Centuries*. New York, NY: Springer-Verlag; 2001:105-110.

Wainer H, Velleman P. Statistical graphics: mapping the pathways of science. *Annual Review of Psychology*. 2001;52:305-335.

Wang JC, Nuccion SL, Feighan JE, Cohen B, Dorey FJ, Scoles PV. Growth and development of the pediatric cervical spine documented radiographically. *Journal of Bone and Joint Surgery*. 2001;83A:1212-1218.

Weyman AE, Butler A, Subhiyah R, Appleton C, Geiser E, Goldstein SA, King ME, Kaul S, Labovitz A, Picard M, Ryan T, Shanewise J. Concept, development, administration, and analysis of a certifying

examination in echocardiography for physicians. *Journal of the American Society of Echocardiology*. 2001;14:158-168.

---

## 2000 PUBLICATIONS

Bowles LT. The evaluation of teaching. *Medical Teacher*. 2000;22:221-224.

Bowles LT, Melnick DE, Nungester RJ, Golden GS, Swanson DB, Case SM, Dillon GF, Henzel TR, Orr NA, Thadani RA. Review of the score-reporting policy for the United States Medical Licensing Examination. *Academic Medicine*. 2000;75:426-431.

Calisias AM, Clyman SG, Fan YY, Stevens RH. Exploring alternative models of complex patient management with artificial neural networks. *Advances in Health Sciences Education Theory and Practice*. 2000;2000:23-41.

Case SM, Swanson DB, Ripkey DR. Setting standards for written exams by mail: an application of the Hofstee methods. In: Melnick DE, ed. *Evolving Assessment: Protecting the Human Dimension: Proceedings of the Eighth International Ottawa Conference on Medical Education and Assessment, July, 1998*. Philadelphia, PA: National Board of Medical Examiners; 2000:162-168.

Clauser BE. Further discussion of SP checklists and videotaped performances. *Academic Medicine*. 2000;75:315-316.

Clauser BE. Recurrent issues and recent advances in scoring performance assessments. *Applied Psychological Measurement*. 2000;24:310-324.

Clauser BE, De Champlain AF, Nungester RJ. Applying sequential testing strategies to performance assessments of clinical skills. In: Melnick DE, ed. *Evolving Assessment: Protecting the Human Dimension: Proceedings of the Eighth International Ottawa Conference on Medical Education and Assessment, July, 1998*. Philadelphia, PA: National Board of Medical Examiners; 2000:226-233.

Clauser BE, Harik P, Clyman SG. The generalizability of scores for a performance assessment scored with a computer-automated scoring system. *Journal of Educational Measurement*. 2000;37:245-262.

De Champlain AF. Further discussion of SP checklists and videotaped performances. *Academic Medicine*. 2000;75:316-317.

De Champlain AF, Fletcher EA, Macmillan MK, Klass DJ, Margolis MJ. Assessing the reliability of post encounter note scores in a large-scale standardized patient examination: comparing the consistency of medical chart abstractors and physicians. In: Melnick DE, ed. *Evolving Assessment: Protecting the Human*



*Dimension: Proceedings of the Eighth International Ottawa Conference on Medical Education and Assessment, July, 1998.* Philadelphia, PA; 2000:421-427.

De Champlain AF, Macmillan MK, Margolis MJ, Klass DJ, Lewis E, Ahearn S. Modeling the effects of a test security breach on a large-scale standardized patient examination with a sample of international medical graduates. *Academic Medicine.* 2000;75 (10 Suppl):S109-S111.

De Champlain AF, Margolis MJ, King AM, Klass DJ. Investigating halo effects in a nationally administered standardized patient examination. In: Melnick DE, ed. *Evolving Assessment: Protecting the Human Dimension: Proceedings of the Eighth International Ottawa Conference on Medical Education and Assessment, July, 1998.* Philadelphia, PA: National Board of Medical Examiners; 2000:400-405.

Dillon GF, Case SM, Melnick DE, Nungester RJ, Swanson DB. Setting standards on the United States Medical Licensing Examination. In: Melnick DE, ed. *Evolving Assessment: Protecting the Human Dimension: Proceedings of the Eighth International Ottawa Conference on Medical Education and Assessment, July, 1998.* Philadelphia, PA: National Board of Medical Examiners; 2000:466-474.

Dillon GF, Walsh W. Using performance data to set standards: practical impact and the perception of judges. *CLEAR Exam Review.* 2000;XI(1):15-18.

Featherman CM, Case SM. Using the Rasch model to analyze examination data: an alternative measurement methodology. In: Melnick DE, ed. *Evolving Assessment: Protecting the Human Dimension: Proceedings of the Eighth International Ottawa Conference on Medical Education and Assessment, July, 1998.* Philadelphia, PA: National Board of Medical Examiners; 2000:155-162.

Fletcher EA, De Champlain AF, Klass DJ, Macmillan MK. Surveying reactions of medical chart abstractors and physicians to the scoring process of post-encounter notes for and NBME standardized patient examination. In: Melnick DE, ed. *Evolving Assessment: Protecting the Human Dimension: Proceedings of the Eighth International Ottawa Conference on Medical Education and Assessment, July, 1998.* Philadelphia, PA: National Board of Medical Examiners; 2000:906-907.

Hatala R, Case SM. Examining the influence of gender on medical students' decision making. *Journal of Women's Health and Gender Based Medicine.* 2000;9:617-623.

Henzel TR, Golden GS. Structural complexity of test items for computer-based testing. *CLEAR Exam Review.* 2000;11(2):18-23.

Henzel TR, LaDuca A, Wemple KG. Reflecting physician/patient encounters in the design of medical licensure examinations. In: Melnick DE, ed. *Evolving Assessment: Protecting the Human Dimension: Proceedings of the Eighth International Ottawa Conference on Medical Education and Assessment.* Philadelphia, PA: National Board of Medical Examiners; 2000:874-875.



- Johnson D, Dillon GF, Henzel TR. The post licensure assessment system. *Journal of Medical Licensure and Discipline*. 2000;86:116-122.
- King AM, Carr BA, Downing BK, Klass DJ. A description of National Board of Medical Examiners' training processes for standardized patient licensing examinations. In: Melnick DE, ed. *Evolving Assessment: Protecting the Human Dimension: Proceedings of the Eighth International Ottawa Conference on Medical Education and Assessment*. Philadelphia, PA: National Board of Medical Examiners; 2000:386-392.
- Klass DJ. Reevaluation of clinical competency. *American Journal of Physical Medicine and Rehabilitation*. 2000;79:481-486.
- Klass DJ, De Champlain AF, Fletcher EA, King AM, Macmillan MK. Development of a performance-based test of clinical skills for the United States Medical Licensing Examination. In: Melnick DE, ed. *Evolving Assessment: Protecting the Human Dimension: Proceedings of the Eighth International Ottawa Conference on Medical Education and Assessment*. Philadelphia, PA: National Board of Medical Examiners; 2000:77-84.
- LaDuca A, De Champlain AF, Sample L. Diagnostic assessment of practicing doctors: computer simulation of patient management skills. In: Melnick DE, ed. *Evolving Assessment: Protecting the Human Dimension: Proceedings of the Eighth International Ottawa Conference on Medical Education and Assessment*. Philadelphia, PA: National Board of Medical Examiners; 2000:209-214.
- Luchins DJ, Klass DJ, Hanrahan P, Qayyum M, Malan R, Raskin-Davis V, Fichtner CG. Computerized monitoring of valproate and physician responsiveness to laboratory studies as a quality indicator. *Psychiatric Services*. 2000;51:1179-1181.
- Luecht RM, Nungester RJ. Computer-adaptive testing. In: van der Linden WJ, Glas CAW, ed. *Computerized Adaptive Testing*. Boston, MA: Kluwer; 2000.
- Macmillan MK, De Champlain AF, Klass DJ. Assessing the comparability of checklist scores across standardized patients using traveling patients. In: Melnick DE, ed. *Evolving Assessment: Protecting the Human Dimension: Proceedings of the Eighth International Ottawa Conference on Medical Education and Assessment*. Philadelphia, PA: National Board of Medical Examiners; 2000:779-780.
- Macmillan MK, Fletcher EA, De Champlain AF, Klass DJ. Assessing post-encounter note documentation by examinees in a field test of a nationally administered standardized patient test. *Academic Medicine*. 2000;75(suppl 10):112-114.
- Margolis MJ, De Champlain AF, Klass DJ. Setting standards for a performance-based assessment of physicians' clinical skills. In: Melnick DE, ed. *Evolving Assessment: Protecting the Human Dimension:*

*Proceedings of the Eighth International Ottawa Conference on Medical Education and Assessment*. Philadelphia, PA: National Board of Medical Examiners; 2000:407-412.

Martz AP, Gessaroli ME, Swanson DB, De Champlain AF. Equating standardized patient cases using structural equation modeling. In: Melnick DE, ed. *Evolving Assessment: Protecting the Human Dimension: Proceedings of the Eighth International Ottawa Conference on Medical Education and Assessment*. Philadelphia, PA: National Board of Medical Examiners; 2000:413-420.

Mislevy R, Chang HH. Does adaptive testing violate local independence? *Psychometrika*. 2000;20:149-165.

Newble D, Swanson DB. Improving the quality of a multidisciplinary test of clinical competence: a longitudinal study. In: Melnick DE, ed. *Evolving Assessment: Protecting the Human Dimension: Proceedings of the Eighth International Ottawa Conference on Medical Education and Assessment*. Philadelphia, PA: National Board of Medical Examiners; 2000:376-380.

Orr NA, Clyman SG. Computer-based case simulation by the National Board of Medical Examiners. In: Melnick DE, ed. *Evolving Assessment: Protecting the Human Dimension: Proceedings of the Eighth International Ottawa Conference on Medical Education and Assessment*. Philadelphia, PA: National Board of Medical Examiners; 2000:943-944.

Ripkey DR, Case SM, Swanson DB, Fincher R. Third-year ambulatory experiences of U.S. students: implications for USMLE Step 2 performance. In: Melnick DE, ed. *Evolving Assessment: Protecting the Human Dimension: Proceedings of the Eighth International Ottawa Conference on Medical Education and Assessment*. Philadelphia, PA: National Board of Medical Examiners; 2000:124-128.

Ross LP, Clauser BE, Clyman SG. The validity of expert judgment for scoring performance assessments: are all judges evaluating the same trait? In: Melnick DE, ed. *Evolving Assessment: Protecting the Human Dimension: Proceedings of the Eighth International Ottawa Conference on Medical Education and Assessment*. Philadelphia, PA: National Board of Medical Examiners; 2000:393-399.

Ross LP, De Champlain AF, Margolis MJ. Examining fairness issues for a large-scale standardized patient examination using structural equation modeling. In: Melnick DE, ed. *Evolving Assessment: Protecting the Human Dimension: Proceedings of the Eighth International Ottawa Conference on Medical Education and Assessment, July, 1998*. Philadelphia, PA: National Board of Medical Examiners; 2000:787.

Scheuneman JD, Clyman SG, Fan YY. An investigation of the properties of computer-based case simulation. *Advances in Health Sciences Education Theory and Practice*. 2000;5:11-22.

Sirotkin A, Fomin Y, Case SM, Jozefowicz R. Implementing an interinstitutional clinical vignette MCQ test in Russia: a first experience. In: Melnick DE, ed. *Evolving Assessment: Protecting the Human Dimension :*

*Proceedings of the Eighth International Ottawa Conference on Medical Education and Assessment*. Philadelphia, PA: National Board of Medical Examiners; 2000:328-329.

Thadani RA, Swanson DB, Galbraith RM. A preliminary analysis of different approaches to preparing for the USMLE Step 1. *Academic Medicine*. 2000;75(suppl 10):40-42.

Winward ML, Ripkey DR, Case SM, Morrison C. Performance of foreign medical graduates on the clinical science component of the United States Medical Licensing Examination: initial and ultimate pass rates. In: Melnick DE, ed. *Evolving Assessment: Protecting the Human Dimension: Proceedings of the Eighth International Ottawa Conference on Medical Education and Assessment*. Philadelphia, PA: National Board of Medical Examiners; 2000:67-74.

---

## 1990 TO 1999

### Back to the Bedside: 1990s

Articles on the use of standardized patients in assessment were common in the 1990s, and the focus of publications turned to the clinical skills important to the practice of clinical medicine, as well as new methods for directly evaluating those constructs in authentic ways. Indeed, it seemed that testing circled back to the skills that had been assessed with the bedside oral exam.

## 1999 PUBLICATIONS

Bowles LT. USMLE and end-of-life care. *Journal of Palliative Care*. 1999;2:3-4.

Case SM, Bowles LT, Melnick DE. Response to editorial on USMLE exam. *Academic Physician and Scientist*. 1999;1999:3-4.

Case SM, Hatala R, Blake J, Golden GS. Does sex make a difference? Sometimes it does and sometimes it doesn't. *Academic Medicine*. 1999;74(10 Suppl):S37-S40.

Chang HH, Ying Z. a-Stratified multistaged computerized adaptive testing. *Applied Psychological Measurement*. 1999;23:211-222.

Chen S, Ankenmann R, Chang HH. A comparison of item selection rules at the early stages of computerized adaptive testing. *Applied Psychological Measurement*. 1999;23:211-222.

Clauser BE, Clyman SG, Swanson DB. Components of rater error in a complex performance assessment. *Journal of Educational Measurement*. 1999;35:29-45.

Clauser BE, Nungester RJ. Considerations in adjusting cut-scores for certification and licensure decisions. *CLEAR Exam Review*. 1999;X(2):18-23.

Clauser BE, Swanson DB, Clyman SG. A comparison of the generalizability of scores produced by expert raters and automated scoring systems. *Applied Measurement in Education*. 1999;12:281-299.

Clyman SG, Melnick DE, Clauser BE. Computer-based simulations from medicine: assessing skills in patient management. In: Tekian A, McGuire CH, McGahie WC, eds. *Innovative Simulations for Assessing Professional Competence*. Chicago, IL: University of Illinois Department of Medical Education; 1999:29-41.

De Champlain AF, Macmillan MK, King AM, Klass DJ, Margolis MJ. Assessing the impact of intra-site and inter-site checklist recording discrepancies on the reliability of scores obtained in a nationally administered standardized patient examination. *Academic Medicine*. 1999;74(10):S52-S54.

De Champlain AF, Macmillan MK, Margolis MJ, Klass DJ, Nungester RJ, Schimpfhauser F, Zimmerstrom K. Modeling the effects of security breaches on a large-scale standardized patient examination. *Academic Medicine*. 1999;74(10 Suppl):S49-S51.

Friedman BD, Klass DJ, Boulet JR, De Champlain AF, King AM, Pohl SA, Gary NE. The performance of foreign medical graduates on the National Board of Medical Examiners (NBME) standardized patient examination prototype: a collaborative study of the NBME and the Educational Commission for Foreign Medical Graduates (ECFMG). *Medical Education*. 1999;33:439-466.

Gordon M, Lewandowski L, Keiser S. The LD label for relatively well-functioning students: A critical analysis. *Journal of Learning Disabilities*. 1999;32(6):485-490.

Keiser S. Testing and measurement issues: understanding equal access in the context of the American with Disabilities Act (ADA). *CLEAR Exam Review*. 1999;10(1):17-18.

Macmillan MK, De Champlain AF, Klass DJ. Using tagged items to detect threats to security in a nationally administered standardized patient examination. *Academic Medicine*. 1999;74(suppl 10):55-57.

Mazor K, Clauser BE, Cohen A, Alper E, Punnaire M. The dependability of students' ratings of perceptors. *Academic Medicine*. 1999;74(suppl 10):19-21.

Melnick DE. Evaluation - telling students what to learn. In: *Perspektiven des Medizinstudiums*. St Ingbert, Germany: Rohrig Universitats Verlag; 1999:113-135.

Ripkey DR, Case SM, Swanson DB. Identifying students at risk for poor performance on USMLE Step 2. *Academic Medicine*. 1999;74(suppl 10):45-48.

Scoles PV, Thompson GH. Part XXXI: Bone and joint disorders. In: Behrman R, Kliegman R, Jenson H, eds. *Nelsons Textbook of Pediatrics*. Philadelphia, PA: WB Saunders; 1999.

Swanson DB, Clauser BE, Case SM. Clinical skills assessment with standardized patients in high-stakes tests: a framework for thinking about score precision, equating and security. *Advances in Health Sciences Education Theory and Practice*. 1999;4:67-106.

---

## 1998 PUBLICATIONS

Brinkerhoff L, Dempsey K, Jordan C, Keiser S, McGuire J. Guidelines for documentation of attention-deficit/hyperactivity disorder for adolescents and adults. Consortium on ADHD Documentation; 1998.

- Clauser BE. Review: Educational Measurement: Origins, Theories and Explications. *Journal of Educational Measurement*. 1998;35:273-275.
- Clauser BE, Mazor KM. Using statistical procedures to identify differentially functioning test items (ITEMS Module). *Educational Measurement: Issues and Practice*. 1998;17(1):31-44.
- Clauser BE, Ross LP, Fan YY, Clyman SG. A comparison of two approaches for modeling expert judgment in scoring a performance assessment of physicians' patient management skills. *Academic Medicine*. 1998;73(10 Suppl):S117-S119.
- De Champlain AF, Clauser BE, Margolis MJ, Klass DJ. Assessing decision consistency with a sequentially administered large-scale standardized patient examination: a Monte Carlo investigation. *Academic Medicine*. 1998;73(10 Suppl):S78-S80.
- De Champlain AF, Macmillan MK, Margolis MJ, King AM, Klass DJ. Do discrepancies in standardized patients' checklist recording affect case and examination mastery-level decisions? *Academic Medicine*. 1998;73(10 Suppl):S75-S77.
- Dillon GF. Testing and measurement issues: the role of survey data in a testing program. *CLEAR Exam Review*. 1998;IX(1):20-22.
- Golden GS. Commentary: Apgar scores as predictors of chronic neurologic disability. *Pediatrics*. 1998;102:262-264.
- Golden GS. Neurology and neuromuscular disorders. In: Finberg L, ed. *Saunders Manual of Pediatric Practice*. Philadelphia, PA: WB Saunders; 1998.
- Golden GS. Neurologic symptoms. In: Finberg L, ed. *Saunders Manual of Pediatric Practice*. Philadelphia, PA: WB Saunders; 1998.
- Gorden M, Keiser S. Accommodations in Higher Education under the Americans with Disabilities Act (ADA). New York, NY: Guilford Press; 1998.
- Gordon M, Keiser S. Clinical psychology, higher education, and the Americans with Disabilities Act (ADA). *Independent Practitioner*. 1998;18:193-198.
- Gordon M, Murphy K, Keiser S. Attention deficit disorder (ADHD) and test accommodations. *The Bar Examiner*. 1998;67(4):26-36.
- Hadadi A, Leucht RM. Some methods for detecting and understanding test speededness on timed multiple-choice tests. *Academic Medicine*. 1998;73(suppl 10):47-50.

- Luecht RM. Computer-assisted test assembly using optimization heuristics. *Applied Psychological Measurement*. 1998;22:224-236.
- Luecht RM. A reaction to: "Moderating possibly irrelevant multiple mean score differences on a test of mathematical reasoning.". *Journal of Educational Measurement*. 1998;35:223-225.
- Luecht RM, Hadadi A, Swanson DB, Case SM. A comparative study of a comprehensive basic sciences test using paper-and-pencil and computerized formats. *Academic Medicine*. 1998;73(suppl 10):51-53.
- Luecht RM, Nungester RJ. Some practical examples of computer-adaptive sequential testing. *Journal of Educational Measurement*. 1998;35:229-249.
- Mazor K, Hambleton RK, Clauser BE. Multidimensional DIF analysis: the effects of matching on unidimensional subtest scores. *Applied Psychological Measurement*. 1998;22:357-367.
- Ripkey R, Swanson DB, Case SM. School-to-school differences in Step 1 performance as a function of curriculum type and use of Step 1 in promotion/graduation requirements. *Academic Medicine*. 1998;73(10 Suppl):S16-S18.
- Scheuneman JD, Fan YV, Clyman SG. An investigation of the difficulty of computer-based case simulations. *Medical Education*. 1998;22:150-158.
- Scheuneman JD, Subhiyah R. Evidence for the validity of a Rasch model technique for identifying differential item functioning. *Journal of Outcome Measurement*. 1998;2(1):33-42.
- Subhiyah R, Morrison C. Computerized adaptive testing: an introduction to basic concepts. *Perspectives on Physician Assistant Education*. 1998;9(2):23-26.
- Wang T, Zeng L. Item parameter estimation for a continuous response model using an EM algorithm. *Applied Psychological Measurement*. 1998;22:333-344.
- 

## 1997 PUBLICATIONS

- Bowles LT. Genes and the environment: thoughts for medical education. *Journal of Cancer Education*. 1997;12:34-39.
- Bowles LT. Emergency medicine: a status report. *Academic Emergency Medicine*. 1997;4:647-648.



Bowles LT. Samuel C. Harvey lecture: Genes and the environment: thoughts for medical education. *Journal of Cancer Education*. 1997;12:34-39.

Carson JD. Current legal climate and candidates with disabilities. In: Mancell EL, Bashook PG, Dockery JL, eds. *Legal Issues in Specialty Board Certification*. Chicago, IL: American Board of Medical Specialties, Research and Education Foundation; 1997:47-56.

Case SM. Assessment truths that we hold as self-evident and their implications. In: Scherpbeir AJJA, van der Vleuten CPM, Rethans JJ, van der Steeg AFW, eds. *Advances in Medical Education*. Dordrecht, The Netherlands: Kluwer; 1997:2-6.

Case SM, Ripkey DR, Swanson DB. The effects of psychiatry clerkship timing and length on measures of performance. *Academic Medicine*. 1997;72(10 Suppl):S34-S36.

Case SM, Swanson DB. The use of computerized testing for students on clinical rotation. *The Neurology Clerkship: Innovative Methods of Evaluating Students. Proceedings of the 49th Annual Meeting of the American Academy of Neurology*. 1997;123:3-16.

Case SM, Swanson DB, Ripkey DR, Bowles LT, Melnick DE. Preliminary descriptive analyses of the performance of U.S. citizens attending foreign schools on USMLE Step 1 and 2. In: Scherpbeir AJJA, van der Vleuten CPM, Rethans JJ, van der Steeg AFW, eds. *Advances in Medical Education*. Dordrecht, The Netherlands: Kluwer; 1997:135-138.

Clauser BE, Margolis MJ, Clyman SG, Ross LP. Development of automated scoring algorithms for complex performance assessments: a comparison of two approaches. *Journal of Educational Measurement*. 1997;34:141-161.

Clauser BE, Margolis MJ, Ross LP, Nungester RJ, Klass DJ. Regression-based weighting of items on standardized patient checklists. In: Scherpbeir AJJA, van der Vleuten CPM, Rethans JJ, van der Steeg AFW, eds. *Advances in Medical Education*. Dordrecht, The Netherlands: Kluwer; 1997:420-423.

Clauser BE, Nungester RJ. Setting standards on performance assessment of physicians' clinical skills using contrasting groups and receiver operating characteristic curves. *Evaluation and the Health Professions*. 1997;20:215-238.

Clauser BE, Ross LP, Clyman SG, Rose KM, Margolis MJ, Nungester RJ, Piemmer TE, Chang L, El-Bayoumi G, Malakoff GL, Pincetl PA. Development of a scoring algorithm to replace expert rating for scoring a complex performance-based assessment. *Applied Measurement in Education*. 1997;10:345-358.

Clauser BE, Ross LP, Luecht RM, Nungester RJ, Clyman SG. Using the Rasch model to equate alternate forms for performance assessments of physicians' clinical skills. In: Scherpbeir AJJA, van der Vleuten

CPM, Rethans JJ, van der Steeg AFW, eds. *Advances in Medical Education*. Dordrecht, The Netherlands: Kluwer; 1997:416-419.

Clauser BE, Ross LP, Nungester RJ, Clyman SG. An evaluation of the Rasch model for equating multiple forms of a performance assessment of physicians' patient management skills. *Academic Medicine*. 1997;72(10 Suppl):S76-S78.

Clyman SG, Melnick DE, Clauser BE. Computer based case simulation by the National Board of Medical Examiners of the United States. *Proceedings of the Boerhaave Conference "Toetsing in de Basisopleiding"*. 1997:133-147.

De Champlain AF, Klass DJ. Assessing the factor structure of a nationally administered standardized patient examination. *Academic Medicine*. 1997;72(10 Suppl):S88-S90.

De Champlain AF, Margolis MJ, King AM, Klass DJ. Standardized patients' accuracy in recording examinees' behaviors using checklists. *Academic Medicine*. 1997;72(10 Suppl):S85-S87.

De Champlain AF, Tang KL. CHIDIM: a FORTRAN program to assess the dimensionality of binary item responses based on McDonald's nonlinear factor and analysis model. *Educational and Psychological Measurement*. 1997;57:174-178.

Dillon GF, Henzel TR, Walsh W. The impact of postgraduate training on an examination for medical licensure. In: Scherpbeir AJJA, van der Vleuten CPM, Rethans JJ, van der Steeg AFW, eds. *Advances in Medical Education*. Dordrecht, The Netherlands: Kluwer; 1997:146-148.

Dillon GF, Marcus LA, Walsh W. The usefulness of test-performance feedback in preparing to repeat the USMLE Step 3 examination. *Academic Medicine*. 1997;72(10 Suppl):S94-S96.

Edelstein RA, Clyman SG. Computer-based simulations as adjuncts for teaching and evaluating complex medical skills. In: Scherpbeir AJJA, van der Vleuten CPM, Rethans JJ, van der Steeg AFW, eds. *Advances in Medical Education*. Dordrecht, The Netherlands: Kluwer; 1997:327-329.

Fan YY, Clyman SG, Clauser BE, Piemme TW, Chang L, El-Bayoumi J, Malakoff GL. A comparison of conjoint analysis with other approaches to model physician policies in scoring complex performance-based assessment. In: Scherpbeir AJJA, van der Vleuten CPM, Rethans JJ, van der Steeg AFW, eds. *Advances in Medical Education*. Dordrecht, The Netherlands: Kluwer; 1997:149-151.

Featherman CM. BIBSTEPS Rasch model computer program version 2.67: software review. *Applied Psychological Measurement*. 1997;21:279-284.

Fincher RE, Case SM, Ripkey DR, Swanson DB. Comparison of ambulatory knowledge of third-year students who learned in ambulatory settings with that of students who learned in inpatient settings. *Academic Medicine*. 1997;72(10 Suppl):S130-S132.

Furman GE, Colliver JA, Galofre A, Reaka MA, Robbs RS, King AM. The effect of formal feedback sessions on test security for a clinical practice examination using standardized patients. In: Scherpbeir AJJA, van der Vleuten CPM, Rethans JJ, van der Steeg AFW, eds. *Advances in Medical Education*. Dordrecht, The Netherlands: Kluwer; 1997:433-436.

Furman GE, Colliver JA, Galofre A, Reaka MA, Robbs RS, King AM. The effect of formal feedback sessions on test security for a clinical practice examination using standardized patients. *Advances in Health Sciences Education Theory and Practice*. 1997;2:3-7.

Glew RH, Ripkey DR, Swanson DB. Relationship between students' performances on the NBME comprehensive basic science examination and the USMLE Step 1: a longitudinal investigation at one school. *Academic Medicine*. 1997;72:1097-1102.

Greenburg AG, Case SM, Golden GS, Melnick DE. Core clinical content on Step 2 of the USMLE: using surgery as an example. In: Scherpbeir AJJA, van der Vleuten CPM, Rethans JJ, van der Steeg AFW, eds. *Advances in Medical Education*. Dordrecht, The Netherlands: Kluwer; 1997:34-36.

Hark LA, Iwomoto C Mel, Young EA, Morgan SL, Kushner R, Hensrud DD. Nutrition coverage on medical licensing examinations in the United States. *American Journal of Clinical Nutrition*. 1997;65:568-571.

Klass DJ. Valuing communication. *Medical Encounter*. 1997;13(1):2-3.

Klass DJ, Fletcher EA, Macmillan MK, King AM, Carr BA, Downing BK. Incorporating measures into a performance test of clinical competence using standardized patients. *Medical Encounter*. 1997;13(1):12-16.

LaDuca A. Diagnostic assessment of physicians' continued competence: a new role for the NBME. *CLEAR Exam Review*. 1997;8(2):19-22.

LaDuca A, Leone-Perkins M, De Champlain AF. Evaluating continuing competence of physicians through multiple assessment modalities: the physicians' continued competence assessment program (PCCAP). *Academic Medicine*. 1997;72:457-458.

Luecht RM. Multidimensional computerized adaptive testing in a certification or licensure context. *Applied Psychological Measurement*. 1997;20:389-404.

Luecht RM, De Champlain AF, Nungester RJ. Maintaining content validity in computerized adaptive testing. In: Scherpbeir AJJA, van der Vleuten CPM, Rethans JJ, van der Steeg AFW, eds. *Advances in Medical Education*. Dordrecht, The Netherlands: Kluwer; 1997:366-369.

Nungester RJ, Clauser BE, Clyman SG. An evaluation of the Rasch Model for equating multiple forms of a performance: a physicians' patient management skills. *Academic Medicine*. 1997;72(suppl 10):76-78.

Page GG, Bandaranayake RC, Case SM, Dauphinee WD, Norcini JJ, Stern ST, Swanson DB. Curriculum design. In: Davis WK, Jolly BC, Page GG, Rothman AI, White BC, eds. *Moving Medical Education from the Hospital to the Community: Report of the Seventh Cambridge Conference on Medical Education*. Ann Arbor, MI: University of Michigan Medical School; 1997:5-31.

Pangaro LN, Worth-Dickstein H, Macmillian MK, Klass DJ, Shatzer JH. Performance of "standardized examinees" in a standardized-patient examination of clinical skills. *Academic Medicine*. 1997;72:1008-1011.

Ripkey DR, Case SM, Swanson DB. Predicting performance on the NBME surgery subject test and USMLE Step 2. *Academic Medicine*. 1997;72(suppl 10):31-33.

Scheuneman JD. Testing and measurement issues: potholes on the road to computer-based testing. *CLEAR Exam Review*. 1997;VIII(1):19-24.

Scheuneman JD, Clyman SG. An investigation of the properties of computer-based simulations. In: Scherpbeir AJJA, van der Vleuten CPM, Rethans JJ, van der Steeg AFW, eds. *Advances in Medical Education*. New York, NY: Kluwer; 1997:184-186.

Scheuneman JD, Grima A. Characteristics of quantitative word items associated with differential performance for female and black examinees. *Applied Measurement in Education*. 1997;10:199-319.

Swanson DB, Case SM. Assessment in basic science in instruction: directions for practice and research. *Advances in Health Sciences Education Theory and Practice*. 1997;2:71-84.

Swanson DB, Case SM, Ripkey DR, Melnick DE, Bowles LT, Gary N. Performance of examinees from foreign schools on the basic science component of the United States Medical Licensing Examination. In: Scherpbeir AJJA, van der Vleuten CPM, Rethans JJ, van der Steeg AFW, eds. *Advances in Medical Education*. 1997:187-190.

Swanson DB, Case SM, van der Vleuten CPM. Strategies for student assessment. In: Boud D, Felett G, ed. *The Challenge of Problem-Based Learning, rev.ed.* London, UK: Kogan Page; 1997:269-282.

Zeng L. Implementation of marginal Bayesian estimation with four-parameter-beta prior distributions. *Applied Psychological Measurement*. 1997;21:143-156.

---

## 1996 PUBLICATIONS

Case SM, Ripkey DR, Swanson DB. The relationship between clinical science performance in 20 medical schools and performance on Step 2 of the USMLE licensing examination. 1994-95 validity study group for USMLE Step 1 and Step 2 pass/fail standards. *Academic Medicine*. 1996;71(10 Suppl):S31-S33.

Case SM, Swanson DB, Becker DF. Verbosity, window dressing and red herrings: do they make a better test item? *Academic Medicine*. 1996;71(10 Suppl):S28-S30.

Case SM, Swanson DB, Ripkey DR. Relationship between achievement in basic science coursework and performance on 1994 USMLE Step 1 test administration. 1994-95 validity study group for USMLE Step 1/2 Pass/Fail Standards. *Academic Medicine*. 1996;71(1 Suppl):S28-S30.

Case SM, Swanson DB, Ripkey DR, Bowles LT, Melnick DE. Performance of the class of 1994 in the new era of USMLE. *Academic Medicine*. 1996;71(10 Suppl):S91-S93.

Clauser BE, Nungester RJ, Mazor MK, Ripkey DR. A comparison of alternative matching strategies for DIF detection in tests that are multidimensional. *Journal of Educational Measurement*. 1996;33:202-214.

Clauser BE, Nungester RJ, Swaminathan H. Improving the matching for DIF analysis by conditioning on both test score and educational background variable. *Journal of Educational Measurement*. 1996;33:453-464.

Clauser BE, Swanson DB, Clyman SG. The generalizability of scores from a performance assessment of physicians' patient management skills. *Academic Medicine*. 1996;71(10 Suppl):S109-S111.

Dillon GF. The expectations of standard setting judges. *CLEAR Exam Review*. 1996;VII(2):22-26.

Gessaroli ME, De Champlain AF. Using an approximate chi-square statistic to test the number of dimensions underlying the responses to a set of items. *Journal of Educational Measurement*. 1996;33:157-149.

Golden GS. Developmental disabilities. In: Bradley WG, Daroff RB, Fenichel GM, Marsden CD, eds. *Neurology in Clinical Practice*. Boston, MA: Butterworth-Heinemann; 1996:1483-1492.

Golden GS. Fainting and syncope. In: Berg BO, ed. *Principles of Child Neurology*. New York, NY: McGraw-Hill; 1996:197-302.

Gruppen LD, Grum CM, Fincher RE, Parenti C, Cleary LM, Swaney J, Case SM, Swanson DB, Woolliscroft JO. Multi-site reliability and validity of a diagnostic pattern recognition knowledge and assessment instrument. *Academic Medicine*. 1996;71(10 Suppl):S65-S67.

LaDuca A. Assessing clinical competence and the continuing challenge of validity. In: *Trends in Medical Education Conference*. Zaragoza, Spain: Archivos de la Facultad de Medicina Zaragoza; 1996:34-36.

Leone-Perkins ML, Dillon GF, Walsh W. Examinee perceptions of the usefulness of performance feedback on an examination for medical licensure. *Academic Medicine*. 1996;71(suppl 10):88-90.

Melnick DE. The experience of the National Board of Medical Examiners: success seems always just over the horizon. In: *Computer-based Examination for Board Certification*. Evanston, IL: American Board of Medical Specialties; 1996:11-120.

Morrison C. Predicting academic performance in college: an investigation of the utility of the graded response model and the partial credit model for scaling first course grades. In: Engelhard G, Wilson M, ed. *Objective Measurement - Theory into Practice, Vol 3*. Norwood, NJ: Ablex; 1996:45-64.

Moser GR. Choosing the right NOS for intranet application development: UNIX vs NT. *InternetWork*. 1996;7(12):37.

Norman GR, Swanson DB, Case SM. Conceptual and methodological issues in studies comparing assessment formats. *Teaching and Learning in Medicine*. 1996;8:208-216.

Ripkey DR, Case SM, Swanson DB. A "new" item format for assessing aspects of clinical competence. *Academic Medicine*. 1996;71(suppl 10):34-36.

Ross LP, Clauser BE, Margolis MJ, Orr NA, Klass DJ. An expert-judgment approach to setting standards for a standardized-patient examination. *Academic Medicine*. 1996;71(suppl 10):4-6.

Swanson DB, Bowles LT. Letter to the editor. *Evaluation and the Health Professions*. 1996;19:412-419.

Swanson DB, Bowles LT. Legal vulnerability of the United States Medical Licensing Examination. *Evaluation and the Health Professions*. 1996;19:412-422.

Swanson DB, Case SM, Koenig J, Killian CD. Preliminary study of the accuracies of the old and new medical college admission tests for predicting performance on USMLE Step 1. *Academic Medicine*. 1996;71(suppl 1):25-27.



Swanson DB, Case SM, Luecht RM, Dillon GF. Retention of basic science information by fourth-year medical students. *Academic Medicine*. 1996;71(suppl 10):80-82.

Templeton B. Reply to Swanson and Bowles. *Evaluation and the Health Professions*. 1996;19:420-422.

Templeton B. USMLE Step 1 Examination - legal vulnerability. *Evaluation and the Health Professions*. 1996;19:131-147.

---

## 1995 PUBLICATIONS

Bowles LT. Assessment - new skills, new approaches, new opportunities beyond standardized testing. In: *Proceedings of the Sixth Ottawa Conference on Medical Education*. Toronto, ON; 1995:5-8.

Bowles LT. A worthy search - the development of the key-features concept. *Academic Medicine*. 1995;70:89-90.

Bowles LT. Barriers and opportunities. In: *Changing Medical Education*. Washington, DC: Institute of Medicine; 1995:45-48.

Bowles LT. Recommendations for emergency medicine [comment]. *Annals of Emergency Medicine*. 1995;25:234-235.

Case SM. "New" evaluation techniques in the era of the primary care agenda. *CREOG and APGO Annual Meeting Syllabus*. 1995:11-18.

Case SM, Swanson DB. Principles of writing extended matching items. In: *Proceeding of the Annual Academy of Neurology*. 1995:15-22.

Case SM, Swanson DB. Principles of writing multiple choice questions. In: *Proceeding of the Annual Academy of Neurology*. 1995:23-32.

Case SM, Swanson DB. Validity of scores on the U.S. licensing examination for predicting performance on the dermatology certifying examination. In: *Proceedings of the Sixth Ottawa Conference on Medical Education*. Toronto, ON; 1995:384-386.

Case SM, Swanson DB, Ripkey DR. Relationship between achievement in the clinical science clerkships and performance on Step 2 of the USMLE licensing examination. In: *Proceedings of the Sixth Ottawa Conference on Medical Education*. Toronto, ON; 1995:113-115.



Cizek CJ, Webb LC, Kalohn JC. The use of cognitive taxonomies in licensure and certification test development: reasonable or customary. *Evaluation and the Health Professions*. 1995;18:77-91.

Clauser BE, Clyman SG, Margolis MJ, Ross LP. Are fully complementary models appropriate for setting standards on performance assessments of clinical skills? *Academic Medicine*. 1995;71(1 Suppl):S90-S92.

Clauser BE, Orr NA, Clyman SG. Models for making pass/fail decisions for performance assessments involving multiple cases. In: *Proceedings of the Sixth Ottawa Conference on Medical Education*. Toronto, ON; 1995:239-242.

Clauser BE, Subhiyah R, Nungester RJ, Ripkey DR, Clyman SG, McKinley D. Scoring a performance-based assessment by modeling the judgments of experts. *Journal of Educational Measurement*. 1995;32:397-415.

Clyman SG, Melnick DE, Clauser BE. Computer-based case simulations. In: Mancall EL, Bashook PG, ed. *Assessing Clinical Reasoning: the Oral Examination and Alternative Methods*. Chicago, IL: American Board of Medical Specialties; 1995:139-150.

Crocker PRE, Bouffard M, Gessaroli ME. Measuring enjoyment in youth sport settings: a confirmatory factor analysis of the physical activity enjoyment scale. *Journal of Sport and Exercise Psychology*. 1995(17):200-205.

Dawson B, Iwamoto CK, Ross LP, Nungester RJ, Swanson DB, Volle RL. Performance on the NBME Part I examination [letters and reply]. *Journal of the American Medical Association*. 1995;273:617-618.

Finkbiner R, Fletcher EA, Orr NA, Klass DJ. Question format and scoring methods for standardized patient interstation exercises. In: *Proceedings of the Sixth Ottawa Conference on Medical Education*. Toronto, ON; 1995:343-345.

Fletcher EA, Klass DJ. The National Board of Medical Examiner's standardized patient project update. *Medical Encounter*. 1995;11(2):4-5.

Fletcher EA, Klass DJ, Clauser BE, Errichetti A, Finkbinder RG, King AM, Orr NA, Ross LP. NBME standardized patient project update. In: *Proceedings of the Sixth Ottawa Conference on Medical Education*. Toronto, ON; 1995:684.

Golden GS. Attention deficit disorder. In: Robertson MM, Eapen V, ed. *Movement and Allied Disorders in Childhood*. West Sussex, UK: John Wiley & Sons; 1995:57-67.

Golden GS. Neurological manifestations of congenital heart disease. In: Aminoff MJ, ed. *Neurology and General Medicine, 2d ed*. New York, NK: Churchill Livingstone; 1995:67-75.

Grum CM, Woolliscroft JO, Case SM, Swanson DB, Ripkey DR. Impact of block assignments on development of diagnostic skills in a medicine clerkship. In: *Proceedings of The Sixth Ottawa Conference on Medical Education and Assessment*. 1995:467-470.

Klass DJ. Review of "The Certification and Recertification of Doctors: Issues in the Assessment of Clinical Competence.". *Teaching and Learning in Medicine*. 1995;7:246.

Klass DJ, Clauser BE, Fletcher EA, Finkbiner R, Errichetti A, King AM, Orr NA, Ross LP. Progress in developing a standardized patient test of clinical skills at the National Board of Medical Examiners: prototype two. In: *Proceedings of the Sixth Ottawa Conference on Medical Education and Assessment*. Toronto, ON; 1995:324-326.

LaDuca A. Setting performance standards for licensing examinations: standardized patients and the professional perspective. In: *Proceedings of the Sixth Ottawa Conference on Medical Education and Assessment*. Toronto, ON; 1995:348-350.

Mazor KM, Kanjee A, Clauser BE. Using logistic regression and the Mantel-Haenszel with multiple ability estimates to detect differential item functioning. *Journal of Educational Measurement*. 1995;32:131-144.

Orr NA, Clauser BE, Ross LP, Clyman SG. A comparison of pass/fail decisions made with CBX and NBMCE Comprehensive Part II. In: *Proceedings of the Sixth Ottawa Conference on Medical Education*. Toronto, ON; 1995:197-200.

Primak ME, Kheyfets BL. A modification of the inscribed ellipsoid method. *Mathematical and Computer Modeling*. 1995;21(11):69-76.

Ripkey DR, Case SM. The hare versus the tortoise: do those who complete tests quickly do better or worse? In: *Proceedings of the Sixth Ottawa Conference on Medical Education*. Toronto, ON; 1995:288-290.

Scheuneman JD. Development of performance assessments for use in professional certification and licensing. *CLEAR Exam Review*. 1995;VI(2):20-24.

Swanson DB, Case SM. Item difficulty and discrimination by item format on Part 1 (Basic Sciences) and Part II (Clinical Sciences) of U.S. licensing examinations. In: *Proceedings of the Sixth Ottawa Conference on Medical Education*. Toronto, Ontario; 1995:285-287.

Woolliscroft JO, Swanson DB, Case SM, Ripkey DR. Monitoring the effectiveness of the clinical curriculum: use of a cross-clerkship exam to assess development of diagnostic skills. In: *Proceedings of the Sixth Ottawa Conference on Medical Education*. Toronto, Ontario; 1995:476-478.

---

## 1994 PUBLICATIONS

Becker DF, Forsyth RA. Gender differences in mathematics problem solving and science: a longitudinal analysis. *International Journal of Educational Research*. 1994;21:407-416.

Case SM. The use of imprecise terms in examination questions: how frequent is frequently? *Academic Medicine*. 1994;69(10 Suppl):S4-S6.

Case SM, Bowmer I. Licensure and specialty board certification in North America: background information and issues. In: Newble DI, Jolly B, Wakefield R, eds. *The Certification and Recertification of Doctors: Issues in the Assessment of Clinical Competence*. New York, NY: Cambridge University Press; 1994:19-27.

Case SM, Swanson DB, Ripkey DR. Comparison of items in five-option and extended matching formats for assessment of diagnostic skills. *Academic Medicine*. 1994;69(10 Suppl):S1-S3.

Clauser BE. Book review: Differential Item Functioning. *Journal of Educational Measurement*. 1994;31:88-92.

Clauser BE, Hambleton RK. Review of Holland, PW and Wainer H, eds: Differential Item Functioning. *Journal of Educational Measurement*. 1994;31:88-92.

Clauser BE, Mazor KM, Hambleton RK. The effects of score group width on the Mantel-Haenszel procedure. *Journal of Educational Measurement*. 1994;31:67-78.

Clauser BE, Ross LP, Fletcher EA, Klass DJ, Finkbiner RG, King AM. Differential item functioning in checklist items from a standardized patient-based examination. *Academic Medicine*. 1994;69(10 Suppl):S72-S74.

Clyman SG, Berksey A. Processing examinee free-text entries and authoring tools for patient care simulations. In: *Proceedings of the Educational Testing Service Conference on Natural Language Processing Techniques and Technology in Assessment and Education*. 1994:73-79.

Dauphinee D, Case SM, Fabb W, McAvoy P, Saunders N, Wakeford R. Standard setting for recertification. In: Newble DI, Jolly B, Wakefield R, eds. *The Certification and Recertification of Doctors: Issues in the Assessment of Clinical Competence*. New York, NY: Cambridge University Press; 1994:210-215.

Dawson B, Iwamoto CK, Ross LP, Nungester RJ, Swanson DB, Volle RL. Performance on the National Board of Medical Examiner's Part I examination by men and women of different race and ethnicity. *JAMA*. 1994;272:674-679.

deLalmerens-Pratt M, Golden GS. Teamwork in medical settings. In: Garner HG, Orelove FP, ed. *Teamwork in Human Services: Models and Applications across the Life Span*. Boston, MA: Butterworth-Heinemann; 1994:159-177.

Fitzgerald JT, Wolf FM, Davis WK, Barclay ML, Bozynski ME, Chamberlain KR, Clyman SG, Shope TC, Woolliscroft JO, Zelenock GB. A preliminary study of the impact of case specificity on computer-based assessment of medical student clinical performance. *Evaluation and the Health Professions*. 1994;17:307-321.

Fletcher EA, Klass DJ, Clauser BE, Errichetti A, Finkbiner R, King AM, Orr NA, Ross LP. NBME standardized patient project update. *The Sixth Ottawa Conference on Medical Education*. 1994:684.

Garibaldi RA, Trontell MC, Waxman H, Holbrook JH, Kanya DT, Khosbin S, Thompson J, Casey M, Subhiyah R, Daidoff F. The In-Training Examination in Internal Medicine. *Annals of Internal Medicine*. 1994;121:117-123.

Golden GS. The role of evaluation on behavioral science training. *Annals of Behavioral Science and Medical Education*. 1994;1:19-25.

Grum CM, Case SM, Swanson DB, Woolliscroft JO. Identifying the trees in the forest: characteristics of students who demonstrate disparity between knowledge and diagnostic-recognition skills. *Academic Medicine*. 1994;69(10 Suppl):S66-S68.

King AM, Perkowski-Rogers LC, Pohl HS. Planning standardized patient programs: case development, patient training and costs. *Teaching and Learning in Medicine*. 1994;6:6-14.

Klass DJ. Audience questions and panelists responses: defining an agenda for validation research for professional licensure and certification examinations. *Evaluation & Health Professions*. 1994;17:236-241.

Klass DJ. "High stakes" testing of medical students using standardized patients. *Teaching and Learning in Medicine*. 1994;6:28-32.

Klass DJ, Clauser BE. Evaluating clinical skills - getting it right slowly. *Archives of Pediatrics and Adolescent Medicine*. 1994;148:133-134.

LaDuca A. Defining an agenda for validation research for professional licensure and certification examinations. *Evaluation & the Health Professions. Special issue*. 1994;17(2).

LaDuca A. Introduction. *Evaluation & the Health Professions*. 1994;17(131-132).

LaDuca A. Validation of professional licensure examinations. *Evaluation & the Health Professions*. 1994;17:178-197.

Lopez S. No internetwork is an island. *Internetwork*. 1994;5(9):45.

Mazor KM, Clauser BE, Hambleton RK. Identification of nonuniform differential item functioning using a variation of the Mantel-Haenszel procedure. *Educational and Psychological Measurement*. 1994;54:284-291.

Newble D, Dauphinee D, Dawson B, MacDonald M, Mulholland H, Page G, Swanson DB, Thomson A, van der Vleuten CPM. Guidelines for assessing clinical competence. *Teaching and Learning in Medicine*. 1994;6:213-220.

Scheuneman JD, Bleistein CA. Item bias. In: *International Encyclopedia of Education*. 2nd ed. New York, NY: Pergamon Press; 1994;5:3034-3051.

van der Vleuten CPM, Newble D, Case SM, Holsgrove G, McCann B, McRae C, Saunder N. Methods of assessment in certification. In: Newble DI, Jolly B, Wakefield R, eds. *The Certification and Recertification of Doctors: Issues in the Assessment of Clinical Competence*. New York, NY: Cambridge University Press; 1994:105-125.

Woolliscroft JO, Howell JD, Patel BP, Swanson DB. Resident-patient interactions: the humanistic qualities of internal medicine residents assessed by patients, attending physicians, program supervisors and nurses. *Academic Medicine*. 1994;69:216-223.

---

## 1993 PUBLICATIONS

Bowles LT. Commentary: use of NBME and USMLE scores. *Academic Medicine*. 1993;68:778.

Case SM. Written assessment in the 1990's: some biased opinions from the USA. In: *Proceedings of the National Symposium on the Changing Context of Assessment in Medicine in Australia*. 1993:33-40.

Case SM, Becker DF, Swanson DB. Performances of men and women on NBME Part I and Part II: the more things change. *Academic Medicine*. 1993;68(10 Suppl):S25-S27.

Case SM, Swanson DB. Validity of NBME Part I and Part II scores for selection of residents in orthopedic surgery, dermatology, and preventive medicine. In: Gonnella J, Hojat M, Erdmann J, Veloski J,

eds. *Assessment Measures in Medical School, Residency and Practice*. New York, NY: Springer; 1993:101-114.

Case SM, Swanson DB. Extended matching items: a practical alternative to free-response questions. *Teaching and Learning in Medicine*. 1993;5:107-115.

Clauser BE, Clyman SG. A contrasting group's approach to standard setting for performance assessments of clinical skills. *Academic Medicine*. 1993;69(10 Suppl):S42-S44.

Clauser BE, Mazor KM, Hambleton RK. The effects of purification of the matching criterion on the identification of DIF using the Mantel-Haenszel procedure. *Applied Measurement in Education*. 1993;6:269-279.

Clauser BE, Piemme TE, Clyman SG, Ripkey DR, Orr NA. A comparison of pass/fail classification made with scores from the NBME standardized patient examination and Part II examination. *Academic Medicine*. 1993;68(10 Suppl):S7-S9.

Clauser BE, Subhiyah R, Piemme TE, Clyman SG, Ripkey DR, Nungester RJ. Using clinician ratings to model score weights for a computer-based simulation performance assessment. *Academic Medicine*. 1993;68(10 Suppl):S64-S66.

Fahn S, Bruun RD, Caine E, Cohen DJ, Comings DE, Como PG, Canneally PM, Goetz C, Golden GS, Jankovic J, Kurlan R, LeWitt P, Pauls D, Riddle MA, Shapiro AK, Singer HS. Definitions and classification of tic disorders. *Archives of Neurology*. 1993;50:1013-1016.

Golden GS. The national childhood vaccine injury act: an update. *Contemporary Pediatrics*. 1993;10(10):96-105.

Golden GS. Tics and Tourette syndrome. In: Burg FD, Ingelfinger JR, Wald ER, eds. *Gellis & Kagan's Current Pediatric Therapy 14*. Philadelphia, PA: WB Saunders; 1993:26-28.

Golden GS. Treatment of attention deficit hyperactivity disorder. In: Kurlan R, ed. *Handbook of Tourette's Syndrome and Related Tic and Behavioral Disorders*. New York, NY: Marcel Dekker; 1993:423-430.

Hambleton RK, Clauser BE, Mazor KM, Jones RW. Advances in the detection of differentially functioning test items. *European Journal of Psychological Assessment*. 1993;9:1-18.

LaDuca A, Melnick DE. Status of the USMLE Step 3 Examination. *Federation Bulletin*. 1993;80:38-41.



Swanson DB, Case SM, Waetcher D, Veloski JJ, Hasbrouck C, Friedman M, Carline J, MacLaren C. A preliminary study of the validity of pass/fail standards for USMLE Step 1 and 2. *Academic Medicine*. 1993;68(suppl 10):19-21.

---

## 1992 PUBLICATIONS

Becker DF, Forsyth RA. An empirical investigation of Thurstone and IRT methods of scaling achievement tests. *Journal of Educational Measurement*. 1992;29:341-354.

Becker DF, Swanson DB, Case SM, Nungester RJ. Results of the initial administration of the NBME comprehensive Part I and Part II examinations. *Academic Medicine*. 1992;67(10 Suppl):S16-S18.

Bowles LT. Evaluation for medical licensure. *Federation Bulletin*. 1992;79(4):54-62.

Case SM, Becker DF, Swanson DB. Relationship between scores on NBME basic science tests and the first administration of the newly designed NBME Part I examination. *Academic Medicine*. 1992;67(10 Suppl):S13-S15.

Case SM, Samph T, Templeton T, Best AM. Comparison of observation-based and chart-based scores derived from standardized patient encounters. In: Harden RM, Hart IR, Mulholland H, eds. *Approaches to the Assessment of Clinical Competence: Fifth Ottawa Conference*. Dundee, UK: Centre for Medical Education; 1992:471-475.

Case SM, Swanson DB. Assessment of diagnostic SP-based exams. In: Hart I, Harden RM, Des Marchais J, eds. *Current Developments in Assessing Clinical Competence*. Montreal, Canada: Can-Heal Publications; 1992:220-225.

Case SM, Swanson DB, Woolliscroft JO. Assessment of diagnostic pattern recognition skills in medicine clerkship using a written test. In: Harden RM, Hart IR, Mulholland H, eds. *Approaches To the Assessment of Clinical Competence : Fifth Ottawa Conference*. Dundee, UK: Centre for Medical Education; 1992:452-458.

Cernius V, Errichetti AM, Kociunas R, Saunders E, Suslavicius A. A comparative exploration of identity consciousness and goals of Vilnius University (Lithuania) and Temple University (Philadelphia) education students. In: Cernius V, ed. *Mokytojo Pagalbininkas (Teacher's Helper)*. Kaunas, Lithuania: Littera Universitatis Vytauti Magni; 1992.

Clyman SG, Klass DJ. Standardized patients and computer simulations in the assessment of physicians. *Proceedings of the 1992 ETS Invitational Conference*. 1992:9-17.



Dillon GF, Clyman SG. The computerization of clinical science examinations and its effect on the performances of third-year medical students. *Academic Medicine*. 1992;67(10 Suppl):S66-S68.

Downing SM. True-false, alternative-choice, and multiple-choice items. *Educational Measurement: Issues and Practice*. 1992;11(3):27-30.

Julian ER, Orr NA. Psychometric issues in the use of simulations and work samples as examinations. *CLEAR Exam Review*. 1992;3(2):22-25.

Klass DJ, Fletcher EA, King AM, Durinzi DM, Nungester RJ, Clauser BE, Ripkey DR. Developing a standard patient test of clinical skills at the National Board of Medical Examiners. In: Harden RM, Hart IR, Mulholland H, eds. *Approaches to the Assessment of Clinical Competence: Fifth Ottawa Conference*. Dundee, UK: Centre for Medical Education; 1992:58-70.

Klass DJ, LaDuca A, Barrows HS, Yu NV. Planning and blueprinting clinical practice examinations. *Academic Medicine*. 1992;67(suppl 10):76.

Kopelow ML, Schnabl GK, Hassard TH, Tamblyn RM, Klass DJ. Assessing practicing physicians in two settings using standardized patients. *Academic Medicine*. 1992;67(suppl 10):19-21.

Mazor K, Clauser BE, Hambleton RM. The effect of sample size on the functioning of the Mantel-Haenszel statistic. *Educational and Psychological Measurement*. 1992;52:443-451.

Piemme TE, Pincetl PS, Malakoff GL, Clyman SG, Julian ER, Case SM, Swanson DB, Cotton KE, el-Bayoumi J, Change L. Use of expert judgment to validate a scoring algorithm in assessing performance on computer simulations. In: *Proceedings of the Seventh World Congress on Medical Informatics, MEDINFO*. 1992:1128-1133.

Piemme TE, Pincetl PS, Malakoff GL, Clyman SG, Julien ER, Case SM, Swanson DB. Validity of an algorithm for scoring computerized patient simulations. In: Harden RM, Hart IR, Mulholland H, eds. *Approaches to the Assessment of Clinical Competence: Fifth Ottawa Conference*. Dundee, UK: Centre for Medical Education; 1992:694-699.

Pincetl PS, Malakoff GL, Clyman SG, Julian ER, Piemme TE. Comparison of computer simulations, multiple-choice testing and faculty observation in the assessment of clinical competence. In: Harden RM, Hart IR, Mulholland H, eds. *Approaches to the Assessment of Clinical Competence: Fifth Ottawa Conference*. Dundee, UK: Centre for Medical Education; 1992:700-705.

Reznick R, Baumber J, Cohen R, Chakmers A, Swanson DB. An objective structured clinical examination for licensure. In: Harden RM, Hart IR, Mulholland H, eds. *Approaches to the Assessment of Clinical Competence: Fifth Ottawa Conference*. Dundee, UK: Centre for Medical Education; 1992:71-77.

Reznick R, Smee S, Rothman A, Chalmers A, Swanson DB, Dufresne L. An objective structured clinical examination for the licentiate: Report of the Pilot Project of the Medical Council of Canada. *Academic Medicine*. 1992;67:487-494.

Sutnick AI, Ross LP, Wilson MP. Assessment of clinical competencies by the Foreign Medical Graduate Examination in the Medical Sciences. *Teaching and Learning in Medicine*. 1992;4:150-155.

Swanson DB, Case SM. Trends in written assessment: a strangely biased perspective. In: Harden RM, Hart IR, Mulholland H, eds. *Approaches to the Assessment of Clinical Competence : Fifth Ottawa Conference*. Dundee, UK: Centre for Medical Education; 1992:38-53.

Swanson DB, Case SM, Melnick DE, Volle FL. Impact of the USMLE Step I on teaching and learning of the basic biomedical sciences. *Academic Medicine*. 1992;67:553-556.

Swanson DB, Haynes R, Killian C, Regan M, Stillman P, Case SM. Validity of undergraduate college GPAs and MCAT scores for predicting performance on a clinical skills examination. In: Harden RM, Hart IR, Mulholland H, eds. *Approaches to the Assessment of Clinical Competence : Fifth Ottawa Conference*. Dundee, UK: Centre for Medical Education; 1992:465-470.

Woolliscroft JO, Swanson DB, Case SM. Validity of extended matching and short answer response formats with pattern recognition items. In: Harden RM, Hart IR, Mulholland H, eds. *Approaches to the Assessment of Clinical Competence: Fifth Ottawa Conference*. Dundee, UK: Centre for Medical Education; 1992:459-464.

---

## 1991 PUBLICATIONS

Christensen C, King AM, Fetzer B. Medical students' reactions to AIDS: influence of patient characteristics on hypothetical treatment decisions. *Teaching and Learning in Medicine*. 1991;3:138-142.

Frisbie DA, Becker DF. An analysis of textbook advice about true-false tests. *Applied Measurement in Education*. 1991;4:67-83.

Gottesman LE, Peskin E, Kennedy KM. Research and program experience in residential care facilities: implications for mental health services to elderly and middle-aged clients. In: Light E, Lebowitz BD, ed. *The Elderly with Chronic Mental Illness*. New York NY: Springer; 1991:229-245.

Gottesman LE, Peskin E, Kennedy KM, Mossey J. Implications of a mental health intervention for elderly mentally ill residents of residential care facilities. *International Journal of Aging and Human Development*. 1991;32:229-245.

Iwamoto CK, Volle RL. Performance on the National Board of Medical Examiners (NBME) Part I and the pharmacology subtest 1986-1990. *The Pharmacologist*. 1991;33(4):279-281.

Julien ER, Wright BD. Distinguishing between shared and unique employee needs. In: Wilson M, ed. *Objective Measurement: Theory into Practice*. Norwood NJ: Ablex Publishing; 1991.

Klass DJ. Standardized patients in clinical assessment: experience at Southern Illinois University and the University of Manitoba. *Federation Bulletin*. 1991;78(2):35-43.

Nungester RJ, Dillon GF, Swanson DB, Orr NA, Powell RD. Standard setting plans for the NBME Comprehensive Part I and Part II examinations. *Academic Medicine*. 1991;66:429-433.

Orr NA, Nungester RJ. Assessment of constituency opinion about NBME examination standards. *Academic Medicine*. 1991;66:465-70.

Page G, Case SM, Macguire T, Swanson DB. Selecting and implementing standard setting procedures. *Academic Medicine*. 1991;66(suppl 10):85.

Rettie CS. Evaluating the "at risk" physician. *Federation Bulletin*. 1991;78:365-371.

Ross DW, Melnick DE. An inventory of the personal computers for students' use at 143 U.S. and Canadian medical schools. *Academic Medicine*. 1991;66:232-234.

Stillman PL, Swanson DB, Regan MB. Clinical skills of foreign medical graduates: Letter to editor and response. *Annals of Internal Medicine*. 1991;115:158-159.

Stillman PL, Swanson DB, Regan MB, Philbin MM, Nelson VE, Ebert T. Assessment of clinical skills of residents utilizing standardized patients - a follow-up study and recommendations for application. *Annals of Internal Medicine*. 1991;115:393-401.

Swanson DB, Case SM, Kelley PR, Lawley JL, Nungester RJ, Powell RD, Volle RL. Phase-in of the NBME comprehensive Part I examination. *Academic Medicine*. 1991;66:443-444.

Swanson DB, Case SM, Nungester RJ. Validity of NBME Part I and Part II scores in prediction of Part II performance. *Academic Medicine*. 1991;66(suppl 10):7-9.

Swanson DB, Case SM, van der Vleuten CPM. Strategies for student assessment. In: Boud D, Feletti GI, ed. *The Challenge of Problem-Based Learning*. London UK: Kagan Page Limited; 1991:260-273.

Tamblyn RM, Klass DJ, Schnabl GK, Kopelow ML. The accuracy of standardized patient presentation. *Medical Education*. 1991;25:100-109.

Tamblyn RM, Klass DJ, Schnabl GK, Kopelow ML. Sources of unreliability and bias in standardized patient rating. *Teaching and Learning in Medicine*. 1991;3:74-85.

Volle RL. Nicotine and ganglion-blocking drugs. In: Smith CM, Reynard AM, ed. *Textbook of Pharmacology*. Philadelphia PA: W.B. Saunders; 1991:119-126.

Wheat JR, Killian CD, Melnick DE. Reevaluation of medical education. A behavioral model to assess health promotion/disease prevention instruction. *Evaluation and the Health Professions*. 1991;14:305-318.

---

## 1990 PUBLICATIONS

Clyman SG. Medical schools testing computer-based exam. *Computer News for Physicians*. 1990.

Clyman SG, Orr NA. Status report on NBME computer-based testing. *Academic Medicine*. 1990;65:235-241.

Cotten KE, Lawley JL. In *Service to Medicine: A Special Review*. Philadelphia PA: National Board of Medical Examiners; 1990.

Dawson-Saunders B, Feltovich PJ, Coulson RL, Steward DE. Survey of medical school teachers to identify basic biomedical concepts medical students should understand. *Academic Medicine*. 1990;65:448-454.

Dawson-Saunders B, Iwamoto CK, Volle RL. Performance on the National Board of Medical Examiners (NBME) Part I and the Pharmacology Subtest 1986-1989. *The Pharmacologist*. 1990;34(4):224-229.

Klass DJ. Performance-based assessment: plans of the National Board of Medical Examiners. *GEA Correspondent*. 1990;3(1):3-4.

Klass DJ, Abrahamowicz M, Tamblyn RM, Ramsey JO, Kopelow ML. Detecting and correcting for rater-induced differences in standardized patient tests of clinical competence. *Academic Medicine*. 1990;65(suppl):55-56.

Klass DJ, Tamblyn RM, Schanbl GK, Kopelow ML. Factors associated with the accuracy of standardized patient presentation. *Academic Medicine*. 1990;65(suppl):25-26.

LaDuca A, Engel JD, Wigton R, Blacklow RS. A social judgment theory perspective on clinical problem-solving. *Evaluation & the Health Professions*. 1990;13:63-78.

Melnick DE. Computer-based simulation: state of the art. *Evaluation and the Health Professions*. 1990;13:104-120.

Nungester RJ, Dawson-Saunders E, Kelley PR, Volle RL. Score reporting on NBME examinations. *Academic Medicine*. 1990;65:723-729.

Swanson DB. Issues in assessment of practical skills in medicine. *Professions Education Researcher Quarterly*. 1990;12(2):3-6.

Swanson DB, Case SM, Stillman PL, Regan MB, McCahan J, Feinblatt J, Smith SR. An assessment of the clinical skills of fourth-year students at four New England medical schools. *Academic Medicine*. 1990;65:320-326.

Swanson DB, Dillon GF, Ross LP. Setting content-based standards for National Board exams: initial research for the comprehensive Part I examination. *Academic Medicine*. 1990;65(suppl 10):17-18.

Swanson DB, Stillman PL. Use of standardized patients for teaching and assessing clinical skills. *Evaluation and the Health Professions*. 1990;13:79-103.

Swanson DB, Van der Vleuten CPM. Assessment of clinical skills with standardized patients: state of the art. *Teaching and Learning in Medicine*. 1990;2:58-76.

Volle RL. Standardized testing of patient management skills. *Clinical Orthopaedics and Related Research*. 1990;257:47-51.

---

## 1980 TO 1989

### Leaning into Technology: 1980s

In the mid-1980s, staff moved toward writing advanced psychometric and measurement theory publications. NBME staff addressed computer-based testing with research and publications on the use of computers, not only to administer traditional multiple-choice tests, but also on the application of artificial intelligence and expert systems to assess the clinical reasoning of medical students.

## 1989 PUBLICATIONS

Haladyna TM, Downing SM. Validity of a taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education*. 1989;2:51-78.

Haladyna TM, Downing SM. Taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education*. 1989;2:37-50.

Volle RL. Licensure examinations - today and tomorrow. *Federation Bulletin*. 1989;76:35-39.

Volle RL. Single examination route to licensure: the National Board perspective. *Federation Bulletin*. 1989;76:355-364.

---

## 1988 PUBLICATIONS

Clyman SG, Melnick DE. Computer-based simulations in the evaluation of physicians' clinical competence. *Machine Mediated Learning*. 1988;2:257-369.

Grosse ME, Wright BD. Psychometric characteristics of scores on a patient management problem test. *Educational and Psychological Measurement*. 1988;48:297-305.

Julian ER, Wright BD. Using computerized patient simulations to measure the clinical competence of physicians. *Applied Measurement in Education*. 1988;1:299-318.

LaDuca A, Engel JD, Chovan JD. An exploratory study of physicians' clinical judgment: an application of social judgment theory. *Evaluation & the Health Professions*. 1988;7:178-200.

Melnick DE, Clyman SG. Computer-based simulations in the evaluation of physicians' clinical competence. *Machine-Mediated Learning*. 1988;2:257-269.

Swanson DB, Webster GD, Shea JA, Norcini JJ, Grosso LJ. Strategies in comparison of methods for scoring patient management problems: use of external criteria to validate scores. *Evaluation and the Health Professions*. 1988;11:231-248.

Volle RL. Using National Board of Medical Examiners scores in selection of residents [editorial]. *Journal of the American Medical Association*. 1988;259:266.

Volle RL. The National Board of Medical Examiners scores in selection of residents. *Journal of the American Medical Association*. 1988;259:266.

Volle RL. The National Boards of the future. *Resident & Staff Physician*. 1988;34:63-64.

---

## 1987 PUBLICATIONS

Maatsch JL, Huang RR, Downing SM. Examiner assessments of clinical performance: what do they tell us about clinical competence. *Evaluation and Program Planning*. 1987;10:13-17.

Melnick DE. Clinical simulations - Pygmalion revisited? In: Stead WW, ed. *Symposium on Computer Applications in Medical Care (SCAMC)*. 1987:7-9.

Swanson DB, Norcini JJ, Grosso LJ. Assessment of clinical competence: written and computer-based simulations. *Assessment and Evaluation in Higher Education*. 1987;12:220-246.

---

## 1986 PUBLICATIONS

Asper SP, Levit EJ. Residencies for foreign medical graduates. [letter]. *New England Journal of Medicine*. 1986;314:1324.

Giannini G, Engel JD. On the meaning of scores derived from patient management problems. *Evaluation & the Health Professions*. 1986;9:103-120.

Grosse ME. Scores based on dangerous responses to multiple-choice items. *Evaluation & the Health Professions*. 1986;9:459-466.

Grosse ME, Wright JD. Setting, evaluating, and maintaining certification standards with the Rasch model. *Evaluation & the Health Professions*. 1986;9:459-466.

LaDuca A, Staples WI, Templeton B. Item modeling procedure for constructing content-equivalent multiple-choice questions. *Medical Education*. 1986;20:53-56.

Vaughan VC. Eponyms. Letter to the editor. *Journal of the American Medical Association*. 1986;255:1879.



Vaughan VC. In reply. Letter to the editor. *Journal of the American Medical Association*. 1986;256:1295-1296.

---

## 1985 PUBLICATIONS

Carson JD. Cheating on licensing examinations - a legal perspective. *Federation Bulletin*. 1985;72:35-42.

Carson JD. The price of cheating on licensing exams. *Resident & Staff Physician*. 1985;31:155-158, 160.

Case SM. Awarding the gold star: a primer on certification examinations. Special Issue. *Diabetes Educator*. 1985;11:47-51.

Fabrey LJ, Case SM. Further support for changing multiple-choice answers. *Journal of Medical Education*. 1985;60:488-490.

Grosse ME, Wright BD. Validity and reliability of true-false tests. *Educational and Psychological Measurement*. 1985;45:1-13.

Hubbard HP, Levit EJ. *The National Board of Medical Examiners: the First Seventy Years*. Philadelphia PA: National Board of Medical Examiners; 1985.

Norman GR, Swanson DB, Muzzin LJ, Williams RG. Simulation in health sciences education. *Journal of Instructional Development*. 1985;8:11-17.

---

## 1984 PUBLICATIONS

Andrew BJ. Implications of computer testing. In: Lloyd JS, ed. *Computer Applications in the Evaluation of Physician Competence*. Chicago IL: American Board of Medical Specialties; 1984:31-34.

Campbell AB, Glazer DL. Recertification: toward the development of standards for assuring continued competence. *Journal of Allied Health*. 1984;13:252-262.

Case SM, Fabrey LJ, Andrew BJ. Clinical skills needed during early residency. *Resident & Staff Physician*. 1984;20:29-35pc.

Dillon GF. The new FLEX and the old "75". *Federation Bulletin*. 1984;71:214-216.

Erviti VF, Fabrey LJ, Andrew BJ. Computerized medical audit to assess residents' performance in ambulatory care. In: Lloyd JS, ed. *Computer Applications in the Evaluation of Physician Competence*. Chicago, IL: American Board of Medical Specialties; 1984:95-101.

Fabrey LJ, Case SM, Andrews BJ. Assessment of clinical skills in US medical schools. *Journal of Educational Measurement*. 1984;59:957-959.

Jewett RE, Jones JJ, Lawley JL. Graphic presentation of examination content. In: Lloyd JS, ed. *Computer Applications in the Evaluation of Physician Competence*. Chicago IL: American Board of Medical Specialties; 1984:65-71.

Jones JJ, Lawley JL. The test item libraries of the National Board of Medical Examiners. In: *Computer Applications in the Evaluation of Physician Competence*. Chicago IL: American Board of Medical Specialties; 1984:61-63.

Kelley PR, Schumacher CF. The Rasch model: its use by the National Board of Medical Examiners. *Evaluation & the Health Professions*. 1984;7:443-454.

LaDuca A, Taylor DD, Hill IK. The design of a new physician licensure examination. *Evaluation & the Health Professions*. 1984;7:115-140.

Vu NU, Neufeld VR, Andrew BJ, Norcini JJ, Stillman P. Symposium: technical considerations and establishing standards for scoring clinical performance in simulated clinical encounters. *Research in Medical Education*. 1984;23:383-390.

---

## 1983 PUBLICATIONS

Andrew BJ. The limitations of written examinations for licensure. *Federation Bulletin*. 1983:35-42.

Carson JD. Challenges to the integrity of the licensing examination process. *The Bar Examiner*. 1983;52:4-10.

Carson JD. Doctors convicted on criminal charges in connection with licensing examination. *Federation Bulletin*. 1983;70:200-201.

Case SM, Fabrey LJ, Andrew BJ. Critical clinical procedures: a survey of residents. *Research in Medical Education*. 1983;22:160-165.

Daeschner C, Templeton B. FLEX task force I update. *Federation Bulletin*. 1983;70:291-294.

Griffin JB, Hill K, Jones JJ, Keeley KA, Krug R. Evaluating alcoholism and drug abuse knowledge in medical education: a collaborative project. *Journal of Medical Education*. 1983;58:859-863.

Jewett LS, MacDonald M, Templeton B. Evaluating communication skills of physicians: four methods of measurement. *Research in Medical Education*. 1983;22:101-106.

Wesner ME. Test center management. *Federation Bulletin*. 1983;70:116-122.

---

## 1982 PUBLICATIONS

Burg FD, Lloyd JS, Templeton B. Competence in medicine. *Medical Teacher*. 1982;4:60-64.

Saffran M, Kennedy WB, Kelley PR. Retention of knowledge of pharmacology by U.S. and Canadian medical students. *Trends in Pharmacological Sciences*. 1982.

Templeton B, MacDonald M. Use of interactional analysis in assessing physician trainee interpersonal skills. In: Lloyd JS, ed. *Evaluation of Noncognitive Skills and Clinical Performance*. Chicago IL: American Board of Medical Specialties; 1982:155-167.

Vaughan VC. Priorities in changing times. In: *Conference Journal*. Philadelphia PA: Delaware Valley Association for the Education of Young Children; 1982:21-24.

Vaughan VC, Ellis EG. Importance of the Primer for pediatric residents and students. *Journal of the American Medical Association*. 1982;248:2584-2585.

---

## 1981 PUBLICATIONS

Chase RA. Paper-and-pencil examinations - what they can do and cannot do. *Surgery*. 1981;89:771-772.

Hubbard JP. A call for action. *Medical Teacher*. 1981;3:85-86.

Kennedy WB, Kelley PR, Saffran M. Use of NBME examinations to assess retention of basic science knowledge. *Journal of Medical Education*. 1981;56:162-173.

Saffran M, Kennedy WB, Kelley PR. Use of National Board examinations to estimate retention of biochemistry. *Biochemical Education*. 1981;9(3):97-99.

Templeton B. Council on Medical Education and Career Development. *American Journal of Psychiatry*. 1981;134:563-568.

---

## 1980 PUBLICATIONS

Andrew BJ. Customized examinations from the National Board. *Trends in Medical Education*. 1980;24(1):1-2.

Andrew BJ. Can professional competence be measured? *New Directions For Program Evaluation*. 1980;6:39-52.

Burg FD. Objectives of recertification. *Continuing Medical Education Newsletter*. 1980;9(2):5-11.

Downing SM. Assessment of clinical competence on the emergency medicine specialty certification examination. *Annals of Emergency Medicine*. 1980;9:554-556.

Erviti VF, Templeton B, Bunce JV, Burg FD. The relationships of pediatric resident recording behavior across medical conditions. *Medical Care*. 1980;18:1020-1031.

Fabrey LJ, Tjosvold D, Johnson DW. Effects of controversy and defensiveness on cognitive perspective taking. *Psychological Reports*. 1980;47:1043-1053.

Holden WD, Levit EJ. Medical education, licensure and the National Board of Medical Examiners. *New England Journal of Medicine*. 1980;303:1357-1360.

Hubbard JP. Reminiscences and reflections. *Medical Teacher*. 1980;2:279-283.

Levit EJ. Lifelong physician competence. *Journal of the Florida Medical Association*. 1980;67:755-765.

Templeton B. Progress report on the Comprehensive Qualifying Evaluation Program. *Federation Bulletin*. 1980;67:35-38.

Tjosvold D, Fabrey LJ. Motivation for perspective-taking: effects of interdependence on interest in learning others' intentions. *Psychological Reports*. 1980;46:755-765.

Tjosvold D, Johnson DW, Fabrey LJ. Effects of controversy and defensiveness on cognitive perspective-taking. *Psychological Reports*. 1980;47:1043-1053.

Vaughan VC. Introduction. In: Bierman CW, Pearlman DS, ed. *Allergic Diseases of Infancy, Childhood, and Adolescence*. Philadelphia, PA: WB Saunders; 1980.

Vaughan VC. Meeting the health needs of children in the 80's. In: *Conference Journal*. Bryn Mawr PA: Delaware Valley Association for the Education of Young Children; 1980:33-38.

---

## 1970 TO 1979

### Leaning into Technology: 1970s

In the 1970s, staff continued to author materials on topics such as clinical medicine and the role of the NBME in the medical landscape; staff writing expanded to include new NBME exams for various health professionals (ie, pediatricians, physician assistants) and post-licensure assessment. A noteworthy and well-referenced book, *Measuring Medical Education: The Tests and the Experience of the National Board of Medical Examiners*, by John Perry Hubbard and Charles Frederick Schumacher, was published in 1971, with a second edition in 1978.

## 1979 PUBLICATIONS

Brazelton TB, Vaughen VC. *The Family: Setting Priorities*. New York, NY: Science and Medicine Publishers; 1979.

Burg FD, Grosse ME, Kay CT. A national self-assessment program in internal medicine. *Annals of Internal Medicine*. 1979;90:100-107.

Hubbard JP, Ball MJ, Burg FD. Hospital information systems: from the perspective of continuing medical education and individual assessment of physician performance. In: Shannon RH, ed. *Hospital Information Systems: An International Perspective on Problems and Prospects*. Amsterdam, Holland: North-Holland Publishing Company; 1979:341-374.

Levit EJ. Boards, cover art and ethics [letter]. *The New Physician*. 1979;28(6).

Samph T, Templeton B. *Evaluation in Medical Education: Past, Present, Future*. Cambridge MA: Ballinger; 1979.

Templeton B. The National Board of Medical Examiners and independent assessment agencies. In: Samph T, Templeton B, ed. *Evaluation in Medical Education: Past, Present, Future*. Cambridge, MA: Ballinger; 1979.

Templeton B. Forecasts for evaluation in medical education. In: Samph T, Templeton B, ed. *Evaluation in Medical Education: Past, Present, Future*. Cambridge, MA: Ballinger; 1979.

Vaughan V, McKay RJ, Behrman RE. *Nelson Textbook of Pediatrics*. Philadelphia PA: WB Saunders, Inc; 1979.

Vaughan VC. The patient management problems as an evaluative instrument. *Pediatrics in Review*. 1979;1:67-76.

## 1978 PUBLICATIONS

Burg FD, Schumacher CF. Objective tests as measures for medical certification. *Federation Bulletin*. 1978;65:331-339.

Holden WD, Levit EJ. Migration of physicians from one specialty to another: a longitudinal study of U.S. medical school graduates. *Journal of the American Medical Association*. 1978;239:205-209.

Hubbard JP. *Measuring Medical Education: The Tests and the Experience of the National Board of Medical Examiners*. 2nd ed. Philadelphia, PA: Lea & Febiger; 1978.

Hubbard JP. The five hundred year Jubilee celebration - the University of Uppsala and profiled continuing education. *Transactions and Studies of the College of Physicians*. 1978;45:185-195.

Hubbard JP. Profiled continuing education. *Transactions and Studies of the College of Physicians of Philadelphia*. 1978(45):190-195.

Levit EJ, Holden WD. Specialty board certification rates: a longitudinal tracking study of U.S. medical school graduates. *Journal of the American Medical Association*. 1978;239:407-412.

Schumacher CF. The effect of open vs. closed book testing on performance on a multiple-choice examination in pediatrics. *Pediatrics*. 1978;61:256-261.

Vaughan VC. Effect of maternal sedation on mother-infant bonding. In: Kumar S, Rathi M, ed. *Perinatal Medicine*. New York, NY: Pergamon Press; 1978.

Weinberg E, Bell AI. Performance of United States citizens with foreign medical education on standardized medical examinations. *New England Journal of Medicine*. 1978;299:858-862.

Willian MK, Weinberg E, Burnett RD, Olsted RW. The pediatric nurse associate - a model of collaboration between medicine and nursing. *New England Journal of Medicine*. 1978;298:740-741.

---

## 1977 PUBLICATIONS

Andrew BJ. The use of behavioral checklists to evaluate physical examination skills. *Journal of Medical Education*. 1977;52:589-591.

Andrew BJ, Miller RE. View Box exercises for teaching problem solving. *American Journal of Roentgenology*. 1977:271-272.



Bowler FL, Brading PL, Burg FD, Finestone AL, Hubbard JP. A practice related educational program. *Journal of the American Medical Association*. 1977;237:1346-1349.

Brading PL, Bowler FL, Burg FD, Finestone AJ. A practice-related educational program. *JAMA*. 1977;237:1346-1349.

Burri AT, Schumacher CF, Vorkauf H. Feasibility of using an American national board examination for the evaluating of Swiss candidates for licensure. *Medical Education*. 1977;11:276-284.

Chase RA. What to do about the incompetent physicians. *Federation Bulletin*. 1977;64:163-179.

Chase RA, Burg FD. Reexamination/recertification. Measurement of professional competence and relation to quality of medical care. *Archives of Surgery*. 1977;112:19-25.

Dowaliby FJ. The effect of certain rater roles on confidence in physician's assistant ratings. *Journal of Medical Education*. 1977;52:914-919.

Erviti VF. Development of a medical record audit for continuing medical education. *Research in Medical Education*. 1977;16:85-90.

Merchant FT, Kelley PR. Performance of current graduates of United States medical schools on FLEX. *Federation Bulletin*. 1977;64:340-352.

Miller RE, Andrew BJ. View box exercises for teaching problem solving in radiology. *American Journal of Roentgenology*. 1977;128:271-272.

Templeton B. Medical audits and recertification: prospects and problems. *Federation Bulletin*. 1977;64:293-304.

Templeton B, Erviti VF, Bunce JV, Burg FD. Pediatric residents: assessing their performance via chart audit. *Resident and Staff Physician*. 1977.

---

## 1976 PUBLICATIONS

Andrew BJ, Hecht JT. A preliminary investigation of two procedures for setting examination standards. *Educational and Psychological Measurement*. 1976;36:45-50.

Burg FD. Continuing education and recertification: a critical link. *International Journal of Radiation Oncology Biology Physics*. 1976;13:323-327.

- Burg FD, Brownlee R, Wright F, Levine H, Daeschner C, Vaughan V, Anderson J. A method and process for defining competency in pediatrics. *Journal of Medical Education*. 1976;51:824-828.
- Chase RA. The National Board of Medical Examiners. In: Purcell E, ed. *Recent Trends in Medical Education*. New York, NY: Josiah Macy Jr. Foundation; 1976:225-241.
- Chase RA. Proliferation of certification in medical specialties: productive or counterproductive? *New England Journal of Medicine*. 1976;294:497-499.
- Dowaliby FJ. Rater-ratee relationships as related to rater confidence for different domains of competence. *Research in Medical Education*. 1976:103-107.
- Dowaliby FJ, Andrews BJ. Relationship between clinical competence ratings and examination performance. *Journal of Medical Education*. 1976;51:181-188.
- Erviti VF, Bermes E, Forman D. Statistics, normal values and quality control. In: Tietz N, ed. *Fundamentals of Clinical Chemistry*. Philadelphia PA: Saunders; 1976.
- Guerin RO, Smilansky J. The accuracy of absolute minimal acceptable performance levels for multiple-choice examinations. *Journal of Medical Education*. 1976;51:416-417.
- Ludwig H, Noe J, Chase RA. Interactive data analysis. *Computers & Industrial Engineering*. 1976;1:47-56.
- Samph T. Observer effects on teacher verbal classroom behavior. *Educational Psychology*. 1976;68:736-741.
- Samph T, Brodner B, Richman J. Health Systems Agencies Public Accountability Checklist. Washington DC: US Department of Health, Education and Welfare; 1976.
- Templeton B. Recertification: A Look at the Issues. *Task Force on Recertification, Report No. 76*. New York: Group for the Advancement of Psychiatry; 1976.
- Templeton B. Medical accountability and medical education. *Journal of Laboratory and Clinical Medicine*. 1976;88:525-527.
- Templeton B, Erviti VF, Bunce JV, Burg FD. Training medical record abstractors to assure high inter-rater reliability. *Proceedings of the Fifteenth Annual Conference on Research in Medical Education Research*. 1976:108-113.

## 1975 PUBLICATIONS

- Andrew BJ. The effects of patient simulations on actors. *Journal of Medical Education*. 1975;50:87-89.
- Andrew BJ. Interviewing and counseling skills: techniques for their evaluation. *Journal of the American Dietetic Association*. 1975;66:576-580.
- Andrew BJ, Glazer D. Physician's assistant certifying examination [letter]. *Journal of the American Medical Association*. 1975;234:1118.
- Bell AI, Mayer S. Sexism in ratings of personality traits. *Personnel Psychology*. 1975;28:239-249.
- Burg FD. Planning a competency-based approach to recertification. *Federation Bulletin*. 1975;62:280-289.
- Burg FD, Schumacher CF. Computerization of a patient management problem examination to prevent "retracing.". *British Journal of Medical Education*. 1975;9:281-285.
- Chase RA ed. Surgery in the United States. A Summary of the Study on Surgical Services for the United States. Chicago, IL: American College of Surgeons and American Surgical Association; 1975.
- Dustan H, Blumenthal S, Templeton B. Education of physicians in high blood pressure, performance characteristics, learning objectives and evaluation approaches. *Circulation*. 1975;51:9-27.
- Guerin RO. A quasi-simplex analysis of a Piaget-based hierarchy. *Science Education*. 1975;59:273-281.
- Hubbard JP. Objective evaluation of medical education. *Journal of the Irish College of Physicians and Surgeons*. 1975;5(1).
- Levit EJ. The role of graduate training programs in assessing physician competence from the licensure and certification point of view. In: *The Role of the Pediatric Program Director in Board Certification*. Philadelphia, PA: American Board of Pediatrics; 1975:19-28.
- Schumacher CF, Burg FD, Taylor WC. Computerization of a patient management problem examination to prevent "retracing.". *British Journal of Medical Education*. 1975;9:218-285.
- Smith DE. Evaluation in the continuum of medical education: the role of examinations. *AHME Journal*. 1975;8:8-12.
- Smith DE. Recertification for the medical specialist. *Federation Bulletin*. 1975;62:361-369.

Waddell W, Kelley PR, Suter E, Levit EJ. Effectiveness of an international health elective as measured by NBME Part II. *Journal of Medical Education*. 1975;51:468-472.

---

## 1974 PUBLICATIONS

Andrew BJ. The evaluation of physical examination skills: techniques for direct observation and their reliability. *Proceedings of Thirteenth Annual Conference on Research in Medical Education Research*. 1974:30-35.

Andrew BJ. A Methodology for the Development of Examinations to Assess the Proficiency of Health Care Professionals. Philadelphia PA: NBME; 1974.

Andrew BJ. First national certifying examination for primary care physician's assistants. *Federation Bulletin*. 1974;61:298-303.

Andrew BJ. What spells success in PA test - practical experience proves the answer in first certifying exam. *Medical World News*. 1974(April).

Burg FD. Foundation for Evaluating the Competency of Pediatricians. Chicago IL: American Board of Pediatrics; 1974.

Carmichael H, Templeton B, Small S, Kelley PR. Results of the 1972 APA self-assessment program. *American Journal of Psychiatry*. 1974;131:658-661.

Erviti VF, Gordon D, Suvanich S, Schwartz M, Martinez C. The serum calcium level and its significance in hyperthyroidism. *American Journal of Medical Science*. 1974;268:31-36.

Erviti VF, Scott M. Research design and statistics in an undergraduate physical therapy curriculum. *Physical Therapy*. 1974;54:256-259.

Guerin RO, Doran R. An analysis of several instruments measuring "nature of science" objectives. *Science Education*. 1974;58:321-329.

Guerin RO, Doran R, Cavaleri J. Assessment of awareness of environmental problems. *Journal of Environmental Science*. 1974;5:14-18.

Guerin RO, Doran R, Sarnowski A. The effect of perceptual preference of students on their performance on pictorial test items. *Science Education*. 1974;58:161-169.

Hubbard JP. Proposals for changes in National Board examinations. *The Physiologist*. 1974;17:149-154.

Levit EJ, Sabshin C, Mueller C. Trends in graduate medical education and specialty certification: a tracking study of U.S. medical school graduates. *New England Journal of Medicine*. 1974;290:545-549.

McGehee E, Levit EJ, Clark J, Coppola E, Gonnella J. The Philadelphia County Medical Society self-evaluation examination. *Journal of Medical Education*. 1974;49:993-994.

Samph T. Teacher behavior and the reading performance of below-average achievers. *Journal of Educational Research*. 1974;67:268-270.

Samph T. Open education students in transition. *Elementary School Journal*. 1974;75:37-41.

Samph T, Sayles F. A validation study of RACE (Racial Attitude and Cultural Expression). Final Report. Washington DC: US Department of Health Education and Welfare, National Institute of Education; 1974.

Templeton B. Multiple-choice testing in psychiatry. In: Muslin H, Thurnblad R, Templeton B, McGuire C, eds. *Evaluative Methods in Psychiatry*. Washington DC: American Psychiatric Association; 1974.

Templeton B. Evaluating the quality of care. In: Muslin H, Thurnblad R, Templeton B, McGuire C, eds. *Evaluative Methods in Psychiatry*. Washington DC: American Psychiatric Association; 1974.

Templeton B, Harless W. The potential of computer-based simulation of the clinical encounter (CASE) for evaluation of undergraduate psychiatric education. In: Muslin H, Thurnblad R, Templeton B, McGuire C, eds. *Evaluative Methods in Psychiatry*. Washington DC: American Psychiatric Association; 1974.

Templeton B, Hubbard JP. The future of medical education and its implication for psychiatry. In: Usdin G, ed. *Psychiatry: Education and Image*. New York: Brunner/Mazel; 1974.

---

## 1973 PUBLICATIONS

Andrew BJ. National examination program for the certification of physician's assistants. *Federation Bulletin*. 1973;60:189-196.

Andrew BJ. Technique for the assessment of pharmacy students' skills in patient interviewing. *American Journal of Pharmaceutical Education*. 1973;37:290-299.

Hubbard JP. Evaluation, certification and licensure in medicine. *Journal of the American Medical Association*. 1973;225:401-406.

Hubbard JP. The future of medical education and its implication for psychiatry. In: Usdin G, ed. *Psychiatry: Education and Image*. New York: Brunner/Masel, Inc; 1973:105-131.

Levit EJ. A national program for certification of assistants to primary care physicians. In: Lippard VW, Purcell EF, ed. *Intermediate-level Health Practitioners*. Josiah Macy, Jr. Foundation; 1973:170-179.

Schumacher CF. Validation of the American Board of Internal Medicine written examination: a study of the examination as a measure of achievement in GME. *Annals of Internal Medicine*. 1973;78:131-135.

---

## 1972 PUBLICATIONS

Andrew BJ. An approach to the construction of simulated exercise in clinical problem-solving. *Journal of Medical Education*. 1972;47:952-958.

Burg FD, Wright FH. 1971 pretest of the American Board of Pediatrics. *Pediatrics*. 1972;50:462-465.

Burg FD, Wright FH. Evaluation of pediatric residents and their training programs. *Journal of Pediatrics*. 1972;80:183-189.

Kelley PR, Sutnick AI, Knapp D. The English language and the FMG. *Journal of Medical Education*. 1972;47:434-439.

Senior J, Jones NA, Olafson RP, Sutin J. Evaluation of clinical competence: the crux of FLEX. *Federation Bulletin*. 1972;59:303-329.

---

## 1971 PUBLICATIONS

Hubbard JP. Self-assessment programs: summation of objectives. *Federation Bulletin*. 1971;58:69-75.

Hubbard JP. Self-education and self-assessment as a new method for continuing medical education. *Archives of Surgery*. 1971;103:422-424.

Hubbard JP. To change or not to change: a dilemma for the National Board of Medical Examiners. *Journal of the American Medical Association*. 1971;217:1698.

Kelley PR. The numbers game. *Federation Bulletin*. 1971;58:233-240.

Kelley PR, Matthews JH, Schumacher CF. Analysis of the oral examination of the American Board of Anesthesiology. *Journal of Medical Education*. 1971;46:982-988.

Levit EJ. Problems in the evaluation of graduate medical education. In: Gilbert JAL, ed. *Proceedings of*

*Conference on Evaluation in Medical Education*. Edmonton, AB; 1971:111-120.

Schumacher CF. How to tell one figure from another or when is 75 not 75 per cent. *Federation Bulletin*. 1971;58:221-232.

---

## 1970 PUBLICATIONS

Hubbard JP, Levit EJ, Barnett O, Goldfinger SE, Dineen J. Computer-based evaluation of clinical competence. *The Bulletin*. 1970:502-505.

Kelley PR, Stumpe AR, Levit EJ. Four-year study of the internship in the United States Air Force hospitals: an objective measurement of gain in clinical competence. *Military Medicine*. 1970;135:537-545.

---



## 1960 TO 1969

### NBME Hits Its Stride: 1960s

In the 1960s, more publications began to focus on graduate medical education. John P. Hubbard, MD, (president of the NBME from 1949 to 1974) and Edithe J. Levit, MD, (president from 1977 to 1986) were prolific authors. Additional works appeared on evolving testing modalities, and the staff wrote about new NBME offerings such as mini tests and in-training exams.

### 1969 PUBLICATIONS

Dubois AB, Memir P, Schumacher CF, Hubbard JP. Graduate medical school education in basic sciences. *Journal of Medical Education*. 1969;44:1035-1043.

Levit EJ. Evaluation of learning in graduate education. *Journal of Neurosurgery*. 1969;30:348-352.

Levit EJ. In-training evaluation of learning: objective measurement of the product and process of graduate medical education. *Archives of Dermatology*. 1969;99:342-349.

Levit EJ. Testing physicians with motion pictures. *Industrial Photography*. 1969:21-22, 61-65.

---

### 1968 PUBLICATIONS

Baue AE, Schumacher CF, Welch JS, Hubbard JP. Special report: In-training evaluation of surgical residents. *Journal of Surgical Research*. 1968;8:341-344.

Hubbard JP. Role of the National Board of Medical Examiners. *Alabama Journal of Medical Sciences*. 1968;5:441-445.

Hubbard JP. Additional methods of testing fitness to practice. *Federation Bulletin*. 1968;55:151-159.

Hubbard JP. Uniformity of qualifications for medical practices and states' rights. *Federation Bulletin*. 1968;55:2-10.

Hubbard JP. The Federation Licensing Examination and the testing of clinical competence. *Federation Bulletin*. 1968;55:151-159.

Levit EJ. The use of motion pictures in evaluation of fitness to practice. *Federation Bulletin*. 1968;55:142-150.

Schumacher CF. Scoring, analysis and reporting. *Federation Bulletin*. 1968;55:160-165.

---

### 1967 PUBLICATIONS

Hubbard JP, Furlow LT, Matson DD. An in-training examination for residents as a guide to learning. *New England Journal of Medicine*. 1967;276:448-451.

Levit EJ. Comments regarding further study of graduate training in neurosurgery. *Journal of Neurosurgery*. 1967;27:385-387.

Levit EJ. Use of the National Board "minitest" for evaluation of curriculum change. *Journal of Medical Education*. 1967;42:930-934.

Levit EJ. The use of motion pictures in testing the clinical competence of physicians. *Annals of the New York Academy of Sciences*. 1967;142:449-54.

Saunders RH. Use of a special examination designed with specific reference to curriculum content. *Journal of Medical Education*. 1967;42:935-937.

---

## 1966 PUBLICATIONS

Templeton B. Two brothers, identical twins discordant for chronic imprisonment. *Diseases of the Nervous System*. 1966;7(suppl 7):5-10.

---

## 1965 PUBLICATIONS

Hubbard JP. The present position of the National Board of Medical Examiners. *Journal of the American Medical Association*. 1965;192:132-136.

Hubbard JP, Levit EJ, Schumacher CF, Schnabel TG. Objective evaluation of clinical competence: new techniques used by the National Board of Medical Examiners. *New England Journal of Medicine*. 1965;272:1321-1318.

---

## 1964 PUBLICATIONS

Levit EJ, Schumacher CF, Hubbard JP. The internship - an evaluation of input and output. *Journal of the American Medical Association*. 1964;189:299-305.

Schumacher CF. Factor-analytic study of various criteria of medical school accomplishment. *Journal of Medical Education*. 1964;39:192-196.

Schumacher CF. Personal characteristics of students choosing different types of medical careers. *Journal of Medical Education*. 1964;39:278-288.

---

## 1963 PUBLICATIONS

Hubbard JP. Programmed testing in the examinations of the National Board of Medical Examiners. In: *Invitational Conference on Testing Problems*. Princeton NJ: Educational Testing Service; 1963:49-63.

Hubbard JP. Walter L. Bierring, MD and the National Board of Medical Examiners. *The Journal-Lancet*. 1963;83:474-478.

Levit EJ, Schumacher CF, Hubbard JP. Effect of characteristics of hospitals in relation to the caliber of interns obtained and the competence of interns after one year of training. *Journal of Medical Education*. 1963;38:909-919.

Schumacher CF. Interest and personality factors as related to choice of a medical career. *Journal of Medical Education*. 1963;38:932-942.

---

## 1962 PUBLICATIONS

Richardson FM, Clemens WV, Ludwug GD, Hubbard JP. The Delaware medical seminars experiment. *GP*. 1962;25:165-173.

---

## 1961 PUBLICATIONS

Cornfeld D, Hubbard JP. Four-year study of the occurrence of beta-hemolytic streptococci in 64 school children. *New England Journal of Medicine*. 1961;264:211-215.

Cornfeld D, Hubbard JP, Harris TN. Epidemiologic studies of streptococcal infection in school children. *American Journal of Public Health*. 1961;51:242-249.

Schumacher CF. The 1960 medical school graduate. *Journal of Medical Education*. 1961;36:398-406.

---

## 1960 PUBLICATIONS

Elsom KA, Hubbard JP, Schor S, Clark TW. Periodic health examination: nature and distribution of newly discovered disease in executives. *JAMA*. 1960;172:5-10.

Hubbard JP. Practices and pitfalls in the early detection and control of heart disease in children. *Journal of Pediatrics*. 1960;56:544-550.

Hubbard JP. Teaching of preventive medicine reflected by results of National Board examinations. *Journal of Medical Education*. 1960;35:644-651.

Hubbard JP, Clemens WV. Comparative evaluation of medical schools. *Journal of Medical*

*Education*. 1960;35(2):134-141.

Schumacher CF. Studies of MCAT as a predictor of medical school achievement. *The Scalpel of the Alpha Epsilon Delta*. 1960;Winter(46-51).

---

## 1923 TO 1959

### The Formative Years: 1920s to 1950s

Between the 1920s and the 1940s, staff publications focused on establishing the value of NBME examinations, explaining how the content of the exams was chosen and why the assessments were useful. The bedside oral examination was a common assessment at that time. In the 1950s, staff members began publishing in a wider variety of journals. Staff physicians wrote books and articles on the latest developments in clinical medicine.

## 1950s PUBLICATIONS

Hubbard JP. Training the general practitioner for his job in public health. *Pennsylvania Medical Journal*. 1951;54:1139-1143.

Hubbard JP. Why National Board examinations? *Journal of the Student American Medical Association*. 1952;April.

Hubbard JP, Cowles JT. A comparative study of essay and objective examinations for medical students. *Journal of Medical Education*. 1952;29.

Hubbard JP, Mitchell AM, Poole ML, Rogers AM. The family in the training of medical students. *Journal of Medical Education*. 1952;27(1):10-18.

Hubbard JP. The National Board of Medical Examiners. *Journal of Medical Education*. 1953;28(1):85.

Hubbard JP. Observation of the family in the home. *Journal of Medical Education*. 1953;28(7):26-30.

Cowles JT, Hubbard JP. Validity and reliability of the new objective tests. *Journal of Medical Education*. 1954;29(6):30-34.

Hubbard JP, Cowles JT. Comparative study of student performance in medical schools using National Board examinations. *Journal of Medical Education*. 1954;29:27-37.

Hubbard JP. The inside story of your family's health. *Public Health News*. 1955;36(10).

Hubbard JP. New methods of examining in medicine. *Journal of the Indian Medical Profession*. 1955(2):7.

Hubbard JP. Prevention of first-attack rheumatic fever. *Annals of Internal Medicine*. 1955;43:504-510.

Hubbard JP, Clark DW. Preventive medicine and the Colorado Springs Conference. *Journal of Medical Education*. 1956;31:151-156.

Levit EJ, Nodine JH, Perloff WH. Progesterone-induced porphyria. *American Journal of Medicine*. 1957;22:831-833.

Cornfeld D, Hubbard JP, Werner G, Weaver R. Streptococcal infection in a school population: preliminary report. *Annals of Internal Medicine*. 1958;49:1305-1319.

Werner G, Cornfeld D, Hubbard JP, Rake G. Study of Streptococcal infection in a school population: laboratory methodology. *Annals of Internal Medicine*. 1958;49:1320-1330.

Hubbard JP. Medical examinations around the world. *Harvard Medical Alumni Bulletin*. 1959.

---

## 1940s PUBLICATIONS

Elwood ES. The National Board of Medical Examiners as related to medical licensure. *Federation Bulletin*. 1941;27:324-332.

Rodman JS. Part III of the National Board of Medical Examiners - its character and purpose. *Federation Bulletin*. 1946;32.

Hubbard JP. The role of the teaching hospital in child care: demands in pediatric education. *Journal of the American Medical Association*. 1949;24:373-378.

---

## 1930s PUBLICATIONS

Elwood ES. State board and National Board relations. *Federation Bulletin*. 1932;18:117-124.

Rodman JS, Elwood ES. Comments on National Board examinations. *Federation Bulletin*. 1936;22:196-204.

Gross RE, Hubbard JP. Surgical ligation of a patent ductus arteriosus: report of the first successful case. *Journal of the American Medical Association*. 1939;112:729-731.

---

## 1920s PUBLICATIONS

Elwood ES. National Board of Medical Examiners changes title of certificate holders. *Federation Bulletin*. 1923;9:133-136.

Rodman JS. National Board of Medical Examiners. *JAMA*. 1924;82:814-815.

---

