Challenges and Opportunities to Enhance the Clinical Reasoning Skills of Medical Students: Lessons Learned from the Advanced Clinical Transactions Pilot Program

Latha Chandran
Frederick Schiavone
Maryam Navaie
John Chen

Follow this and additional works at: https://scholars.fhsu.edu/alj

Part of the Educational Leadership Commons, Higher Education Commons, and the Teacher Education and Professional Development Commons

Recommended Citation
Chandran, Latha; Schiavone, Frederick; Navaie, Maryam; and Chen, John (2009) "Challenges and Opportunities to Enhance the Clinical Reasoning Skills of Medical Students: Lessons Learned from the Advanced Clinical Transactions Pilot Program," Academic Leadership: The Online Journal: Vol. 7 : Iss. 1 , Article 10.
Available at: https://scholars.fhsu.edu/alj/vol7/iss1/10

This Article is brought to you for free and open access by FHSU Scholars Repository. It has been accepted for inclusion in Academic Leadership: The Online Journal by an authorized editor of FHSU Scholars Repository.
Academic Leadership Journal

BACKGROUND

According to the Association of American Medical Colleges (AAMC) Task Force on the Clinical Skill Education of Medical Students, “skillful performance in the act of medical care is fundamental to the delivery of quality professional service to those who seek the care of a physician” (Nutter and Whitcomb 2000). One of the areas emphasized by AAMC’s Task Force was the analytic component of basic clinical methods, namely the ability of medical students to diagnose clinical problems including differential reasoning and problem identification. A subsequent report of the Ad Hoc Committee of Deans commissioned by AAMC recommended that curriculum reform target the entire continuum of medical education, with emphasis on undergraduate programs (Ad Hoc Committee of Deans, 2004).

Medical school faculty must foster high-quality patient care while simultaneously assessing the clinical skills and reasoning of students in order to promote their independent functioning in the clinical setting. A growing concern noted across medical faculty is that students may complete their undergraduate medical studies without sufficient understanding of how simple clinical tasks are done with real patients. These concerns have been validated by results from numerous studies that have reported many medical students express a lack of confidence and competence to critically evaluate patient needs based on data, a factor that has been linked to problems in performance and patient safety (Radcliffe and Lester 2003; Seabrook 2004; Chumley et al. 2005; Patey et al. 2007).

There is no clear-cut solution in the evidence. It has been well acknowledged that clinical reasoning pathways are influenced by the clinician’s knowledge base, the contextual presentation of the patient, and the clinician’s experience (Schmidt et al., 1990; Cunnington et al., 1997; Eva 2005; Norman 2005; Bowen 2006; Eva et al. 2007). Although numerous clinical reasoning frameworks have been proposed, the accumulated evidence suggest that the process of clinical reasoning is complex and likely to include a combination of analytic and non-analytic approaches (Bowen 2006). Researchers have identified multiple key components of clinical reasoning, such as: (a) content specificity (medical knowledge); (b) pattern recognition; (c) mental matrices based on Bayesian probabilities; (d) decision trees; and (e) causal reasoning in the form of multiple “if-then” rules. The literature has also concisely depicted the hierarchical progression of diagnostic reasoning: (1) intake assessment of the patient’s story; (2) data acquisition through past medical history, physical examination, laboratory testing, imaging results, and other forms of information; (3) accurate problem representation, (4) generation of comparative hypotheses and illness scripts to aide in differential diagnosis generation, (5) diagnosis determination using non-analytic (e.g., pattern recognition of clinical presentation) and analytic reasoning strategies (Bowen, 2006). The complexity of clinical reasoning makes it difficult for medical educators to generate a simple solution to the problem.

To date, despite a growing body of research on clinical diagnostic reasoning, there is little data on curricula designed to enhance the clinical reasoning skills of undergraduate medical students. In a relatively recent effort to support medical education reform initiatives during the final two years of
undergraduate medical school, the AAMC joined forces with the New York Academy of Medicine and other leading sponsors to nationally launch the Enhanced Clinical Transactions (ECT) initiative. This initiative aimed to improve the clinical transaction skills of undergraduate medical students (Corbett, 2004). Through its ECT project, the AAMC lent strategic support to selected Schools of Medicine to develop creative, yet practical, approaches to enhancing the clinical diagnostic reasoning skills of their undergraduate student body. Following a nationally competitive review process, six academic medical institutions were awarded grants to pilot innovative programs under the ECT initiative including the State University of New York at Stony Brook (SUNY Stony Brook), Harvard University, Case Western Reserve University, Vanderbilt University, University of Washington, and Syracuse University. As one of the national award recipients, the Office of the Dean in the School of Medicine at SUNY Stony Brook designed and implemented the Advanced Clinical Transactions (ACT) Pilot Program targeting third and fourth year medical students.

The purpose of this article is to build on the dearth in the literature by providing an overview of the ACT Pilot Program components, its primary results, and unexpected yet valuable secondary findings. In addition, insightful lessons learned pertaining to the challenges and opportunities related to implementing programs to enhance clinical reasoning skills are presented to illuminate future practice implications for teaching and learning in medicine.

DESCRIPTION
Overview of the ACT Pilot Program
Program Goals, Objectives, Design and Components
The primary goal of the ACT Pilot Program was to enhance the undergraduate medical education experiences of third and fourth year medical students related to conducting proper clinical transactions. The program’s focus centered on enhancing the students’ ability to: (1) accurately capture comprehensive and sequential patient medical histories, (2) conduct thorough and informed clinical examinations, (3) engage in sequential clinical reasoning using relevant clinical and laboratory data; (4) demonstrate a sophisticated understanding of the pathophysiology of each abnormal symptom and the relation to clinical presentation; and (5) develop humanistic and patient-centered attitudes in the delivery of care. The program provided an opportunity to augment students’ skills beyond the standard medical curriculum at SUNY Stony Brook through the application of case-base demonstrations, simulations, computer-assisted instruction and integration of physical findings into the overall clinical diagnostic process. Approval from the Institutional Review Board (IRB) at SUNY Stony Brook was obtained prior to implementing the ACT Pilot Program.

The ACT Pilot Program was designed by Master Clinical Instructors at SUNY Stony Brook School of Medicine who were senior medical school faculty with in-depth subject content expertise and extensive years of clinical practice experience. These faculty were selected to develop education modules through a competitive intramural review process that awarded them modest monetary stipends for their participation. The program employed multiple instruction modalities including: (a) didactic lectures, (b) case-based learning through vignettes, (c) computer-assisted instruction, (d) bedside teachings, and (f) real as well as simulated patient encounters. In addition, students participated in four-hour small group “pull out” sessions twice monthly during clerkship rotations devoted to the study of advanced physical diagnosis skills and clinical decision-making. The pull out sessions represented a dramatic change in the way that third-year clerkships were generally organized at SUNY Stony Brook. These sessions provided a systematic and centralized structure in which to: (a) enhance students’ skills; (b) reinforce
critical techniques in history taking, clinical reasoning and physical examination; (c) present opportunities for students to address problematic issues or concerns; and (d) ensure increased faculty time and availability for students. Embedded in all of the sessions was an instructional emphasis on clinical decision making pathways to facilitate proper diagnostic discernment.

The key components of the ACT Pilot Program included nine clinical training modules that were taught within nine sessions, supplementing the core medical curriculum at SUNY Stony Brook. These modules focused on the following subject areas: musculoskeletal (family medicine module); cardiology and pulmonary (two internal medicine modules); obstetrics (OBGYN module); growth and development as well as child abuse in a developmental context (two pediatric modules); abdominal examination (surgery module); and functional assessment as well as mental status assessment in older adults (two geriatrics modules). Upon development of modules by Master Clinical Instructors, subsequent selection and finalization ensued by a faculty review committee focused on incorporating advanced clinical reasoning in SUNY Stony Brook’s undergraduate medical curriculum.

Study Population, Metrics and Data Collection
At the completion of the second year of medical school, students at SUNY Stony Brook utilize a lottery system that permits the selection of clerkship disciplines (core requirements plus electives), the order in which clerkships are completed, and clerkship sites (at SUNY Stony Brook or at pre-determined affiliated clinic sites). Given these system restrictions, all students who received clerkship rotations at SUNY Stony Brook were automatically enrolled in the ACT Pilot Program (ACT program participants). Students were not made aware that they were receiving augmented clerkship training; that is, their participation in the ACT Pilot Program was blind, with students assuming that all training was “standard” practice. Students who completed their clerkships at affiliated sites rather than at SUNY Stony Brook served as the comparison group (non-ACT participants). Given the dynamic clerkship selection process at SUNY Stony Brook, in essence, there were nine different ACT participant and non-participant groups (i.e., one for each ACT session) since some students considered as ACT participants for one module may have served as non-ACT participants for another module if they completed one rotation at SUNY Stony Brook and another elsewhere. Randomization of ACT and non-ACT participants was not feasible given SUNY Stony Brook’s medical school policy governing clerkship selection and rotations processes.

For each of the ACT Program’s training modules (except the growth and development module for which no testing was done due to time constraints), data were collected specific to clinical performance for various tasks using short essays or standard checklists containing multiple Likert scale schematics appropriate for the skill being performed by the student and evaluated by the Master Clinical Instructor. For each of the ACT sessions, assessments of clinical reasoning skills encompassed analysis of checklists and short essay answers (rated by two trained physician investigators and co-authors, LC and AM) that captured proficiency in performing a patient history and physical examination, appropriate data acquisition (e.g., lab testing and radiology) and interpretation, accurate problem assessment, generation of comparative hypothesis for differential diagnosis, diagnosis determination using pattern recognition of clinical presentation and analytic reasoning strategies, as well as patient communication and professionalism. For example, in assessing a student’s history taking skills, instructors rated skill performance based on multiple questions using a three-point scale that ranged from “not done” to “done”. When evaluating data interpretation skills, a five-point scale was used ranging from “worst” to “best”. When assessments were based on
standardized patients (SPs), a four-point scale was utilized by the SPs ranging from “task omitted” to “task completed”. Score ranges and interpretations were determined, a priori, by two trained physician investigators blinded to the ACT participation status of the students. Inter-rater reliability was over 90%.

When disagreement was present between the two physician raters, resolution was reached by averaging disparate scores to generate the final score. In addition, data specific to gender, age, duration of time lapsed between completion of a specific ACT session and related Objective Structured Clinical Examination (OSCE), and the number of ACT sessions participated in by each student were gathered.

Statistical Analysis
Clinical performance scores were computed for each ACT training module separately and as an aggregated measure that represented a summary (overall) performance score. Based on a task’s relative importance in the clinical diagnostic reasoning process, Master Clinical Instructors assigned weighting coefficients on an a priori basis. In several modules, weight assignments were allocated as follows: 1 negative point was given for any significant clinical error made by students, 1 point was given for each required task that would support diagnosis differentiation, or 2 points were given for top priority tasks without which the ability to render a differential diagnosis or an appropriate management strategy would be severely compromised.

OSCE scores were used to compare primary programmatic outcomes of each ACT session between ACT Program participants and non-ACT Participants. To allow for meaningful comparisons across various ACT sessions, standardized deviation scores, calculated as [(raw score – mean)/ standard deviation], were computed. For each of the eight ACT sessions (covering nine training modules), means of the standardized scores for ACT Program participants were compared with non-ACT participants for each ACT session using two-sample t-tests. The overall significance of all ACT sessions was computed using a weighted inverse chi-square method for correlated significance tests (Makambi, 2003).

Multiple linear regression analyses were performed for OSCE scores related to each of the ACT sessions to further evaluate programmatic effects, adjusted for potential confounders, including gender, age, duration (days) between completion of a specific ACT session and the OSCE exam, and whether the students participated in other ACT sessions. Given the lack of randomization, National Board of Medical Examiners (NBME) Subject Examination Scores and United States Medical Licensing Examination Step I scores for 2005 and 2006 classes were examined to evaluate for possible quality differences between ACT Program participants and non-participants. For scores from each year, two sample t-tests were utilized to evaluate the score difference between medical students who completed clerkship rotations at SUNY Stony Brook, where the ACT sessions were implemented in 2006, and those who completed their clerkship rotations at other affiliated sites.

The impact of ACT sessions on module-specific outcomes was also examined. For the surgery abdominal pain module, whether ACT Program participants received higher scores from the SP, as compared to non-ACT Program participants, was evaluated. For the geriatrics functional assessment module, a comparison of ACT Program participants and non-ACT participants relative to performing a “get up and go” test of mobility as well as scores on the patient’s Activities of Daily Living (ADLs) assessment were examined. For the geriatrics mental status module, additional analyses explored whether ACT Program participants were more likely to perform a depression screening and receive a higher score on their performance of a Mini-Mental Status Exam (MMSE), as compared to non-ACT
participants. Furthermore, the OB module was divided into a delivery skills sum score and a problem solving score for the areas that involved arrest of dilation, interpretation of fetal heart tracing and post partum management in a patient with fetal distress and postpartum hemorrhage. For the pediatric child abuse module, additional analyses included whether ACT Program participants were more likely to identify child abuse in their differential diagnosis as well as whether they received higher scores when examining SPs, compared to non-ACT participants. For the pulmonary module, these included a history and physical sum score and a management sum score. In addition, supplemental analyses were performed to examine whether a higher history and physical score was associated with a higher management score, independent of ACT program participation status.

For comparison of module-specific outcomes between the two groups, two-sample t-tests for continuous measures and Fisher’s exact tests were used for categorical measures. To adjust for important covariates, multiple linear regressions and multiple logistic regressions were implemented for continuous and categorical outcomes, respectively. Adjusted mean differences with standard errors for continuous outcomes and odds ratios with 95% confidence intervals for categorical outcomes were computed, as appropriate.

EVALUATION
Primary Findings of the ACT Pilot Program
Table 1 provides descriptive characteristic comparisons between ACT Program participants and non-ACT participants for each of the ACT sessions. For the majority of the sessions, there were no significant differences between the two groups in gender or age distribution. However, for modules in abdominal pain and cardiology, ACT participants were significantly older (p<.05) and less likely to be women. For the pulmonary module, non-ACT participants tended to be older than ACT Program participants (p<.05). Furthermore, for both of the geriatric modules, a significantly higher proportion of ACT Program participants were women, compared to non-ACT participants (p<.05).
Comparisons of OSCE scores for each of the ACT Pilot Program sessions between ACT Program participants and non-ACT participants are depicted in Table 2. Crude comparisons (unadjusted) suggested significant differences between the two groups in the geriatric skills modules. However, aggregate crude performance scores indicated no significant difference in OSCE scores between the two groups (p = 0.12). After adjustments for potential confounders, no significant differences were observed in OSCE scores for any of the modules between the two groups.
Table 3 provides summary comparisons of National Board of Medical Examiners Subject Shelf examination scores and USMLE Step I scores for SUNY Stony Brook students from the class of 2005, the year preceding the ACT Pilot program and the class of 2006 who participated in the ACT Pilot program. Since system restrictions at SUNY Stony Brook did not make randomization feasible among the study population, Step I scores between the groups were used as surrogate indicators to assess comparability between the two groups in their basic science knowledge. NBME subject examination scores were used as a surrogate of their clinical knowledge and problem solving abilities. There were no significant differences in Step I scores among students who attended SUNY Stony Brook versus those who engaged in clerkship rotations at other affiliate clinic locations in either year. For the Internal Medicine and Pediatrics board examinations, the class without the ACT program had significantly higher scores among those students who attended SUNY Stony Brook (p=0.03 and p=0.04, respectively). Interestingly, after the ACT program, there were no differences in the shelf exam scores between ACT Program participants and non-ACT participants even in Internal Medicine or Pediatrics.
Table 4 details the results of supplemental stratified outcomes for selected ACT Program modules. In the pulmonary module, history and physical score did not appear to be associated with total management score (p= 0.18). The crude analysis for the geriatric module specific to performance of the “get up and go test”, total scores in the MMSE and the SP score in the child abuse case were statistically significant (p=.003, 0.05, and 0.02, respectively), however the confidence interval was very wide, suggesting a high degree of imprecision. Not surprisingly, after adjustments for covariates, no statistically significant differences between the groups were observed for this assessment. Additionally, for the Obstetrics module, there was a highly significant difference in the delivery performance skills of the ACT group compared to the control group, even after adjusting for confounding variables (p=0.02), while no differences were found in the problem solving score between the two groups.
Secondary Findings of the ACT Pilot Program

In addition to the ACT Pilot Program’s primary results, two unexpected secondary findings provided valuable insights that served as the basis for proposed changes in SUNY Stony Brook’s medical school policy and curriculum. Both of these secondary findings had strong implications for teaching clinical diagnostic reasoning skills. Our first unexpected finding was that although our Master Clinical Instructors had prepared to teach our medical students “advanced” clinical reasoning skills, their repeated empirical observations soon revealed that most of the students were unable to adequately demonstrate “basic” clinical reasoning; that is, students were ill prepared to differentiate diagnoses during simulated as well as live patient encounters. Although faculty observations at SUNY Stony Brook were similar to other reports in the literature (Elnicki et al., 2003; Groves et al., 2003; Poncelet and O’Brien, 2008), this finding was met with general surprise across our faculty, irrespective of clinical specialty area.

The clear disconnect between the expectations of Master Clinical Instructors regarding the third and fourth year medical students’ proficiency levels in clinical reasoning and the students’ actual competency levels was another significant finding of the ACT study. The general expectation among the faculty was that students had acquired basic clinical reasoning by the completion of their third year, a presumptive belief given observed student performances. Although our faculty were surprised by the extent of disparity between their perceptions and actual medical student ability, this finding also has been reported by past researchers (Prince et al., 2000; Makoul, 2000; Radcliffe and Lester, 2003; Chumley et al., 2005; Prince et al., 2005; Bowen, 2006; Whipple et al., 2006; O’Brien et al., 2007).

These secondary findings heightened concerns about the lack of sufficient emphasis on teaching introductory clinical reasoning skills earlier in the medical education curriculum. The medical faculty began exploring various factors that could have contributed to inadequate understanding of diagnostic pathways and differential clinical reasoning among third and fourth year medical students. By consensus, it was acknowledged that a lack of structured curricular emphasis on clinical reasoning was likely to be the leading contributor to the observed inadequacies. As described in detail elsewhere

---

### Table 4. Comparative results of ACT program sessions on module-specific outcomes

<table>
<thead>
<tr>
<th>ACT Session</th>
<th>Outcome Description</th>
<th>ACT</th>
<th>Non-ACT</th>
<th>Unadjusted (s.e.)</th>
<th>Adjusted (s.e.)</th>
<th>P-value</th>
<th>Adjusted (s.e.)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abdominal Pain</td>
<td>Total SP score (n)</td>
<td>41</td>
<td>33</td>
<td>38.3 (1.46)</td>
<td>40.5 (1.55)</td>
<td>0.31</td>
<td>-0.18 (0.32)</td>
<td>0.96</td>
</tr>
<tr>
<td>Geriatric skills Functional</td>
<td>ADL total (n)</td>
<td>36</td>
<td>40</td>
<td>7.7 (0.21)</td>
<td>7.5 (0.24)</td>
<td>0.55</td>
<td>0.15 (0.39)</td>
<td>0.70</td>
</tr>
<tr>
<td>“Get up and go” done %</td>
<td>No</td>
<td>0 (0.0)</td>
<td>9 (22.5)</td>
<td>22.02 (1.23, 393.59)</td>
<td>0.003*</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>36 (100)</td>
<td>31 (77.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geriatric skills Psychiatric</td>
<td>MMSE total (n)</td>
<td>36</td>
<td>40</td>
<td>27.8 (.05)</td>
<td>24.6 (1.17)</td>
<td>0.05*</td>
<td>0.19 (2.10)</td>
<td>0.93</td>
</tr>
<tr>
<td>“depression scale” done %</td>
<td>No</td>
<td>7 (19.4)</td>
<td>13 (32.5)</td>
<td>1.99 (0.69, 5.75)</td>
<td>0.30</td>
<td>2.29 (0.50, 10.62)</td>
<td>0.29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>29 (80.6)</td>
<td>27 (67.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obstetrical skills and</td>
<td>Delivery performance</td>
<td>27</td>
<td>46</td>
<td>17.6 (0.87)</td>
<td>15.6 (0.52)</td>
<td>0.03*</td>
<td>2.80 (1.17)</td>
<td>0.02*</td>
</tr>
<tr>
<td>reasoning</td>
<td>Problem solving</td>
<td>27</td>
<td>46</td>
<td>23.5 (1.17)</td>
<td>23.3 (1.05)</td>
<td>0.91</td>
<td>-0.34 (2.12)</td>
<td>0.88</td>
</tr>
<tr>
<td>Child abuse in a developmental context</td>
<td>Total SP score (n)</td>
<td>43</td>
<td>36</td>
<td>43.5 (1.07)</td>
<td>39.6 (1.28)</td>
<td>0.02*</td>
<td>3.85 (2.21)</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>Abuse %</td>
<td>No</td>
<td>5 (12.5)</td>
<td>3 (8.1)</td>
<td>0.62 (0.14, 2.79)</td>
<td>0.72</td>
<td>0.21 (0.62, 2.28)</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>35 (87.5)</td>
<td>34 (91.9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pulmonary</td>
<td>Total management (n)</td>
<td>36</td>
<td>44</td>
<td>9.1 (0.37)</td>
<td>9.0 (0.42)</td>
<td>0.84</td>
<td>-0.22 (0.83)</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>H&amp;P total (n)</td>
<td>36</td>
<td>44</td>
<td>12.8 (0.40)</td>
<td>12.7 (0.34)</td>
<td>0.79</td>
<td>0.03 (0.81)</td>
<td>0.97</td>
</tr>
</tbody>
</table>

*For continuous outcome variables (Total SP score, ADL score, MMSE total, Delivery performance, Problem solving, Total management, H&P score, mean (s.e.)); For categorical outcome variables (“Get up and go” done, “depression scale” done, Abuse); crude odds ratio (95% CI); Adjusted for gender, age, duration (days) between ACT session and OSCE exam date, and whether student participated in other ACT sessions; s.e. = standard error; for continuous variables: adjusted mean difference and for categorical variables, adjusted odds ratio estimate (95% CI); *Denotes statistical significance at p<0.05

---

**Table 4. Comparative results of ACT program sessions on module-specific outcomes**

- **Abdominal Pain**: Total SP score for ACT was 38.3 (1.46) compared to 40.5 (1.55) for Non-ACT. The p-value was 0.31, indicating no significant difference.
- **Geriatric skills Functional**: ADL total (n) for ACT was 7.7 (0.21) compared to 7.5 (0.24) for Non-ACT. The p-value was 0.55, indicating no significant difference.
- **Geriatric skills Psychiatric**: MMSE total (n) for ACT was 27.8 (.05) compared to 24.6 (1.17) for Non-ACT. The p-value was 0.05*, indicating a significant difference.
- **Obstetrical skills and reasoning**: Delivery performance (n) for ACT was 17.6 (0.87) compared to 15.6 (0.52) for Non-ACT. The p-value was 0.03*, indicating a significant difference.
- **Child abuse in a developmental context**: Total SP score (n) for ACT was 43.5 (1.07) compared to 39.6 (1.28) for Non-ACT. The p-value was 0.02*, indicating a significant difference.
- **Pulmonary**: Total management (n) for ACT was 9.1 (0.37) compared to 9.0 (0.42) for Non-ACT. The p-value was 0.84, indicating no significant difference.

---

**Secondary Findings of the ACT Pilot Program**

In addition to the ACT Pilot Program’s primary results, two unexpected secondary findings provided valuable insights that served as the basis for proposed changes in SUNY Stony Brook’s medical school policy and curriculum. Both of these secondary findings had strong implications for teaching clinical diagnostic reasoning skills. Our first unexpected finding was that although our Master Clinical Instructors had prepared to teach our medical students “advanced” clinical reasoning skills, their repeated empirical observations soon revealed that most of the students were unable to adequately demonstrate “basic” clinical reasoning; that is, students were ill prepared to differentiate diagnoses during simulated as well as live patient encounters. Although faculty observations at SUNY Stony Brook were similar to other reports in the literature (Elnicki et al., 2003; Groves et al., 2003; Poncelet and O’Brien, 2008), this finding was met with general surprise across our faculty, irrespective of clinical specialty area.

The clear disconnect between the expectations of Master Clinical Instructors regarding the third and fourth year medical students’ proficiency levels in clinical reasoning and the students’ actual competency levels was another significant finding of the ACT study. The general expectation among the faculty was that students had acquired basic clinical reasoning by the completion of their third year, a presumptive belief given observed student performances. Although our faculty were surprised by the extent of disparity between their perceptions and actual medical student ability, this finding also has been reported by past researchers (Prince et al., 2000; Makoul, 2000; Radcliffe and Lester, 2003; Chumley et al., 2005; Prince et al., 2005; Bowen, 2006; Whipple et al., 2006; O’Brien et al., 2007).

These secondary findings heightened concerns about the lack of sufficient emphasis on teaching introductory clinical reasoning skills earlier in the medical education curriculum. The medical faculty began exploring various factors that could have contributed to inadequate understanding of diagnostic pathways and differential clinical reasoning among third and fourth year medical students. By consensus, it was acknowledged that a lack of structured curricular emphasis on clinical reasoning was likely to be the leading contributor to the observed inadequacies. As described in detail elsewhere.
(Schiavone and Chandran, forthcoming), to address this shortcoming, efforts are presently underway to modify the medical curriculum at SUNY Stony Brook to further enhance clinical reasoning instruction.

CONCLUSIONS
Insightful Lessons Learned
One of the objectives of undergraduate medical curricula is to provide students with the knowledge, skill, and attitudes required for entering specialty training. The goal is that tomorrow’s physicians master the elementary skills that are vitally important for daily clinical work. However, how best to achieve this goal remains unclear. Skills in teaching and communicating have become increasingly recognized as important competencies for medical trainees (Makoul, 2001; Fisher et al., 2007; Accreditation Council for Graduate Medical Education, 2008). Despite this recognition, past studies have shown that there are often insufficient opportunities in medical school curricula to adequately discuss attitudes, strategies, and skills for dealing with many difficult situations that arise during patient encounters (Shapiro et al., 2006; Wu et al., 2006). In fact, many undergraduate medical students feel unprepared to begin clerkships and find the conversion to clinical learning particularly stressful (Radcliffe and Lester, 2003; Seabrook, 2004; Sanders et al., 2004; Chumley et al., 2005; Wu et al., 2006;). In response, academic medical institutions have employed a variety of strategies to improve the physical diagnostic skills and critical clinical reasoning aptitudes of their students (Elnicki et al., 2003; Fagan et al., 2003; Eva et al., 2007). Nevertheless, there is little data on the content of curricula designed to enhance the diagnostic reasoning skills of medical students, challenges and opportunities for strategic program implementation within the confines of natural settings (i.e., academic medical institutions), and the impact of such programs.

Although a demonstration program, the results of the ACT project provide some promise that modest effects on improving the clinical reasoning performance of medical students can be achieved. However, due to several implementation and methodological limitations, significant differences in overall performance scores between ACT participants and non-ACT participants were not attained. These limitations alluded to important lessons with relevance to teaching and learning in academic medicine which merit further discussion.

Programmatic Challenges for Enhancing Clinical Reasoning Skills of Medical Students
In carrying out the ACT Pilot Program, several implementation and methodological challenges were encountered by the authors. The main program implementation challenges were: (a) motivating the medical school faculty to become engaged as Master Clinical Instructors given conflicting priorities; and (b) standardizing evaluator assessments of clinical reasoning skills across modules per clinic site. The salient lessons gained were that faculty were most responsive to becoming motivated to design ACT training modules when modest monetary incentives were offered to them. It is common knowledge that faculty in academic institutions rarely receive incentives or sufficient recognition for their teaching excellence. In fact, it has been shown that teaching and changing curricula are perceived by faculty as requiring more time for education and less time for those areas that contribute to academic recognition and financial rewards (Gruppen et al., 2003). Thus, after an initially poor response rate to faculty solicitation encouraging their involvement in the ACT Program, it became evident that such requests, albeit from the leadership at SUNY Stony Brook’s School of Medicine, were inadequate means to securing faculty engagement. Hence, an important lesson is that program budgets need to account for such expenses to increase the likelihood of success during the implementation phase.
Another key take-away lesson was that system limitations negatively impacted the feasibility of ensuring standardization in skills assessments by evaluators across clinical modules. The ACT sessions were created and implemented by different faculty for the various clerkships creating some inherent variability in student training. The lack of significant programmatic impact between the two study groups could be attributed, in part, to this limitation given performance scores were the key outcome measure of interest. This need for increased standardization of medical education interventions has also been advocated in the literature (Colliver, 2003).

In addition to program implementation challenges, several methodological challenges were encountered including: (a) limitations in identifying appropriate comparison groups, (b) inability to use randomization in the study design given system and policy restrictions, and (c) insufficient statistical power to detect measurable differences. In accordance with policies instituted by the School of Medicine at SUNY Stony Brook, medical students are able to select the location of their clerkships. Therefore, participants in the ACT Pilot Program had self-selected to undergo their clerkships at SUNY Stony Brook for each of the ACT sessions. Given that system restrictions did not permit randomization, the modules were implemented to groups of students throughout the academic year in order to gain enough subjects. Students who received the ACT sessions early on in their third year may have had different baseline skills and knowledge relevant to clinical reasoning from those whose sessions were held later on in their training. Thus, determination of true comparability between ACT Program participants and non-ACT participants with confidence is limited. There also may have been meaningful differences between the two groups in terms of motivations to select SUNY Stony Brook relative to other affiliated clinic sites for clerkship rotations and site-level differences in clerkship trainings. In addition, in the intervention, there were many factors that changed including students, physician trainers, clerkship sites, and training modules. Hence, it is difficult to ascertain which programmatic component may have had the strongest pull to make a difference in enhancing clinical reasoning skills. Moreover, small sample sizes limited power for more robust analytic applications.

In hindsight, it may have been a better approach to have matched ACT participants with a comparable group of non-ACT participants using relevant characteristics. In the absence of randomization, to ensure comparability between the two groups, a single cohort of students could be followed through all ACT modules and compared to a matched cohort of students at non-SUNY clerkship sites. As one strategic approach to increase the study’s sample size (and thus statistical power), multiple intervention sites could have been identified and compared with non-intervention sites with sites serving as the unit of analysis rather than the students themselves. The feasibility of implementing these potential strategies would have to be examined given added layers of complexity in implementation and analyses.

Opportunities for Enhancing Clinical Reasoning Skills of Medical Students
Medical education is structured as a sequence of stages, whereby more advanced levels are built on the foundations laid in the preceding levels (Benbassat et al., 2005; Benbassat and Bauman, 2007). Based on the secondary findings of the ACT Pilot Program, challenges faced by medical students were apparent in their interactions with Master Clinical Instructors in all dimensions of clinical transactions including technical, interpersonal, and interpretive. An understanding of clinical diagnostic pathways appeared to pose difficulties for students in far greater intensity than initially anticipated by the medical faculty. Unexpected disparities in clinical competency perceptions between the ACT Program’s Master Clinical Instructors and actual abilities of medical students provided some empirical
evidence that students struggle in their transition from a more passive clinical learning setting in classrooms to a more active, hands-on approach in clerkships. In addition, teachings about basic clinical reasoning pathways were de-contextualized and not well integrated into the medical curriculum, further exacerbating difficulties faced by students. Deficiencies were prevalent in high-value practical skills such as real-time documentation and subsequent data interpretation. Core clinical skills including interviewing, physical exam, communication skills, medical ethics, and professional responsibility required additional emphasis and repetitive introduction to crystallize new learning among medical students. The most functional opportunity to address this disconnect at SUNY Stony Brook was the formal incorporation of clinical reasoning pathway instruction into the University’s medical education curriculum.

One innovative solution to facilitate documentation, advancement in acquiring clinical exposure to enhance diagnostic differentiation skills, and monitoring student and faculty interactions at SUNY Stony Brook has been the introduction of a standardized assessment tool named Clinical Skills Passport (aka Passport). This tool formally captures the interpretive and procedural skill sets required for solid clinical reasoning among students in their clerkship trainings. The Passport tracks 91 clinical domains that include procedural and technical skills such as venipuncture and bladder catheterization as well as interpretive skills such as analysis of electrocardiograms and blood gases. Students are required to bring their Passports to patient encounters and clinical assessors (e.g., Master Clinical Instructors) are required to complete the tool with students, an approach to teaching and learning in medicine that has shown promising preliminary results in other similar competency-based curricula (Denton et al., 2006; Madigosky et al., 2006).

Another approach to enhancing clinical reasoning skills that surfaced in designing the ACT Pilot Program was the use of multiple instructional modalities. The ACT Program’s instructional modalities included interactive problem-based learning, simulated exercises, standardized patients using actors, and real patient encounters through mini-clinical evaluation exercises. Past studies have examined the strength (S), weaknesses (W), opportunities (O), and threats (T) (i.e., SWOT) for each of these educational options (Charlin et al., 2000; Goldstein et al., 2005; McGovern et al., 2006; Peltier et al., 2007; Palmer and Devitt, 2007; Steadman et al., 2006). The general consensus among medical educators is that a hybrid combination of these instructional modalities is best, with emphasis on improving knowledge, skill, and self-confidence. The experiences of this study’s authors suggest the following stepwise progression for teaching clinical reasoning mastery is advisable: (1) build knowledge base, (2) teach patient interviewing skills, (3) provide hands-on instruction specific to physical examination skills, (4) teach data interpretation and results documentation skills, and (5) build communication and patient counseling skills related to negotiating care plan development.

Practice Implications for Teaching and Learning in Medicine Insights derived from the planning, design, implementation, and results generation phases of the ACT Pilot Program have several practice implications that illuminate the potential for future efforts to impact clinical diagnostic reasoning in academic medicine.

Specifically, the authors pose the following recommendations:

• Findings from the ACT Program concur with past studies that suggest student difficulties in mastery of clinical diagnostic reasoning skills may be more complex than clerkship directors perceive (O’Brien et al., 2007). Thus, medical schools should consider presenting learning experiences that introduce pathways for clinical reasoning and differential diagnosis early in their curriculum. Since learning in
clerkships is an active and experiential process, success depends, in part, on the students’ capacity for reflective practice and accurate self-assessment (Eva, 2005; Eva et al., 2007). Therefore, integration of relevant knowledge domains and skills occurs through opportunities to perform and improve on sets of highly differentiated tasks; deliberate mixed practice and feedback should be strongly encouraged.

• To promote ample opportunities to practice clinical reasoning skills and diagnosis identification among undergraduate medical students, medical schools should consider utilizing a combination of instructional modalities. Instructional options should include a hybrid of multi-media computerized patient simulation (Gordon et al., 2004; Hanna and Fins, 2006; Feldman et al., 2006), standardized patients (Fletcher et al., 2004), case-based small group feedback involving Master Clinical Instructors or Clinical Scholars (Srinivasan et al., 2007), web-based interactive teaching (Kerfoot et al., 2006), and live patient encounters.

• Use of learning communities to enhance medical school education has recently gained favorable review (Rosenbaum et al., 2007). Hence, in addition to learning from medical faculty, peer-to-peer mentoring communities could be established so that support could be gained by more junior students from upper classmen.

• Academic institutions should prioritize developing a cadre of teaching faculty whose main responsibility would be the education of students. Medical schools should also institute initiatives that ensure that students are exposed during their clinical education to members of clinical faculty who are recognized to be outstanding clinicians and clinician educators. Investing in faculty development and academic vitality cannot be successfully achieved without instituting organizational policies and procedures that encourage and reward teaching by faculty (Muller and Irby, 2006; Steinert et al., 2006). Therefore, leadership support needs to be secured to champion such undertaking.

• Establishing specific learning objectives as well as rigorous formative and summative evaluations to ensure that students are acquiring the knowledge, skills, attitudes, and values deemed necessary at their stage of learning is imperative. Students need to perform in a developmentally appropriate manner to integrate complex tasks in their clinical repertoire to ensure the provision of high quality and safe patient care. Without thoughtful measurement that permits the application of analytic rigor, it would be difficult for demonstration programs to gain sufficient traction to be adopted into the mainstream curriculum at medical schools, a challenge faced by investigators undertaking the ACT Pilot Program. Despite the study’s limitations, this article contributes to the void in the clinical reasoning literature on several levels. First, it provides an overview of the ACT Pilot Program’s key components and implementation procedures thereby informing the literature on a curriculum designed to enhance clinical reasoning skills. Second, the ACT Program’s primary as well as unexpected secondary findings are highlighted providing a quantitative overview of programmatic impact as well as more qualitative results from group observations by Master Clinical Instructors. Third, the article describes programmatic and methodological challenges encountered during the ACT Program, with translatable implications for future programs intending to enhance clinical diagnostic reasoning skills of medical students. Lastly, possible solutions related to implementing future programs as well as practical considerations are presented to illuminate future directions for teaching and learning in academic medicine.

ACKNOWLEDGEMENT
The authors gratefully acknowledge by the American Association of Medical Colleges, New York Academy of Medicine (NYAM), Arthur Vining Davis Foundation, and Josiah Macy, Jr. Foundation for generously funding, in part, the ACT Pilot Program under the NYAM designated grant number 0505250001. The authors also express their appreciation to Drs. Ronna Chan and Beronie
Richardson as well as Jessica Smith, all from Advance Health Solutions, LLC for their contributions to manuscript conceptualization, development and editing.

REFERENCES


Fletcher, K., D. Stern, C. White, L. Gruppen, M. Oh, and V. Cimmino. 2004. The physical examination of patients with abdominal pain: The long term effect of adding standardized patients and small group
feedback to a lecture presentation. Teaching and Learning in Medicine 16 (2):171-174.
Muller, J., and D. Irby. 2006. Developing educational leaders: The teaching scholars program at the University of California San Francisco School of Medicine. Academic Medicine 81 (11):959-964.

Please forward all future correspondence to:

Primary Correspondent

Latha Chandran, MD, MPH
Associate Dean for Academic and Faculty Affairs
Vice Chair for Education, Department of Pediatrics
Office of the Dean, School of Medicine
Stony Brook University Hospital and Medical Center
Health Sciences Center, L4-175
Stony Brook, NY 11794
Tel: (631) 638-2005
Fax: (631) 444-7865
E-mail: latha.chandran@stonybrook.edu

Alternate Correspondent:

Maryam Navaie, DrPH
Senior Vice President
Advance Health Solutions, LLC
7660 Fay Avenue, Suite H530
La Jolla, CA 92037
Tel: (858) 456-3592
Tel: (866) 330-0596 (toll free)
Fax: (858) 876-0132
E-mail: mnavaie@advancehealthsolutions.com

VN:R_U [1.9.11_1134]