Assessing Effectiveness of a Problem-Based Learning Curriculum in Teaching Clinical Reasoning Skills

DANE M. CHAPMAN, MD, PhD[1,2,3]
JUDITH G. CALHOUN, PhD, MBA[1]
ADRIAN P. VAN MONDFRANS, PhD[2]
WAYNE K. DAVIS, PhD[1]

Institutional Affiliations:

1. *University of Michigan Medical School, Department of Medical Education, Ann Arbor, MI*
2. *Brigham Young University, McKay School of Education, Department of Instructional Psychology and Technology, Provo, UT*
3. *University of Missouri-Columbia, School of Medicine, Center for Clinical Reasoning and Procedural Competency, Department of Emergency Medicine, Columbia, MO*

Abstract

Background:
Problem-based learning has been advocated in teaching clinical reasoning, yet it is unclear how to best measure clinical reasoning skills using this approach.

Objective:
To evaluate the efficacy of free-inquiry, virtual patient simulations compared to menu-driven, branching, written patient simulations in assessing data-gathering and clinical decision-making skills during a four-week, post-second year problem-based learning curriculum.

Methods:
Experimental, multiple-groups pretest-posttest control group and quasi-experimental pretest-posttest control group designs were used to evaluate expected improvements in clinical reasoning over the four-week curriculum. All post-second year medical students (N=191) were required to participate in the problem-based learning curriculum and complete a written patient simulation at the end of each week. Ninety-seven of the 191 medical students volunteered without compensation to participate in an additional 6.5-hour virtual patient simulation elective during the problem-based curriculum and were randomly assigned to one of the four weeks. Simulation elective students completed three virtual patient practice simulations with feedback and three assessment simulations (one pretest and two posttests) measuring nineteen dependent variables from four clinical reasoning constructs: proficiency, efficiency, thoroughness and strategy.

Results:
Multivariate, repeated-measures, factorial ANOVA statistics revealed a significant problem-based learning curriculum effect upon the second virtual patient simulation posttest along all four clinical reasoning constructs: proficiency (p = .03), efficiency (p = .01), thoroughness (p = .00), and strategy (p = .01). For three of the four constructs (proficiency, efficiency and strategy), no significant differences among the four weeks were found on multivariate analyses of the virtual patient pretest, suggesting a combined problem-based curriculum and virtual patient simulation practice enhancement of clinical reasoning competency. These enhancements were not detected by the written patient simulations.

Conclusions:
A four-week, problem-based learning curriculum can significantly enhance clinical reasoning competency and three hours of virtual patient simulation practice augments that effect. Results also support the utility of free-inquiry, virtual patient simulations in teaching and assessing clinical reasoning competency.

Keywords:
problem-based learning, clinical reasoning, competency assessment, virtual patient simulations, medical students

Introduction

In response to public pressure for greater accountability from the medical profession, medical schools and residency programs are undergoing major transformations to ensure students and residents learn what they need to know to become competent physicians.[1,2] Coupled with this accountability response, undergraduate and graduate medical education programs and certifying boards have stepped up efforts to assure physician competency through outcomes measurement.[3-4] While few would argue the importance of an outcomes approach to competency assessment, uncertainty exists on how to best teach and evaluate clinical reasoning skills of data gathering (i.e. history taking, physical examinations and selecting and interpreting diagnostic tests) and diagnostic and therapeutic decision making.[3-5]

Problem-based learning has been advocated for teaching clinical reasoning skills. Initially described by Neufeld and Barrows,[6] problem-based learning (PBL) has been implemented to varying degrees within the more conventional medical school curriculum. Of the eighty percent of U.S. medical schools that report using PBL, 45 percent report fewer than 10 percent PBL preclinical contact hours.[7] In a meta-analysis of 20 years of PBL experience (1972-1992) compared with conventional students, PBL graduates demonstrated gaps in their cognitive knowledge base and used a more backward rather than forward clinical reasoning approach characteristic of expert clinicians.[8] More recently, Hoffman et al.[9] using an outcomes approach, describe their ten year experience using PBL at the University of Missouri-Columbia School of Medicine from 1993-2006. They report significantly higher USLME Step 1 and Step 2 scores for their PBL students compared to first-time examinees nationally. These gains in performance appear to continue into residency as program directors note superior performance of the school's PBL graduates.[9] However, they do not report how well PBL graduates did on the USMLE Step 3 Computer-based Case Simulation (CCS) exam designed to measure an examinees' approach to clinical management, including diagnosis, treatment and monitoring.[9]

Uncertainty remains regarding the best method for measuring PBL effectiveness in teaching clinical reasoning.[3-5] Based upon a review of the past thirty-five years of professional medical assessment, new, more reliable modalities are recommended for assessing clinical reasoning and expert judgment among other professional attributes.[10] Fifteen years ago, the Accreditation Council for Graduate Medical Education (ACGME) initiated the "ACGME Outcome Project" to shift the focus of residency program requirements and accreditation from process-oriented assessment to an assessment of educational outcomes of resident and residency program performance as a basis for accreditation.[11] A tidal wave of change has come as a result of that initiative including the ACGME core competencies,[12] the ACGME toolbox,[13] and controversy over the psychometric inadequacy of the tool box instruments[13] and whether the core competencies themselves represent valid measurement constructs.[14]

Fueled by this massive push in educational outcomes assessment and attendant controversy, a critical need exists for psychometrically proven assessment modalities. Despite increasing popularity of computer-based or virtual patient simulations in medical schools and residencies, their role in teaching and assessing clinical skills and diagnostic and therapeutic decision making requires greater clarificaiton.[10,15-18] To maximize instructional validity (i.e. that the test actually measures what it purports to measure), an optimal virtual patient simulation will be high fidelity—meaning it will faithfully simulate the actual physician-patient encounter. It will also be free-inquiry—meaning users can access data freely without menus or other branching limitations, and without cues.[16-18] Rather than text or verbal descriptions of physical exam and diagnostic test findings, actual visual and auditory responses will be provided, such as visual cues for skin rashes, cardiac and respiratory sounds, and digital images for electrocardiographs (EKG's) and radiographs.

This study aims to evaluate whether: (1) high-fidelity, free-inquiry virtual patient (VP) simulations compared to menu-driven, branching written patient (WP) simulations are better in detecting the impact of a four week PBL curriculum (PBLC) upon clinical reasoning (CR) skill acquisition of post-second year medical students; and (2) three hours of VP simulation practice with feedback would further impact CR skill acquisition beyond the PBL curriculum alone. It was hypothesized that VP compared to WP simulations would be more sensitive (instructionally valid) in detecting significant improvements in clinical reasoning competency over the four-week PBLC, and that VP simulation practice would increase CR skill acquisition over PBLC alone.

Methods

Study Design:

We used an experimental, multiple-groups, pretest-posttest control group design[19] to assess the effect of the required four-week PBL curriculum and three hours of VP simulation practice (independent variables) upon nineteen previously established clinical reasoning skill indices[20] (dependent variables). The multiple-groups experimental pretest-posttest design was chosen to control for the influence of historical events and non-PBLC related maturation throughout the four weeks of the PBLC. Random assignment of students to groups made it unlikely that a consistent effect throughout the four weeks of the PBLC would be due to historical or maturational events unrelated to the PBL curriculum itself. As an additional control, a second analysis using a quasi-experimental pretest-posttest control group design[19] compared PBLC-only students with those students who also experienced VP simulation practice and assessment in addition to the PBL curriculum.

Study Setting and Population:

The study qualified for institutional review board (IRB) exemption as a curriculum innovation project and was conducted at the University of Michigan Medical School. The entire medical school class of post-second year medical students (N=191) was required to take the PBL curriculum between the preclinical and clinical years of medical school. Ninety-seven of the 191 post-second-year medical students volunteered without compensation to participate in a virtual patient simulation elective during the required four-week PBL curriculum and formed the VP elective group. The remaining 94 students formed the PBLC-only control group. The entire class of 191 students underwent random assignment to each week of the four week PBLC stratified on VP elective participation, gender and minority status such that of the 46-47 students randomly assigned to each week of the PBLC, 23-25 were randomly assigned to the VP elective.

Problem-based Learning Curriculum:

The PBL curriculum was a required, four-week block from May 4th to June 1st between the second and third years of medical school. It was "designed to provide a relevant practical clinical problem solving experience for students about to enter the clinical phase, with emphasis on the philosophy of lifetime learning and independent and group thought in the practice of clinical problem solving and diagnosis."[21] It consisted of expert presentations on the practice and concepts of clinical problem solving and physical diagnosis with faculty and students participating in both lecture and small group discussions. The PBLC sequence was originally developed around eight organ-system themes with each of the four weeks of the PBLC focusing upon two major organ systems: Week 1, cardiovascular and pulmonary; Week 2, gastrointestinal and renal; Week 3, neurologic and musculoskeletal; and Week 4, endocrine and reproductive.[22] The objectives of the PBLC were to: (1) bridge the gap between the basic science and clinical practice years of medical school, (2) review patient cases and provide practice in clinical problem solving and diagnosis, (3) emphasize clinical assessment skills, (4) provide mini-elective opportunities such as the VP elective, and (5) introduce medical students to the hospital setting during a clinical week.

During the mornings of their clinical week, PBLC students saw patients and functioned as part of the hospital ward team. All 191 students were expected to attend regularly scheduled afternoon dialectic sessions on Monday, Wednesday and Friday afternoons from 1:00 to 4:00 PM throughout the four-week PBLC. Completion of a weekly WP simulation was required of all PBLC students during the Friday afternoon dialectic session. VP students experienced their VP elective week during a week other than their clinical week. The 23-25 VP students randomly assigned to each week of the PBLC underwent further random assignment into one of four treatment subgroups.

Virtual Patient Simulation Elective:

The 6.5 hour VP elective included two sessions (3.0 and 3.5 hours) on Monday through Wednesday, Tuesday through Thursday or Wednesday through Friday mornings, during which students worked through six VP simulations: one 60-minute pretest (Case 1: Cardiology), three 60-minute practice simulations with corrective feedback (Case 2: Pediatric Endocrinology, Case 3: Infectious Disease and Case 4: Pulmonary), and two 45-minute posttests (Case 5: Pulmonary and Case 6: Cardiology). No corrective feedback was provided for pretest or posttest assessment simulations. VP students were randomly assigned to work individually or in groups of three during practice simulations. All VP students completed their pretest and posttest simulations as individuals.

Written Patient Simulations:

A different WP simulation was administered each week to all PBLC students during the Friday afternoon dialectic session. The WP simulations were developed by the University of Michigan Medical School following a menu-driven and branching format used by the National Board of Medical Examiners.[23] Each WP simulation began with an "opening scene" which introduced the patient and presented the chief complaint, the setting and the students' role. Each opening scene was followed by five sections representing steps in the workup, diagnosis and management of the patient. These sections were: A) History, B) Physical Examination, C) Diagnostic Studies, D) Differential and Principal Diagnoses, and E) Therapeutic Procedures. The first three sections incorporated 15-40 decision options. Some options were appropriate or indicated, others were inappropriate or contraindicated, and still others were optional or neutral.

Virtual Patient Simulations:

The multi-problem, network-based VP simulations used in this study simulated the actual physician-patient encounter with high fidelity and free inquiry, without menu-cueing or branching limitations. They were derived from content experts, had documented critical actions for data gathering and diagnostic elements, and could be altered to prevent corrective feedback during the assessment cases. Following an "opening scene," users assumed the role of physicians and moved to and from history, physical examination, diagnostic study, diagnosis and treatment sections without menu-driven cueing or branching limitations.[20]

The VP simulations were not the ultimate virtual patient, however, as artificial intelligent responses to all history, physical exam and diagnostic test inquiries were provided as text, and not virtual touch, sound, or images.

Assessments and Procedure:

WP Simulation: An individual score for each WP simulation was represented by a percentage of the correct points for each of the five sections. A "problem-solving index" was used as an overall performance score on each WP simulation and consisted of the average of section scores across the five sections.[22-23] Concurrent, criterion-referenced and instructional validities for the WP simulations have been previously reported.[22] The WP simulation assessments were administered on the Friday of each PBLC week simultaneously to VP elective and PBLC-only groups in a standard lecture hall of the University of Michigan Medical School.

VP Simulation: To standardize transcript scoring, coding regulations were developed using sample VP simulation transcripts. Case-specific VP simulation scoring protocols provided a summary of those expert-recommended critical inquiries that had been made. Diagnosis sections were independently scored by two individuals using case-specific coding regulations that identified acceptable synonyms for diagnoses. Transcripts were scored by at least one rater who was blinded to pretest-posttest classification, and inter-rater agreement was consistently high (r>.90). While therapeutic and management plans were also computer-scored, these were ignored for the purposes of this study.

The VP simulation pretest (Case 1: Cardiology) and both posttests (Case 5: Pulmonary and Case 6: Cardiology) were used to measure the effect of the four-week PBL curriculum upon the nineteen clinical reasoning competency indices, while controlling for any VP simulation pretest-posttest differences due

TABLE 1: Mathematical Descriptions of Nineteen Clinical Reasoning Performance Indices Derived for Use in Multi-Problem Virtual Patient Simulations

Index	Abbreviation	Description[a]
Proficiency		
History Taking	HTP	(Obtained CHT/Total CHT) X 100
Physical Examination	PEP	(Obtained CPE/Total CPE) X 100
Diagnostic Tests	DTP	(Obtained CDT/Total CDT) X 100
Correct Diagnoses	DP	(Obtained CD/Total CD) X 100
Program Solving Index	PSI	(HTP + PEP + DTP + DP) / 4
Proficiency Index	PI	(Obtained CHT + CPE + CDT) X 100 / (Total CHT + CPE + CDT)
Efficiency		
History Taking	HTE	(CHT Obtained/HTT) X 100
Physical Examination	PEE	(CPE Obtained/PET) X 100
Diagnostic Tests	DTE	(CDT Obtained/DTT) X 100
Thoroughness		
History Taking	HTT	Total HT
Physical Examination	PET	Total PE
Diagnostic Tests	DTT	Total DT
Total Data-Gathering	TDG	(HTT + PET + DTT)
Diagnosis	DT	Total D
Strategy		
History Taking	HTS	[HTT/(HTT+PET + DTT)] X 100
Physical Examination	PES	[PET/(HTT+PET + DTT)] X 100
Diagnostic Tests	DTS	[DTT/(HTT+PET + DTT)] X 100
Focused Strategy Index	FSI	(HH + PP+ DD +1) / (HP + HD + PH + PD + DH + DP + 1)
Invasiveness/Cost Index	ICI	[DTT/(HTT + PET)] X 100

[a] *Symbol Key: HT*= history taking inquiries, *PE*= physical examination inquiries, *DT*= diagnostic test inquiries, *D*= diagnoses indicated, *C*= critical inquiry or diagnosis (e.g. CHT=critical history taking inquiries), *HH*= history to history transition, *PP*= physical exam to physical exam transition, *DD*= diagnostic test to diagnostic test, *HP*= history to physical exam, HD= history to diagnostic test, *PH*= physical exam to history, *PD*= physical exam to diagnostic test, *DH*= diagnostic test to history, and *DP*= diagnostic test to physical exam transition.

to the instructional effect of the three VP practice simulations. The VP simulation pretest was administered during the first VP Session on Monday, Tuesday or Wednesday of each of the four PBLC weeks. The two VP simulation posttests were administered during the second VP session on Wednesday, Thursday or Friday. VP simulation assessments were administered in the Taubman Health Sciences Library Learning Resource Center. Proctors were available to respond to VP simulation computer interface or program questions.

VP Simulation Clinical Reasoning Indices: CR performance scores along nineteen predetermined CR indices[20] (dependent variables) were used to assess the effect of PBLC and VP simulation practice over the four weeks of the PBLC. The nineteen CR competency indices were previously validated and are represented as mathematical descriptions under their corresponding clinical reasoning construct: proficiency, efficiency, thoroughness, and strategy (See Table 1).

Clinical reasoning proficiency referred to how effectively critical data were gathered and correct diagnoses made. The CR proficiency indices were: percent of critical data-gathering inquiries obtained for history (history proficiency), physical examination (physical examination proficiency), and diagnostic tests (diagnostic test proficiency); percent of correct diagnoses made (diagnosis proficiency); Problem-Solving Index (PSI)—an average of data-gathering and decision-making proficiencies; and Proficiency Index (PI)—the percent of data-gathering critical information obtained.

Clinical reasoning efficiency was defined as the percentage of data-gathering inquiries that were critical in making the diagnosis of a patient's problem(s). Higher scores represented greater efficiency in making medical inquiries. Clinical reasoning efficiency indices included history, physical examination and diagnostic test efficiencies.

Clinical reasoning thoroughness reflected the frequency of data-gathering inquiries made or diagnoses indicated. Clinical reasoning thoroughness indices included: total number of history inquiries (history thoroughness), physical examination inquiries (physical examination thoroughness), and diagnostic test inquiries (diagnostic test thoroughness); total number of history, physical examination and diagnostic test inquiries combined (total data-gathering thoroughness); and total number of diagnoses hypothesized at the completion of each simulated case (diagnosis thoroughness).

Clinical reasoning strategy referred to the cognitive strategies used to arrive at correct diagnoses. It reflected individual preference for certain data-gathering techniques (e.g. to use either a focused inquiry approach or a "shot-gun" or haphazard approach). CR strategy indices included: percent of total data-gathering inquiries that relate to history taking (history strategy), physical examination (physical examination strategy), or diagnostic test (diagnostic test strategy); Focused Strategy Index—the standardized proportion of data-gathering inquiry transitions of similar type (e.g. history to history) to all other combinations of possible inquiry transitions from one type of inquiry to another (e.g. history to physical examination, diagnostic test to history, etc.), where high scores reflect a more focused and systematic data-gathering approach; and Invasiveness/Cost Index—the standardized proportion of diagnostic test inquiries (relatively invasive and costly) to the sum of history-taking and physical examination inquiries (relatively non-invasive and less costly), where higher scores reflect a more invasive and costly data-gathering approach.

Data Analysis:

BMDP multivariate, repeated measures, factorial ANOVA statistics were used to determine any pretest-posttest differences in CR constructs due to VP simulation practice while controlling for week of the PBL curricululm (independent and control variables). If a significant multivariate effect was observed, then univariate ANOVA statistics were used to test for differences along the nineteen clinical reasoning dependent variables.[20] Univariate, factorial ANCOVA statistics were used to compare VP elective and PBLC-only groups over PBL curriculm weeks 1 through 4. (http://www.statistical-solutions-software.com/bmdp-statistical-software/bmdp/).

Results

Complete data for 184 of the 191 PBL curriculum students were available for comparing VP elective and PBLC-only groups over the four weeks of the PBL curriculum. The VP elective students (N=97) appeared to be representative of the entire medical school class (N=191) as VP students did not differ significantly from other class members in ethnicity, sex, or prior clinical experience on the hospital wards. Approximately two-thirds of the VP enrollees had never previously participated in computer-based instruction and almost one-fourth had never interacted with a computer in any capacity at the time of the original study, making a selection bias, which favored students who were more comfortable with using computers for learning, unlikely.

Computer transcripts (N=486) were generated by 81 VP medical students with complete data, and documented student-computer interactions for 243 hours of practice and 202 hours of assessment. Outcome performance scores along nineteen predetermined CR indices[20] (dependent variables) were derived from the 243 hard-copy VP simulation assessment transcripts (one pretest: Case 1; and two posttests: Case 5 and Case 6) and are represented as means and standard deviations (SD) (See Table 2).

TABLE 2: Pretest-Posttest Means (SD) for Nineteen Clinical Reasoning Indices Over Four Weeks of the Problem-Based Learning Curriculum (N=81)[a]

Index	Week	Case 1 Pretest (Cardiology)	Case 5 Posttest (Pulmonary)	Case 6 Posttest (Cardiology)
Proficiency				
History Taking	1	38.1 (24.5)	57.7 (17.4)[c]	44.3 (15.0)
	2	50.0 (26.2)	55.9 (19.6)	52.8 (13.8)
	3	48.5 (25.7)	60.3 (19.4)	51.2 (14.5)
	4	44.3 (23.0)	51.7 (20.1)	48.2 (16.2)
Physical Examination	1	46.4 (27.2)	65.1 (19.7)[c]	52.4 (20.5)[cd]
	2	54.8 (17.0)	60.3 (13.4)	62.8 (19.3)
	3	57.3 (24.6)	72.6 (17.6)	62.3 (13.9)
	4	46.6 (20.8)	71.2 (18.7)	70.9 (19.2)
Diagnostic Tests	1	47.6 (23.6)	49.0 (10.6)[b]	42.3 (13.9)[bd]
	2	58.3 (31.0)	55.1 (17.0)	56.6 (13.6)
	3	69.1 (18.8	52.1 (7.0)	54.2 (20.4)
	4	61.4 (25.3)	50.6 (11.4)	57.1 (15.5)
Correct Diagnoses	1	28.6 (26.4)	55.9 (22.2)[c]	30.9 (20.8)[b]
	2	39.7 (27.1)	44.0 (23.6)	35.7 (18.7)
	3	29.4 (20.0)	47.0 (17.4)	26.5 (10.7)
	4	48.5 (26.7)	40.9 (19.7)	27.3 (17.1)
Program Solving Index	1	40.2 (13.1)	56.9 (8.6)[c]	42.5 (9.1)[e]
	2	50.7 (17.0)	53.8 (11.6)	52.0 (9.8)
	3	51.1 (12.0)	58.0 (10.9)	48.6 (8.1)
	4	50.2 (14.8)	53.6 (10.4)	50.9 (9.1)
Proficiency Index	1	44.0 (14.5)	55.6 (8.9)[b]	45.2 (7.8)[e]
	2	54.4 (18.0)	56.4 (12.5)	56.3 (9.2)
	3	58.3 (12.8)	59.1 (10.0)	54.6 (10.1)
	4	50.7 (16.0)	54.5 (11.5)	56.2 (9.3)
Efficiency				
History Taking	1	7.3 (4.4)	20.4 (8.7)[c]	26.3 (12.2)[c]
	2	8.8 (5.3)	20.2 (9.8)	22.7 (7.6)
	3	8.2 (4.0)	19.1 (7.8)	26.7 (7.5)
	4	8.4 (5.9)	17.8 (8.4)	24.9 (14.2)
Physical Examination	1	18.5 (9.2)	18.8 (6.7)	25.3 (10.9)[cd]
	2	21.3 (10.6)	19.6 (11.5)	24.7 (12.1)
	3	15.3 (5.9)	14.5 (8.3)	17.2 (8.4)
	4	15.4 (6.2)	16.3 (12.7)	19.8 (6.5)
Diagnostic Tests	1	40.4 (23.8)	45.1 (18.9)[c]	33.8 (9.6)
	2	32.3 (25.2)	38.8 (16.7)	32.2 (12.4)
	3	34.3 (14.2)	46.1 (13.0)	40.0 (18.2)
	4	34.5 (25.0)	40.3 (14.6)	39.0 (13.1)

[a] Repeated-Measures multivariate factorial ANOVA for Pretest - Posttest Comparisons, controlling for significant PBLC Week Effect for 81 medical students with complete data, significant pretest-posttest effect; [b]p ≤ 0.050; [c]p ≤ 0.010; significant PBLC Week effect [d]p ≤ 0.050; [e]p ≤ 0.010; [f]standardized T-Score mean=50, SD=10

Index	Week	Case 1 Pretest (Cardiology)	Case 5 Posttest (Pulmonary)	Case 6 Posttest (Cardiology)
Thoroughness				
History Taking	1	21.9 (8.5)	25.4 (11.2)[b]	20.8 (11.4)
	2	24.5 (10.8)	25.8 (11.0)	25.4 (8.8)
	3	23.3 (9.2)	27.8 (11.8)	20.3 (7.3)
	4	24.9 (10.5)	26.0 (10.6)	22.6 (11.7)
Physical Examination	1	10.0 (5.0)[d]	12.1 (5.2)[cd]	11.7 (5.5)[ce]
	2	11.5 (4.7)	12.8 (7.0)	14.9 (6.8)
	3	15.2 (5.0)	18.1 (6.7)	18.4 (3.4)
	4	12.8 (6.2)	17.4 (8.9)	18.9 (5.8)
Diagnostic Tests	1	5.5 (2.6)[e]	8.2 (3.0)[b]	11.9 (4.2)[cd]
	2	9.0 (4.4)	11.0 (4.8)	18.2 (8.0)
	3	9.8 (5.3)	8.6 (2.8)	13.6 (5.9)
	4	9.7 (4.7)	9.7 (3.3)	14.2 (5.6)
Total Data-Gathering	1	37.4 (11.1)[d]	45.7 (14.1)[c]	44.3 (12.4)[cd]
	2	45.0 (13.2)	49.6 (17.3)	58.5 (14.9)
	3	48.2 (11.7)	54.5 (15.4)	52.3 (9.5)
	4	47.4 (12.8)	53.0 (17.4)	55.8 (15.1)
Diagnosis	1	2.5 (1.1)[e]	3.0 (0.8)	4.3 (1.2)[c]
	2	3.7 (1.3)	3.3 (1.5)	5.8 (2.5)
	3	2.8 (0.8)	3.4 (1.1)	4.7 (1.5)
	4	3.1 (1.2)	2.6 (0.9)	4.5 (1.6)
Strategy				
History Taking	1	58.1 (11.3)	54.2 (11.3)	44.7 (16.3)[c]
	2	52.8 (12.8)	51.6 (8.6)	43.7 (12.9)
	3	47.5 (10.8)	50.3 (10.5)	38.3 (10.2)
	4	51.3 (12.8)	48.1 (8.9)	38.2 (13.3)
Physical Examination	1	26.4 (8.6)	26.6 (7.9)[e]	26.6 (10.6)[be]
	2	26.2 (9.6)	24.8 (6.7)	25.1 (8.6)
	3	31.4 (7.1)	33.0 (8.4)	36.0 (8.2)
	4	26.6 (9.2)	31.3 (9.0)	34.2 (6.6)
Diagnostic Tests	1	15.5 (7.1)	19.3 (8.0)[d]	28.7 (12.0)[c]
	2	20.9 (10.1)	23.6 (9.4)	31.2 (11.9)
	3	21.0 (11.5)	16.7 (6.0)	25.7 (8.4)
	4	22.0 (13.8)	20.5 (9.5)	27.6 (13.4)
Focused Strategy Index[f]	1	49.0 (8.4)	50.8 (9.3)[c]	47.7 (8.7)[c]
	2	49.6 (8.6)	47.3 (10.4)	50.3 (12.5)
	3	51.9 (11.7)	51.4 (7.9)	50.3 (7.8)
	4	52.6 (12.7)	50.7 (12.2)	51.0 (10.1)
Invasiveness/Cost Index[f]	1	45.8 (4.7)	48.5 (7.5)[d]	50.3 (9.7)[c]
	2	50.0 (7.7)	52.8 (9.1)	52.1 (10.3)
	3	50.5 (9.4)	45.9 (4.9)	47.6 (6.0)
	4	52.2 (13.4)	50.0 (9.7)	50.2 (12.5)

[a] Repeated-Measures multivariate factorial ANOVA for Pretest - Posttest Comparisons, controlling for significant PBLC Week Effect for 81 medical students with complete data, significant pretest-posttest effect; [b]$p \leq 0.050$; [c]$p \leq 0.010$; significant PBLC Week effect [d]$p \leq 0.050$; [e]$p \leq 0.010$; [f]standardized T-Score mean=50, SD=10

Detecting Problem-Based Learning Curriculum Impact Using VP Simulation Assessments:

Multivariate, repeated measures, factorial ANOVA revealed a significant week of PBLC effect for the second posttest (Case 6: Cardiology) across all four clinical reasoning constructs: proficiency (Wilks' λ=.61, F(18,170)=1.80; p = .03), efficiency (Wilks' λ=.73, F(9,153)=2.38; p = .01), thoroughness (Wilks' λ=.54, F(12,164)=3.53; p = .00), and strategy measures (Wilks' λ=.73, F(15,168)=2.38; p = .01) (See Table 2). A significant week of PBLC effect was also noted for the first posttest (Case 5: Pulmonary) along thoroughness (Wilks' λ=.66, F(12,164)=2.36; p = .00), and strategy constructs (Wilks' λ=.61, F(15,168)=2.21; p = .01); and for the pretest (Case 1: Cardiology) along the thoroughness construct (Wilks' λ=.59, F(12,164)=2.99; p = .00) (See Table 2). Since clinical reasoning competency differences over the weeks of the PBLC were noted for each of the four clinical reasoning constructs on the second posttest only, follow-up univariate ANOVAs were performed on the second posttest only.

Univariate ANOVA resulted in significantly higher proficiency posttest scores among the four groups (PBLC week 1 through 4) in physical exam (F(3,65)=3.25, p = .027), diagnostic tests (F(3,65)=3.96, p = .012), Problem-Solving Index (F(3,65)=4.56, p = .006) and Proficiency Index (F(3,65)=6.66, p = .000), but not in history taking or diagnostic accuracy proficiency indices (see Table 2, Figure 1). Univariate ANOVA of efficiency scores detected a significant decrease rather than increase in physical examination efficiency over the four-week PBLC (F(3,65)=3.51, p = .020), suggesting that even though students were making more overall physical examination inquiries, they were not obtaining proportionally more critical physical examination findings (See Figure 1). Univariate ANOVA demonstrated significantly higher posttest thoroughness scores over the four weeks of the PBLC in physical examination (F(3,65)=6.65, p = .000), diagnostic study (F(3,65)=3.45, p = .021), and total data gathering thoroughness (F(3,65)=3.69, p = .016), but not in history thoroughness (See Figure 2). Univariate ANOVA of strategy measures revealed a tendency to use proportionally more physical examination inquiries compared to history taking or diagnostic study inquiries over the four-week PBLC (F(3,65)=7.30, p = .000) (See Figure 2).

FIGURE 1: Clinical Reasoning Proficiency and Efficiency of Preclinical Medical Students (N=81) Using a Virtual Patient Simulation Over Four Weeks of the PBLC (*ANOVA, F Test)

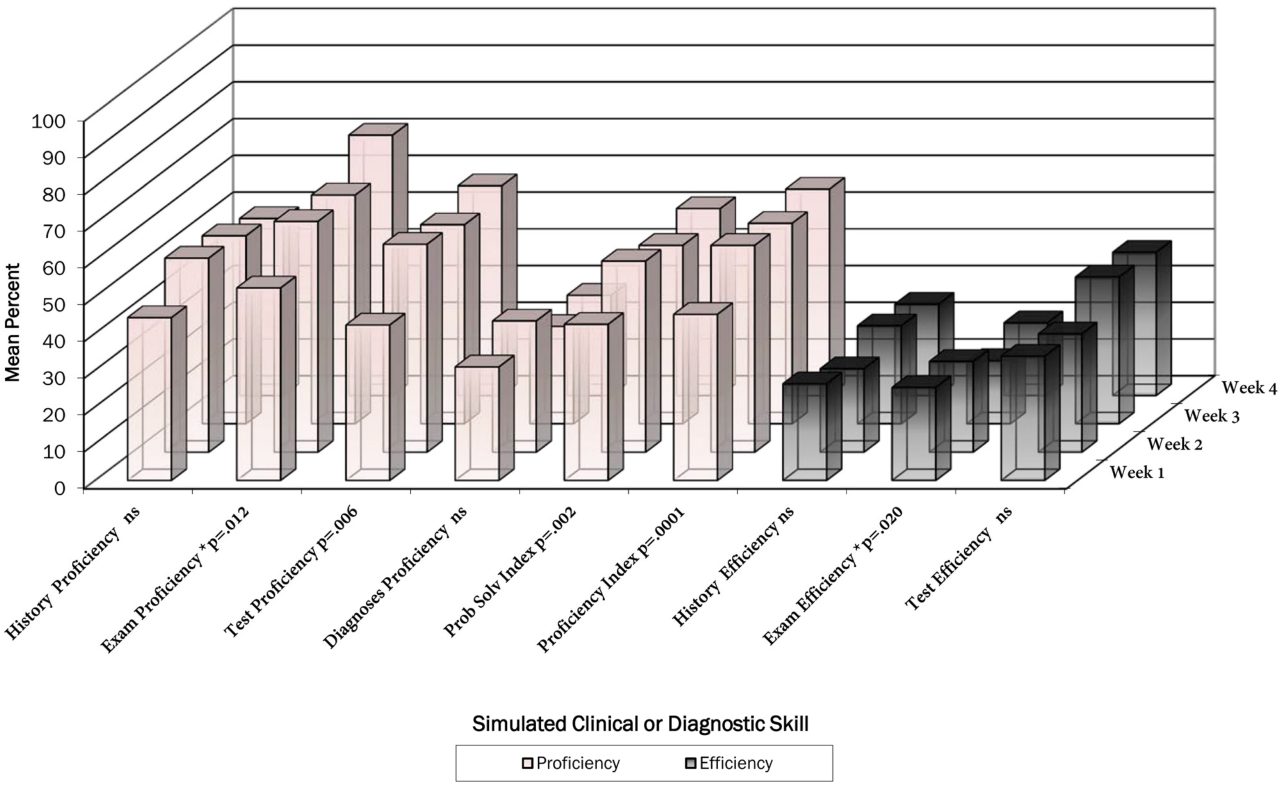
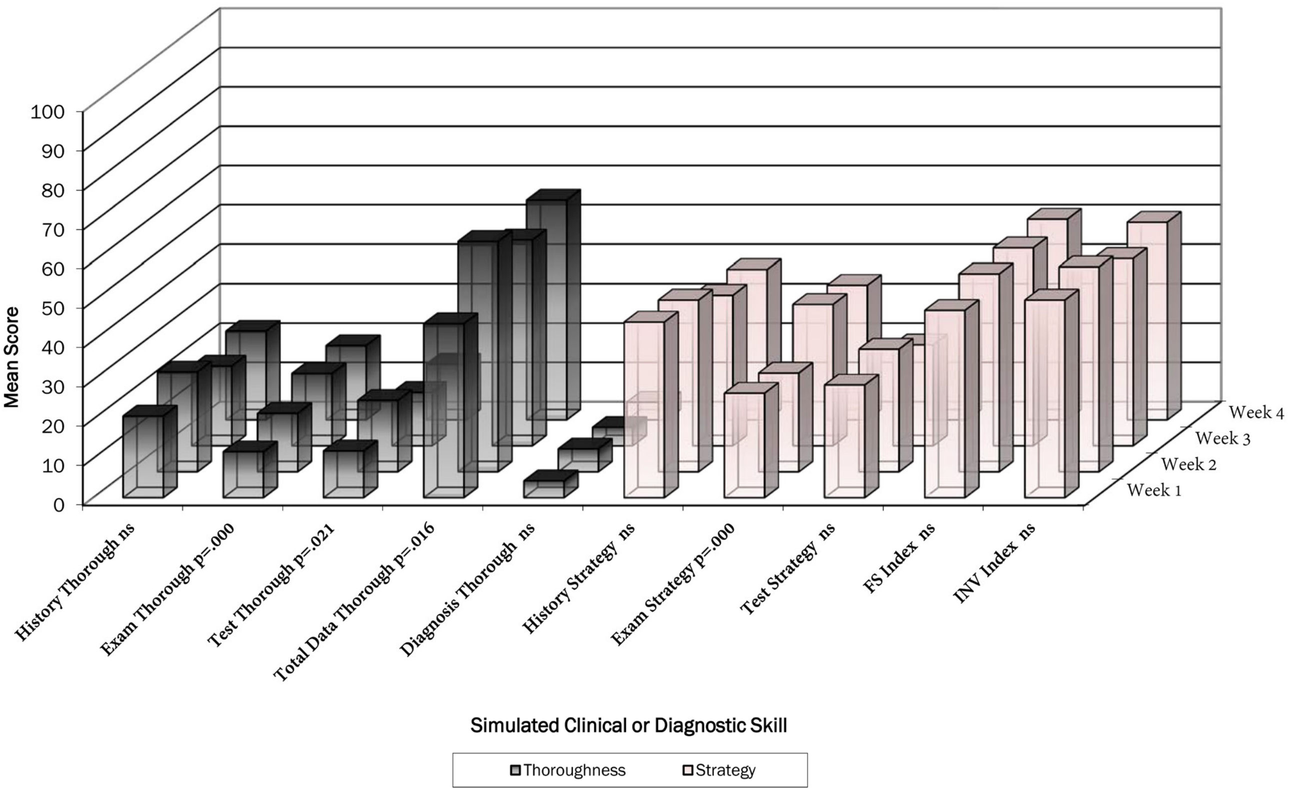


FIGURE 2: Clinical Reasoning Thoroughness and Strategy of Medical Students (N=81) Using a Virtual Patient Simulation Over Four Weeks of the PBLC (*ANOVA, F Test)



Detecting Problem-Based Learning Curriculum Using WP Simulation Assessments:

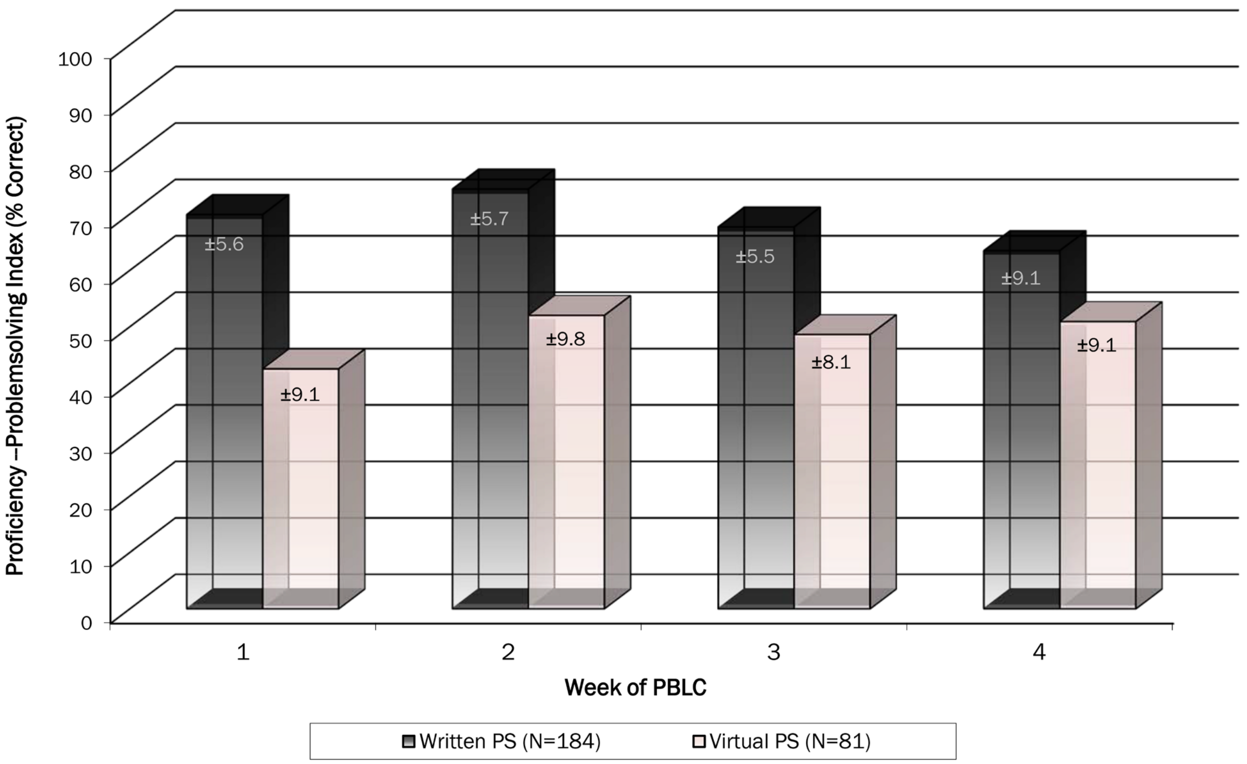
A strong PBL curriculum week effect was noted on all six WP simulation proficiency measures including History Taking (F(3,735)=55.32; p=.00), Physical Examination (F(3,735)=147.9; p = .00), Diagnostic Examination (F(3,735)=46.78; p = .00), Principal and Differential Diagnoses (F(3,735)=70.33; p = .00), Therapeutic Procedures (F(3,735)=138.9; p = .00), and Problem-Solving Index (F(3,735)=99.83; p = .00). However, rather than demonstrating improved clinical reasoning performance over each of the four weeks of the PBL curriculum, medical student (N=184) proficiency scores tended to decrease over the four weeks of the PBL curriculum (See Figure 3); suggesting WP simulation assessments were of varying difficulty level and were likely confounding our results.

Using multivariate ANOVA statistics, no differences in VP elective and PBLC-only performance were found for any of the six WP simulation proficiency measures over any of the four weeks of the PBL curriculum. To better control for any confounding pretreatment differences among students who enrolled in the VP elective and those who did not, and to increase the precision of comparisons between the VP elective and PBLC-only groups, a repeat analysis was done using univariate factorial analysis of covariance (ANOVA) statistics in a pretest-posttest control group design with the pretest (Week 1 WP simulation) being treated as the covariate. Comparisons were made between VP elective and PBLC-only groups for weeks 2, 3 and 4 of the PBL curriculum. Using this approach, no significant differences were noted between the VP elective and PBLC-only groups for any of the six WP simulation proficiency measures for any week of the PBL curriculum (p>.05, ANCOVA).

Discussion

The fact that both WP and VP simulation scores for CR constructs varied significantly over the four weeks of the PBLC supports the PBL curriculum impact on CR competency. Since no differences were found between VP elective and PBLC-only students, groups were likely equal in clinical reasoning proficiency during each week of the PBLC. If three hours of VP simulation practice was truly enhancing CR competency as measured by VP simulation assessments (See Table 2), such improvements were not being detected by the weekly WP

FIGURE 3: Comparison of Problem-Solving Index over Four Weeks of Problem-Based Learning Curriculum (PBLC) for Written and Virtual Patient Simulations (PS)



simulations. As such, the free-inquiry VP simulations appear to be a more sensitive measure of CR proficiency compared to the WP simulation assessments. Interestingly, expected gains in clinical reasoning competency outcomes over the four weeks of the PBL curriculum were also detected by VP simulations but not WP simulation assessments—again supporting VP simulations as more sensitive in measuring improvements in CR competency.

It should be noted that the WP simulations were limited to measuring CR proficiency and not the other three CR constructs of efficiency, thoroughness and strategy measured by the VP simulations. It is understandable that a significant PBLC week effect might not be detected by the VP pretest if CR proficiency, efficiency and strategy skills were not being taught in the PBL curriculum. However, the first posttest (Case 5: Pulmonary) likewise did not detect a PBLC week effect for proficiency and efficiency constructs, but did detect a strong pretest-posttest VP simulation practice effect where three hours of VP practice significantly improved CR proficiency, efficiency, thoroughness and strategy scores (See Table 2). Improvements in the second posttest (Case 6: Cardiology) and not the pretest (Case 1: Cardiology) suggest a strong combined effect of the PBL curriculum and the VP elective that does not appear to be case content related as both the pretest and second posttest were from the same content domain of cardiology.

These results demonstrate the utility of free-inquiry virtual patient simulations in assessing clinical reasoning competence as the second VP simulation posttest detected expected improvements in clinical reasoning proficiency, efficiency, thoroughness, and strategy over the four-week PBL curriculum. The four clinical reasoning constructs each provide a unique view; and, the results of this study suggest that they should be considered together in assessing overall clinical reasoning competency. For example, a thorough and complete medical record is desired whenever possible, and high scores along the thoroughness construct would seem to be desirable. However, when time is limited, physicians must be able to discriminate between critical and non-critical information, and high thoroughness scores may also reflect indiscriminate data gathering and listing of medical problems as potential diagnoses.[24] Such problem solvers would tend to score low along the efficiency construct. In this example, the thoroughness and efficiency constructs taken together provide a more accurate view of overall clinical reasoning competency.

Clinical reasoning allows physicians to move from positions of clinical uncertainty to points where the medical literature can offer guidance.[25] Greater understanding of the clinical reasoning process can potentially improve patient care by helping medical students and clinicians recognize the cognitive processes underlying their decision-making errors. Both data gathering and data integration have been found to be sources of error in diagnostic decision making.[26] Computer-based diagnostic decision consultation has been found to positively influence diagnostic decision making in clinicians and students, with a larger impact upon students.[27-28] Life-sized and web-based computer simulations are also gaining increasing acceptance in PBL curricula, and may be useful in assessing clinical skills and diagnostic and therapeutic decision making.[15,29]

This study has limitations. It was conducted nearly three decades ago as part of a PhD dissertation,[30] and was never formally published. With recent developments in the ACGME core competencies,[11] new accreditation system (NAS) and milestones,[1] the a priori development and validation of CR constructs with a scoring rubric using free-inquiry VP simulations has greater relevance now than thirty years ago. VP simulations have changed in some ways that might impact study results. Still, similar or greater limitations exist for live simulated patients and the USMLE® step 3® computer-based exam.[31] Although this study was conducted at a single institution at a single point in time some years ago, more than half of a large medical school class participated, and results would be expected to generalize to other post-second year medical students with similar aptitudes and experiences. It is less clear whether results would generalize to medical students in their clinical years or to residents and physicians.

In summary, medical schools and residency programs are undergoing major transformations to ensure physician competency through outcomes measurement. This study demonstrated that even a four-week PBL curriculum can significantly impact acquisition of clinical skill and diagnostic decision-making competency, and that adding just three hours of virtual patient practice significantly augments that effect. The instructional (construct) validity of the nineteen clinical reasoning proficiency, efficiency, thoroughness, and strategy indices[20] is again suggested as they detected expected changes in clinical reasoning competencies over the four-week PBL curriculum. It seems clear that free-inquiry virtual patient simulations will have an increasingly important role in clinical reasoning outcomes assessment in the future.

Received for publication February 20, 2013. Revisions received: February 26, 2013. Accepted: February 26, 2013. Available online: February 27, 2013.

Presented in part at the Society for Academic Emergency Medicine (SAEM) 25th Annual Meeting, San Antonio, Texas, May 21-24, 1995; American Educational Research Association (AERA) Education in the Professions Annual Meeting, New York City, NY, April 10, 1996.

Contributions of Authors: Conception (DMC, JGC, APV, WKD), design (DMC, JGC), analysis and interpretation (DMC, JGC), drafting (DMC, JGC, WKD), and revising (DMC) the article. DMC takes responsibility for the paper as a whole.

Acknowledgements: The authors are indebted to Robert H. Bartlett, MD and Richard G. Judge, MD, course directors of the Problem-Based Learning Curriculum, for their enthusiastic endorsement of the medical problem solving computer simulation elective; the ninety-seven University of Michigan medical students who enrolled in the elective; the Taubman Medical Library Learning Resource Center staff for their essential role in making the computer simulation elective an overwhelming success, namely: Cindy A. McBride, Patricia Tomlin, Ellen S. Hoffman, Achla B. Karnani, Chris M. Chapman, Denise M. Axelrod, Patricia W. Martin, Dennis M. Poupart, Jackie E. McKenzie, Lois L. Katon, Mark D. Heidt, William A. Vicini, and Sharon J. Love; and, Fredric M. Wolf, PhD, Elaine M. Hockman, PhD, Nancy P. Allen, PhD, C. Victor Bunderson, PhD, Russell T. Osguthorpe, PhD and Sally H. Cavanaugh, PhD for their insightful review and recommendations to improve this work; and, to Becky Bluett and Christine Downs for their assistance in preparing the manuscript.

Funding: In part by a University of Michigan curriculum innovation grant.

Conflicts of Interest: No personal, commercial, political, academic or financial conflict of interest was reported by any of the authors. DMC is editor-in-chief of JCRPC.

Ethical Approval: Authors report the study qualified for IRB exemption.

Correspondence: Dane M. Chapman, MD, PhD, Director, Center for Clinical Reasoning and Procedural Competency, Department of Emergency Medicine, University of Missouri-Columbia School of Medicine, 1 Hospital Drive, Columbia MO, 65201, e-mail: <chapmandan@health.missouri.edu>

References

1. Nasca TJ, Philibert I, Bringham T, Flynn TC. The next GME accreditation system-Rationale and benefits. NEJM Special Report. Available online: http://www.acgme-nas.org/nejm-report.html (accessed 26 February 2013).
2. Chapman DM, Hayden S, Sanders AB et al. Integrating the Accreditation Council for Graduate Medical Education core competencies into the Model of the Clinical Practice of Emergency Medicine. *Acad Emerg Med.* 2004;11(6):674-685.
3. Van Gessel E, Nendaz MR, Vermeulen B Junod A, Vu NV. Development of clinical reasoning from the basic sciences to the clerkships: a longitudinal assessment of medical students' needs and self-perception after a transitional learning unit. *Med Educ.* 2003;37:966-974.
4. Vernon DTA, Blake RL. Does problem-based learning work? A meta-analysis of evaluative research. *Acad Med.* 1993;68:550-563.
5. Elstein AS. Clinical problem solving and decision psychology: Comment on "The epistemology of clinical reasoning". Acad Med. 2002;75-S134-6.
6. Neufeld VR, Barrows HS. The "McMaster philosophy": an approach to medical education. *J Med Educ.* 1974;49:1040-50.
7. Kinkade S. A snapshot of the status of problem-based learning in U.S. Medical Schools, 2003-2004. *Acad Med.* 2005;80:300-301.
8. Albanese MA, Mitchell S. Problem-based learning: a review of literature on its outcomes and implementation issues. *Acad Med.* 1993 Jan;68(1):52-81.
9. Hoffman K et al. Problem-Based Learning Outcomes: Ten Years of Experience at the University of Missouri-Columbia School of Medicine. *Acad Med.* 2006 July;81(7):617-25.
10. Epstein RM, Hundert EM. Defining and assessing professional competence. *JAMA* 2002;287(2):226-235.

11. Swing SR. Assessing the ACGME general competencies: general considerations and assessment methods. *Acad Emerg Med.* 2002 Nov;9(11):1278-88.
12. ACGME/ABMS Joint Initiative: Attachment/Toolbox of Assessment Methods. Version 1.1. September 2000. Available online: http://www.partners.org/Assets/Documents/Graduate-Medical-Education/ToolTable.pdf (accessed 25 February 2013).
13. Green ML, Holmboe E. Perspective: the ACGME toolbox: half empty or half full? *Acad Med*. 2010 May;85(5):787-90.
14. Lurie SJ, Mooney CJ, Lyness JM. Commentary: pitfalls in assessment of competency-based educational objectives. *Acad Med.* 2011 Apr;86(4):412-14.
15. Murray D, Boulet J, Ziv A, Woodhouse J, Kras J, McAllister J. An acute care skills evaluation for graduating medical students: a pilot study using clinical simulation. *Med Educ.* 2002;36:833-841.
16. Friedman CP. The marvelous medical education machine or how medical education can be unstuck in time. *Acad Med*. 2000;75(10):S137-142.
17. Harless W. CASE: A computer-assisted simulation of the clinical encounter. *J of Med Educ.* 1971;46:443-448.
18. Hawkins R, Gaglione MM, LaDuca T, Leung C, Sample L, Gliva-McConvey G, Liston W, De Champlain A, Ciccone A. Assessment of patient management skills and clinical skills of practicing doctors using computer-based case simulations and standardized patients. *Med Educ.* 2004;38:958-968.
19. Campbell DT, Stanley JC 1963. E*xperimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally College Publishing Company.
20. Chapman DM, Calhoun JG, Van Mondfrans AP, Davis WK. Measuring clinical reasoning competency using a virtual patient model. *JCRPC.* 2013;1(1):1-9.
21. The University of Michigan Medical School Bulletin. 1982; pp. 11-14.
22. Wolf FM et al. Concurrent and criterion-referenced validity of patient management problems. Proceedings of the Twenty-Second Annual Conference on Research in Medical Education, 1983. Washington, D.C.: Association of American Medical Colleges, 115-121.
23. Allen NP et al. The development of patient managment problems to assess clinical problem-solving knowledge and abilities of second-year medical students (Abstract). 1982. Innovations in Medical Education Exhibits. Washington, D.C.: Association of American Medical Colleges, p. 6.
24. Voytovich AE, Rippey RM, Copertino L. Scorable problem lists as measures of clinical judgment. *Eval Health Prof.* 1980;3:159-171.
25. Kempainen RR, Migeon MB, Wolf FM. Understanding our mistakes: a primer on errors in clinical reasoning. *Med Teach.* 2003;25(2):177-181.
26. Gruppen LD, Wolf FM, Billi JE. Information gathering and integration as sources of error in diagnostic decision making. *Med Decis Making.* 1991;11(4):233-9.
27. Berner ES, Webster GD, Shugerman AA et al. Performance of four computer-based diagnostic systems. *NEJM.* 1994;330(25);1792-96.
28. Friedman CP, Elstein AS, Wolf FM et al. Enhancement of clinicians' diagnostic reasoning by computer-based consultation. *JAMA*. 1999;282:1851-56.
29. Kamin C, O'Sullivan P, Deterding R, Younger M. A comparison of critical thinking in groups of third-year medical students in text, video, and virtual PBL case modalities. *Acad Med.* 2003;78(2):204-211.
30. Chapman DM. Teaching and evaluating clinical reasoning through computer-based patient management simulations. *Dissertation Abstracts International*. 1985;46:784-B.
31. USMLE Bulletin of Information. Available online: http://www.usmle.org/bulletin/overview (accessed 25 February 2013).