

## Advances in Medical Education

# Advances in Medical Education

*edited by*

A.J.J.A. Scherpbier

C.P.M. van der Vleuten

J.J. Rethans

A.F.W. van der Steeg

*Maastricht University, The Netherlands*



**SPRINGER-SCIENCE+BUSINESS MEDIA, B.V.**

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 978-94-010-6048-6    ISBN 978-94-011-4886-3 (eBook)  
DOI 10.1007/978-94-011-4886-3

---

*Printed on acid-free paper*

All Rights Reserved  
©1997 Springer Science+Business Media Dordrecht  
Originally published by Kluwer Academic Publishers in 1997  
Softcover reprint of the hardcover 1st edition 1997  
No part of the material protected by this copyright notice may be reproduced or  
utilized in any form or by any means, electronic or mechanical,  
including photocopying, recording or by any information storage and  
retrieval system, without written permission from the copyright owner

## Preface

About 550 registrants from 51 different countries attended the Seventh Ottawa Conference on Medical Education and Assessment in Maastricht. We received 525 abstracts for the conference, divided in thematic poster sessions and platform presentations. Organising the conference was an honour and we tried to meet the high standards of a friendly and relaxed atmosphere which has characterized previous Ottawa conferences.

During and after the conference about 250 papers were submitted for publication in the conference proceedings, leaving us little time for a post-conference depression. Despite the large number of papers, the editors have attempted to review and edit the papers as carefully as possible. Occasionally, however, correspondence exceeded reasonable deadlines, preventing careful editing of a small number of the papers. Although we felt that our editorial task was not quite finished, we nevertheless decided to include these papers. We thank the many authors for their enthusiastic and prompt response to - occasionally tedious - editorial suggestions and requests. We are sure that this collective effort has resulted in a book that will make an important contribution to the field of medical education.

The editors want to thank Jocelyn Flippo-Berger whose expertise with desk top publishing and perseverance was a great help.

Albert Scherpbier  
Cees van der Vleuten  
Jan-Joost Rethans  
Lideke van der Steeg

Maastricht, September 1997

## Table of Contents

### Key-Note Addresses

Assessment Truths that We Hold as Self-Evident and Their Implications  
S.M. Case

2

Adult Learning Theory, Prizes and Motivation: Fundamental Learning Issues  
T.J. David, L. Patel

7

Summary of the Conference  
G.R. Norman

11

The Student as Quality Controller  
C.J.A.H. Prince, K. Visser

15

Does Clinical Performance Correlate with Physicians' Scores on Written Examinations?  
P.G. Ramsey

19

Strategies for Effecting Change in Teaching Practices: A Review of Current Models  
L.A. Wilkerson, D. Irby

23

### Aims and Objectives

Northerners and Southerners: Two Approaches to Defining Core Content in Medical Education  
J.G. Bligh, D.J. Prideaux

32

Core Clinical Content of Step 2 of the USMLE: Using Surgery as an Example  
A.G. Greenburg, S.M. Case, G.S. Golden, D.E. Melnick

34

Determining Teaching Objectives for the Family Medicine Clerkship in Medical School: A National Delphi Survey  
U. Milman, M. Alperin, R. van Raalte, S. Reis

37

Core Skills in Administering and Prescribing Drugs  
P.A. O'Neill, T. Dorman

40

Cross-Section of the Utrecht Medical Curriculum  
J.A. Smal

43

### Community Based Programmes

Community Health Issues of the New PBL Undergraduate Medical Curriculum at the University of Geneva: A Description  
P. Chastonay, H. Stalder, F. Mottu, A. Rougemont, T. Perneger, A. Morabia, P. Huber, N.V. Vu

46

A Student Evaluation of Community Based Education in General Practice  
S.L. Hartley, J.H.S. Fuller, A. Berlin

49

Qualitative LEAP Project

A. Hatim Ricardo, R.S. Salas-Perea

53

The Implementation of a Multi-Institution, Multi-Disciplinary Project to Develop Community Undergraduate Medical Education

E.R. McLennan, M. Ricketts, S.L. Hartley, K. Hoffmann, D. Gill, A. Berlin, J.H.S. Fuller

55

### Continuing Medical Education

Interactive Continuing-Education Sessions: An Effective Way to Change Strategies?

E. Crone-Kraaijeveld, L.W.A. van Suijlekom-Smit, H.J. Dokter

60

Can the Effect of Continuing Medical Education on Physician Performance Be Measured?

A.J. Finestone, R.R. Smedley, A.I. Meyer

63

Academic Skills Training for Young General Practitioners: the New London Academic Training Scheme (LATS)

G.K. Freeman, F. Smith, J.H.S. Fuller

65

Attendance at an Academic Meeting: Does It Change Clinical Practice

C.R. Hutchison, M.D. McKee, D. Yoo

68

Increasing Physician Involvement in Cholesterol-Lowering Practices: the Role of Knowledge, Attitudes and Perceptions

K.V. Mann, E.A. Lindsay, R.W. Putnam, D.A. Davis

71

Continuing Medical Education in University Hospital Centers Becomes Strategic

C. Segouin, E. Rusch, M. Reninger

75

Continuous Built-in Evaluation of Participants' Motivation in PME and CME

S. Wilm, M. Erbe

78

### Curriculum Development

Design of an Integrative Course and Assessment Method The CIP (Comprehensive Integrative Puzzle)

R. Ber

84

A Universal Scheme of Analysis of the Educational Processes and its Application to Health Sciences: Initial Premises and Hypotheses A. Boczkowski, J.A. Indulski	87	<b>Examining Examinations</b>	
Experience with the Process of Curriculum Innovation D.A. Cook, A.H. Walji	91	Maintaining and Enhancing Key Decision-Making Skills from Graduation into Practice: An Exploratory Study G. Bordage, C.A. Brailovsky, T. Cohen, G. Page	128
Development of a Longitudinal Behavioural Medicine Continuity Curriculum for Third Year Medical Students R.A. Edelstein, V.M. Ranganath	94	Detection of Gender Differences in High-Stakes Performance SP-Based Examinations in Family Medicine C.A. Brailovsky, P. Grand'Maison, F. Miller, P. Rainsberry	131
Curriculum Information Management System: Structure and Process of Analysis with CATs P.K. Fullagar, E.R. Petrusa	97	Preliminary Descriptive Analyses of the Performance of US Citizens Attending Foreign Medical Schools on USMLE Steps 1 and 2 S.M. Case, D.B. Swanson, D.R. Ripkey, L.T. Bowles, D.E. Melnick	135
Problem-Based Learning: Always Small or Sometimes Large? A.J. Gercama, R.O.B. Gans	101	Applying Learning Taxonomies to Test Items: Is a Fact an Artifact? J.P.W. Cunningham, G.R. Norman, J.M. Blake, W.D. Dauphinee, D.E. Blackmore	139
Developing a Curriculum for Personal and Professional Development J. Gordon, S. Dunn	104	The Risks of Thoroughness: Reliability and Validity of Global Ratings and Checklists in an OSCE J.P.W. Cunningham, A.J. Neville, G.R. Norman	143
A Cross-Cultural Curriculum for Medical Students M. Hanson, M.L. Russell, A. Robb	106	The Impact of Postgraduate Training on an Examination for Medical Licensure G.F. Dillon, T.R. Henzel, W.P. Walsh	146
Innovations in Medical Education: A Comparison of Canadian and European Medical Schools B. Klemen, D. McSherry	109	A Comparison of Conjoint Analysis with Other Approaches to Model Physician Policies in Scoring Complex Performance-Based Assessment Y.V. Fan, S.G. Clyman, B.E. Clauser, T.W. Piemme, L. Chang, J. El-Bayoumi, G.L. Malakoff	149
The Teaching of Anatomy in the Integrated Medical Curriculum as Practised at the School of Medical Sciences, Universiti Sains Malaysia M. Othman	111	A Comparison of Diagnostic Pattern Recognition Skills to Performance on a "High Stakes" OSCE P.H. Harasym, N.G. Mohtadi, H. Henningsmoen	152
A Patient-Based Workshop in Anatomy for First Year Medical Students M.C.E. O'Driscoll	114	Knowledge on Skills: What is in the Name? A.W.M. Kramer, J.C.M. Bloemen, J.J.M. Jansen	156
Responses of Female and Male Medical Students to Case-Based Learning Programme in Anatomy P.V. Peplow	118	The Swiss Way to Score Multiple True-False Items: Theoretical and Empirical Evidence R. Krebs	158
The Use of Novel Educational Techniques in the Teaching of Molecular Cell Biology J. Szeberényi, N. Nusser, G. Boglári, L. Komáromy, M. Pap, Á. Sebok, G. Sétáló jr., A. Tigyí	121	Retention and Forgetting of Biological Facts and Concepts, Learnt for the First Basic Science Exam, over a Two Year Period R. Krebs, J.J. Guilbert, R. Hofer, R. Bloch	162
A New Approach to the Teaching of Clinically Relevant Anatomical Sciences and Compassionate Medicine "Clinical Anatomy - The Structure of the Human Body" A.H. Walji, D. Begg, G. Buzzell, B. Stevenson	124	Spatial Abilities and Learning Modes in Anatomy Beginners M. Lischka, G. Gittler	166
		Examiner Differences in the Mini-CEX J.J. Norcini, L.L. Blank, G.K. Arnold, H.R. Kimball	170
		Repeated Assessment of Medical Knowledge During the Two Years of Clerkships M.J. Quaak, J.A. Smal	173

Performance of Examinees from Foreign Schools on the Clinical Science Component of the United States Medical Licensing Examination D.R. Ripkey, S.M. Case, D.B. Swanson, D.E. Melnick, L.T. Bowles, N.E. Gary	175	A Qualitative Study of Faculty Techniques for Giving Feedback to Interns Following an Observed Standardized Patient Encounter A.W. Frye, M.A. Hollingsworth, A. Wymer, A. Hinds	216
Item Analysis in Clinical Encounter Scoring J. Roma, R. Gispert	179	Improving Clinical Teaching Skills Using Interaction Analysis J.H. George	220
The Use of Global Ratings in OSCE Station Scores A.I. Rothman, D.E. Blackmore, W.D. Dauphinee, R.K. Reznick	181	Assessment of Needs and Perceived Competence of Novice Clinical Teachers D. Gill, M. Ricketts, E.R. McLennan	223
An Investigation of the Properties of Computer-Based Case Simulations J.D. Scheuneman, S.G. Clyman, Y.V. Fan	184	Multiple Station Exams and Teaching Exercises (MSEs) for Teaching about Teaching E. Krajic Kachur	226
Performance of Examinees from Foreign Schools on the Basic Science Component of United States Medical Licensing Examination D.B. Swanson, S.M. Case, D.R. Ripkey, D.E. Melnick, L.T. Bowles, N.E. Gary	187	Development of a Medical Education Journal Club E. Krajic Kachur	228
Analysis of the Use of Global Ratings by Standardized Patients and Physicians M. Tann, G.E. Amiel, A. Bitterman, R. Ber, R. Cohen	191	Assessing the Implementation of Teaching Portfolios in the Academic Medical Center L.S. Lee, R.C. Froh, E.R. Petrusa	231
Comparing Reliabilities of True/False and Short-Answer Questions in Written Problem Solving Tests Th.J. ten Cate	193	Faculty Development and Primary Health Care Education in China H. Liu, D.H. Zhou, D.L. Fink, J.E. Rodnick, H.Z. Yuan, K. Lin, L.Y. Wang, F. Lu, A.I. Sutnick	234
Overcoming Difficulties in the Introduction of a Summative Assessment of Clinical Competence in a Brazilian Medical School L.E.A. Tronconi, M.L.V. Rodrigues, C.E. Piccinato, J.F.C. Figueiredo, L.C. Peres, A.R.L. Cianflone	197	Long-Term Impact of Faculty Development Workshops: A Pilot Study L. Nasmith, A. Saroyan, Y. Steinert, N. Daigle, E.D. Franco	237
Medical Student Perceptions of the Accuracy of Various New Clinical Evaluation Methods R.P. Usatine, R.A. Edelstein, A. Yajima, S.J. Slavin, M.S. Wilkes	200	Tutor Training for Problem-Based Learning: The Experience of a Large Medical School D.A. Price	241
<b>Faculty Development</b>		Planning Frameworks and Staff Participation: Changing to a Graduate Entry, Problem-Based Medical Course D.J. Prideaux	244
Re-engineering the Specialty Consult Paradigm: Extending the Generalist and Maximizing the Specialist Through Educational Reform B.A. Chernof, J.J. Guterman, M. Heckman, M. Esquivel, C.S. Hodgson	204	Determining the Change in Staff Attitudes during Training in Preparation for Radical Change from a Traditional Curriculum to Problem-Based Learning D.C.M. Taylor, E.A. Jump	247
Developing Staff: Developing Stations A Dental Objective Structured Clinical Examination (OSCE) Activity Day J.E.C. Davis, H. Fry, A. Jones	207	A Prospective Analysis Instrument for a Medical Education Centre Y. Urdaneta	249
The Training and Education of Osteopathic Practitioners to Become External Examiners in Clinical Competence D.R. Edwards	210	<b>Family Medicine</b>	
Educational Training of Medical Teachers: Recent Developments in the Netherlands C.R.M.G. Fluit, A.J.J.A. Scherpbier, J.C.M. Metz	213	A Treasure Discovered by Dutch Students in Canada; Results of a Student Exchange Program in Family Medicine R.J.M. Boot, S.E.M. Houben, M. Wijma, G.J. Dinant	252
		The Family Medicine Clerkship: Comparison of Two Programs in Different Continents R.J.M. Boot, M. Tarrant, J.M.H. Op 't Root	256

Defensibility of Competency Assessments of Practising Family Physicians J.P.W. Cunnington, E. Hanna, J. Turnbull, T. Kaigas, G.R. Norman	259	Is There a Difference in the Clinical Experience of Students with Diverse Career Interests? A Comparison Study within a Family Medicine Clerkship M. Tarrant, W. Woloschuk, H. Mandin	302
Innovations in Physician Performance Enhancement-Communication Skills T. De Cloedt, G. Schnabl	263	Use of Immediate Video Debriefing During Practice Teaching Visits C.A. Thompson	305
Training of GP Trainers in Flanders (Belgium) A.G.M. Derese, D.R. Leroy	265	Gaining an Appreciation of Differing Ethnic Influences on General Practice in Western Australia C.A. Thompson, Training Program, RACGP	308
How Students Learn from Ambulatory Preceptors: Critical Incidents R.M. Epstein, D. Cole, B. Gawinski, N. Ruddy, S. Lee	268	Final-Year Students' Experiential Learnings in Family Medicine at the University of Cape Town D. Whittaker	311
Alternative Routes to Vocational Training for General Practitioners in Australia G. Holsgrove, B. Jolly, A. Jones, L. Southgate	272	Perceived Learning Opportunities: Students' Expectations of a Family Medicine Clerkship W. Woloschuk, M. Tarrant, H. Mandin	313
A Tool for Assessing the Readiness of Family Physician Practices to Adopt Prevention Strategies P.A. Jennett, K. Premkumar, W.R. Elford, L. Meadows, R. Renger	275		
Competence of Trainees for GP at the Start of Their Training P. Ket, H. Düsman, L.H.C. Tan, E. Schadé, R.P.T.M. Grol, C.P.M. van der Vleuten	278	<b>Information Technology</b>	
Undergraduate Medical Education in Family Medicine in Slovenia M. Kolsek	282	Knowledge and Attitudes towards Computers among Graduates from Developing Countries A.M. Ahmed, G. Guerra, J. Giaconi, F. Piccinno	318
Educational Objectives and Requirements of an Undergraduate Clerkship in General Practice: To Which Extent do General Practice Teachers Achieve These? F.M.J.G. Martens, C.P.M. van der Vleuten, R.P.T.M. Grol, J.M.H. Op 't Root, H.F.J.M. Crebolder, J.J. Rethans	284	The Effectiveness of Interactive Multimedia as an Instructional Aid for Learning Basic Clinical Skills and Knowledge M. Broudo, M. White, D. Rodenburg, R. Arseneau, A. Chalmers, J. Wright, H. Mizgala, D. Lirenman	321
The Comprehensive Assessment Project in Continuing Medical Education of GPs P. Ram, J.J. Rethans, R.P.T.M. Grol, C.P.M. van der Vleuten	287	Computer Based Simulations as Adjuncts for Teaching and Evaluating Complex Clinical Skills R.A. Edelstein, S.G. Clyman	327
Hospital Clinicians and General Practitioners Working Together to Teach Medical Students M. Rickets, E.R. McLennan, P.G. Kopelman	290	On-Line Evaluation of Teaching and Courses: The Methods, Practice, and Politics D.A. Lloyd	330
Constructing a Congruent Curriculum for the Vocational Training of GPs E. Runia	292	Disease-Class Specific, Computer Adaptive Testing on the World Wide Web F.J. Papa, D.G. Aldrich, R.C. Stone, R.E. Schumacker	333
Improving the Quality of Tutorials in General Practice by Means of Instruction and Quality Assessment W. Slort, M. de Haan, G. van Stavereen, P.L. Schoonheim	295	The Effects of Visual Genetics on the Learning of Students in a Problem Based Curriculum B. Schutte, T de Goeij, W. de Grave, A.M. Koehorst	336
Evaluation of the Impact of Clinical Skills Teaching on General Practice J.A. Spencer, A.L. Robinson, A. Corradine, D.D. Smith	299	The Internet as a Cross-Disciplinary Learning Environment M.L. Walker	339
		Teaching Skills for Accessing and Interpreting Information from Systematic Reviews/Meta-analyses, Practice Guidelines, and the Internet F.M. Wolf, J.G. Miller, L.D. Gruppen, W.D. Ensminger	342



## Innovations in Assessment

### Testing Psychiatric Diagnostic Recognition Skills with Video Vignettes

J.B. Battles, D.J. Crowder, D.A. Miller, J.A. Sadler, P.C. Mohl, D.D. McIntire

348

### A Method for Developing Interdisciplinary Examinations to Test Application of Knowledge by Students in a PBL Curriculum

M.G. Camp

351

### The Diagnostic Schema Questionnaire: A New Tool to Assess a Specific Dimension of Clinical Competence

B. Charlin, C. Leduc, C.A. Brailovsky, D. Blouin, R. Hivon

354

### The Contribution to Final Clinical Competence of Preparing and Presenting a Dissertation within an Undergraduate Course in Osteopathy

D.R. Edwards

357

### The Use of a Neural Net to Reproduce and Assess a Physician's Clinical Diagnostic Skills

P.H. Harasym, F.J. Papa, R. Stone, D. Aldrich

359

### Developing Written Assessments for Use in Advanced Life Support Courses

G. Holsgrove, G. Bowler, M. Elzubeir

363

### Maintaining Content Validity in Computerized Adaptive Testing

R.M. Luecht, A. De Champlain, R.J. Nungester

366

### Assessment of Self Directed Learning

P. McCrorie

370

### Developing In-Course and Final Assessments in a New Dental Curriculum

C. Mercer, G. Holsgrove

374

### Development of Clinical Reasoning Exercises in a Problem-Based Curriculum

A.J. Neville, J.P.W. Cunningham, G.R. Norman

377

### 'Disease Class-Specific, Typicality-Graded' Test Case Vignettes: Towards 'Construct-Referenced' Assessments of Diagnostic Performance

F.J. Papa, R.C. Stone, D.G. Aldrich, R.E. Schumacker, P.H. Harasym

380

### Fair Reliable Essay Examinations for a Problem Based Curriculum

J.R. Philp, A.W. Frye, L.D. Cariaga-Lo, B. Richards, E. Bradley

383

### Critical Thinking in Medical School Exams

T.M. Scott, L.L. Chafe

387

### The Assessment of Higher Level Cognitive Objectives in Undergraduate Medical Students Using a Clinical Audit Project

F. Sullivan, J. Morrison

390

### Medical Students' Ability to Self Assess Knowledge and Skill Levels: Findings from One Class of Seniors

C. White, J.T. Fitzgerald, W.K. Davis, L.D. Gruppen, G. Regehr, M.A. McQuillan, M.L. Barclay, T.J. Bergstrom, K.R. Chamberlain, A.J. Zweifler

395

### Clinical Skills Assessment at NEOUCOM: A Description

L.F. Wilkinson, L.P. Pethel

397

## OSCE

### A Combined Structured Oral Examination (SOE) and OSCE for the Assessment of Clinical Competence

G.E. Amiel, M. Tann, M. Krausz, A. Bitterman, R. Cohen

402

### An OSCE Trial in Japan: Feasibility and Correlation with a Written Test

N. Ban, M. Hatao, J. Ohtaki, T. Saito, K. Matsueda, Y. Izumi, M. Kusumoto, S. Hosoda

404

### OSCE Stations for the Longitudinal Assessment of Cancer Screening and Detection

J.B. Battles, S.J. Sprankell, L. Baker, M. Becker, G. Camp, C. Browne, S. Morell, J.R. Bowling, J. Shores, E. Buck, C.O. Frates, J.B. Green, G. Smith, B.C. Pence, A. Weinberg, L. Laufman, B.U. Phillips

407

### Assessing the Adequacy of the Post-Encounter Written Scores in Standardized Patient Exams

J. Boulet, M. Friedman Ben-David, R. Hambleton, W.P. Burdick, A. Ziv

410

### Reproducibility of Emergency Medicine Standardized Patient Cases from Philadelphia to Barcelona

W.P. Burdick, R. Pujol-Farriols, J.M. Martinez-Carretero, R. Descarrega, M. Friedman Ben-David, J. Boulet

413

### Using the Rasch Model to Equate Alternate Forms for Performance Assessments of Physician's Clinical Skills

B.E. Clauser, L.P. Ross, R.M. Luecht, R.J. Nungester, S.G. Clyman

416

### Regression-Based Weighting of Items on Standardized Patient Checklists

B.E. Clauser, M.J. Margolis, L.P. Ross, R.J. Nungester, D.J. Klass

420

### A Large-Scale Study of the Reliabilities of Checklist Scores and Ratings of Components of Interpersonal and Communication Skills Evaluated on a SP Examination

D.S. Cohen, J.A. Colliver, R.S. Robbs, M.H. Swartz

424

### Clinical Assessment in an Academic Environment: To Do or Not To Do?

D.H. Davis, B. Hill

427

### Academic Assessment of Clinical Skills Inferences of Predictive Validity

D.H. Davis, B. Hill

430

The Effect of Formal Feedback Sessions on Test Security for a Clinical Practice Examination using Standardized Patients G.E. Furman, J.A. Colliver, A. Galofré, M.A. Reaka, R.S. Robbs, A. King	433	The Comprehensive Objective Examination (COE) D. Massey, L. Papineau	480
The Quebec Licensing OSCE: Modifications and Improvements over 6 Years of Experience P. Grand'Maison, C.A. Brailovsky, J. Lescop	437	Improving Reliability and Validity of the Clinical Examination for Fellowship of the Royal Australasian College of Physicians N.S. Paget	482
Reliability, Number of Stations, and Examination Length in an Objective Structured Clinical Examination L.D. Gruppen, W.K. Davis, J.T. Fitzgerald, M.A. McQuillan	441	Development of the Catalan OSCE of Internal Medicine R. Pujol-Farriols, A. Nogueras-Rimblas, M. Vilardell-Tarrés, C. Blay-Pueyo, J. Roma-Millan, J.M. Martinez-Carretero	485
Construct Validity of "High Stakes" OSCE Scores P.H. Harasym, N.G. Mohtadi, H. Henningsmoen	443	Using Peer Comparisons to Evaluate Combined Residency Training in Pediatrics, Psychiatry, and Child and Adolescent Psychiatry J. Schowalter, C.P. Friedman, S.C. Scheiber, D. Juul	488
OSCA: Assessing to Unite Theory and Practice J.M. Kelly	446	The Relationship Between Age of Candidates and Performance in the FRACGP Examination N.A. Spike, P.C. Veitch	490
An Evaluation of OSCEs for Final-year Students E.M. Long, J.A. Martin	448	Remediation for the Intern with Deficiencies in Clinical Competence S. Willcock, J. Gordon	494
Implementing a Multiple Station Clinical Examination: An International Collaboration J.M. Martinez-Carretero, J. Roma-Millan, R. Pujol-Farriols, C. Blay-Pueyo, A. Nogueras-Rimblas, C.A. Brailovsky, P. Grand'Maison	452		
<b>Clinical Skills Assessment (CSA) in Undergraduate Medical Education in Catalonia, 1995</b>		<b>Problem Based Learning</b>	
J.M. Martinez-Carretero, F. Borrell-Carrió, C. Blay-Pueyo, J. Roma-Millan, R. Pujol-Farriols, J.M. Gomez-Saez, L. Masana-Marin, J. Sanchis-Aldas	454	The New Preclinical Medical Curriculum at the University of Geneva: Processes of Selecting Basic Medical Concepts and Problems for the PBL Learning Units A. Baroffio, J.P. Giacobino, B. Vermeulen, N.V. Vu	498
An OSCE for Licensure: The Canadian Experience R.K. Reznick, D.E. Blackmore, W.D. Dauphinee, S.M. Smee, A.I. Rothman	458	Implementation of an Integrated Educational Curriculum at the Medical Faculty of Lausanne R. Bonvin, B. Waeber, C. Perret, J.P. Wauters	501
Preparing Physician Examiners for a High Stakes, Multi-Site OSCE S.M. Smee, D.E. Blackmore	462	An Introduction of Certain Basic Principles in Kinesiology: Incorporation into a Problem-Based Setup M. Chakravarty, F.N.L. Engmann	503
<b>Postgraduate Training and Assessment</b>		The Rome Parallel Track (CPR) M. Fioravanti, P. Gazzaniga, A. Marcheggiano, J. Osborn, T. Renda, A. Torsoli, V. Ziparo and CPR Group	506
A Self-Assessment Study of Competency Levels in the Postgraduate School of Public Health A. Boczkowski	466	Macroanalysis of Integration of the Basic Sciences in a Problem-Based Learning Curriculum P.G. Iatridis, W.M. Anderson, D. Gupta, N.W. Hudson, V. Hoftiezer, P. Bankston, J.C. Vanden Berge, S.F. Echtenkamp, B.G. Kennedy, S.P. Sivam, C.F. Marfurt, R. Dziarski, W. Baldwin, S.D. Vaughn	509
An Innovative CME Cardiology Program for General Practitioners D. Feldman, J. Sirkin, M.K. O'Brien, D. Cohen, D.H. Novack	470	The Source of Final Exams (Internal versus External) and their Impact on Learning in Biochemistry in a PBL Curriculum P.G. Iatridis, S.D. Vaughn, W.M. Anderson, D. Gupta, N.W. Hudson	513
Assessment of Junior Doctors' Clinical Competence: A Feedback Loop J. Gordon, I. Rolfe, S. Pearson, S. Willcock	474		
Can You Teach an Old Doc New Tricks? E. Hanna, J. Premi, J. Turnbull	477		

Promoting Rational Prescribing in Problem-Based Curriculum V.S. Mathur, R.P. Sequeira, K.A. Jassim	516	The Evaluation of the Practice Quality Review Component of the Maintenance of Professional Standards Program of the Royal Australasian College of Physicians D.I. Newble, N.S. Paget	567
The Synthesis Unit: A Complementary Approach to Integration in a Problem-Based Learning Curriculum A. Perrier, Y. Dunant, J. Ferrero, N.V. Vu	519	Ten Year Observation of Problem Based Learning: A Guide to Future Evaluations? J.R. Philp, A. LeMaster, C.M. Lane	571
A Qualitative Comparison of Student and Faculty Effectiveness as PBL Tutors K.E. Sardeson, A. Haramati, M. PreJean	523	'Happiness Index' to Health Gain: Using Evaluation of Teaching to Benefit Patient Care J. Pitts, C. Coles, D. Percy, P. White	575
Problem-Based Learning Using Ambulatory Patients T.K. Sen Gupta, R.B. Hays	525	Modifying the Categorisation of Validated Criteria to Assess Teaching Performance in the Tutor-Led, Task Orientated Small Group M.E. Preston-Whyte, R. McCulloch, R.C. Fraser	577
Teaching Ethics using a Small-Group, PBL Format J.W. Tysinger, L.K. Klonis, J.M. Wagner, J.Z. Sadler, J.B. Battles	528	Evaluating the Outcomes of Changes in Undergraduate Surgery Education: Interns' Views on Their Preparedness for Surgical Rotations D.J. Prideaux, S.M. Henry-Edwards, V.R. Marshall	581
The Redesigned Undergraduate Medical Curriculum at the University of Geneva N.V. Vu, C.R. Bader, J.D. Vassalli	532	Some Curriculum Implications of fewer In-Patients in Teaching Hospitals A.I. Rothman, A. Rachlis	585
<b>Programme Evaluation</b>		Educational Audit into Medical Practice: A Quality Approach R.S. Salas-Perea, A. Hatim Ricardo	588
Models for Predicting Career Patterns of Graduates of Combined Baccalaureate-MD Degree (CD) Programs L. Arnold, T.L. Willoughby, L.C. Epstein, B. Jones, G. Xu	538	Evaluation of Clinical Teaching Processes: The Development of a Questionnaire I. Sukcharoen	590
Recording Patient Encounters to Verify the Clinical Curriculum of a Family Medicine Clerkship J.B. Battles, L.M. Snell, E. Washington, J.A. Bedford	542	The Effect of Health Reform on Surgical Clerks' Experience and Performance G.L. Warnock, P. Craig, T.O. Maguire, E.N. Skakun	594
Factors which Influence the Outcome of Student Evaluation of Teaching D.A. Cook, L.M. Brown, E.N. Skakun	545	An Effective Educational Model for Tuberculosis in the Ambulatory Care Clinic: Documentation of Clinical Competence and Skills S.K. Willsie, T.L. Willoughby, G.A. Salzman, S.C. Hamburger	597
Self-Directed Learning (SDL): Changing Perceptions H. Fry, A. Jones	548	<b>Reasoning and Learning</b>	
A Problem-Based Approach to Non-Redundant Resident Education T.L. Johns, J.R. Philp	551	Sherlock Holmes and Clinical Reasoning: Empiric Research on a Methodology to Teach Clinical Reasoning W.R. Ayers	600
A Comparative Study of Medical Curriculum Outcomes: Opinions of the Graduates of a Traditional Curriculum and an Innovative Curriculum R. Lewkonja, J. Bamber, D. Gupta A.H. Walji	554	The Structure of Medical Knowledge Reflected in Clinicians' Estimates of the Probabilities of Signs/Symptoms within Diseases P.H. Harasym, F.J. Papa, R.E. Schumacker	602
An Evaluation of the Horizontal Research Elective at The University of Calgary H. Mandin, W. Woloschuk, P. Leco	557	The Effect of Clinical History on Physician's ECG Interpretation Skills R.A. Hatala, G.R. Norman, L.R. Brooks	608
Evaluating the Success of a Medical School: The Beer-Sheva Experiment C.Z. Margolis, S. Glick, S. Warshawsky, M. Friedberg, B. Gvili, S. Burger	560		
Evaluation of the Effectiveness of Clinical Skills Teaching in the Community E. Murray, B. Jolly, M. Modell	564		

The Effect of ECG Instructional Materials on Medical Students' Reasoning Strategy R.A. Hatala, G.R. Norman, J.P.W. Cunnington, L.R. Brooks	611	Optimizing the Input of Physician Examiners in Setting Standards for a Large Scale OSCE: Experience with Part II of the Qualifying Examination of the Medical Council of Canada W.D. Dauphinee, D.E. Blackmore, S.M. Smee, A.I. Rothman, R.K. Reznick	656
Reflective Practice and Clinical Placements J.M. Kelly	615	Student and Faculty Ratings of Task Difficulty in Predicting Performance on a Clinical-Skills Exam J.T. Fitzgerald, L.D. Gruppen, C. White, W.K. Davis, M.L. Barclay, T.J. Bergstrom, K.R. Chamberlain, M.A. McQuillan, A.J. Zweifler	659
Student Thought Processes Evoked by Multiple Choice and Constructed Response Items T.O. Maguire, E.N. Skakun, O.H. Triska	618	The Use of Critical Stations to Determine Clinical Competency in a "High Stakes" OSCE P.H. Harasym, N.G. Mohtadi, H. Henningsmoen	661
Use of a "Scheme" Approach to Problem Solving during a Family Medicine Clerkship H. Mandin, W. Woloschuk, M. Tarrant, P.H. Harasym	622	Relative and Absolute Standards in Assessing Medical Knowledge A.M.M. Muijtjens, R.J.I. Hoogenboom, G.M. Verwijnen, C.P.M. van der Vleuten	665
Expert-Novice Differences in the Use of History and Visual Information from Patients G.R. Norman, L.R. Brooks, J.P.W. Cunnington, V. Shali, M. Marriott, G. Regehr	625	Competence Judgments as an Approach to Standard Setting for Multiple Choice Items E.N. Skakun, T.O. Maguire	668
A Comparison of Expert Clinician and Expert Student Reasoning on Multiple Choice Items O.H. Triska, T.O. Maguire, E.N. Skakun	629	Variations on a Theme: Scoring the Clinical Skills Assessment L.F. Wilkinson, N.M. Cuevas	671
Medical Students' Learning Strategies in Problem-Based Learning and Traditional Courses R.A. Woodhouse, M.D. Delva, A.F. Hadwin, R.V. Birtwhistle, J.R. Kirby, C. Knapper	632		
<b>Selection</b>		<b>Skills Training</b>	
Experience With a Selection Method for Admittance to the Universidad Nacional De Cuyo Medical Sciences School A. Binia, A.S. Ciccarelli, A.M. Ortiz	636	Do Medical Students Believe that Communication Skills can be Learned? A. Cushing, E. Najberg, P. Hajek	676
Beliefs in Personal Control, Styles of Learning and the Prediction of Academic Performance of Student Health Professionals P. Eachus	638	Design and Development of the New Preclinical Practice Skills (CPS) Program at the University of Geneva P. Huber, A. Perrier, J.F. Balavoine, M. Archinard, D. Lefebvre, N.V. Vu	679
The Effect of Changing Selection Procedures and the Pre-Entry Experience of Intending Students of Osteopathy D.R. Edwards	643	The Effectiveness of a Technical Skills Course for Surgical Residents C.R. Hutchison, M. Brown, G. Regehr, D.J. Backstein, J.J. Murnaghan, R.K. Reznick	682
The Reliability of a Personal Statement as a Selection Tool in the Admissions Process M. Nayer, J. Howe	646	Development of Interviewing Skills in a Problem-Based Medical Curriculum D.M. Kaufman, B.L. Prime-Walker, K.V. Mann, P. Andreou	685
<b>Setting Standards</b>		Pediatric Simulated Scenarios in Undergraduate Medical Teaching-Exposure to the Under Fives M. Kavanagh, K.D. Blake	689
Using an Algebraic Inversion of the Dependability Index to Determine Pass/Fail Criteria for Clinical Performance Examinations J.B. Battles, D.D. McIntire, J.L. Carpenter, J.M. Wagner	652	An Innovative Course Combining the Acquisition of Communication Skills with First Year Medical Student's Perception of their Psychosocial Role in Primary Care M. Perleth, J. Collatz, G.C. Fischer	692

The Social History: A Vanishing Portion of the Medical Interview C.A. Pfeiffer, H. Madray, A.J. Ardolino, J.L. Willms	695	Sequential Testing in a Standardized Patient Based Test F.H.M. van Vollenhoven, S.J. van Luijk, A.M.M. Mulijtjens	734
<b>Standardized Patients</b>		Cognitive Strategies used in Standardized Patients' Checklist Recording Y. Wang, G. Gliva-McConvey	737
Using Standardized Patients in the Time-in/Time-out Interviewing Technique M. Cantrell, J.K. Heard, M. Lindsey, E. Cleveland	700	<b>Teaching and Learning</b>	
Training Standardized Patients to Provide Feedback for Ethics Teaching S.M. Dyer, S.M. Smee, V. Sweeney, A. Browne, S. Clifford	703	Educating Medical Students on Cultural and Societal Issues in the Health Care of Diverse Population Groups L.D. Cariaga-Lo	740
Volunteers and Standardized Patients in Health Sciences Education: Our Experience in Fostering a Valuable Community Resource B. Fletcher, D. Morales	706	Differences Between First and Fourth Year Medical Students in their Attitudes Toward Primary Care B.G. Ferrell	742
The Incorporation of Standardized Patients into a Continuing Education Conference on Child Abuse C.W. Gowen, G. Gliva-McConvey	709	Comparing a Four-Week Block Clerkship to a Twelve-Week Longitudinal Experience in Family Medicine B.G. Ferrell, D.L. Camp	744
Health Care Perceptions: Do They Change for the Standardized Patient? N.J. Rubin, E.B. Philp, J.A. Hartman	711	Moral Development During Surgical Residency Training J.H. George	747
Advancing the Use of Standardized Patients: A Workshop for the Consortium of Thai Medical Schools S.M. Smee, V. Sumawong	714	The Effects of Trainings During the Vocational Training for General Practitioners on the Quality of Trainee's Consultation Behaviour with Standardized Patients J. Goedhuys	749
Using Standardized Patients to Assess Clinical Competence in the Primary Care of Pediatric Cancer Patients S.J. Sprankell	717	Presentation Skills: a Course for Students on Voice Production and Confidence-Building L.E. Gray, P. McCrorie, A. Cushing	753
Using Standardized Patients to Develop Clinical Competence in the Delivery of Bad News S.J. Sprankell	719	Comparing Two Formats for Practising Electrocardiograph Interpretation L.D. Gruppen, C.M. Grum, J.T. Fitzgerald, M.A. McQuillan	756
Using SPs for Teaching and Assessing Clinical Skills for Encounters with Patients with Problems of Substance Abuse, Pain Control, and AIDS B. Stimmel, J.A. Colliver, D.S. Cohen, L. Smith, M.H. Swartz	722	Inter-Disciplinary Clinical Teaching R.B. Hays, P.R. Mudge, P.C. Veitch, J. Harte, J.T. La Brooy	759
The Validity of Standardized Patient Assessment Using Faculty-Physician Global Ratings as the Gold-Standard Criterion M.H. Swartz, J.A. Colliver, C.L. Bardes, R. Charon, E.D. Fried, S. Moroff	725	Designing Basic Medical Education in Nijmegen R.S.G. Holdrinet, B. Oeseburg, J.A. Bulte, R.R.M. Leunissen	762
The Use of Standardized Patients in a First-Year Problem-Based Learning Introduction to Clinical Medicine Course J.W. Tysinger, J.M. Wagner, L.K. Klonis, J.Z. Sadler	728	Using Literature to Learn About and From the Elderly P.M. Jucovy	765
Predicting Performance Through Clinical Skills Assessment with Standardized Patients M. Vadillo, R. Pujol-Farriols, J. Roma, J.M. Martinez-Carretero	732	An Environment for the Development and Presentation of Computerized Cases to Learn Diagnostic Reasoning A.M. Koehorst, C.F.M. Ronteltap	769
		Interprofessional Teaching for Clinical Skills P.G. Kopelman, M.J. Nicol, D.S. Freeth	774
		Presentation Skills; a Student Learning-Activity R.C. Ladage, E. Crone-Kraaijeveld	777

Projects - Are They Worth All the Hassle? P. McCrorie, G. Harris, S.G. Welch	780	The Influence of Neuro-Cognitive Modelling and Philosophy on Issues of Curriculum and Assessment in Medical Education R.C. Stone, F.J. Papa, D.G. Aldrich	801
Video-Based Supplemental Instruction (VSI) as an Integral Part of an Academic Support and Development Program P.P.C. Nel, A.A. Beyliefeld, M.M. Nel	784	What Students Want from Their Pastoral Care System D.C.M. Taylor	803
New Horizons in Surgical Teaching: Undergraduate Medical Teaching in Day Surgery M.C.E. O'Driscoll, G.E. Rudkin, V.M.Carty, G.J. Maddern	787	Addressing Telephone Medicine in Medical School Education R.P. Usatine, S.J. Slavin, M.S. Wilkes	805
Mentoring Undergraduate Medical Students: A Study of Organizational Aspects P.A. Poldre, A. Rachlis	790	Teaching Clinical Decision Making and Cost- Effectiveness in Medical School: A Simulated Back Pain Case R.P. Usatine, S.J. Slavin, M.S. Wilkes	807
Learning About Learning: A Course for Incoming Medical and Dental Undergraduates J. Reidy, M. Wilde, G. Holsgrove	793	Implementation of a "Well Physician" Course for Students: A Formative Evaluation W. Woloschuk, J. Wright, M. McBean, H. Mandin	809
Critical Thinking, Evidence-Based Medicine and Clinical Practice T.M. Scott	795	The Impact of Curriculum Change on Student Stress and Workload W. Woloschuk, H. Mandin, P.H. Harasym	812
Comparing Medical Knowledge of Osteopathic Trainees in DO and MD Programs: A Random Effect Meta-Analysis L. Shen	798	Authors Index	815
		Subject Index	819

# Key-Note Addresses

# Assessment Truths that We Hold as Self-Evident and Their Implications

S.M. Case

National Board of Medical Examiners, USA

## Summary

*This paper discusses six truths related to assessment, highlighting at first the fact that there seem to be few real truths in assessment; most “truths” are at best half-truths (Truth #1). Nevertheless, it is important to read the literature (Truth #2), because strong relationships between today’s ‘new’ assessment methods and the methods of our predecessors make it important to note lessons of the past. The method used for assessment provides no guarantee of what is measured, nor how well it is measured (Truth #3). Therefore, it makes no sense to hawk one method over another as if the method alone provides a guarantee of anything. One reason that the quality of assessment is important is that evaluation drives student learning (Truth #4). Teachers should use the power of assessment to drive student learning in the direction they want it to go. The truth (#5) that it takes many cases/items/raters to generate a reproducible score has stimulated an ongoing debate between educators and measurement experts: this truth is of relatively little importance to educators who might not want to extrapolate beyond the particular performance that is observed, but of paramount importance to measurement experts who want to generalize across tasks, across patients, and across raters. In order to view a score as a measure of general clinical competence, you need to sample from the larger domain to which you want to extrapolate. The final truth (#6) is that no single method can assess competence fully. Each method has something unique to offer; it is ridiculous to become such a zealot of one method that all other forms of assessment are avoided.*

*It is a real pleasure to have the opportunity to discuss the lofty topic of “truth”, particularly within the area of assessment. This paper presents a few truths and discusses their implications. The first truth can be viewed as an introduction to the topic*

## Truth number 1: There are few truths in assessment

If I were to ask each of you to generate a list of truths related to the topic of assessment, I am sure that you could do so. If I asked you to limit the list to include only those that are so true that you personally believe they are self-evident, I am sure you would have little difficulty. If I then asked you to further limit your list to include only those truths that almost everyone would endorse, you will experience part of what I have been thinking about for the last few months.

In searching for truths, I re-read some articles that were piled in corners of my office and I compiled some writings of people who I think are very wise. In addition, several colleagues

stopped by to suggest that I include their favourite truth; the difficulty was that I didn’t think most of them were true, which leads me to my first quote:

*“There are no whole truths; all truths are half-truths.”*

*Alfred North Whitehead, 1953  
(Dialogues of Alfred North Whitehead)*

## Truth number 2: It is important to read the literature. Corollary- Performance assessment is not new

Those of us who have been around for awhile have observed a circular, or at least a pendular, evolution in assessment. The first assessments undoubtedly involved oral examinations. From the time of God and Adam, oral examinations have represented a cross-cultural, time-honored form of assessment. However, over the centuries, concerns arose about the subjectivity inherent in oral exams, and multiple-choice questions (MCQs) were invented. While MCQs were heralded as promoting objectivity in assessment, the features of MCQs that made them so popular, have turned out to be the same features that have made them so objectionable. As a result, coming full-circle, we are faced with rekindled interest in new and improved versions of those oral examinations.

Many old testing concepts are being marketed today under bright new labels. I don’t want to focus on the lack of real change, and I don’t want to present a view that is too cynical; nevertheless, I recommend that any of you who believes they have invented a new form of assessment look back into the literature, many of today’s new techniques have been around for decades under old labels. For example, simulations probably had their origins in war games where they were used for both training and for assessment. Within medicine, so called “programmed patients” were certainly used in the 1960s; we used generalizability analyses to discover that a dozen or more cases were needed in the 1970s; and the need for many cases or questions was known in the context of oral and essay exams 100 years before that. The relationship between “new” methods and the methods of our predecessors becomes obvious as you read the literature; there are many lessons to be learned from the past.

When you read the literature, you will find that some famous people were famous for a reason: they had remarkable judgement and knowledge about test development and measurement. For example, sound advice was provided 25 years ago in a chapter by Fitzpatrick and Morrison on performance testing.<sup>1</sup> They quote



a publication written 40 years ago by R. W. Highland, which provides advice on whether to assess process or product.<sup>2</sup> Highland wrote that we should think about measuring process:

1. if the steps in a procedure can be specified;
2. if the extent to which the examinee deviates from accepted procedure can be measured accurately and objectively;
3. if much or all of the evidence needed to evaluate performance is to be found in the way the performance is carried out, not in what is present at the end of the performance;
4. if an ample number of persons are available to observe, record, and score the procedures used during the performance.

On the other hand, he wrote, you should score performance in terms of *product* if (1) the product can be measured accurately and objectively, etc.

Highland's book entitled "A guide for use in performance testing in Air Force technical schools" also provides excellent advice regarding exam security within the performance testing arena.<sup>2</sup> I suspect that knowledge of his work would have been useful to many test developers in the audience today.

The underlying point that lessons can be learned from the past generalizes beyond the field of medicine and assessment, as illustrated in the following quote:

*"Most of the change we think we see in life is due to truths being in and out of favor."*

*(Robert Frost, The Black Cottage, 1914)*

### **Truth number 3: The method used for assessment does not directly affect test quality, nor does it determine the component of competence measured by the test**

For this truth, I am indebted to Christine McGuire, who made a similar point at the Ottawa Conference in 1986, in her so-called Myths paper.<sup>3</sup>

Everyone in the audience today would agree that MCQs can be a truly horrible form of assessment. MCQs can be ambiguous, and they can focus on completely irrelevant trivia. On the other hand, although many would argue that MCQs only measure knowledge, most would agree that good MCQ tests can provide an excellent indication of what students know (a characteristic that should not be dismissed lightly). Those with a higher regard for MCQs go much further; they believe MCQs can be a form of simulation, albeit less realistic than many performance-based tests. These advocates believe that well-developed MCQs assess *application* of knowledge, including interpretation and synthesis of data, and therefore can provide a good indication of examinee's clinical judgement.

On the other hand, many people endorse the use of oral exams as the best method for determining what a student really knows; some believe this is the only method for determining whether the student's thought processes "are in order". While it is not politically correct today to challenge the measurement derived from performance testing, I must report that I have seen appalling oral examinations. The cases are often selected based on conven-

ience. Bedside orals, for example, involved cases that were selected based on who was sick that day. Some, but not all, floundering examinees received "help" from the patient -- the patients were "professional" in the sense that they had been worked up so many times that they knew what the doctors were supposed to do and what their diagnosis was. If they liked the examinee, these patients were known to whisper advice to those who seemed to miss something important.

Attempts to standardize the procedure resulted in scripted oral exams. However, despite careful scripting of the oral exam cases, the examiners often veered from protocol. Like the patients in the bedside orals, some provided hints to some of the examinees. Other examiners focused on asking irrelevant questions which could do nothing more than assess knowledge of isolated facts. In one oral that I observed, the examiner asked "What is this procedure called in Chicago?"; he then followed up with: "Who invented it?"

While I have only mentioned MCQs and oral examinations, this truth generalizes to other methods. It makes no sense to hawk one method over another as if the method alone provides a guarantee of anything. Unfortunately, there are examples of shoddy test development directed at all assessment methods; the method itself provides no guarantee of what is measured, nor how well it is measured.

Initially, beliefs about a new assessment method are often viewed as heresy; ultimately, they might be viewed as superstitions; the dangerous period is in-between, when they are commonly viewed as gospel:

*"It is the customary fate of new truths to begin as heresies and to end as superstitions."*

*(Thomas Huxley, The Coming of Age of The Origin of Species, 1880)*

### **Truth number 4: Evaluation drives student learning**

Everyone believes in this truth. People believe that, if you give true/false tests that assess recall of isolated facts, students will go off and memorize facts. On the other hand, if you assess students based on their ability to take a history and do a physical exam, they will go off and try to get some clinical experience. There is even research to support this truth.<sup>4</sup>

The debate really centers on whether this is a good or a bad consequence of testing. Some believe this consequence is so bad that they go to elaborate lengths to avoid what they call the "steering effect" of tests. One of the strategies used by these people is to avoid testing entirely: there are a few schools in the US which have essentially no tests, and students proceed under a pass/fail system. A second strategy is to avoid targeting the test to the curriculum. There are a few schools that administer a comprehensive exam a few times a year to all students; students at all levels take the same test. The rationale is that, because the exam blueprint includes all of medicine, students cannot possibly go off and study for the test.

Those on the other side of the debate want to deliberately use the power of assessment to drive student learning in a particular

direction. For example, some faculty are concerned that clinical clerkships cover content as isolated topics. Because they want students to think in an interdisciplinary way, they develop interdisciplinary tests that students must pass for promotion or graduation. Students then learn the clerkship-specific material in order to earn good marks within the clerkship; they also learn the material that would be required to answer the interdisciplinary questions in order to progress.

If your tests reflect what you want students to learn, it is appropriate to view tests as a component of instruction. Testing time should only be viewed as taking time away from instruction time if you don't view the test as driving student learning in an appropriate direction. Of course, each item should pass the 'who cares' test. It doesn't make sense to construct tests that ask irrelevant questions.

The implications of this truth are to develop a clear map of where you want the curriculum to go; drive student learning in the same direction; and, finally, develop assessments in which teaching to the test is a valid use of instructional time.

*"The ideal condition would be, I admit, that men should be right by instinct; but since we are all likely to go astray, the reasonable thing is to learn from those who can teach."*

*(Sophocles, Antigone, 442 B.C.)*

### **Truth number 5: It takes a lot of cases/items/raters to generate a reproducible score**

I think this truth is self-evident. All of us know someone who took an essay or oral examination where the examinee was asked about the only topic he didn't know. Many of us also know someone who took an oral examination given by a notoriously harsh rater or by a rater who was otherwise unfair. Ratings are affected by almost everything related to examinees: age, gender, race, clothes, height, weight, speech. One of the problems is that these variations across raters are idiosyncratic, not everyone is bothered by the same things. This variation makes it difficult to statistically control bias, or even to adequately investigate it. As a result, measurement experts advocate using a lot of cases and raters, in the hope that things will "even out".

Issues of sampling don't commonly arise with written tests: it's harder for a student to complain about a written test that included the only 100 topics he didn't know, or about the scoring machine that's made by one company instead of another.

Appreciation of this truth has been present for decades if not centuries. In 1845, Horace Mann listed eight reasons to use written standardized exams instead of oral exams.<sup>5</sup> These reasons included: "they are more thorough than older forms of examination"; "they take away all possibility of favouritism"; "they enable all to appraise the ease or difficulty of the questions." He also wrote that "it is clear that the larger the number of questions put to a scholar, the better is the opportunity to test his merits. If but a single question is put, the best scholar in the school may miss it." He went on to note: "Each question is a

partial test, and the greater the number of questions, therefore, the nearer does the test approach to completeness." The focus on large samples is not a new fad.

The controversy surrounding this 'truth' centers on the extent to which it is important. Measurement experts are certain that you need 15 or 20 cases to produce a valid exam; faculty sometimes use only one. I have heard faculty lamenting the current emphasis on generalizability because it requires sampling across so many things: across items or cases, across raters, across patients, across sites. They believe that all this sampling is just a lot of smoke getting in the way of the real assessment they want to do. The debate has continued in a series of articles and then letters to the editor in recent issues of *Educational Measurement: Issues and Practices*. Davison cleverly identified the underlying problem: teachers and measurement specialists do not always share the same beliefs about what is truth.<sup>6</sup>

In thinking about assessment, teachers tend to focus on the questions or the tasks that students must perform and the student's answer to the questions or performance on the tasks. From their perspective, a more authentic task is *necessarily* a better assessment than a less authentic task, and the need to evaluate the reliability and validity of scores is NOT self-evident. Measurement specialists, on the other hand, are more likely to focus on outcome (i.e., the score that provides an indication of student ability), rather than on the answer to individual questions or performance on individual tasks. From their perspective, standards of reliability and validity must be applied to scores.

This is a situation where both sides may be correct. The rules depend on the purpose of the assessment and the level of generalization. The teachers are completely correct that you don't have to worry about generalizability IF you don't want to generalize beyond the test itself. If you want to know whether or not a student can do a focused history and physical examination on this particular patient, under these particular circumstances, with one particular rater, then you can get by with an assessment based on a single patient. On the other hand, if you want to generalize across tasks, you must sample tasks; if you want to generalize across patients, you must sample patients; if you want to generalize across raters, you must sample raters. In other words, if you want to extrapolate, you need to sample from the larger domain to which you want to extrapolate. It's that simple. The test developer needs to decide if the test is being used to make important, relatively long-term decisions regarding individuals OR if the test is being used as part of ongoing instruction and is used largely to guide short-term instructional decisions. As long as assessments are limited to instructional uses that are under the control of individual teachers, technical issues may play a relatively minor role. But, if you want to generalize beyond the particular sample you have used, the game has changed and you must follow the measurement rules.

*"Take nothing on its looks; take everything on evidence. There's no better rule."*

*(Charles Dickens, Great Expectations, 1859)*

## Truth number 6: No single method can assess competence fully

The purpose of assessment within medicine is to determine the extent to which an examinee can use the appropriate knowledge, skills, and judgement across the range of encounters that are within his area of practice. Life would be easy if we could observe each examinee's performance over the full range of encounters and could evaluate that performance in a way that everyone would agree with the evaluation. The best we can do is evaluate a sample; and a sample, by definition, is limited, nonrepresentative, and constrained because behaviour is observed under conditions that are at least somewhat different from those found in practice. In addition, no single assessment method is a panacea. There are inevitable trade-offs that must be recognized from one method to the next.

Kane developed a model for comparing the quality of scores derived from various assessment methods.<sup>7</sup> He noted that interpreting a score as a measure of an examinee's competence requires a chain of at least three inferences: evaluation, generalization, and extrapolation. Evaluation involves deciding whether the observed performance is good, bad, or indifferent: in effect, scoring the performance. This transformation from examinee actions to scores is deceptively difficult. For example, it is difficult to reward alternative and equally correct approaches to patient care. It is also difficult to achieve a balance between efficiency and thoroughness. The second link in the chain is generalization, which involves generalizing from the results of the observed performance to the universe; this link brings us back to the issues of sampling. Finally, extrapolation involves moving the conclusion about an examinee's performance from the assessment context, which is always artificial, to expected performance in actual practice. Kane illustrates the use of this model by comparing the relative benefits of three methods: direct observation of performance in real practice; simulations; and objective tests in terms of these three inferences.

Kane notes that, in real life situations, evaluation is difficult. We know that any face-to-face interaction between rater and examinee is necessarily less anonymous and more personal than an objective test. Bias is of considerable concern. To improve agreement across raters, rating criteria must be developed. In order to ensure that the rating criteria are applicable to the range of experiences that might occur in real life, they must be general and therefore subject to much judgement when applied. The second inference, generalization, is also a serious problem: testing is inconvenient and expensive, so it's uncommon to have a sample that is broad enough or representative enough. On the other hand, because assessment is based on observation in the real world, the third inference, extrapolation, is taken care of.

Simulations were developed to limit the problems inherent in evaluating real life performance. By definition, simulations limit the degree of fidelity. As a result, evaluation is easier with simulations than real life because scoring criteria can be developed; the more the simulation is constrained, the easier it is to evaluate the performance. Generalization is also less of a problem in simulations than in real life, because you can evaluate over a larger predefined domain. Extrapolation is more of a problem

than real life, since no matter how authentic a simulation is, it is still a simulation and examinees will not behave as they would in real life. Any simulation is a simulation - it's not real.

Finally, with objective tests, evaluation is relatively straightforward. We really do know how to generate a scoring key that experts agree on; items with no correct answer are dropped. Generalization is also no problem; we can sample 50 to 100 "patients" in an hour, and scores from one set of items are highly predictive of scores from a second set of items. The problem arises with extrapolation. Even those who believe MCQs are a form of simulation don't believe they are an exact match for real life.

The balance between fidelity and good measurement has been a theme in the measurement literature for decades. The specificity of the task affects the generalizability of the construct: as you standardize the assessment, you gain objectivity, but, as you limit and standardize the situation to ensure comparability of scores, you reduce the likelihood that the task will represent a realistic complex problem. The closer the assessment is to real life, the more scoring problems and ambiguity you will face. As Kane said, it's not that we haven't been clever enough, it's inevitable. Each method has something unique to offer; it is ridiculous to become such a zealot of one method that all other forms of assessment are avoided.

*"Science is the search for truth; it is not a game in which one tries to beat his opponent."*

*(Linus Pauling, No More War, 1958)*

## Conclusions

Assessment has been a focus of attention for teachers and learners for centuries. Despite the attention given to assessment, there are remarkably few truths that come to mind, and most of these have been known for generations. Nevertheless, it is important to keep in mind a few fundamentals when developing an assessment instrument and a program of evaluation. It is clear that assessment should be designed to enhance student learning and provide support for instruction. Because assessment drives student learning, it is important that assessment focus on major educational goals. It should also promote application of knowledge and integration across courses and clerkships. The extent to which psychometric properties are important depends on the purpose of the test - issues of reliability and validity are of increasing concern as the stakes of the testing is increased. Finally, no assessment method is a panacea. Generally, use of a blend of methods is appropriate. Content sampling and tasks posed by the test are more important than the method per se.

## References

1. Fitzpatrick R, Morrison EJ. Performance and product evaluation. In: Thorndike RL, editor. Educational Measurement. Washington, DC: American Council on Education, 1971:237-70.
2. Highland RW. A guide for use in performance testing in Air Force technical schools. Lowry Air Force Base, Colorado: Armament Systems Personnel Research Laboratory, 1955.

3. McGuire C. Written methods for assessing clinical competence. In: Hart IR, Harden RM, editors. *Further developments in assessing clinical competence*. Montreal: CanHeal Publications, 1987:46-58.
4. Newble D, Jaeger K. The effect of assessment and examinations on the learning of medical students. *Medical Education* 1983;17:165-71.
5. Mann H. A description of a survey of the Grammar and Writing Schools of Boston in 1845. Quoted in *Measurement Problems; Program Statistics Research Technical Report no. 92-20*. Princeton, New Jersey: Educational Testing Service, 1992:15.
6. Davison ML. How should performance assessments be evaluated? The role of scores. Letter to the editor. *Educational Measurement: Issues and Practice* 1996;15:37.
7. Kane MT. The assessment of professional competence. *Evaluation and the Health Professions* 1992;15:163-82.