SPECIAL ISSUE
September 2012
Supplement to Military Medicine, Volume 177, Number 9
Celebrating 40 Years Of Medical Education At The Uniformed Services University Of The Health Sciences: Lessons Learned From The Long-Term Career Outcome Study (LTCOS)

Guest Editors

Shalini Reddy, MD
Associate Professor of Internal Medicine
Section of Hospital Medicine
University of Chicago Pritzker School of Medicine

Dario M. Torre, MD, MPH, PhD
Associate Professor of Medicine
Drexel University College of Medicine

John R. Boulet, PhD
Associate Vice President for Research and Data Resources
Foundation for Advancement of International Medical Education and Research

Cees van der Vleuten, PhD
Chair, Department of Educational Development and Research and
Scientific Director School of Health Professions Education
Maastricht University

Peer Reviewers

Brian V. Reamy, MD
Mark Stephens, MD
Matt Eberly, MD
Thomas Grau, MD
Temple Ratcliffe, MD
Brian Neubauer, MD
Scott DePaul, MD
Gregory J. Argyros, MD
Alyssa Perroy, MD
Charles Magee, MD, MPH

Jeffery La Rochelle, MD, MPH
Jim Balcius, MS, MPH
Galen Barbour, MD
Lynn Byars, MD, MPH
Douglas Maurer, MD
William Watson, MD, PhD
Woodson Jones, MD
Mariya A. Yukhymenko, MA
Paul Hemmer, MD, MPH
Born in Travail, Survival Under a Cloud: The Birth, Improbable Survival and Notable Success of the Uniformed Services University of the Health Sciences (1972–2012)
Val G. Hemming

40 Years of Military Medical Education: An Overview of the Long-Term Career Outcome Study (LTCOS)
Steven J. Durning, Anthony R. Artino Jr., Ting Dong, David F. Cruess, William R. Gilliland, Kent J. DeZee, Aaron Saguil, Donna M. Waechter, John E. McManigle

BEFORE MEDICAL SCHOOL

Alternate List Matriculants: Outcome Data From Those Medical Students Admitted From the Alternate List
Dinchen A. Jardine, Ting Dong, Anthony R. Artino Jr., Steven J. Durning, Donna M. Waechter, Kent J. DeZee, David F. Cruess, William R. Gilliland

Exploring the Relationship Between Self-Reported Research Experience and Performance in Medical School and Internship
Ting Dong, Steven J. Durning, William R. Gilliland, Donna M. Waechter, David F. Cruess, Kent J. DeZee, Margaret Calloway, Anthony R. Artino Jr.

Identifying Themes Within a Medical School Admission Committee’s Reviews of Applicants
William R. Gilliland, Anthony R. Artino Jr., Donna M. Waechter, John E. McManigle, David F. Cruess, Margaret Calloway, Bonnie C. Arze, Steven J. Durning

Relationship Between Admissions Committee Review and Student Performance in Medical School and Internship
William R. Gilliland, Ting Dong, Anthony R. Artino Jr., Donna M. Waechter, David F. Cruess, Kent J. DeZee, John E. McManigle, Steven J. Durning

DURING MEDICAL SCHOOL

Writing, Self-Reflection, and Medical School Performance: The Human Context of Health Care
Mark B. Stephens, Brian V. Reamy, Denise Anderson, Cara Olsen, Paul A. Hemmer, Steven J. Durning, Simon Auster

Development and Initial Validation of a Survey to Assess Students’ Self-Efficacy in Medical School
Anthony R. Artino Jr., Ting Dong, Kent J. DeZee, William R. Gilliland, Donna M. Waechter, David F. Cruess, Steven J. Durning

Instructional Authenticity and Clinical Reasoning in Undergraduate Medical Education: A 2-Year, Prospective, Randomized Trial
Steven J. Durning, Ting Dong, Anthony R. Artino Jr., Jeffery LaRochelle, Louis N. Pangaro, Cees van der Vleuten, Lambert Schwirch

Relationship Between OSCE Scores and Other Typical Medical School Performance Indicators: A 5-Year Cohort Study
Ting Dong, Aaron Saguil, Anthony R. Artino Jr., William R. Gilliland, Donna M. Waechter, Joseph Lopreaiuto, Amy Planagan, Steven J. Durning

The Association Between Specialty Match and Third-Year Clerkship Performance
Aaron Saguil, Erin K. Balog, Matthew N. Goldberg, Ting Dong, Anthony R. Artino Jr., Christopher M. Zahn, Jessica T. Survey, E. Matthew Ritter, David R. Wellin, Laura B. Ramsay, Gerald Ming, Steven J. Durning

Guest Editorial: Reflections on LTCOS Projects During Medical School
Dario M. Torre

AFTER MEDICAL SCHOOL

Postinterview Communication Between Military Residency Applicants and Training Programs

Leadership Success and the Uniformed Services University: Perspectives of Flag Officer Alumni
Ting Dong, Steven J. Durning, William R. Gilliland, Kent J. DeZee, Donna M. Waechter, John E. McManigle, David F. Cruess, Sharon K. Willis, Anthony R. Artino Jr.

Where are They Now? USU School of Medicine Graduates After Their Military Obligation is Complete
Kent J. DeZee, Steven J. Durning, Ting Dong, Anthony R. Artino Jr., William R. Gilliland, Donna M. Waechter, John E. McManigle, Aaron Saguil, David F. Cruess, John R. Boulet

Using Functional Neuroimaging Combined With a Think-Aloud Protocol to Explore Clinical Reasoning Expertise in Internal Medicine
Steven J. Durning, John Graner, Anthony R. Artino Jr., Louis N. Pangaro, Thomas Beckman, Eric Holmboe, Terrance Oakes, Michael Roy, Gerard Riedly, Vincent Capaldi, Robert Walter, Cees van der Vleuten, Lambert Schwirch

Guest Editorial: Secrets of Success in Medical Training
Cees van der Vleuten

The Long-Term Career Outcome Study (LTCOS): What Have We Learned From 40 Years of Military Medical Education and Where Should We Go?
Steven J. Durning, Anthony R. Artino Jr., Ting Dong, David F. Cruess, William R. Gilliland, Kent J. DeZee, Aaron Saguil, Donna M. Waechter, John E. McManigle
But for the will of one persistent legislator, the Uniformed Services University of the Health Sciences (USU) and its F. Edward Hébert School of Medicine would only be a footnote in the history of the U.S. Department of Defense. Congressman F. Edward Hébert (D-Louisiana) began championing the concept of a national health university near the end of World War II. Hébert found few sympathetic legislators and government leaders willing to support his legislative efforts to establish the proposed university. Besides a lack of Congressional enthusiasm, other formidable barriers to his proposal included initial opposition by the American Medical Association (AMA) and the Association of American Medical Colleges (AAMC). He persisted. Between 1961 and 1971, he repeatedly submitted bills to the House to fund a federal military medical school. In 1971, Hébert became the Chairman of the House Armed Services Committee. In that year the Committee was considering legislation to create a Department of Defense Health Professions Scholarship Program (HPSP) aimed at providing a reliable source of young physicians for military service. Hébert combined his military medical school proposal with the HPSP legislation. With some arm twisting by Congressman Hébert and Secretary of Defense Melvin Laird, the combined bill passed both the House and the Senate and was signed into law by President Richard Nixon on September 21, 1972 (Public Law 92-496).

Despite passage of the law founding the school, opposition to the federal medical school continued both in the White House and Congress. However, with pressure from Hébert the funds were released permitting the selection of a Presidentially appointed and Senate-approved Board of Regents (BOR) and the initiation of a process to plan, staff and build the new medical school. The BOR, led by Chairman Dr. David Packard, first met in May 1973. By 1974, Dr. Anthony Curreri was selected as President of the new USU and Dr. Jay P. Sanford as Dean of the new medical school. The BOR, in October 1973, selected the National Naval Medical Center grounds as the site for the school’s campus. Ground was broken for the first building in 1975 and the beautiful complex of buildings was largely completed by 1980. The medical school was named for Hébert after his death. The Charter Class of 1980, comprised of 32 students, matriculated in 1976 (28 men and 4 women). Dr. Sanford was appointed University President in 1981 in addition to retaining his duties as Dean. During these early years, compelling issues were resolved regarding the content and structure of the medical school’s curriculum, receipt of accreditation by the Liaison Committee for Medical Education (LCME), the development and accreditation of the required graduate studies programs and the relationship of the medical school to the uniformed services (Army, Navy, Air Force, and the Public Health Service).

Political opposition to the university persisted. Newly elected President Jimmy Carter sought closure in 1977. Throughout the 1980s, additional serious congressional and administrative proposals for university termination were studied and debated. Critics, including the Office of Management and Budget, regularly charged that the cost of training physicians in a government owned medical school was prohibitively expensive. One threatening study by the Office of Personnel Management led to the resignation and retirement of President and Dean Jay P. Sanford in 1990. Throughout Dr. Sanford’s tenure as Dean and President, excellent teaching and research faculty were recruited, curricula were developed and implemented, quality students were recruited, trained, and graduated, and the medical school and graduate schools took their place among LCME-accredited and Middle States Association for Higher Education-accredited American educational institutions.

The Persian Gulf War (August 1990–February 1991) was the first serious military test of USU-trained physicians. Many graduates were deployed to the Gulf. Their outstanding performance revealed to their military commanders the relevance and quality of the USU medical graduate’s military and joint service medical training. Despite the evidence of the utility of military-trained physicians in wartime, serious efforts were continued to close the school. After the 1992 election, the President-Elect organized the National Partnership for Reinventing Government. One item on the lengthy list of actions, proposed by the Partnership, was closure of the University. Between 1993 and 1998, the Clinton Administration proposed, in the annual Department of Defense budget, to decrement the USU budget by one-quarter and recommended the school not to matriculate a freshman class the next year. However, the University could not be closed by presidential directive as it was established by public law. Each year, from 1993 through 1999, a supportive Congress added funds to the Defense Budget to sustain the University. During these years medical organizations including the AMA, the AAMC and
multiple specialty societies actively supported survival of the University and its important role in providing specially trained military physicians for the Department of Defense. These were challenging times for the new University President, Dr. James Zimble, and the medical school’s administrators, staff, and students. Throughout the 1990s, it was uncertain from year to year whether the school could survive. During these trying years the University began, both retrospectively and prospectively, to collect and collate data on the performance and careers of the school’s graduates and to use the objective data to demonstrate the substantial contributions made by the university’s graduates and faculty to military medicine as military officers, medical practitioners, researchers, teachers, careerists, and leaders. The Long-Term Career Outcome Study (LTCOS) is an important continuation of that effort.

Throughout the University’s 40 years of existence, exemplary faculty-directed biomedical research and teaching programs have developed and prospered which have advanced many scientific disciplines critical to military medicine and important to national and international health such as trauma care, wound healing, biomedical imaging, infectious diseases, microbiology and immunology, genetics, cancer research, human immunodeficiency virus, infection acquired immunodeficiency syndrome, human behavior, disaster response, and sports medicine. In 1993, the Graduate School of Nursing was added to USU and matriculated its first 10 specialty extended care nursing students. Noteworthy educational successes for USU have been major contributions to physician and nurse education including innovative methods and tools for teaching, student evaluation, and the use of standardized patient simulations. A substantial group of University faculty members have won national honors for contributions in medical student, physician, and nurse education. Several faculty members have been recruited from the University to assume direction of a number of prestigious national medical organizations including the National Board of Medical Examiners, the AMA, and the Educational Commission for Foreign Medical Graduates, the AAMC, and several specialty societies. Several of the medical school’s clinical departments and their faculties have for many years been in the forefront of American medical education research as exemplified by the LTCOS project and its team members. Congressman Hébert and the American people have been well rewarded for their investment in the USU experiment!

Three recent unfortunate events affirm Congressman Hébert’s wisdom in promoting the founding of a military health university and medical school for the United States. The World Trade Center attack on September 11, 2001 followed by the Allied Forces invasion of Afghanistan on October 7, 2001 began a strenuous test for the medical services of the armed forces. The challenges of providing medical care to deployed troops at war were expanded when the Allies invaded Iraq on March 26, 2003. Through the more than 10 years of continuous combat by U.S. and allied forces, thousands of military physicians have been deployed to battle zones. Career military medical officers trained at USU have provided enlightened leadership, stability and state of the art medical knowledge and skills to serve the hundreds of thousands of soldiers, sailors, and airmen placed in harm’s way.

I have been associated with the University’s effort to provide, not just physicians but military physician officers, for the U.S. military for the past 32 years. Through the herculean efforts of hundreds of civilian and military faculty and staff members and the leadership of Dr. Packard, Dr. Curreri, Dr. Sanford, Dr. Zimble, Dr. Rice, and many others, the University has more than filled its promise.

This supplement provides further evidence of the wisdom of the school’s founding, its serious contributions to American medical education and its promise for the future of military medicine.
ABSTRACT In 2005, the Long-Term Career Outcome Study (LTCOS) was established by the Dean, F. Edward Hébert School of Medicine, Uniformed Services University of the Health Sciences (USU). The original charge to the LTCOS team was to establish an electronic database of current and past students at USU. Since its inception, however, the LTCOS team has broadened its mission and started collecting and analyzing data on a continuous basis for the purposes of program evaluation and, in some cases, research. The purpose of this commentary is to review the history of the LTCOS, including details about USU, a brief review of prior LTCOS work, and progress made since our last essay on LTCOS efforts. This commentary also provides an introduction to the special issue, which is arranged as a series of articles that span the medical education continuum (i.e., before, during, and after medical school). The relative balance of articles in each phase of training represents the LTCOS team’s efforts to address the entire continuum of medical education.

INTRODUCTION
The Long-Term Career Outcome Study (LTCOS) was established by the Dean, F. Edward Hébert School of Medicine (SOM), Uniformed Services University of the Health Sciences (USU), in 2005. We previously reported on the progress of this initiative in Military Medicine. In the current article, we review the history of the LTCOS, including details about USU, a brief review of prior LTCOS work, progress made since our last essay publication on LTCOS efforts and finally an overview of the articles in this special issue.

BRIEF HISTORY OF LTCOS
The original charge to the LTCOS team was to establish an electronic database of current and past students at USU. The team was initially comprised of a small cadre of faculty from the Department of Medicine and Preventive Medicine and Biometrics and members of the Dean’s staff. The first LTCOS research activity entailed the development of an alumni survey in 2005 to 2006 to assess outcomes of USU graduates in terms of leadership positions, operational accomplishments, awards, and academic landmarks. After the success of this alumni survey, which yielded important findings for USU, the Dean of SOM provided essential direction and support to extend the work of the LTCOS group beyond this one-time survey. In particular, the LTCOS group broadened its vision and began collecting and analyzing data on an ongoing basis for the purposes of program evaluation and, in some cases, research. This included describing a wider scope of long-term outcomes, as well as considering “before” and “during” medical school outcomes. Borrowing from the program evaluation and quality assurance literatures, our projects have followed this thread: before medical school (i.e., admissions), during medical school (course, clerkship, or entire medical school curriculum), and after medical school investigations. It is important to note that these program evaluation and research efforts could not have been accomplished without the Dean’s support through the Henry M. Jackson endowment fund.

Over the years, the LTCOS team has grown into a group of core individuals (the authors of this article). The LTCOS team members span several departments, including Medicine, Preventive Medicine and Biometrics, and Family Medicine. LTCOS members have roles and responsibilities which are quite diverse, including deans, a vice chair, program directors, a course director, teaching faculty, and a research associate. We are involved in preclinical, clinical, and graduate medical education, as well as in graduate programs in fields other than medicine. Further, we have a broad spectrum of expertise, spanning medicine, educational research, and statistics. Only one individual (our research associate; T. Dong) is assigned full time to LTCOS work. The remaining members of the LTCOS team perform a variety of other jobs for the university; the common “thread” of the team is an interest in improving how we select (before), educate (during), and follow (after) our graduates. Ultimately, the goal is to improve what we do at the medical school level.

OUR MEDICAL SCHOOL
USU’s F. Edward Hébert SOM is the nation’s only federal medical school. USU’s MD graduates matriculate as officers in the uniformed services and receive pay and benefits commensurate with O-1 rank while enrolled in medical school. Students have a 10-year service obligation of which at least 7 must be on active duty. We encourage the interested reader to review the article in this special issue on the history of the establishment of USU.

We currently matriculate roughly 170 students per year. The number of students per service is consistent with the size...
of the respective service, with Army students comprising over one-third of a class, followed by Navy, Air Force, and Public Health Service medical students.

During the time of the LTCOS studies described in this special issue, USU had a traditional curriculum consisting of 2 years of preclinical education followed by 2 years of clinical education (i.e., clerkship rotations). All USU students complete their preclinical training in Bethesda, Maryland, at our local campus. USU students then go on to complete clerkship rotations in military teaching hospitals throughout the United States. These include major military hospitals in California, Hawaii, Louisiana, Maryland, Nebraska, Ohio, Texas, Virginia, and Washington.

PRIOR LTCOS STUDIES
An important purpose of this special issue is to provide a single source of LTCOS work published to date. Accordingly, we will briefly review several studies previously published in Military Medicine2 and elsewhere. We encourage readers to review the details of each study. Nonetheless, we hope this review of our prior work will “set the stage” for the follow-on, original research studies presented in this issue.

We believe this prior work, which has included a number of collaborations, helped to launch LTCOS into the international arena in medical education. We will first briefly describe articles reported in a prior LTCOS “progress report” essay in 2010, entitled “LTCOS: Where We’ve Been and Where We Hope to Go.”2 We then discuss articles published (or in press) subsequent to this report, but before this special issue. For the former, we will begin by discussing single-institution studies and will end with LTCOS multi-institutional work.

For the first single-institution (USU) study, we collected reliability and validity evidence for a program director’s evaluation form completed by program directors at the end of the first postgraduate year (internship; PGY-1) of our graduates.4 In this work, we found high feasibility (80% response rate overall), good internal consistency reliability (measured by Cronbach’s α), and several lines of validity evidence to include expected correlations with grade point average and U.S. Medical Licensing Examination (USMLE) scores. Finally, an exploratory factor analysis suggested the survey data could be explained by two factors: expertise (knowledge) and professionalism.4 This evaluation form serves as a main outcome measure for our medical school for both program evaluation purposes, such as accreditation and research studies. Qualitative data from this program director’s evaluation form also yielded important information. In particular, the qualitative comments were found to be a more sensitive marker of poor performance than the numeric scores on this form.5

After we assessed the reliability and validity of USU’s program director evaluation form, we focused on identifying predictors of below-average performance, as indicated by low ratings on this form. Appearance before the student promotions committee (SPC; a committee that reviews the records of students experiencing academic difficulty, for any reason, and provides recommendations to the Dean’s office on how to assist the student) during medical school was associated with below-average performance during internship; however, SPC appearance was not a specific indicator of such performance.6 In other words, the majority of students who eventually had one or more below-average ratings on the program director form during internship appeared before the SPC; however, 75% of students who appeared before the SPC did not receive any below-average ratings during internship.6

We then explored markers of poor professionalism and knowledge during internship (the two factors measured by our program director evaluation form).7 We found that lower third-year clerkship grades were associated with low program director ratings for both of these factors, and, furthermore, that low scores on the USMLE predicted low knowledge only.7 Next, we investigated the prevalence, causes, and consequences of having a major life crisis at USU.8 Students who reported having a major life crisis during medical school had lower grade point average, USMLE Step 1 and 2 scores, and lower program director ratings than students who did not report such a crisis during medical school.8 These latter two articles7,8 were completed by a physician who enrolled in USU’s Master of Public Health program. It is worth noting that this special issue also contains articles coauthored by a medical student and a resident.

The multi-institutional LTCOS studies (i.e., those which involved multiple medical schools) conducted in the early years of the project included explorations into student career choice.9 We also participated in a commentary on longitudinal databases in medical education research.10

A number of studies have been conducted since the 2010 Military Medicine review. These studies largely followed the reported plans for next steps outlined in the 2010 essay. From the before medical school phase, we explored self-reported clinical experience upon matriculation to USU;11 Our findings suggest that applicant’s self-reported clinical experience before medical school is a poor predictor of student performance in medical school and internship.11 We also found that application essays were not associated with future performance in medical school.12

Our during medical school LTCOS-related studies have explored emotion and motivation and their association with performance in a preclinical course.13 Members of the LTCOS team have also conducted a series of studies that explored instructional authenticity in a preclinical course and its effect on preclinical and clinical performance.14,15 One additional study in this 2-year, prospective, randomized experiment is described in this issue.16

In multi-institutional LTCOS-related studies, we have observed a number of interesting findings, including the prevalence of burnout and suicidal ideation,17 as well as burnout and serious thoughts of dropping out of medical school18 across multiple medical schools, including USU. We also found that burnout was associated with poor professional conduct.19 Finally, we found that pass/fail grading
was more favorable to medical students’ well-being than assigning letter grades.\textsuperscript{20}

Our after or postmedical school studies have explored the association of interest group participation during medical school and choice of residency.\textsuperscript{21} This work has also looked at how context impacts physician decision making, using both quantitative and qualitative research methods.\textsuperscript{22,23} From this work, we are learning that cognitive load theory, which refers to limits in working memory, seems to play an important role in how context impacts physician performance. A number of teaching and practice implications are outlined in this work.\textsuperscript{22,23} Multi-institutional work in the after phase has included studies that compare those who considered internal medicine (IM) as a specialty during medical school but ultimately chose another residency with those who selected IM for residency and those who never considered IM as a residency choice.\textsuperscript{24} views of medical students about primary career fields over time,\textsuperscript{25} and the problems that program directors inherit in terms of medical student well-being.\textsuperscript{26}

In this latter work, we found that a high percentage of fourth-year medical students reported distress before even entering residency (burnout = 49\%, depressive symptoms = 38\%, and low mental quality of life = 34\%).\textsuperscript{26} Further, we have also described communications between residency program personnel and applicants (potential residents) in the civilian match\textsuperscript{27} that occur after being interviewed by a program but before being selected for residency training (i.e., the match) so that comparisons could be made with our military system, which is an article described in this issue.\textsuperscript{28} In the article by Jena,\textsuperscript{27} we suggest ways to improve these communications.

Notably, the articles described above have been published in a variety of top-tier medical education journals, including articles in the \textit{Journal of the American Medical Association}, \textit{Annals of Internal Medicine}, \textit{Academic Medicine}, \textit{Archives of Internal Medicine}, \textit{Medical Education}, and \textit{Advances in Health Sciences Education}. Readers are encouraged to review these articles for a more in-depth understanding of the breadth and depth of the LTCOS team’s work.

\section{OVERVIEW OF THE SPECIAL ISSUE}

This \textit{Military Medicine} special issue represents an attempt by the LTCOS team, under the vision and support of our Dean, to assess USU’s medical school program from a variety of research “lenses.” There is also a quality assurance component to this work that is not specifically described in this special issue.

LTCOS research has spanned medical student selection and education, and it has included both short- and long-term outcomes. Indeed, we have arranged the special issue into these three previously mentioned phases in medical education: before medical school, during medical school, and after medical school. The relative balance of articles in each “phase” represents our efforts to address the entire continuum of medical education. The research articles presented in this issue followed the typical peer-review process. Specifically, articles were blinded and then sent to at least two peer reviewers. These reviewers rendered their opinions regarding acceptance or rejection, and these reviews were subsequently sent to the guest editors. The guest editors then rendered final decisions for each article and provided feedback to the authors on their decisions. Consistent with established standards, reviewers and editors did not render opinions on articles where they were an author. Finally, accepted article were revised before final publication in the journal.

We are very grateful for the work of our reviewers and our guest editors (both groups are listed on the inside cover of this issue). Their dedication, time, and expertise have made this special issue possible.

\section*{REFERENCES}


Alternate List Matriculants: Outcome Data From Those Medical Students Admitted From the Alternate List

LT Dinchen A. Jardine, MC USN*; Ting Dong, PhD†; CDR Anthony R. Artino Jr., MSC USN‡; Steven J. Durning, MD, PhD†; Donna M. Waechter, PhD†; LTC Kent J. DeZee, MC USA†; David F. Cruess, PhD†; COL William R. Gilliland, MC USA (Ret.)†

ABSTRACT
Background: Medical schools are increasing class size to meet future health care needs for our nation. This may lead to more students being accepted from an alternate list (vs. primary acceptances). Given these trends, performance outcomes were compared for alternate list matriculants and primary acceptances. Our hypothesis was that those students accepted from an alternate list would perform equally to the primary acceptances on these outcomes.

Method: We compared medical school performance of students who received a primary recommendation of “accept” and compared them to those who received a recommendation of “alternate” over a 10-year period. Given the small sample size of this alternate list group (N = 23), descriptive statistics are reported. Results: No consistent differences between alternate and primary acceptance matriculants in terms of cumulative medical school grade point average, United States Medical Licensing Examination (USMLE) Step 1 scores and USMLE Step 2 Clinical Knowledge scores were found. Only three alternates (13.0%) were presented to student promotion committee compared to 17.2% for matriculants who were primary acceptances. Three alternates were required to repeat a year (average percentage of 8.7%) compared to 5.6% of matriculants who were primary acceptances.

Conclusions: This observational study provides some reassurance that as long as the qualifications of the applicant pool remain adequate, admissions policies that provide for alternate list acceptances may not produce poorer performing students, at least by our current outcome measures.

INTRODUCTION
Recently, the Association of American Medical Colleges suggested that the United States must urgently increase medical school and graduate medical education capacity to meet the needs of the nation in 2015 and beyond.1 The predicted shortage of physicians in all subspecialties will be 91,500 in 2020.3 Fortunately, first-time applicants to medical school have increased from 27,190 in 2004 to 32,654 in 2011, reaching an all-time high.2 Over the same period, the number of available positions has increased from 16,648 to 19,2302 in 137 accredited U.S. schools.3

In terms of addressing this physician shortage, the trend of increasing medical school applicants is encouraging. However, many well-qualified students are accepted by more than one medical school, which means that they must also decline acceptances at other schools. Given these declinations, admissions committees must choose more applicants to fill the allotted positions. Many schools, like the Uniformed Services University of the Health Sciences (USU), have alternate lists for students who may then make the “second cut” after the primary acceptance list has been exhausted.

In the present study, six performance outcomes were compared for those students who were primary accepts vs. those who were selected from the alternate list at one U.S. medical school. These outcomes were chosen because they are likely to be measured at other medical schools and data for each outcome were available for all matriculants from alternate lists analyzed in this study. We hypothesized that those students accepted from an alternate list would perform equally to the primary acceptances on these outcomes.

METHODS
Admissions Process
The admissions committee at USU consists of 18 voting members and a Chair who votes only if there is a tie. The process at USU includes three levels of review:

During the primary review, each application is reviewed by admissions office staff to determine if the applicant has met basic required eligibility and academic requirements. Qualified applicants are asked to submit supplementary application materials (progress to secondary review). During the secondary review, admissions committee members independently evaluate the credentials and recommend either an invitation for interview or rejection. Therefore, invitation by the admissions committee to interview is only extended to the most promising students and that decision is largely based on academic credentials and other information contained within the American Medical College Application Service application. Each applicant has two separate interviews that are conducted by uniformed faculty members
(retired or active duty) or members of the senior class. After personal interviews are conducted, each candidate’s folder is reviewed by three committee members who then each provide a tertiary review. The committee reviews all of the available data to include personal statements, undergraduate and/or postgraduate grade point average (GPA), Medical College Admission Test (MCAT) scores, demographic data, letters of recommendation, interview score sheets, and any other information contained in their admissions packet. The full committee then makes one of three recommendations: primary acceptance, alternate, or nonacceptance.

Since declination at USU over the past 5 years has ranged from 36 to 43%, offers of primary acceptances are sent to more than 260 students in order to fill a class of 170. If the class is still not filled with applicants offered acceptance, then applicants from an alternate list are rank ordered and considered for acceptance.

**Study Participants**

School performance of students who received a recommendation of “accept” and those who received a recommendation of “alternate” over a 10-year period (classes of 2001–2010) were compared.

**ANALYSIS**

Since the number of alternates was small, especially compared to the number of primary acceptance matriculants, and they were spread over six of the 10 classes, descriptive rather than inferential statistics are reported and include standard deviations for the matriculants for each variable by matriculation year to help determine differences between alternates and primary acceptance matriculants given these limitations with direct comparison. Six variables were compared—medical school cumulative GPA using a standard 4-point letter grade system, United States Medical Licensing Examination (USMLE) Step 1 score, USMLE Step 2 Clinical Knowledge (CK) score, USMLE Step 2 Clinical Skills (CS), number of students presented to the student promotions committee (SPC), and number of students who had to repeat a year (or more) in medical school. This study was approved by USU’s Institutional Review Board.

**RESULTS**

The distributions of the matriculants from the primary acceptance and alternate lists, which includes standard deviations for matriculants, for all the classes are presented in Table I. There were 23 alternates and 966 regular matriculants in total, and by graduation year as follows: 2001 = 2, 2003 = 4, 2005 = 2, 2008 = 5, 2009 = 1, and 2010 = 9. For the classes of 2002, 2004, 2006, and 2007, there were no matriculants from an alternate list. Therefore, data from the six classes in which students from an alternate list matriculated were analyzed.

The average medical school cumulative GPAs of those from an alternate list were very close (and in some cases...
even higher for the classes of 2005, 2009, and 2010) than their primary acceptance counterparts. Similarly, there was no consistent pattern of differences between alternate and primary acceptance matriculants on the USMLE Step 1 and USMLE Step 2 CK examinations. With only one exception, the differences in medical school cumulative GPA, USMLE Step 1 scores, and USMLE Step 2 CK scores between those matriculants from an alternate list and primary acceptance list in six classes were less than 1 standard deviation in all comparisons. The one exception was comparing the USMLE Step 2 CK score for the Class of 2009 in which the one matriculant from an alternate list scored more than 1 standard deviation higher than matriculants from the primary acceptance list. The pass rate of the USMLE Step 2 CS was almost 100% with our sample (Note: this examination was not required for the class of 2001 and 2003). None of the matriculants from the alternate list failed USMLE Step 2 CS. In comparison, there were two primary acceptances (one in the class of 2005 and one in the class of 2008) who failed the USMLE Step 2 CS.

For the six classes under study, only three alternates (13.0%) in the class of 2008 were presented to the SPC. The average percentage of presentation to the SPC was 17.2% for matriculants who were primary acceptances. There were three alternate students required to repeat a year in medical school: one alternate in the class of 2001 and two in the class of 2008, which gave an average percentage of 8.7% for the alternates. By comparison, the average percentage of matriculants who were primary acceptances required to repeat a year was 5.6%.

**DISCUSSION**

Taken together, the matriculants from an alternate list and the primary acceptance list did not appear to differ in terms of their medical school GPA, USMLE Step 1 score, and USMLE Step 2 CK score. In addition, there were three alternates in the class of 2008 who presented to the SPC and two alternates who had to repeat a year.

This data suggest that our alternate selection process leads to medical student graduates who are similar to primary acceptance matriculants. This data suggest that as long as the applicant pool remains robust, matriculants from an alternate list have similar success in medical school and can serve as an additional source of applicants to fill the expanding class size at some schools. Obviously, it would be useful to be able to predict how many primary acceptances would be needed to fill a medical school class. However, that is at least partly dependent on the declination rate which fluctuates at USU. Alternatively, one could increase the number on the list of primary acceptances to fill the entire class and avoid the need for an alternate list altogether. However this would potentially unnecessarily increase the workload of the admissions committee by increasing the time spent reviewing additional applicants.

Hypothetically, matriculants chosen from our alternate list may represent a group that by traditional standards would be an “at risk” academic population since for some reason they were felt not to meet the criteria for primary acceptance. In fact at Nottingham University Medical School, acceptances late in the application process were identified as one predictor of students who they identified as “strugglers,” students who had experienced academic or personal difficulties during the course of their medical school studies. However, the findings of our study seem to refute that.

Our results were similar to previous studies in that there were few differences in our outcome measures between the two groups. Two prior studies attempted to quantify the outcome of students accepted from an alternate list but these studies are limited by sample size in using only a single year’s data. Both studies found little significant difference in the matriculant rate of graduation or licensure board scores. Although the number of matriculants chosen from our alternate list was small, it studied students from 6-year groups that spanned a 10-year period of time. Ours is the first study that examined medical school performance of these students from several year groups.

More recently, Paolo et al studied a larger cohort of matriculants entering from 1997–2003 at the University of Kansas School of Medicine, comparing the main list applicants to those admitted from an alternate list. Similar to our findings, they concluded that outcome measures as defined by standard measures academic performance, graduation rate, delay in graduation were not different between groups; they also concluded that residency match information and residency director ratings were not significantly different between the two groups. This is encouraging as our alternate list processes are different than those of Paolo and, in sum with our and others’ data, suggest that alternate list candidates perform similarly to regular matriculants on traditional performance measures. Given these findings, in total, perhaps medical schools should either place less weight on level of performance on standard application measurements or include other performance measurements at the time of application, such as multiple mini interviews that may provide additional discriminating data to help select the candidates most likely to succeed at a given medical school.

There are several limitations of this study. First, this study was conducted at a single medical school. As a result, this study may not be generalizable to other schools with different admissions processes and who may utilize the “alternate” list in a different way than USU. Second, although the uneven distribution of matriculants from an alternate list corroborated our decision to use descriptive statistics only, the small sample size in each class did not allow reliable estimates of test statistics. Third, the long-term outcome data for this group are incomplete as a significant proportion of students accepted from an alternate list have only recently graduated. Further study will be required to determine if the “catch up” that seems to occur in medical schools holds throughout
residency and into their postgraduate military medical careers. Fourth, this study did not seek to determine how matriculants are chosen from applicants on an alternate list. For example when applicants from an alternate list were studied at the University of Kentucky School of Medicine, some matriculants from their “alternate list” were selected based on traditional determinants of performance such as the Medical College Admission Test scores whereas others were selected based on demographic variable to ensure a more balanced class.9 Although we acknowledge that the ability to draw conclusions from purely observational studies may be limited, our plan is to continue to study the outcomes of those students who are admitted from “alternate” means, which will eventually contribute to our ongoing efforts to determine what, if any, key indicators of future success can be determined from the medical student application including clinical experience,10 program director evaluation forms,11 and application essays.12

Given the cost in both time and resources invested in every potential medical student, identifying those with the greatest likelihood for success is of highest priority. The means that medical schools use to identify students for admission varies widely both within the United States and internationally. Results from this and other studies cited provide evidence that admission policies that provide for alternate list acceptances do not produce poorer performing students, at least by traditional outcomes.

REFERENCES

ABSTRACT
Purpose: To investigate the relationship between self-reported research experience and medical students’ performance in medical school and internship. Methods: We collected data from seven year-groups (1993–1999; N = 1,112) and examined 7 performance outcomes: medical school preclinical grade point average (GPA), medical school clinical GPA, cumulative medical school GPA, U.S. Medical Licensing Examination Step 1 and 2 scores, and scores on a previously validated program director’s survey of intern professionalism and expertise. We then conducted a series of multiple linear regressions to determine the relations between self-reported research experience and our seven outcomes. Results: When compared to those who reported no prior research experience, students who reported research experience performed significantly better on U.S. Medical Licensing Examination Step 1 and had a higher medical school preclinical GPA. However, these same students scored significantly lower on intern professionalism and expertise ratings. Self-reported research experience did not show statistically significant correlations with the other outcome variables. Conclusions: The results from our large, multiyear, cohort study suggest that prior research experience may account for some variance in outcomes in the early stages of medical school education, but that variance explained diminishes considerably as trainees progress into the more senior phases of education. On the other hand, prior research experience may be negatively related to students’ performance in internship. In all cases, however, effect sizes are small.

INTRODUCTION
Medical school admission committees strive to select candidates who are most likely to succeed. Prior research has investigated the impact of both cognitive and noncognitive factors on achievement in medical education.1–3 For example, in a meta-analysis utilizing data from hundreds of studies in this area of inquiry, the authors found that previous academic achievement, such as medical college admission test and grade point average (GPA), accounted for 23% of the variance in performance in undergraduate medical education and 6% of the performance in postgraduate education.3 Although the focus (and findings) of the studies conducted to date have varied, there is general agreement among medical educators that medical schools should broaden the criteria they use to select future doctors.

Self-reported research experience is commonly used by U.S. medical schools as a selection criterion. Many admission committees assume that research experience can reinforce and extend the knowledge gained in coursework. Moreover, by understanding the process of posing questions and investigating those questions systematically, some believe student researchers begin to develop important skills that will benefit them as future physicians. However, a review of the literature revealed no published research to support the positive relation between applicants’ prior research experience and their medical school achievement and internship performance.

In the current study, we explored the association between self-reported research experience and medical trainees’ performance during medical school and internship. We predicted that self-reported research experience would be positively related to higher performance in medical school and internship. As self-reported research experience is commonly used for medical student admission processes, we believe our findings could inform standard admission committee practices in the United States and abroad.

METHODS
Study Context and Participants
The present study was part of the Long-Term Career Outcome Study (LTCOS) conducted at the F. Edward Hebert School of Medicine, Uniformed Services University of the Health Sciences (USU). The LTCOS was created under the direction of the medical school dean, and its purpose is to collect data on past, present, and future students. The overarching goal of the LTCOS is to track student performance through medical school and on into residency education and practice, with an emphasis on evaluating the effectiveness of USU’s admission selection process and curriculum initiatives.

As the United States’ only federal medical school, USU matriculates approximately 170 medical students annually. At the time of this study, USU offered a traditional 4-year curriculum: 2 years of basic science courses followed by 2 years of clinical rotations. Upon matriculation, students are commissioned as officers in the Uniformed Services (Army, Navy, Air Force, and Public Health Service) and after graduation, the majority of graduates train in military-affiliated programs across the country. For the purposes of the current study, we drew our sample from graduates who received their
MD degree from USU between 1993 and 1999 (N = 1,112). This was a purposeful sample as we wanted to track performance through internship. A previous study used the same sample to examine the reliability and validity of a program director’s survey to evaluate medical school graduates’ internship performance.\(^4\) Furthermore, although the data were derived from students who graduated in the 1990s, it is important to note that the demographics of our matriculants and the application process at USU are still largely unchanged, emphasizing many of the same admission procedures, including consideration for self-reported research experience.

**Study Variables and Procedures**

We used two primary sources of data. The first dataset was obtained from the medical school’s admissions office and the university’s registrar; they provided demographic (age, gender) and academic information. The admissions office provided the following information: cumulative undergraduate GPA and open-ended responses from the “work and activities” section of the American Medical College Application Service (AMCAS) application. The registrar provided the following information: average medical school GPA at the end of the first 2 years of study (preclinical GPA), average medical school GPA of the second 2 years of study (clinical GPA), cumulative medical school GPA upon graduation from USU, and U.S. Medical Licensing Examination (USMLE) Step 1 and 2 scores.

The second dataset was obtained from a program director’s survey of trainee performance completed at the end of the intern year (postgraduate year 1, PGY-1). These study variables are described in greater detail below.

**Explanatory Variables**

Self-Reported Research Experience

The AMCAS application, which is currently used by 125 out of 134 accredited medical schools represented by the Association of American Medical Colleges in the United States, includes a “work and activities” section.\(^5\) In this section, students are asked to list any work or extracurricular activities, awards, honors, or publications that they would like to bring to the attention of the medical schools to which they are applying. We focused on whether students reported that they had any research experience (e.g., working on basic science or clinical research, research-related presentations or publications). A dichotomous variable was created with “yes” coded as 1 and “no” coded as 0.

Undergraduate GPA and Gender

We controlled for undergraduate GPA and gender, as GPA and gender have been reported to influence trainee performance.\(^6\)–\(^8\) Each medical student’s undergraduate GPA was converted to a common 4-point scale (range: 0.0–4.0). It included all undergraduate course grades, not including post-baccalaureate or other graduate school grades. For gender, a dichotomous code was created with female students coded as 1 and male students coded as 0.

**Outcome Measures**

**Average Medical School Preclinical GPA**

For each medical student, a medical school GPA was calculated using course grades from the first 2 years. The GPA is a weighted average created by multiplying each course grade by the number of contact hours for the given course, summing the weighted grades across courses, and then dividing the sum by the total number of contact hours. The resulting averages were converted to a common 4-point scale (range: 0.0–4.0).

**Cumulative Medical School GPA**

This outcome measure refers to the GPA calculated based on the entire 4 years of medical school study.

**USMLE Step 1 and 2 Scores**

Students in this sample completed the Step 1 examination after their first 2 years of medical school. The Step 1 examination focuses on understanding the basic sciences relevant to medical education. Students completed the Step 2 examination during their fourth year of medical school. This examination emphasizes clinical knowledge. Scores on the USMLE Step examinations are three digits and range from 140 to 280, with an average of between 200 and 220 and a standard deviation of 20. In the present study, we used students’ scores on their first attempt at each of the step examinations.

Program Director’s Rating Form

The program director’s rating form was completed at the end of the intern year. Items on this form were developed by interdepartmental educators with expertise spanning undergraduate and graduate medical education programs. The survey consisted of 18 items scored on a 5-point Likert scale, where 5 = outstanding, 4 = superior, 3 = average, 2 = needs improvement, 1 = not satisfactory, and 0 = unable to judge (note: scores of 0 were treated as missing data). The feasibility (annual response rate ranged from 72 to 90%), reliability (Cronbach’s \(\alpha\) ranged from 0.93 to 0.96), and validity of this instrument have been reported in a previous study using the same cohort of students.\(^9\) In this previous work, exploratory factor analysis was performed to test the construct validity of the instrument. The results revealed that the 18 survey items loaded onto two constructs, expertise (13 items) and professionalism (5 items). We used these two constructs (PGY-1 Professionalism and PGY-1 Expertise) as separate outcomes in the present study.
**Statistical Analysis**

We conducted multiple linear regression analyses to examine the relations between the explanatory variables and the outcome measures. For each outcome, we started with the saturated model, which included the main effects of the three explanatory variables (self-reported research experience, undergraduate GPA, and gender), the two-way interactions between the variables (self-reported research experience by undergraduate GPA, undergraduate GPA by gender, and self-reported research experience by gender), and the three-way interaction among the variables (self-reported research experience by undergraduate GPA by gender). If the interaction terms turned out to be nonsignificant ($p > 0.05$), they were dropped from the regression model. However, the main effects of the explanatory variables were always kept in the final model because their influences were of particular interest in the present study. All the analyses were conducted using PASW Statistics 18.0 (SPSS, Chicago, Illinois). This study was approved by the USU Institutional Review Board.

**RESULTS**

**Descriptive Statistics and Pearson Correlations**

Of the 1,112 medical students who graduated from USU between 1993 and 1999, we collected 943 complete records (85%) with all variables of interest included. This was the sample used for analysis. The sample included 717 men (76%) and 226 women, and the average age was 24.7 (SD = 1.6). Descriptive statistics and Pearson correlations for the explanatory variables and outcome variables are shown in Table I.

As shown in Table I, there were several statistically significant correlations between the variables. In particular, students’ self-reported research experience was positively correlated with their USMLE Step 1 scores ($r = 0.11$, $p < 0.01$), but negatively correlated with program director’s ratings on professionalism ($r = -0.10$, $p < 0.01$) and expertise ($r = -0.07$, $p < 0.05$). Students’ undergraduate GPA showed positive correlations with their medical school preclinical GPA ($r = 0.17$, $p < 0.01$), cumulative medical school GPA upon graduation ($r = 0.14$, $p < 0.01$), USMLE Step 1 scores ($r = 0.14$, $p < 0.01$), and USMLE Step 2 scores ($r = 0.12$, $p < 0.01$). Gender was also significantly correlated with medical school preclinical GPA ($r = -0.11$, $p < 0.01$) and USMLE Step 1 scores ($r = -0.18$, $p < 0.01$).

As for the correlations among the outcome variables, medical school preclinical GPA was moderately correlated with medical school clinical GPA ($r = 0.51$, $p < 0.01$), professionalism ($r = 0.15$, $p < 0.01$) and expertise ($r = 0.26$, $p < 0.01$), and highly correlated with USMLE Step 1 ($r = 0.76$, $p < 0.01$) and Step 2 ($r = 0.61$, $p < 0.01$) scores. Medical school clinical GPA was moderately correlated with USMLE Step 1 ($r = 0.46$, $p < 0.01$) and Step 2 ($r = 0.49$, $p < 0.01$) scores, professionalism ($r = 0.33$, $p < 0.01$) and expertise ($r = 0.43$, $p < 0.01$), Cumulative medical school GPA was moderately correlated with expertise ($r = 0.37$, $p < 0.01$) and professionalism ($r = 0.24$, $p < 0.01$) and was highly correlated with USMLE Step 1 ($r = 0.73$, $p < 0.01$) and Step 2 ($r = 0.65$, $p < 0.01$) scores. Moreover, expertise and professionalism were highly correlated with one another ($r = 0.77$, $p < 0.01$), and expertise scores were moderately correlated with USMLE Step 1 ($r = 0.24$, $p < 0.01$) and Step 2 ($r = 0.24$, $p < 0.01$) scores. On the other hand, the correlations between professionalism and USMLE Step 1 ($r = 0.10$, $p < 0.01$) and Step 2 ($r = 0.11$, $p < 0.01$) scores were small. Finally, USMLE Step 1 and 2 scores were highly correlated with one another ($r = 0.76$, $p < 0.01$).

**Multiple Linear Regression Modeling**

We built multiple linear regression models for our outcome measures. For all of the outcomes, none of the interaction terms (the three two-way interactions and the one three-way interaction) of the explanatory variables had a significant effect. Thus, only the main effects of the explanatory variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Means</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Self-Reported Research Experience</td>
<td>—</td>
<td>—</td>
<td>0.01</td>
<td>0.05</td>
<td>0.06</td>
<td>—</td>
<td>0.11**</td>
<td>0.02</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>(2) Undergraduate GPA</td>
<td>341.34</td>
<td>29.16</td>
<td>0.03</td>
<td>0.17**</td>
<td>0.04</td>
<td>0.14**</td>
<td>0.14**</td>
<td>0.12**</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>(3) Gender</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.03</td>
<td>0.06</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>(4) Average Medical School Preclinical GPA</td>
<td>2.96</td>
<td>0.48</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.51**</td>
<td>0.92**</td>
<td>0.76**</td>
<td>0.61**</td>
<td>0.15**</td>
</tr>
<tr>
<td>(5) Average Medical School Clinical GPA</td>
<td>3.38</td>
<td>0.34</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.78**</td>
<td>0.46**</td>
<td>0.49**</td>
<td>0.33**</td>
</tr>
<tr>
<td>(6) Cumulative Medical School GPA</td>
<td>3.17</td>
<td>0.35</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.73**</td>
<td>0.65**</td>
<td>0.24**</td>
</tr>
<tr>
<td>(7) USMLE Step 1 Scores</td>
<td>205.24</td>
<td>15.90</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.76**</td>
<td>0.10**</td>
</tr>
<tr>
<td>(8) USMLE Step 2 Scores</td>
<td>202.22</td>
<td>18.21</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.11**</td>
</tr>
<tr>
<td>(9) PGY-1 Professionalism</td>
<td>4.18</td>
<td>0.75</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>(10) PGY-1 Experience</td>
<td>3.87</td>
<td>0.70</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

The explanatory variable, self-reported research experience, was coded as no = 0; yes = 1; the explanatory variable, gender, was coded as male = 0; female = 1. *$p < 0.05$, **$p < 0.01$. 

**TABLE I.** Means, Standard Deviations, and Pearson Correlations for the Explanatory Variables (Variables 1–3) and the Outcome Variables (Variables 4–10) ($N = 943$)
The explanatory variable, gender, was coded as male = 0; female = 1. The numbers in parenthesis are standardized regression coefficients. The explanatory variable, self-reported research experience, was coded as no = 0; yes = 1. The results are displayed in Table II.

After controlling for the students’ undergraduate GPA and gender, self-reported research experience was positively related to students’ USMLE Step 1 scores ($\beta = 0.12, p < 0.001$) and their medical school preclinical GPA ($\beta = 0.06, p < 0.10$). Those students who reported that they had some research experience before entering USU showed advantage on these two measures. However, this positive association seemed to fade away as the students moved into the clinical years of medical school and internship. Self-reported research experience did not have a significant association with medical school clinical GPA or USMLE Step 2 scores. In fact, self-reported research experience was negatively related to both of the program director rating variables; professionalism ($\beta = -0.10, p < 0.05$) and expertise ($\beta = -0.07, p < 0.05$).

Although the undergraduate GPA and gender relations were not the focus of the study, these two variables did yield statistically significant associations with some of the outcomes. Undergraduate GPA was associated with the medical school preclinical GPA ($\beta = 0.16, p < 0.001$), cumulative medical school GPA ($\beta = 0.13, p < 0.001$), USMLE Step 1 scores ($\beta = 0.14, p < 0.001$), and USMLE Step 2 scores ($\beta = 0.10, p < 0.05$). Gender was associated with medical school preclinical GPA ($\beta = -0.13, p < 0.001$) and USMLE Step 1 scores ($\beta = -0.673, p < 0.001$). With prior research experience and undergraduate GPA being equal, a male student was predicted to achieve nearly 7 points higher than a female counterpart on USMLE Step 1. The cumulative medical school GPA was also predicted to be significantly higher for male students ($\beta = -0.05, p < 0.10$) when undergraduate GPA and prior research experience were controlled.

As detailed in Table II, self-reported research experience, undergraduate GPA, and gender together can explain 6% of the variance of USMLE Step 1 scores and 4% of the variance of the medical school preclinical GPA.

## DISCUSSION

Medical school admission committees carefully consider the components in each applicant’s record when making the important decision of admitting (or denying) an applicant to medical school. Prior research experience is commonly considered because it is thought to have a positive impact on students’ performance in medical school study and internship. Because prior research experience is commonly considered, potential matriculants and faculty spend considerable time and resources obtaining and evaluating research experience, respectively. From the results of the present study, we found that those students who reported some type of research experience performed somewhat better on USMLE Step 1 and had a slightly higher medical school preclinical GPA. However, these positive relationships were small and did not persist into the third and fourth years of medical school. Further, self-reported research experience did not predict students’ performance on USMLE Step 2 or medical school clinical GPA. Moreover, in our sample, the existence of prior research experience showed a small, unexpected, negative correlation with program director ratings of professionalism and expertise.

Not surprisingly, undergraduate GPA was moderately associated with students’ USMLE Step 1 and Step 2 performance, medical school preclinical GPA, and cumulative GPA upon graduation. These findings are consistent with prior studies and add validity to our findings. Interestingly, just as self-reported research experience, the initial advantage of having a higher undergraduate GPA seemed to become weaker as time went on. Undergraduate GPA did not have a significant relation to medical school clinical GPA or PGY-1 professionalism and expertise, the two outcomes measured during internship. This may be attributable to the unique, clinical nature of medical education in the later years, which is quite different from traditional post-secondary education. This finding has implications for admission committees who may stringently set cut points for undergraduate GPA when admitting students to their institutions. Our results suggest that this practice may be flawed.

Additionally, in our sample, the male students showed considerably better performance than the female students on USMLE Step 1 and medical school preclinical GPA. Again, gender was not significantly related to medical school clinical GPA, USMLE Step 2, or internship outcomes. In fact, on average, the female students outperformed male students on medical school clinical GPA. The relationship between gender and USMLE scores found in this study is generally consistent with the previous larger-scale studies, again adding valid evidence to our findings.

### TABLE II. Impact of the Explanatory Variables on the Outcome Measures Reflected by Multiple Linear Regression Models

<table>
<thead>
<tr>
<th>Outcome Measures</th>
<th>Regression Coefficients of the Explanatory Variables</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Self-Reported Research Experience</td>
<td>Undergraduate GPA</td>
</tr>
<tr>
<td>Average Medical School Preclinical GPA</td>
<td>0.06 (0.06)*</td>
<td>0.003 (0.16)***</td>
</tr>
<tr>
<td>Average Medical School Clinical GPA</td>
<td>-0.004 (-0.006)</td>
<td>0.000 (0.04)</td>
</tr>
<tr>
<td>Cumulative Medical School GPA</td>
<td>0.01 (0.02)</td>
<td>0.002 (0.13)***</td>
</tr>
<tr>
<td>USMLE Step 1 Scores</td>
<td>3.9 (0.12)***</td>
<td>0.08 (0.14)***</td>
</tr>
<tr>
<td>USMLE Step 2 Scores</td>
<td>0.91 (0.03)</td>
<td>0.06 (0.10)**</td>
</tr>
<tr>
<td>PGY-1 Professionalism</td>
<td>-0.14 (-0.10)**</td>
<td>-0.001 (-0.04)</td>
</tr>
<tr>
<td>PGY-1 Expertise</td>
<td>-0.10 (-0.07)**</td>
<td>0.001 (0.02)</td>
</tr>
</tbody>
</table>

The numbers in parenthesis are standardized regression coefficients. The explanatory variable, self-reported research experience, was coded as no = 0; yes = 1; the explanatory variable, gender, was coded as male = 0; female = 1. *p < 0.10, **p < 0.05, and ***p < 0.001.
The total variance explained in each of the regression models was small and is consistent with prior work.\textsuperscript{1,3} Not surprisingly, there is considerable variance left to be explained, more for the second 2 years of medical school and internship than for the first 2 years of study. The small $R^2$ is also likely attributable to the fact that we had a fairly large sample size in the present study. Notably, research experience did add to the explained variance of a validated program director’s rating form. The reasons for the negative correlation are unclear but may suggest, for example, that these students prefer bench-top research to clinical work and/or may have less tolerance for ambiguity (with ambiguity being quite common in clinical medicine). Using research experience as a criterion for admission could be more helpful for schools with a focus on physician-scientist education. In total, these data suggest that although research experience should be considered by admission committees, its importance in the selection process of medical students may be limited. Of course, we also recognize that our findings and this interpretation should be confirmed by other institutions.

There are several important limitations in our study. First, the number of explanatory variables was small. Some typical medical school admission criteria, such as Medical College Admission Test score, were not included in the analysis because of a large amount of missing data. Second, despite the large sample size, the observations were all derived from a single institution, and thus we need to be careful when generalizing our results to other medical schools and medical education contexts. Accordingly, future work should attempt to replicate our findings with more heterogeneous samples and more diverse medical education settings. Third, our study assessed associations and not causation. And finally, our data were several years old. Nonetheless, it is important to understand that although the data were derived from students who graduated in the 1990s, the application process, including the AMCAS application, which is utilized by most medical schools in the United States, currently still emphasizes many self-reported experiences, including self-reported research experience. Therefore, we believe our findings may be applicable to many of today’s medical schools.

Notwithstanding the limitations of the present study, we believe there were several important strengths. First, data were collected across multiple years and are quite complete (with complete records for 85% of the participants). In particular, the outcome measures were gathered across the medical education continuum from medical school preclinical study to clinical study and on into internship. Additionally, the instrument used to assess performance in internship has evidence of reasonable reliability and validity, based on previous empirical work.\textsuperscript{5}

**CONCLUSION**

This study provides some justification for the practice of considering self-reported research experience as one of many selection criteria. However, our results also suggest that admission committees should use this information with caution. Although some prior experience in doing research may predict better performance in the preclinical phase of medical school, our data suggest that educators should not expect medical school applicants’ self-reported research experience to continue to make positive contributions during the clinical phase of medical school or internship. Further, our results indicate that prior research experience can have a small, negative association with internship performance. Finally, based on our results, it seems that even undergraduate GPA becomes insignificant when one attempts to determine an applicant’s potential in internship.

As is clear from the present study, much variance in medical school and internship performance goes unexplained. As such, medical educators should continue to broaden the criteria they use to select future doctors.\textsuperscript{1,3,9} For example, the multiple mini-interview, a rigorous assessment tool with substantial evidence supporting its reliability and validity, has been adopted, with great success, by more and more medical school admission committees.\textsuperscript{2,10} We believe this tool, as well as other nontraditional assessments of characteristics such as self-regulated learning, motivation, and study skills might provide additional insight into the difficult task of medical student selection.

**REFERENCES**

Identifying Themes Within a Medical School Admission Committee’s Reviews of Applicants

COL William R. Gilliland, MC USA (Ret.); CDR Anthony R. Artino Jr., MSC USN; Donna M. Waechter, PhD; Col John E. McManigle, USAF MC (Ret.); David F. Cruess, PhD; CAPT Margaret Calloway, MC USN; Bonnie C. Arze, MD; Steven J. Durning, MD, PhD

ABSTRACT Background: Admissions committees attempt to select the most qualified applicants based on many cognitive and “noncognitive” factors. Purpose: Identify common themes cited in the admissions committee member summaries of medical school matriculants and determine the relative frequency and importance of these themes. Methods: After reviewing a convenience sample of 150 reviewer comments, 14 qualitative themes were identified. Utterances (thematic word strings) from each of the three reviewer comments for each matriculant for 7 academic years (1989–1996) were then categorized and coded as being positive, negative, or neutral. Intra-rater and inter-rater reliabilities were calculated. Results: Utterances (n = 9299) about 981 matriculants were categorized by theme and sorted as being positive, neutral, or negative. Intra-rater reliabilities were excellent (mean $K_r = 0.98$, range 0.90–1.00). Similarly, inter-rater reliabilities were also excellent (mean $K_i = 0.94$, range 0.55–1.00 and mean $K_i = 0.90$, range 0.08–1.00). Four themes (overall summarizing comments, academic, test scores, and motivation) accounted for more than half (56%) of the utterances. Conclusions: We were able to qualitatively identify themes and provide information about how one committee weighs both cognitive and “noncognitive” factors. Admission committees should consider reexamining their process and potentially expanding, eliminating, or modifying application components.

INTRODUCTION Choosing the best applicants for one’s school—those individuals who are most likely to succeed not only as students but also as clinicians—remains a challenging task for medical school admissions committees. A number of reasons make this task so crucial, not to mention difficult. The first is that there still is an abundance of qualified candidates for a very limited number of positions. In 2009–2010, 42,742 people applied to the 2010 MD degree-granting medical schools in the United States and 18,665 applicants ultimately matriculated.¹ A second reason is that the medical profession and society demands that medical schools select candidates with those academic abilities and humanistic qualities to become physicians or conversely to not accept those candidates who do not. But perhaps the most important reason for a rigorous selection process is that the vast majority of medical school matriculants graduate from medical school. Given that the attrition rate for academic reasons is less than 2%,² matriculation is indeed tantamount to graduation.

Many factors weigh into an admissions committee’s decision to accept or not accept applicants. Although it is relatively easy to obtain information about undergraduate academic and standardized test performance used by medical schools,³ it is difficult to find detailed information about the tools used by admissions committees to assess so-called “noncognitive” qualities. For example, our school, the Uniformed Services University (USU) simply states that our admissions process is a “…three-stage, progressive screening process…the first stage consists of submission of an AMCAS (American Medical College Application Service) statement; the second, the submission of supplementary materials; and the third, personal interviews, which are conducted on campus.”³³ Although the “selection factors” section of our admissions guidelines is meant to be brief, medical school selection is obviously a much more complicated process that weighs many academic and noncognitive factors.

Even less is known about the composition, structure, and the “weight” of the various factors used by medical school admissions committees to determine which applicants to offer an acceptance. In a survey of admissions officers in 1982, the factors considered most important, in descending order of importance, were the undergraduate grade point average (GPA), personal interview, and scores on the Medical College Admission Test (MCAT).⁴ Similarly, a survey sent to admissions officers of U.S. medical schools in 1981 found that the most important selection factors were GPA, interviews, MCAT scores, and letters of references.⁵ More recently, in a survey of admission officers by the Association of American Medical Colleges (AAMC), the five most important factors contributing to an applicant receiving an invitation to an interview were cumulative undergraduate GPA, cumulative science/math GPA, MCAT total scores, and community service, whereas the five most important factors in offering acceptances were the interviewer recommendation, letters of recommendation, cumulative science/math GPA, cumulative undergraduate GPA, community service/volunteerism, and MCAT total scores.⁶

F. Edward Heber School of Medicine, Uniformed Services University of the Health Sciences, 4301 Jones Bridge Road, Bethesda, MD 20814.

The authors are U.S. Government employees. The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Uniformed Services University of the Health Sciences, Department of Defense, or the U.S. Government.
Our admissions committee consists of 18 voting members, four nonvoting members, and a chairman who only votes if there is a tie. Admission committee members are selected from faculty who have conducted admissions interviews in the past, USU alumni, or basic science faculty. They all have faculty appointments and hold an MD, PhD, or combined MD/PhD degree. Committee appointments are for 3 years with the option for additional appointments with approval for the Dean, School of Medicine. As described above, our admission methodology consists of a three-step review process followed by committee deliberation. During the primary review, each application is reviewed to determine if the applicant has met the required eligibility and academic requirements. Qualified applicants are asked to submit supplementary application materials (personal photograph, USU essay, activity update, and military experience form). During the secondary review, two committee members independently evaluate the credentials and recommend either an invitation for interview or rejection. If one of the committee members performing the secondary review recommends an invitation, the applicant is invited for an interview. Therefore, invitation by the admissions committee to interview is only extended to the most promising students, and that decision is largely based on academic credentials, written statements, letters of recommendation, personal attributes, and other information contained in the AMCAS application. After personal interviews are conducted, three voting committee members review each candidate independently and provide a tertiary review. The committee reviews all of the available information to include written statements, undergraduate and/or postgraduate GPA, MCAT scores, demographic data, letters of recommendation, interview score sheets, and any other information contained in their admissions packet. At our institution, we do not differentially “weigh” cognitive and noncognitive factors in the admission process, so that task is left to the tertiary reviewers. (For the purpose of this article, cognitive factors included GPAs and scores on standardized examinations. All other items were considered to be noncognitive factors.) Each tertiary reviewer then summarizes the applicant with a few sentences or short paragraph and assigns them a score between 1 (lowest) and 5 (highest). This is an attempt to quantify the process. All three scores are then totaled, and if an applicant scores an 11 or above, they are presented to the full committee with the reviewer with the highest score selected for presentation of the record. There is no limitation on the length of the reviewer’s comments. It is the content of these reviews that are the subject of this article. The last step in the process is the deliberation phase when a decision is made by majority of the voting members attending the admissions committee meetings.

The purpose of this study was to identify common themes in the tertiary reviewer summaries and determine the relative importance of these themes in accepting applicants to medical school. Indirectly, this approach allowed us to investigate the frequency of the various tools that were cited in the tertiary reviews. Although not addressed in this manuscript, a broader purpose of this work will be to determine predictive validities of these themes as we track the long-term performance of our graduates. We believe that findings from our institution can be generalizable and help streamline the admissions process at other institutions.

**METHODS**

This retrospective study also used a database linking various cognitive and noncognitive measurements with products (or outcomes) known as the Long-Term Career Outcome Study. This database includes extensive information about our graduates and includes over 500 characteristics including demographic, academic, licensing, residency, and professional information on our graduates. Unlike any other U.S. medical school, USU has the unique ability to longitudinally track the performance of our graduates. We are able to track our students over the long-term because our graduates are obligated to serve 7 years in the U.S. Army, U.S. Navy, U.S. Air Force, or U.S. Public Health Service after completion of their graduate medical education. This gives us the opportunity to collect data on our graduates long after graduation and link that information to earlier data from medical school and premedical school. Our institutional review board approved this study.

**RESULTS**

After review of the convenience sample of 150 tertiary reviews, 14 qualitative themes were identified: (1) summarizing...
Identifying Themes

For the 7 academic years included in this study (1989–1996), USU matriculated 1024 students. Of these 1024 matriculants, tertiary comments could be located for 981 matriculants. Therefore, the total number of reviewer comments that were coded was 2943. A total of 9299 utterances were categorized by theme and sorted as being positive, neutral, or negative. Approximately 4% of the words included in the tertiary reviews were uncodable because they did not fit into any of the themes or were comments that could not be sorted as being positive, neutral, or negative, such as being from a “rural background.”

The frequency of the utterances within themes, sorted as being positive, neutral, and negative, is presented in Table I.

Intra-rater reliabilities were calculated for the 42 categories of utterances. Kappa scores ranged from 0.90 to 1.00 with a mean of 0.98. Kappa scores for 12 of the 42 comparisons could not be calculated because of too few entries or asymmetric data. The mean inter-rater reliability between the PI (William R. Gilliland) and one coauthor (Steven J. Durning) was 0.94 (range 0.55–1.00). The mean inter-rater reliability between the PI (William R. Gilliland) and another coauthor (Bonnie C. Arze) was 0.90 (range 0.08–1.00). Kappa scores could not be calculated for 11 categories for the first and 19 for the second inter-rater reliabilities again because there were too few entries.

Four themes (overall summarizing comments, academic, test scores, and motivation) accounted for more than half (55.5%) of the utterances. Approximately 14% of the

<table>
<thead>
<tr>
<th>Themes</th>
<th>Total Utterances</th>
<th>% of Total Utterances</th>
<th>Positive</th>
<th>Neutral</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summarizing Comments</td>
<td>1724</td>
<td>18.5</td>
<td>1414</td>
<td>138</td>
<td>172</td>
</tr>
<tr>
<td>Academic</td>
<td>1413</td>
<td>15.2</td>
<td>881</td>
<td>104</td>
<td>428</td>
</tr>
<tr>
<td>Test Scores</td>
<td>1100</td>
<td>11.8</td>
<td>550</td>
<td>136</td>
<td>414</td>
</tr>
<tr>
<td>Motivation</td>
<td>932</td>
<td>10.0</td>
<td>702</td>
<td>42</td>
<td>188</td>
</tr>
<tr>
<td>Interviews</td>
<td>774</td>
<td>8.3</td>
<td>545</td>
<td>71</td>
<td>159</td>
</tr>
<tr>
<td>Recommendation Letters</td>
<td>728</td>
<td>7.8</td>
<td>602</td>
<td>67</td>
<td>59</td>
</tr>
<tr>
<td>Military Experience</td>
<td>693</td>
<td>7.5</td>
<td>608</td>
<td>8</td>
<td>77</td>
</tr>
<tr>
<td>Medical Experience</td>
<td>623</td>
<td>6.7</td>
<td>411</td>
<td>12</td>
<td>200</td>
</tr>
<tr>
<td>Personal Traits</td>
<td>500</td>
<td>5.4</td>
<td>452</td>
<td>2</td>
<td>46</td>
</tr>
<tr>
<td>Written Statements</td>
<td>244</td>
<td>2.6</td>
<td>148</td>
<td>18</td>
<td>78</td>
</tr>
<tr>
<td>Extracurricular Activities</td>
<td>197</td>
<td>2.1</td>
<td>162</td>
<td>2</td>
<td>33</td>
</tr>
<tr>
<td>Maturity</td>
<td>167</td>
<td>1.8</td>
<td>153</td>
<td>1</td>
<td>13</td>
</tr>
</tbody>
</table>

Four themes (overall summarizing comments, academic, test scores, and motivation) accounted for more than half (55.5%) of the utterances. Approximately 14% of the

<table>
<thead>
<tr>
<th>Theme</th>
<th>Value</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summarizing Comments</td>
<td>Positive</td>
<td>“This fellow is truly formidable, I hope he comes here!”</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>“Would be an average medical student.”</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>“I am concerned about offering a place at USU.”</td>
</tr>
<tr>
<td>Academics</td>
<td>Positive</td>
<td>“Excellent recent academic performance.”</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>“GPA is not super, but okay.”</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>“Undistinguished academic work.”</td>
</tr>
<tr>
<td>Motivation</td>
<td>Positive</td>
<td>“Clear motivation for medicine.”</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>“I am not clear of his motivation to ‘clinical medicine’ or if he is so focused on aerospace medicine after his NASA experience.”</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>“Unsure of passion for medicine.”</td>
</tr>
<tr>
<td>Recommendation Letters</td>
<td>Positive</td>
<td>“Committee letter presents him as an outstanding candidate.”</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>“LORs are good, but very superficial.”</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>“One LOR doubts he would do well in medical school.”</td>
</tr>
<tr>
<td>Military Experience</td>
<td>Positive</td>
<td>“Distinguished career as enlisted Sgt in U.S. Air Force.”</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>“No first-hand experience with the military, but familiar with it through discussion with naval officer.”</td>
</tr>
<tr>
<td>Written Statements</td>
<td>Negative</td>
<td>“Major wrinkle is relative lack of exposure to military.”</td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>“He writes a strong personal statement.”</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>“Statements are acceptable.”</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>“AMCAS statement left me with no clear handle on this applicant.”</td>
</tr>
</tbody>
</table>
Identifying Themes

It may not be surprising that overall summarizing comments such as “this is an outstanding candidate” or “there are no wrinkles in this packet” were the most common theme identified in our study. In fact, in many of our highest rated applicants, the entire tertiary review had only one positive summarizing utterance. It seems that brief summarizing comments may be particularly useful as they allow tertiary reviewers, who are likely to be extremely busy faculty members, to get their “bottom line” across to the rest of the committee quickly and succinctly. Conversely, those applicants who have a mixture of positive, neutral, and negative aspects of their application frequently had the lengthiest tertiary reviews.

Although it may appear that the 14 themes are mutually exclusive, in many cases, they may not be. For example, if one considers a theme such as “motivation,” the source of information on which the tertiary reviewer’s comment is based may have come from one or many parts of the application, including the interview, written statements, or letters of recommendations. Therefore, the contributions of these aspects of the applications may be underrepresented.

The order of the themes in this study varied from the “importance factors” identified in the most recent survey from the AAMC. This is in part because the AAMC asked two separate questions regarding the importance of the various components. Specifically, they asked admission officers to assign “importance ratings” to components that led to an “Invitation to Interview” and “Offer for Acceptance.” This suggests that many schools have a “stepwise” approach to the admission process, relatively emphasizing academic performance and standardized test scores in order for an applicant to get an invitation to interview and relatively emphasizing noncognitive factors when offering acceptances to medical school.

Although it is beyond the scope of this article to discuss the predictive validity of the various components of an application packet, a recent study suggests that the components that correlate best with future performance are academic scores such as GPAs, aptitude tests such as the MCAT, and elements of a multiple mini-interview. The list of components that seem to have little or no predictive validity include personal interviews, letters of reference, personality testing, measures of emotional quotient/intelligence, and situational judgment tests. Prior comprehensive reviews reached similar conclusions about the lack of predictive validity of these various noncognitive tools used in admissions decisions but suggested that there might be ways to modify some of the existing tools in an effort to improve their predictive validities. Therefore, it may not be surprising that academic and standardized tests were two of the most common themes identified in this study. Indeed, in the medical school admission requirements of almost all schools, data about these selection factors are easy to find.

We are particularly interested in determining if there is any correlation of negative utterances in the tertiary reviews and long-term outcomes.

Perhaps the finding that is the most surprising is the relatively low percentage of tertiary reviews that specifically commented on interviews, recommendation letters, and written statements. All three are common components of most medical school applications. Obviously, these components require a great deal of time and effort in the scheduling, organization, implementation, and review by admissions committee personnel and faculty, not to mention considerable time and effort on the part of the applicant.

As the only U.S. military medical school, our institution has a unique mission to train health care professionals in the U.S. Department of Defense and the U.S. Public Health Service. Therefore, it was not surprising that 7.5% of the utterances addressed military service. However, qualities such as leadership and service, which are thought to be critically important traits in the uniformed services, only accounted for 1.2% and 1.0% of utterances, respectively.

According to the AAMC, the majority of accepted applicants at virtually every medical school self-report experience in medically related work and research. However, medical experience accounted for only 6.7% of the utterances.

As with other medical schools, the admissions process is very time consuming for the members. During our 7-month admission cycle, each week the average committee member spends 1 hour screening files, 2 hours reviewing files, and 3½ hours in the committee meeting. Based on the relative frequency of the themes identified in this article, admission committees should reexamine their admissions process and consider expanding, eliminating, or modifying components of their medical school application that may not be influencing their committee decisions or adding any predictive validity.

Many medical schools have attempted to modify and improve some of the components of a traditional admissions
process. For example, applicants to Scottish medical schools sit for a battery of psychometric tests to measure cognitive ability, personality traits, and moral/ethical reasoning (Personal Qualities Assessment), and the developers suggest that it may improve the selection process.\(^1\) The University of British Columbia replaced individual interviews with a panel interview after they found that the interpanel reliability coefficient was superior to inter-rater reliability from individual interviews.\(^2\) In an attempt to minimize the inherent context specificity in the traditional “one-on-one” interview, Eva et al\(^3\) at McMaster created the multiple mini-interview consisting of 10 short objective structured clinical examination-style stations and found this approach cost-effective and feasible, with a reliability of 0.65. It is hoped that innovations such as these will benefit and streamline the process for admissions committees around the world, but it is equally important that innovations such as these be systematically studied to determine if they can benefit committee processes at other schools.

There are several important limitations of the present study. First and foremost, this study was conducted at a single medical school. The admissions process and the composition of the admission committee obviously vary across medical schools. In addition, the noncognitive information that is reviewed by admissions committees at various schools differs. For example, although many schools include an interview in the admissions process, the characteristics of that interview may vary greatly. For instance, some interviews are conducted by a single person or a member of the committee, whereas others are conducted by a panel. Further, interview structures may range from being totally unstructured to interviews where questions are entirely predetermined.

Another limitation concerns the academic years chosen for this study. These years were chosen in part because it will allow us to retrospectively examine the predictive validity of these comments on many of the long-term outcomes spanning medical school, graduate medical training, and a military medical career available to us as part of the Long-Term Career Outcome Study. Additionally, although the years chosen do not represent our most recent graduates, the components of the admission process at our school have not changed substantially in several decades. Tertiary reviewer comments from nonmatriculants were not analyzed since they would be no long-term outcomes on them.

Finally, this article does not address the critical final phase, deliberation, in ours and most other admissions committees. In fact, Elam et al\(^4\) reported that approximately 20\% of votes cast after initial screening were changed following committee deliberation.

**CONCLUSION**

Identifying the qualities that a medical school admissions committee considers in applicants and the tools used to assess these qualities remains a controversial topic in medical education. As noted in the discussion section, it is a very time-consuming process for all schools. In this article, we were able to qualitatively identify, with excellent inter-rater reliabilities, 14 themes present in tertiary reviews. In addition, this article provides information about how one such committee “weights” the various factors they collect in their admission process. Given the findings of this study, it is reasonable for medical schools to examine the components that they use to choose among their many qualified applicants and consider modifying components that could streamline the process, thereby making it more efficient for both committee members and applicants.

**REFERENCES**

1. Association of American Medical Colleges: Medical School Admission Requirements (MSAR\(^5\)). Washington, DC, Association of American Medical Colleges, 2011, p 286.
Relationship Between Admissions Committee Review and Student Performance in Medical School and Internship

COL William R. Gilliland, MC USA (Ret.); Ting Dong, PhD; CDR Anthony R. Artino Jr., MSC USN; Donna M. Waechter, PhD; David F. Cruess, PhD; LTC Kent J. DeZee, MC USA; COL John E. McManigle, USAF MC (Ret.); Steven J. Durning, MD, PhD

ABSTRACT

Purpose: To investigate the association between tertiary reviewer (admissions committee member) comments and medical students’ performance during medical school and into internship. Methods: We collected data from seven year-groups (1993–1999) and coded tertiary reviewer comments into 14 themes. We then conducted an exploratory factor analysis to reduce the dimensions of the themes (excluding the Overall impression theme). Subsequently, we performed Pearson correlation analyses and multiple linear regression analysis to examine the relationship between the factors and seven outcome measures: medical school preclinical grade point average (GPA), medical school clinical GPA, cumulative medical school GPA, U.S. Medical Licensing Examination Step 1 and 2 scores, and scores on a program director’s evaluation measuring intern professionalism and expertise. Results: We extracted seven factors from the 13 themes and found small-to-moderate, significant correlations between the factors, the Overall impression theme, and the outcome measures. In particular, positive comments on Test and Maturity were associated with higher U.S. Medical Licensing Examination Step 1 and 2 scores. Negative comments on Interview and Recommendations were associated with lower ratings of professionalism during internship. Comments on Overall impression were significantly associated with all the outcome measures. Conclusions: Tertiary reviewer comments were weakly associated with performance in medical school and internship. Compared with positive comments, negative comments had stronger associations with medical school and internship performance measures.

INTRODUCTION

Medical school admissions committees seek to enroll the best applicants. As such, many schools have a multistep admissions process. Initial application materials are reviewed, followed by invited interviews. Typically, these invited interviews are followed by a second step: admissions committee member review, whereby one or more members of the admissions committee reviews the entire package and makes comments on the applicant’s suitability for enrollment. The importance of admissions committee member review—referred to as “tertiary review” in the remainder of this article—is apparent; recommendations regarding accepting and rejecting students into the matriculating class often emerge from these tertiary reviews.

Medical educators want to know how reliable these tertiary reviews are in predicting future performance. For example, are such comments associated with students’ performance in medical school and internship? This question guided the conduct of the current study.

In the medical school admissions literature, researchers have focused primarily on the impact of applicants’ previous academic achievement and interview performance, personality traits, emotional intelligence, learning styles, and a few on personal statements and references. There is a dearth of research addressing the issue of the reliability and validity of tertiary reviews.

We hypothesized that tertiary reviewer comments would be associated with medical school and internship performance. Further, we predicted that the associations between tertiary reviewer comments and educational outcomes would be small, given the intervening time (up to 6 years).

METHODS

Study Contexts and Participants

The present study was part of the Long-Term Career Outcome Study conducted at the F. Edward Hébert School of Medicine, Uniformed Services University of the Health Sciences (USU). The study collects data on past, present, and future students to track their performance through medical school, graduate medical training, and the rest of their military career.

USU matriculates approximately 170 medical students annually. After graduation, nearly all of USU students train in military-affiliated programs across the country. At the time of the present study, the university offered a traditional 4-year curriculum: 2 years of basic science courses followed by 2 years of clinical rotations. For the current study, we drew our sample from graduates who received their MD degree from USU between 1993 and 1999 for whom we also had a previously validated internship performance measure (N = 767). These years were chosen in part because it would allow us to retrospectively examine the predictive validity of these comments on many of the long-term outcomes spanning medical school, graduate medical training, and a military medical career.

F. Edward Hébert School of Medicine, Uniformed Services University of the Health Sciences, 4301 Jones Bridge Road, Bethesda, MD 20814.
Coding Scheme of Tertiary Reviewer Comments

We obtained the tertiary reviewer comments from the USU admissions office, and each matriculant was rated by three reviewers. We adopted a two-step procedure to analyze the comments as reported elsewhere. First, the first author (William R. Gilliland) chose, at random, 150 tertiary reviews and used a grounded theory approach to identify themes for utterances. An utterance is a thematic word string, and each tertiary review usually contained multiple utterances. After analyzing the 150 reviews, the three authors discussed the codes and established consensus on 14 themes—Academic, Test, Medical experience, Military experience, Service, Leadership, Interview, Recommendations, Motivation, Maturity, Personal statement, Personal traits, Extracurricular activities, and Overall impression. Next, the three authors coded the remainder of the tertiary reviewer comments independently, using the coding scheme described above, evaluating each coded utterance as positive, negative, or neutral. As a result, there were 42 categories of codes (14 themes crossed with each positive, negative, and neutral category).

The following is an example of one tertiary review with a representative coding sample in parentheses:

**Overall solid (overall impression, positive)—Good letters of recommendation (recommendation letters, positive) and interviews (interviews, positive). Military motivation average and typical for nonmilitary background (military experience, neutral).**

Medical College Admission Test (MCAT) is real soft spot (test, negative).

We checked both intra-rater reliability and inter-rater reliability. Intra-rater reliability was assessed by having the first author (William R. Gilliland) recode an entire year of tertiary reviewer comments and compare them with his prior codes. Kappa scores ranged from 0.90 to 1.00, with a mean of 0.98. Kappa scores for 12 of the 42 categories could not be calculated because there were too few entries or the data were asymmetric. We conducted the inter-rater reliability analysis by having the coders each independently recode an entire year of tertiary reviews and compare their codes with those of the first author. The mean inter-rater reliability between the first author and one co-coder (Steven J. Durning) was 0.94 (range 0.55 to 1.00). The mean inter-rater reliability between the first author and the other co-coder (research assistant) was 0.90 (range 0.08 to 1.00). Kappa scores could not be calculated for 11 categories for the first and 19 for the second inter-rater reliabilities, again because there were too few entries or the data were asymmetric.

Study Variables and Procedures

We calculated the total number of positive and negative comments on each theme for each student. For example, if altogether the three tertiary reviewers gave four positive comments for a student on Test, the student received a score of 4 on Test positive category. If the student’s Motivation was evaluated negatively twice by the reviewers, he or she would receive a score of 2 on Motivation negative category. We also calculated a composite score (positive scores minus negative scores) on each theme for each student.

We acquired the outcome measures from two sources. First, the USU registrar’s office provided the following information: medical school grade point average (GPA) at the end of the first 2 years of study (preclinical GPA), medical school GPA of the second 2 years of study (clinical GPA), cumulative medical school GPA upon graduation from USU, and U.S. Medical Licensing Examination (USMLE) Step 1 and 2 scores. The second dataset was obtained from a program director’s survey of trainee performance completed at the end of the intern year (postgraduate year 1 [PGY-1]). These outcome variables are described in greater detail below.

Medical School GPA (Preclinical, Clinical, and Cumulative)

For each medical student, a medical school preclinical GPA was calculated using course grades from the first 2 years. The GPA is a weighted average created by multiplying each course grade by the number of contact hours for the given course, summing the weighted grades across courses, and then dividing the sum by the total number of contact hours. The resulting averages were converted to a common 4-point scale (range: 0.0–4.0). Similarly, a GPA based on the second 2 years of medical school (clinical) and a cumulative GPA for all 4 years of medical school were calculated.

USMLE Step 1 and 2 Scores

Students in this sample completed the Step 1 exam which focuses on understanding of the basic sciences after their first 2 years of medical school. Students completed the Step 2 exam that focuses on clinical knowledge during their fourth year of medical school. Scores on the USMLE Step exams are three digits and range from 140 to 280, with a mean of between 200 and 220 and SD of 20. In the present study, we used students’ scores on their first attempt at each of the step exams.

Program Director’s Ratings

The program director’s rating form was completed at the end of the intern year. Survey items were developed by interdepartmental educators at USU with expertise spanning undergraduate and graduate medical education programs. The survey consisted of 18 items scored on a 5-point Likert scale, where 5 = outstanding, 4 = superior, 3 = average, 2 = needs improvement, 1 = not satisfactory, and 0 = unable to judge (note: scores of 0 were treated as missing data). The feasibility (annual response rate ranged from 72 to 90%) and reliability (Cronbach’s α ranged from 0.93 to 0.96) of this instrument were reported in a previous study using the same cohort of students. An exploratory factor analysis (EFA) was also performed in a previous study to test the construct validity of the instrument. The results indicated that the 18 survey items loaded onto two constructs: expertise (13 items) and professionalism (5 items). We used these two constructs
(PGY-1 Professionalism and PGY-1 Expertise) as separate outcomes in the present study.

**Statistical Analysis**

The analysis consists of three parts. First, we performed EFA on the 13 themes to reduce the number of dimensions, excluding the Overall impression theme. The reason we excluded the Overall impression theme from the factor analysis is that it was the tertiary reviewers’ summary and did not provide additional information to the other 13 themes. We used the composite score of each theme to conduct the EFA. Second, we examined the Pearson correlations between the tertiary reviewer comments and the students’ performance in medical school and internship. Finally, we calculated the total variance explained by tertiary reviewer comments for each performance measure using multiple linear regression modeling. All the analyses were carried out using PASW Statistics 18.0 (SPSS, Chicago, Illinois). This study was approved by the USU Institutional Review Board.

**RESULTS**

**Descriptive Statistics and Factorization of the Themes**

Descriptive statistics of the 14 themes are shown in Table I. The large SDs, compared to the means, and the wide range indicated considerable variability of students’ scores on the themes, especially on the category of Military experience, Academic, and Test.

For the EFA, we used principle component extraction with varimax rotation, which assumes the independent structure of the factors. The results of the EFA are shown in Table II. The 13 themes of the tertiary reviewers’ comments were reduced to seven factors. For this analysis, the following rule was used: a theme was assumed to load on a particular factor if the absolute value of the loading was greater than 0.50. It should be noted that the purpose of the EFA was only to reduce the number of dimensions of the themes for subsequent analyses. Therefore, some of the criteria of factor selection commonly adopted in survey research are not applicable here, such as the common rule that a factor has to contain at least three items to be retained.

The seven factors that emerged from the EFA were (1) Recommendations and Interview, (2) Personal traits, (3) Academic and Extracurricular activities, (4) Leadership and Medical experience, (5) Test and Maturity, (6) Motivation and Military experience, and (7) Personal statement and Service. The seven factors extracted accounted for 64.14% of the total variance in the 13 themes. Inspection of the communalities revealed the majority of the themes had high extracted communalities (i.e., >0.50; Table II), which indicates that much of the variance of the themes can be explained by the seven factors.

**Pearson Correlation and Multiple Regression Analyses**

We conducted two correlation analyses to investigate the association between the seven factors and the outcome variables measuring the students’ performance during medical school and into internship. The results are shown in Table III. In the first correlation analysis, we focused solely on the negative comments of the tertiary reviewers. It should be noted that a larger value of negative comments on a particular theme or factor means higher frequency of receiving negative comments on this theme or factor from the reviewers as we described our coding scheme earlier. Scores on Recommendation and Interview were negatively correlated with program directors’ ratings on professionalism \((r = -0.12, 95\% CI = [-0.19, -0.05], p < 0.01)\). Scores on Personal traits had a negative correlation with medical school clinical GPA \((r = -0.10, CI = [-0.17, -0.03], p < 0.05)\). Scores on Academic and Extracurricular activities were negatively correlated with students’ medical school preclinical GPA \((r = -0.11, CI = [-0.18, -0.04], p < 0.01)\) and USMLE Step 1 score \((r = -0.13, CI = [-0.20, -0.06], p < 0.01)\). Scores on Test and Maturity had negative correlations with medical school preclinical GPA \((r = -0.14, CI = [-0.21, -0.07], p < 0.01)\), medical school clinical GPA \((r = -0.11, CI = [-0.18, -0.04], p < 0.01)\), cumulative medical school GPA \((r = -0.14, CI = [-0.21, -0.07], p < 0.01)\), and USMLE Step 1 \((r = -0.24, CI = [-0.31, -0.17], p < 0.01)\) and Step 2 score \((r = -0.24, CI = [-0.31, -0.17], p < 0.01)\).

The second correlation analysis focused on the positive comments of the tertiary reviewers. Scores on Motivation and Military experience were negatively correlated with students’ USMLE scores, both Step 1 \((r = -0.11, CI = [-0.16, -0.02], p < 0.01)\) and Step 2 \((r = -0.11, CI = [-0.16, -0.02], p < 0.01)\). Students’ score on Tests and Maturity had a positive correlation with their performance on USMLE Step 1 \((r = 0.10, CI = [0.03, 0.17], p < 0.05)\).

We also included the correlations between the Overall impression theme and the outcome variables at the bottom of
Factor loadings were suppressed if smaller than 0.10.

### TABLE III. Pearson Correlations Between the 7 Factors (and the Overall Impression Theme) and the Outcome Measures and Total Explained Variances ($N = 767$)

<table>
<thead>
<tr>
<th>Theme</th>
<th>Commuinity</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Interview</td>
<td>0.684</td>
<td>-0.02</td>
<td>-0.07</td>
<td>-0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>-0.12**</td>
<td>-0.07</td>
</tr>
<tr>
<td>2 Recommendations</td>
<td>0.649</td>
<td>-0.05</td>
<td>-0.10*</td>
<td>-0.07*</td>
<td>-0.02</td>
<td>-0.05</td>
<td>-0.03</td>
<td>-0.04</td>
</tr>
<tr>
<td>3 Personal Traits</td>
<td>0.614</td>
<td>0.004</td>
<td>-0.03</td>
<td>0.05</td>
<td>0.07</td>
<td>0.06</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>4 Military Experience</td>
<td>0.740</td>
<td>-0.11**</td>
<td>-0.02</td>
<td>-0.09*</td>
<td>-0.13**</td>
<td>-0.07</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>5 Motivation</td>
<td>0.581</td>
<td>0.05</td>
<td>0.04</td>
<td>0.03</td>
<td>0.07</td>
<td>0.06</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>6 Extracurricular Activities</td>
<td>0.576</td>
<td>-0.14**</td>
<td>-0.11**</td>
<td>-0.14**</td>
<td>-0.24**</td>
<td>-0.24**</td>
<td>0.02</td>
<td>-0.02</td>
</tr>
<tr>
<td>7 Academic</td>
<td>0.533</td>
<td>-0.06</td>
<td>-0.09*</td>
<td>-0.05</td>
<td>-0.06</td>
<td>-0.09*</td>
<td>-0.03</td>
<td>-0.06</td>
</tr>
<tr>
<td>8 Medical Experience</td>
<td>0.678</td>
<td>0.04</td>
<td>0.03</td>
<td>0.04</td>
<td>0.07</td>
<td>0.07</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>9 Leadership</td>
<td>0.615</td>
<td>-0.04</td>
<td>-0.03</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.01</td>
<td>-0.005</td>
<td>-0.04</td>
</tr>
<tr>
<td>10 Maturity</td>
<td>0.708</td>
<td>-0.03</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.03</td>
<td>-0.06</td>
<td>-0.09*</td>
<td>-0.03</td>
</tr>
<tr>
<td>11 Tests</td>
<td>0.637</td>
<td>-0.06</td>
<td>-0.09*</td>
<td>-0.05</td>
<td>-0.06</td>
<td>-0.09*</td>
<td>-0.03</td>
<td>-0.06</td>
</tr>
<tr>
<td>12 Service</td>
<td>0.730</td>
<td>0.04</td>
<td>0.03</td>
<td>0.04</td>
<td>0.07</td>
<td>0.07</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>13 Statements</td>
<td>0.593</td>
<td>0.04</td>
<td>0.03</td>
<td>0.04</td>
<td>0.07</td>
<td>0.07</td>
<td>0.02</td>
<td>0.01</td>
</tr>
</tbody>
</table>

**Total R^2**

<table>
<thead>
<tr>
<th>Score on Negative Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>IR</td>
</tr>
<tr>
<td>P</td>
</tr>
<tr>
<td>EV</td>
</tr>
<tr>
<td>XA</td>
</tr>
<tr>
<td>ML</td>
</tr>
<tr>
<td>YT</td>
</tr>
<tr>
<td>SW</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Score on Positive Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>IR</td>
</tr>
<tr>
<td>P</td>
</tr>
<tr>
<td>EV</td>
</tr>
<tr>
<td>XA</td>
</tr>
<tr>
<td>ML</td>
</tr>
<tr>
<td>YT</td>
</tr>
<tr>
<td>SW</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Score on the Overall Impression Theme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Impression</td>
</tr>
<tr>
<td>R^2</td>
</tr>
</tbody>
</table>

I. Interview; R. Recommendations; P. Personal traits; E. Military experience; V. Motivation; X. Extracurricular activities; A. Academic; M. Medical experience; L. Leadership; Y. Maturity; T. Test; S. Service; W. Personal Statement. *p < 0.05; **p < 0.01.

Table III. Students’ score on this theme was positively correlated with their medical school preclinical GPA ($r = 0.14$, CI = [0.07, 0.21], $p < 0.01$), medical school clinical GPA ($r = 0.13$, CI = [0.06, 0.20], $p < 0.01$), cumulative medical school GPA ($r = 0.15$, CI = [0.08, 0.22], $p < 0.01$), and USMLE Step 1 ($r = 0.10$, CI = [0.03, 0.17], $p < 0.05$) and Step 2 ($r = 0.11$, CI = [0.04, 0.18], $p < 0.01$).

To more clearly assess the strength of association between the seven factors and the outcomes, we conducted several multiple linear regressions and reported total model $R^2$ in Table III. Consistent with the results of the correlation analysis, the variances explained by the seven factors were small (<10%) for all the outcome measures. The strongest association existed between scores on negative comments and USMLE scores; 7% of the variance of both Step 1 and Step 2 scores was accounted for by scores on negative comments from tertiary reviewers.

**DISCUSSION**

The purpose of this study was to examine the predictive validity of tertiary reviewer comments. Our results suggest that although tertiary reviewer comments were significantly
associated with our selected outcome variables, which measured medical students’ performance in medical school and internship, the effects were quite small. We highlight the important findings below.

When we consider the negative comments of the tertiary reviewers, students who received more negative comments on Interview and Recommendations had worse evaluations from the program directors on professionalism during internship. This finding has important implications for graduate medical education as it suggests that these tertiary reviewer comments, especially negative ones, may add validity to predicting future performance. Those students who were judged more negatively by the tertiary reviewers on Academic and Extracurricular activities later showed lower medical school preclinical GPA, USMLE Step 1, and cumulative medical school GPA. The frequency of getting negative comments on Test and Maturity was significantly associated with all the outcome variables measuring performance in medical school, again supporting the predictive validity of tertiary reviewer comments. When we consider the positive comments of the tertiary reviewers, those students who received more positive comments on Test and Maturity continued to show better performance on standardized tests in medical school, revealed by USMLE Step 1 and Step 2. However, students who received more positive comments on their Motivation and Military experience performed worse on these two standardized tests.

There are several limitations of our study. First, the observations were from a single institution. Thus, we need to be careful when generalizing our results to other medical students. Accordingly, future work should attempt to replicate our findings with more heterogeneous samples and more diverse medical education settings. Second, our study assessed associations and not causation. Third, some aspects of the tertiary reviewers’ comments could be unique to USU, such as Military experience. Thus, not all of the findings are applicable to civilian medical schools. Nevertheless, this novel study was meant to investigate tertiary reviewer comments and relate these comments to the outcomes of medical students’ performance. We did this across the medical education continuum from medical school into internship, and our data were collected across multiple years, both of which we consider strengths of the present study. Fourth, not every student in the cohort (classes of 1993 to 1999) was included in the study. Finally, as the largest amount of disagreement between coders occurred with neutral comments, we subsequently dropped this category from future analysis, understanding that removing this category represents an important study limitation. This might have added bias to the analysis.

In conclusion, findings from the present study suggest that tertiary reviews are associated with students’ academic performance in medical school. However, these results also suggest that these associations are quite weak. The admissions process at USU and other medical schools is expensive and time-consuming. This study should lead medical school admissions committees to evaluate the various components that they use in their admissions process to see if any components can be dropped or improved upon to help predict performance of their graduates. Nonetheless, compared with frequency of receiving positive comments, frequency of receiving negative comments had stronger associations with medical school performance measures and therefore may be more indicative of future performance.

REFERENCES

Writing, Self-Reflection, and Medical School Performance: The Human Context of Health Care

CAPT Mark B. Stephens, MC USN*; Col Brian V. Reamy, USAF MC (Ret.)†; Denise Anderson, MS‡; Cara Olsen, DrPH§; Col Paul A. Hemmer, USAF MC‖; Steven J. Durning, MD, PhD**; Simon Auster, MD, JD††

ABSTRACT
Introduction: Finding ways to improve communication and self-reflection skills is an important element of medical education and continuing professional development. This study examines the relationship between self-reflection and educational outcomes. Methods: We correlate performance in a preclinical course that focuses on self-reflection as it relates to contextual elements of patient care (Human Context of Health Care), with educational measures such as overall grade point average, clinical clerkship scores, and Medical College Admission Test (MCAT) scores. Results: Student performance in Human Context of Health Care correlated with MCAT-Verbal scores, MCAT-writing sample scores, clerkship grades, and overall medical school grade point average ($R = 0.3, p < 0.001$). Conclusion: Writing and self-reflection skills are often neglected in undergraduate medical curricula. Our findings suggest that these skills are important and correlate with recognized long-term educational outcomes.

INTRODUCTION
Taking a medical history is the most common clinical skill in medicine. Most physicians perform somewhere between 150,000 and 300,000 clinical interviews during their careers. The ability to take a thorough medical history requires competency in interpersonal communication skills and is essential at all levels of medical training. To be successful in their introduction to clinical practice, medical students must be able to communicate effectively with their patients and with their colleagues. Interpersonal communication skills are core professional competencies promoted by both the Accreditation Council on Graduate Medical Education and the Liaison Committee on Medical Education. The ability to successfully communicate with patients and relate this information to peers in the context of a complete medical history requires students to personally master the fundamental medical art of narrative competence.

Narrative competence refers to “the ability to acknowledge, absorb, interpret and act on the stories and plights of others.” Personal mastery is the “personal motivation to continually learn how our actions impact our world.” Despite the fact that communication and self-reflection skills are recognized as important elements of undergraduate medical education, few published findings examine the relationship among communication skills, self-reflection skills, and educational outcomes. Based on our prior work, the present study was specifically designed to explore the hypothesis that student performance in a first-year medical school course that encourages self-reflective writing correlates with premedical school measures (performance on the Medical College Admissions Test [MCAT], admission interviews) and specific postcourse measures (summary grade point average [GPA] at the conclusion of medical school, U.S. Medical Licensing Examination [USMLE] scores, and grades for Family Medicine and Internal Medicine core third-year clerkships) where communication skills are routinely assessed.
METHODS

Institution

The Uniformed Services University of the Health Sciences (USU) is the only accredited federal medical school. Located in Bethesda, MD, USU has an average class size of 170 students. The first two (preclinical) years focus on the basic sciences with integrated coursework included to introduce the principles of clinical medicine. The third and fourth years are traditional clinical rotations. During the third year, students complete required core clerkships in surgery, internal medicine, family medicine, psychiatry, pediatrics, and obstetrics/gynecology.

The Human Context of Health Care (HCHC) is a required course for all first-year medical students at USU (http://www.usuhs.mil/fap/hchc.html). A total of 12 topics are addressed in the course during the academic year. The topics covered are contextual thinking, uncertainty, confidentiality, the doctor–patient relationship, medical error, sexuality, illness and disability, survivorship, alcohol, sexism, racism, and family violence. Students are assigned reading materials and attend a panel presentation during which physicians (or medical students) describe personal experiences relating to the topic-at-hand. Students then write an essay exploring their personal reactions to the panel presentation and assigned reading materials. In this essay, students are challenged to identify individual life experiences, beliefs, and values that led to their reactions to the presented material. Students analyze their reactions, identify their perspective on the issue being addressed, and articulate potential implications for future patient care. Students include a specific plan designed to minimize potential problems or conflicts in the clinical environment as part of their essay.

Student essays are read and evaluated by physician faculty from diverse medical specialties. Each essay is evaluated using a 4-point scale (−1 poor; 0 insufficient analysis; +1 thoughtful; +2 outstanding). Faculty members are given a scoring sheet with writing samples to guide grading of student essays. Students then meet with faculty preceptors the following week in groups of 8 to 10 students. The small group discussions focus on issues raised by the articles, assigned readings, or the panel presentations. Students are graded on a similar 4-point scale based on their contributions to the small group sessions. Course letter grades (A through D) represent a combination of scores for their essays and small group sessions.

Measurements

Following Institutional Review Board approval, data were abstracted for students enrolled at USU from 2003 through 2008. Data fields include MCAT scores (MCAT-Physical Science [PS]; MCAT-Biological Science [BS]; MCAT-Verbal Reasoning [VR]; MCAT-Writing Sample [WS]), summary USU GPA, admission interview scores for the student’s verbal and communication skills (ranked on a 5-point scale), core third-year-clerkship grades for Family Medicine and Internal Medicine, and the final course grade for the HCHC course. Other measures included USMLE Step 1 and Step 2 scores and undergraduate science GPA.

MCAT scores were obtained from the USU Office of Admissions via access to American Medical College Application Service reports. For students who took the MCAT examination more than once, MCAT subsection scores were pooled and the average score was used. Clerkship grades in Internal Medicine and Family Medicine were specifically chosen as intermediate-term outcomes because each of these core clerkships emphasizes patient communication as part of the overall course evaluation. The Family Medicine clerkship is a 6-week rotation with inpatient, outpatient, maternity care, and emergency care experience. The overall clerkship grade is determined by a student’s clinical performance (65%), family case study (5%), National Board of Medical Examiners (NBME) subject examination score (20%), and performance on an objective structured clinical exam that includes a specific case on patient communication (10%). The Internal Medicine clerkship is 12 weeks in length with 6 weeks of inpatient medicine and 6 weeks of ambulatory care. The final internal medicine clerkship grade is determined by the student’s clinical performance (70%) and performance on three end-of-clerkship examinations (30%). USMLE scores were obtained from the USU Office of Student Affairs as was overall student GPA.

Data Analysis

Data were transposed into SPSS version 16 (SPSS; Chicago, Illinois) for statistical analysis. Basic descriptive statistics were used for baseline comparisons. Pearson’s product-moment correlation coefficients (r) were used to describe the strength and direction of the association between pairs of variables. Exploratory factor analysis with maximum likelihood extraction was used to identify strongly correlated sets of educational variables and identify potential underlying latent constructs.

RESULTS

At least partial course and exam data were available for 994 students in the classes of 2003 through 2008 (96%). Table I presents general academic outcomes. Table II presents the correlations between student performance in the HCHC and educational outcomes including USMLE scores, undergraduate science GPA, core clerkship scores in Family Medicine and Internal Medicine, and student interview scores.

Overall, performance in HCHC correlated weakly to moderately with MCAT-VR, MCAT-WS, Internal Medicine clerkship grades, Internal Medicine and Family Medicine subject exam scores, USMLE Step 1 and Step 2 scores, and overall USU GPA (Table II). HCHC grades did not correlate with MCAT-PS, MCAT-BS, undergraduate science GPA, or the final Family Medicine clerkship grade. HCHC grades had a weakly negative correlation with USU admission interview scores. The strongest positive correlation was between HCHC grades and overall USU GPA (r = 0.3; p < 0.001). This is
notable as HCHC represents only 3% of the student’s calculated GPA in our analysis. When the HCHC grade was excluded from the overall USU GPA calculation, the correlation decreased slightly ($r = 0.23$, $p < 0.001$).

An exploratory factor analysis based on complete data available for 453 students (45% of the sample) identified four factors with eigenvalues greater than one. Both orthogonal (varimax) and oblique (promax) rotations resulted in the same number of factors and similar factor loadings. Findings from the orthogonal rotation are presented in Table III. HCHC scores grouped with measures of communication including admission interview communication scores and MCAT-WS (Group 1). Additional variable groupings included within medical school measures: USMLE scores, overall USU GPA, and subject exam scores (Group 2); medical school admission science subject scores; MCAT-PS with MCAT-BS (Group 3); and medical school admission measures of reasoning: MCAT-VR with undergraduate science GPA (Group 4).

**DISCUSSION**

In modern medicine, the “science of curing” has, at times, taken precedence over the “art of healing.”11 In addition to technical and scientific knowledge, medical students must also demonstrate effective interpersonal communication skills. These skills include the ability to hear a patient’s story, interpret what is being told, respond to and report information in a manner that is clinically appropriate. By necessity, healing involves interpersonal communication and self-reflection. Experienced physicians regularly use emotion, perception, and experience to aid in both diagnosis and therapy.12 Each individual physician carries a system of biases, fears, and attitudes that can either help or hinder their ability to provide patient care.13,14 By reflecting on these
emotions and experiences consciously, communication with patients can be improved.11

To date, relatively few studies have addressed the role of self-reflection in undergraduate medical education. Sobral15 showed that exercises in self-reflection during a third-year clinical clerkship did not significantly impact self-reports of reflective ability. Medical students with prior health care experience and students interested in general practice have been shown to score higher on case vignettes requiring self-reflection compared with their peers.16

Our results suggest that student performance in a course that emphasizes self-reflection is associated with end-of-medical school GPA and weakly with performance on sections the USMLE examination. The association with grades from a first-year course emphasizing contextual thinking with medical school GPA and USMLE scores is a new and important finding. Using our model, nearly 10% of the variance in USU GPA was explained by a student’s performance in the HCHC course. We hypothesize that the unique combination of experience, self-reflection, and multimodal communication (written and oral small group discussions) blends elements of cognition and communication that are associated with success in medical school (USU GPA) and national board examination scores. Interestingly, we found a negative correlation between student performance in HCHC and admission interview scores. This is of limited concern, however, because traditional medical school interviews have long been associated with inherent interviewer bias17 and has limited value for predicting medical school performance or patient satisfaction.18

The results of our factor analysis groups HCHC scores with admissions interviews and with the MCAT-WS. This provides a measure of internal validity to our results since these factors specifically relate to verbal and written interpersonal communication skills. The clustering of these noncognitive factors as associated with GPA and USMLE scores supports the inclusion of curricular elements that promote verbal and written communication skills during medical education.

Although course work in medical interviewing is routine at most schools, mandatory evaluation of writing skills is not.19 Writing, in the form of clinical documentation through the medical record, is a primary form of interpersonal communication between professional colleagues in the context of patient care. Our results suggest that not only are writing skills important, they are also associated with broader educational outcomes. This implies that students who struggle in courses like HCHC may be candidates for early educational interventions to improve their communication skills. Early identification of students requiring help with interpersonal communication skills could mitigate future difficulties on clerkship rotations or in clinical practice.

Our study has several important limitations. First, our data are from a single institution. How results from a course like HCHC at USU extend to other universities is not known. Similarly, few (if any) medical schools have a course similar to HCHC. Although schools generally recognize the importance of communication skills in undergraduate medical education, inconsistencies in courses emphasizing self-reflection and writing skills across institutions limit generalizability. Additionally, we examined associations between performance in HCHC and other educational outcomes. This does not equate to causality. The inclusion of other intermediate outcomes, such as scores on other core clerkships or other preclinical courses, would further strengthen our conclusions.

In summary, our findings suggest that writing and self-reflection skills are important elements of undergraduate medical education and that these elements can potentially predict long-term educational success. Medical students should be encouraged to develop self-reflective communication skills across written and verbal domains. Medical students should be encouraged to develop self-awareness and reflective practice skills11 to develop their own sense of professional identity.20 At USU, these domains are incorporated in the design of the HCHC course. Our hope is that skills learned in this course will imbue an inherent sense of professionalism and make students more reflective and contextually competent physicians in the long run.

REFERENCES


MILITARY MEDICINE, Vol. 177, September Supplement 2012 29
Development and Initial Validation of a Survey to Assess Students’ Self-Efficacy in Medical School

CDR Anthony R. Artino Jr., MSC USN; Ting Dong, PhD; LTC Kent J. DeZee, MC USA; COL William R. Gilliland, MC USA (Ret.); Donna M. Waechter, PhD; David F. Cruess, PhD; Steven J. Durning, MD, PhD

ABSTRACT  Self-efficacy is a personal belief in one’s capability to successfully execute the behaviors necessary to attain designated types of performances. Sometimes described as task-specific self-confidence, self-efficacy is a key component in many contemporary theories of motivation and learning. The purpose of this study was to develop a survey for measuring students’ medical skills self-efficacy and to collect reliability and validity evidence for the instrument. A secondary purpose was to explore differences in students’ self-efficacy from year 1 of medical school to year 4. We created 19 survey items based on the 6 core competencies of the Accreditation Council for Graduate Medical Education, and we collected data from 304 medical students. Results from an exploratory factor analysis suggested three interpretable factors: patient care self-efficacy (eight items, Cronbach’s $\alpha = 0.92$), interpersonal skills self-efficacy (three items, Cronbach’s $\alpha = 0.76$), and evidence-based medicine self-efficacy (three items, Cronbach’s $\alpha = 0.79$). We then compared students’ self-efficacy at different stages of training using a one-way multivariate analysis of variance. Consistent with our expectations, we found several statistically significant differences, suggesting students’ self-efficacy increased considerably from year 1 of medical school to year 4, $F(9, 725) = 30.58, p < 0.001$, Wilks’ $\lambda = 0.46$. Using this survey, medical educators and researchers have a psychometrically sound tool for measuring students’ medical skills self-efficacy during undergraduate medical education. Practical implications and future directions are discussed.

INTRODUCTION
Self-efficacy is a personal belief in one’s capability to successfully execute the behaviors necessary to attain designated types of performances. According to Bandura, self-efficacy beliefs lie at the core of human functioning. It is not enough for individuals to possess the requisite knowledge and skills to perform a task; they also must have the conviction that they can successfully perform the required behavior(s) under typical and, importantly, under difficult circumstances. Self-efficacy differs from general self-confidence in that “self-efficacy is context-specific rather than a stable personality trait, and it is therefore thought to have a direct effect on performance in specific contexts.” In the realm of medical education, several scholars have recently discussed the importance of self-efficacy beliefs for the development of medical knowledge and skills proficiency.

The purpose of this study was to develop a quantitative self-report measure of medical skills self-efficacy and to collect reliability and validity evidence for the instrument. A secondary purpose was to explore potential differences in students’ medical skills self-efficacy from year 1 of medical school (MS-1) to year 4 (MS-4). Investigations of this kind are particularly important because students’ academic self-efficacy has been shown to predict cognitive engagement and academic achievement in a variety of educational settings. Moreover, recent research in medical education suggests that medical students’ self-efficacy beliefs have moderately strong links to their current and future academic success. Given that survey research in medical education hinges on the legitimacy of the measurement tools themselves, studying such instruments further informs the field. As such, the ultimate goal of the present study was to produce a psychometrically sound instrument that medical educators and researchers can use to make valid inferences.

THEORETICAL FRAMEWORK
Bandura defined self-efficacy as, “people’s judgments of their capabilities to organize and execute courses of action required to attain designated types of performances.” Self-efficacy is a belief about one’s capability. Thus, self-efficacy beliefs do not always match one’s actual ability in a specific domain. In fact, research findings have suggested that most individuals actually overestimate their capabilities. Bandura made the case, however, that the most useful efficacy judgments are those that slightly exceed one’s actual capabilities as this overestimate can actually increase effort and persistence during challenging times. Although feeling slightly overconfident may be adaptive in most educational contexts, in the clinical setting, it is clear that greatly overestimated self-efficacy beliefs could result in an unsafe environment for patients (e.g., the medical student who overestimates her ability to independently manage a complex clinical case). On the other hand, a medical student who has underestimated his self-efficacy may possess low confidence in his ability to draw blood and, because of this, may be unmotivated to practice the skill or persevere in the face of difficulties experienced during a practice session.
Bandura hypothesized that self-efficacy beliefs influence an individual’s choice of activities, effort, and persistence. People who have low self-efficacy for accomplishing a specific task may avoid it, whereas those who believe they are capable are more likely to become engaged. Moreover, individuals who possess higher levels of self-efficacy are hypothesized to put forth more effort and persist longer in the face of difficulties than those who are uncertain of their capabilities. The tendency for efficacious people to put forth more effort and persist longer is critically important because most personal success, and particularly success in medicine, require persistent effort and prolonged, deliberate practice. As such, low self-efficacy becomes a self-limiting process. In order to succeed, then, individuals need resilient self-efficacy beliefs to meet the unavoidable obstacles of life.

In medical education research, motivational constructs such as self-efficacy have received less emphasis than more traditional measures of academic success, such as scores on standardized tests and observations of clinical performance. However, recent work suggests that self-efficacy beliefs have moderately strong links to students’ current and future performance in medical school. Considering this recent work, our primary purpose was to develop a survey for measuring medical skills self-efficacy. In particular, we sought to create a survey that could be administered to students during any phase of their undergraduate medical training such that growth (or decline) in their medical skills self-efficacy could be assessed.

METHOD

Study Context

This study was part of the larger Long-Term Career Outcome Study conducted at the F. Edward Hebert School of Medicine, Uniformed Services University of the Health Sciences (USU). As the United States’ only federal medical school, USU matriculates approximately 170 medical students annually and, at the time of this study, offered a traditional 4-year curriculum: 2 years of basic science courses followed by 2 years of clinical rotations.

Item Development and Content Validation

Because we ultimately sought to compare students’ medical skills self-efficacy at different phases of medical education, we chose to develop initial survey items based on the six core competencies of the Accreditation Council for Graduate Medical Education (ACGME): patient care, medical and population health knowledge, interpersonal and communication skills, practice-based learning and improvement, professionalism, and systems-based practice.

Following initial item development, we recruited five medical education experts to participate in a content validation in accordance with the guidelines outlined by McKenzie and colleagues. We provided each content expert with the draft items and asked them to consider the extent to which each item addressed the construct of interest (i.e., medical skills self-efficacy). We also asked them to suggest additions to and/or subtractions from the list of draft items and to recommend changes for any items they felt were unclear.

Results from the content validation yielded a 19-item instrument designed to measure students’ medical skills self-efficacy in relation to the attainment of the ACGME’s six core competencies (Table I). All items employed a 5-point, Likert-type response scale (not at all confident, slightly confident, moderately confident, quite confident, and extremely confident).

Participants and Procedures

We drew our sample from all USU students enrolled in the medical school during the 2010–2011 academic year (N = 673). In May 2011, we invited these students, via e-mail, to complete an online survey that assessed their beliefs, emotions, and behaviors in relation to their experiences in medical school, and our 19-item self-efficacy scale was included as part of this larger survey. We gave students approximately 2 weeks to complete the survey, and we sent two follow-up e-mail reminders over the next 4 weeks. Participation in the survey was voluntary, and ethical approval was obtained from the USU Institutional Review Board.

TABLE I. The 19 Items Created From the Item Development and Content Validation Procedures

<table>
<thead>
<tr>
<th>Item (SE)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SE-1</td>
<td>Apply knowledge of normal function to each of the major organ systems?</td>
</tr>
<tr>
<td>SE-2</td>
<td>Effectively manage the uncertainty associated with patient care, such as when the patient has multiple treatment options, each with its own risks and benefits?</td>
</tr>
<tr>
<td>SE-3</td>
<td>Apply knowledge of epidemiology of common diseases, such as heart disease, to reduce disease incidence?</td>
</tr>
<tr>
<td>SE-4</td>
<td>Use effective listening skills when interacting with a patient?</td>
</tr>
<tr>
<td>SE-5</td>
<td>Demonstrate caring when counseling a patient?</td>
</tr>
<tr>
<td>SE-6</td>
<td>Accurately gather essential information from a patient?</td>
</tr>
<tr>
<td>SE-7</td>
<td>Perform a thorough physical exam?</td>
</tr>
<tr>
<td>SE-8</td>
<td>Develop an appropriate differential diagnosis?</td>
</tr>
<tr>
<td>SE-9</td>
<td>Generate a patient-specific treatment plan?</td>
</tr>
<tr>
<td>SE-10</td>
<td>Use information technology to support patient-care decisions?</td>
</tr>
<tr>
<td>SE-11</td>
<td>Work effectively with other health care professionals to provide high-quality patient care?</td>
</tr>
<tr>
<td>SE-12</td>
<td>Improve clinical practice using a systematic approach?</td>
</tr>
<tr>
<td>SE-13</td>
<td>Evaluate evidence from scientific studies relevant to your patients’ health problems?</td>
</tr>
<tr>
<td>SE-14</td>
<td>Stay abreast of relevant scientific advances by reading peer-reviewed medical journals?</td>
</tr>
<tr>
<td>SE-15</td>
<td>Demonstrate sensitivity to patients’ cultural differences?</td>
</tr>
<tr>
<td>SE-16</td>
<td>Balance professional responsibilities with personal responsibilities?</td>
</tr>
<tr>
<td>SE-17</td>
<td>Discuss methods of controlling health care costs?</td>
</tr>
<tr>
<td>SE-18</td>
<td>Practice cost-effective health care delivery that does not compromise quality of care?</td>
</tr>
<tr>
<td>SE-19</td>
<td>Apply high-quality health care in deployed military environments?</td>
</tr>
</tbody>
</table>
Statistical Analyses

Before analysis, we screened the data for accuracy and missing values, and we checked each survey item response pattern for normality. Next, to investigate the representativeness of our sample, we used multivariate analysis of variance (MANOVA) to compare students who completed the survey and those who did not on Medical College Admissions Test (MCAT) scores and cumulative medical school grade point average (GPA). We then conducted an exploratory factor analysis (EFA) to examine the factorial validity of the survey. Next, we subjected each of the subscales identified in the EFA to an internal consistency reliability analysis and computed a mean score for the items associated with a particular subscale (i.e., the variables were unweighted composite scores); we also calculated descriptive statistics for the total sample and the 4 year groups. Finally, we performed a MANOVA to compare students’ self-efficacy scores across the 4 years of medical school, treating each year as a group. All analyses were completed using SPSS 20.0 (IBM, New York, NY).

RESULTS

Of the 673 medical students enrolled during the 2010–2011 academic year, we collected surveys from 304 students (45%). Of those who completed the survey, there were 87 (29%) MS-1 students, 88 (29%) MS-2 students, 64 (21%) MS-3 students, and 65 (21%) MS-4 students. The sample included 223 men (73%) and 81 women, which is representative of the medical student population at USU. Results from the MANOVA comparing those who completed the survey and those who did not revealed no statistically significant differences between the two groups on MCAT scores or cumulative medical school GPA, \( F(2, 633) = 0.72, p = 0.49 \).

Exploratory Factor Analysis

We conducted a principal axis factor analysis with oblique rotation (Oblimin; \( \delta = 0 \)) on the 19 self-efficacy items.\(^{14} \) Oblique rotation methods allow for factors to be correlated, and we assumed that any underlying factors would be related. Evaluation of the correlation matrix indicated that it was factorable: Kaiser–Meyer–Olkin Measure of Sampling Adequacy was 0.91, which is “marvelous” (>0.90) according to Kaiser’s criteria.\(^{15} \) Bartlett’s Test of Sphericity (\( \chi^2 = 3526.70, df = 171, p < 0.001 \)) was significant, indicating that the correlation matrix was not an identity matrix, and all measures of sampling adequacy were deemed sufficient (i.e., >0.60).\(^{15} \)

We determined the number of factors to extract using several criteria, including parallel analysis, examination of the resulting scree plot, and eigenvalues greater than 1.0.\(^{16} \) All three criteria suggested a three-factor solution, with the three factors accounting for 63.2% of the total variance in the items. Inspection of the table of communalities revealed all but one item had high extracted communalities (i.e., >0.40; Table II), indicating that much of the common variance in the items can be explained by the three extracted factors.\(^{15} \) The one exception, SE-16, had a fairly low extracted communality (0.32).

We used several additional rules to determine the number of factors and individual items to retain in the final solution: (a) factors needed to contain at least three items, (b) the absolute value of all factor pattern coefficients needed to be >0.50 on at least one factor, and (c) items with factor pattern coefficients (absolute value) ≥0.30 on more than one factor were dropped (recommendations in Pett et al.\(^{15} \)). The factor pattern and structure coefficients from the principal axis factor analysis are displayed in Table II.

The first factor (extraction eigenvalue = 8.48) included eight items: SE-9, SE-8, SE-2, SE-10, SE-1, SE-12, SE-19, and SE-3. Although item SE-7 loaded highly on Factor 1, it also loaded on Factor 2 and was therefore dropped from the final solution. The second factor (extraction eigenvalue = 1.96) included three items: SE-5, SE-4, and SE-15. Although item SE-6 loaded highly on Factor 2, it also loaded on Factor 1 and was therefore dropped from the final solution. The third factor (extraction eigenvalue = 1.58) also included three items: SE-14, SE-18, and SE-13. The correlation between the first factor and the second factor was 0.36, the correlation between the first factor and the third factor was 0.45, and the correlation between the second factor and the third factor was 0.31.

Factor Labels, Reliability Analysis, and Subscale Creation

Based on our EFA results, we named each of the three factors retained: (a) Factor 1 was labeled patient care self-efficacy, (b) Factor 2 was labeled interpersonal skills self-efficacy, and (c) Factor 3 was labeled evidence-based medicine self-efficacy.

Next, we conducted a reliability analysis on the scores of the items retained in each of the three self-efficacy subscales (Table III). The Cronbach’s \( \alpha \) values were as follows: patient care self-efficacy (eight items) = 0.92, interpersonal skills self-efficacy (three items) = 0.76, and evidence-based medicine self-efficacy (three items) = 0.79. These internal consistency reliabilities were all considered acceptable (i.e., >0.75; see guidelines in Gable and Wolfe\(^{17} \)). Furthermore, inspection of the interitem correlation matrix revealed little redundancy in the items within any of the three subscales (i.e., all correlations were ≤0.70).

Next, we created composite variables to be used in subsequent analyses. These variables were created by computing a mean score for the items associated with a particular subscale. Table IV presents the means and standard deviations of the three self-efficacy variables for the total sample and the 4 year groups (MS-1 to MS-4).

Differences in Self-Efficacy Across Medical School

We performed a MANOVA to compare students’ self-efficacy scores on the three subscales from MS-1 to MS-4, treating each year as a group. The overall comparison indicated that students in different year groups had significantly different scores on the three self-efficacy variables, \( F(9, 725) = 30.58, p < 0.001 \), Wilks’ \( \lambda = 0.46 \). As the overall \( F \)-test was statistically significant, we moved on to additional univariate analysis.
### TABLE II. Results From the EFA With Oblique Rotation (Oblimin; $\delta = 0$) on Self-Efficacy ($N = 304$)

<table>
<thead>
<tr>
<th>Item</th>
<th>Communality</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>SE-9 Generate a patient-specific treatment plan</td>
<td>0.775</td>
<td>0.93 (0.87)</td>
<td>-0.13 (0.19)</td>
<td>0.00 (0.36)</td>
</tr>
<tr>
<td>SE-8 Develop an appropriate differential diagnosis</td>
<td>0.759</td>
<td>0.92 (0.87)</td>
<td>0.00 (0.30)</td>
<td>0.00 (0.32)</td>
</tr>
<tr>
<td>SE-2 Effectively manage the uncertainty associated with patient care, such as when the patient has multiple treatment options, each with its own risks and benefits</td>
<td>0.600</td>
<td>0.75 (0.78)</td>
<td>0.00 (0.26)</td>
<td>0.00 (0.41)</td>
</tr>
<tr>
<td>SE-10 Use information technology to support patient-care decisions</td>
<td>0.587</td>
<td>0.66 (0.72)</td>
<td>0.00 (0.32)</td>
<td>0.00 (0.41)</td>
</tr>
<tr>
<td>SE-1 Apply knowledge of normal function to each of the major organ systems</td>
<td>0.517</td>
<td>0.64 (0.71)</td>
<td>0.13 (0.37)</td>
<td>-0.00 (0.38)</td>
</tr>
<tr>
<td>SE-12 Improve clinical practice using a systematic approach</td>
<td>0.675</td>
<td>0.61 (0.76)</td>
<td>0.00 (0.37)</td>
<td>0.27 (0.57)</td>
</tr>
<tr>
<td>SE-19 Apply high-quality health care in deployed military environments</td>
<td>0.533</td>
<td>0.57 (0.67)</td>
<td>0.00 (0.24)</td>
<td>0.26 (0.51)</td>
</tr>
<tr>
<td>SE-3 Apply knowledge of epidemiology of common diseases, such as heart disease, to reduce disease incidence</td>
<td>0.503</td>
<td>0.55 (0.67)</td>
<td>0.00 (0.27)</td>
<td>0.26 (0.50)</td>
</tr>
<tr>
<td>SE-7 Perform a thorough physical exam</td>
<td>0.511</td>
<td>0.55 (0.62)</td>
<td>0.37 (0.53)</td>
<td>-0.14 (0.22)</td>
</tr>
<tr>
<td>SE-11 Work effectively with other health care professionals to provide high-quality patient care</td>
<td>0.599</td>
<td>0.48 (0.65)</td>
<td>0.24 (0.47)</td>
<td>0.19 (0.47)</td>
</tr>
<tr>
<td>SE-5 Demonstrate caring when counseling a patient</td>
<td>0.608</td>
<td>0.00 (0.28)</td>
<td>0.82 (0.82)</td>
<td>0.00 (0.28)</td>
</tr>
<tr>
<td>SE-4 Use effective listening skills when interacting with a patient</td>
<td>0.611</td>
<td>0.13 (0.37)</td>
<td>0.77 (0.79)</td>
<td>0.00 (0.21)</td>
</tr>
<tr>
<td>SE-6 Accurately gather essential information from a patient</td>
<td>0.536</td>
<td>0.43 (0.56)</td>
<td>0.54 (0.65)</td>
<td>-0.14 (0.22)</td>
</tr>
<tr>
<td>SE-15 Demonstrate sensitivity to patients’ cultural differences</td>
<td>0.402</td>
<td>-0.19 (0.18)</td>
<td>0.50 (0.56)</td>
<td>-0.14 (0.49)</td>
</tr>
<tr>
<td>SE-16 Balance professional responsibilities with personal responsibilities</td>
<td>0.321</td>
<td>-0.19 (0.28)</td>
<td>0.50 (0.45)</td>
<td>0.42 (0.39)</td>
</tr>
<tr>
<td>SE-17 Discuss methods of controlling health care costs</td>
<td>0.597</td>
<td>0.00 (0.45)</td>
<td>0.35 (0.23)</td>
<td>-0.27 (0.76)</td>
</tr>
<tr>
<td>SE-14 Stay abreast of relevant scientific advances by reading peer-reviewed medical journals</td>
<td>0.589</td>
<td>0.15 (0.43)</td>
<td>0.00 (0.35)</td>
<td>0.71 (0.74)</td>
</tr>
<tr>
<td>SE-18 Practice cost-effective health care delivery that does not compromise quality of care</td>
<td>0.668</td>
<td>0.00 (0.57)</td>
<td>0.12 (0.22)</td>
<td>0.66 (0.73)</td>
</tr>
<tr>
<td>SE-13 Evaluate evidence from scientific studies relevant to your patients’ health problems</td>
<td>0.640</td>
<td>0.29 (0.57)</td>
<td>0.00 (0.36)</td>
<td>0.53 (0.70)</td>
</tr>
</tbody>
</table>

Pattern coefficients are presented first, followed by structure coefficients in parentheses. Entries in bold indicate pattern coefficients (absolute values) >0.50 on at least one factor and pattern coefficients (absolute values) >0.30 on only one factor.

### TABLE III. The Items Remaining in Each of the Three Self-Efficacy Subscales

<table>
<thead>
<tr>
<th>Item</th>
<th>Patient care self-efficacy</th>
<th>Interpersonal skills self-efficacy</th>
<th>Evidence-based medicine self-efficacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>SE-1 Apply knowledge of normal function to each of the major organ systems?</td>
<td>(Cronbach’s $\alpha = 0.76$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE-2 Effectively manages the uncertainty associated with patient care, such as when the patient has multiple treatment options, each with its own risks and benefits?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE-3 Apply knowledge of epidemiology of common diseases, such as heart disease, to reduce disease incidence?</td>
<td></td>
<td></td>
<td>(Cronbach’s $\alpha = 0.92$)</td>
</tr>
<tr>
<td>SE-8 Develop an appropriate differential diagnosis?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE-9 Generate a patient-specific treatment plan?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE-10 Use information technology to support patient-care decisions?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE-12 Improve clinical practice using a systematic approach?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE-19 Apply high-quality health care in deployed military environments?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All items were measured on a 5-point, Likert-type confidence scale.
Analyses. Tests of between-subjects effects indicated that class year group was related to patient care self-efficacy, $F(3, 300) = 87.65, p < 0.001$, and evidence-based medicine self-efficacy, $F(3, 300) = 9.69, p < 0.001$. There was no significant difference on interpersonal skills self-efficacy between the year groups. We then conducted Bonferroni multiple comparisons to closely examine the differences between any two groups. As indicated in Table V and Figure 1, students scored

<table>
<thead>
<tr>
<th>Variable</th>
<th>Overall ($N = 304$)</th>
<th>MS-1 ($n = 87$)</th>
<th>MS-2 ($n = 88$)</th>
<th>MS-3 ($n = 64$)</th>
<th>MS-4 ($n = 65$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient Care Self-Efficacy</td>
<td>2.99 (0.86)</td>
<td>2.18 (0.74)</td>
<td>2.97 (0.65)</td>
<td>3.36 (0.61)</td>
<td>3.75 (0.53)</td>
</tr>
<tr>
<td>Interpersonal Skills Self-Efficacy</td>
<td>4.13 (0.65)</td>
<td>4.07 (0.66)</td>
<td>4.02 (0.70)</td>
<td>4.26 (0.58)</td>
<td>4.23 (0.61)</td>
</tr>
<tr>
<td>Evidence-Based Medicine Self-Efficacy</td>
<td>3.02 (0.87)</td>
<td>2.75 (0.88)</td>
<td>2.89 (0.85)</td>
<td>3.20 (0.85)</td>
<td>3.40 (0.74)</td>
</tr>
</tbody>
</table>

All self-efficacy variables were measured on a 5-point, Likert-type confidence scale.

**TABLE V.** Bonferroni Multiple Comparisons for the Three Self-Efficacy Subscales

<table>
<thead>
<tr>
<th>(I) Factor</th>
<th>(J) Factor</th>
<th>Mean Difference (I–J)</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Patient Care</td>
<td>Interpersonal Skills</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lower</td>
<td>Upper</td>
</tr>
<tr>
<td>MS-1</td>
<td>MS-2</td>
<td>-0.76***</td>
<td>0.04</td>
</tr>
<tr>
<td>MS-2</td>
<td>MS-1</td>
<td>0.76***</td>
<td>-0.04</td>
</tr>
<tr>
<td>MS-3</td>
<td>MS-1</td>
<td>-0.46***</td>
<td>-0.19</td>
</tr>
<tr>
<td>MS-3</td>
<td>MS-4</td>
<td>-0.80***</td>
<td>-0.16</td>
</tr>
<tr>
<td>MS-4</td>
<td>MS-1</td>
<td>1.22***</td>
<td>0.15</td>
</tr>
<tr>
<td>MS-4</td>
<td>MS-2</td>
<td>0.46***</td>
<td>0.19</td>
</tr>
<tr>
<td>MS-4</td>
<td>MS-3</td>
<td>-0.35*</td>
<td>0.02</td>
</tr>
<tr>
<td>MS-4</td>
<td>MS-1</td>
<td>1.56***</td>
<td>0.13</td>
</tr>
<tr>
<td>MS-3</td>
<td>MS-2</td>
<td>0.80***</td>
<td>0.16</td>
</tr>
<tr>
<td>MS-3</td>
<td>MS-1</td>
<td>0.35*</td>
<td>-0.02</td>
</tr>
</tbody>
</table>

All self-efficacy variables were measured on a 5-point, Likert-type confidence scale. *$p < 0.05$*, **$p < 0.01$**, ***$p < 0.001$***.

**FIGURE 1.** Differences in students’ medical skills self-efficacy across the 4 year groups.
Our goals in this study were to develop a survey for predictor of cognitive engagement and academic achievement. Instead, medical students likely need both “the skill and the skill” to successfully function in dynamic clinical contexts. Therefore, Bandura and others have suggested that teachers should consider employing instructional practices that not only develop competence but also promote the growth of the necessary accompanying self-efficacy. In medical education, more research is clearly needed to explore the veracity of such a pedagogical approach.

Findings from this study also indicate that students report consistently high interpersonal skills self-efficacy (>4.0 on a 5-point scale) across all 4 years of medical school. Thus, it seems that students in our sample are quite confident in their ability to use effective interpersonal practices with patients. Nonetheless, an important issue is the extent to which these self-efficacy beliefs match reality. That is, are students accurately assessing their own interpersonal skills capabilities? As discussed previously, findings from other educational contexts suggest that students often overestimate their own capabilities. And although a slight overestimation may be adaptive—helping students persist in the face of difficulties—more accurate estimates of capability are also pedagogically important because grossly inaccurate estimates “can work against the benefits of high self-efficacy.” As such, a logical next step in this line of research is to compare students’ self-efficacy beliefs to their actual capabilities. Using this information, we believe medical educators will be in a better position to help students more accurately estimate their own self-efficacy beliefs and potentially improve their clinical performance.

Our study had several limitations, including the single-institution, cross-sectional nature of the study design. Because we did not employ a longitudinal design, care must be taken not to overinterpret our findings, particularly with respect to the differences we observed across the 4 years of medical school. In future work, we plan to include longitudinal collection of these same data. A second limitation is the suboptimal response rate we obtained. Although our comparisons of study participants versus nonparticipants revealed no differences on two performance measures (MCAT scores and cumulative medical school GPA), we cannot rule out the possibility of response bias. Finally, like all self-reports, our survey has reliability and validity limitations, including, for example, social desirably bias. That said, our EFA and reliability analyses suggest that our survey instrument has reasonable psychometric properties.

Although we feel this study is an important first step to understanding changes in self-efficacy across medical school, we also propose that future validation of the instrument include replication in other medical school populations. Additionally, future research should examine other forms of validity evidence, such as convergent, discriminant, and predictive validity, which could ultimately improve the functioning of our self-efficacy survey instrument.

Despite some limitations, results from this study have produced a survey with evidence of reliability and validity in the context of our medical school. Using this survey, medical
educators and researchers have another tool for measuring students’ medical skills self-efficacy. Bandura’s emphasis that to achieve predictive power, measures of self-efficacy should be tailored to specific domains of functioning. Therefore, measuring medical students’ beliefs about global capabilities, such as “clinical skills,” likely has little predictive value. Instead, to properly measure students’ self-efficacy beliefs, assessment tools must clearly define the domain of functioning being assessed. We believe the survey development approach used here—that is, creating items directly related to the ACGME’s six core competencies—may be an effective method for developing survey scales to assess self-efficacy; it may also be an effective approach to measuring other “noncognitive” constructs, such as student interest and anxiety. As illustrated in the present study, by linking students’ confidence beliefs to specific core competencies, we are able to assess if and how these beliefs change across the medical education continuum. From an educational perspective, we feel it is important to assess, and ultimately attempt to positively influence, these self-efficacy trajectories as we strive to develop physicians who persist through difficulties and expend effort to actively update their knowledge and skills throughout their clinical practice.

REFERENCES
Instructional Authenticity and Clinical Reasoning in Undergraduate Medical Education: A 2-Year, Prospective, Randomized Trial

Steven J. Durning, MD, PhD*; Ting Dong, PhD*; CDR Anthony R. Artino Jr., MSC USN*; Maj Jeffery LaRochelle, USAF MC*; COL Louis N. Pangaro, MC USA (Ret.)*; Cees van der Vleuten, PhD†; Lambert Schuwirth, MD, PhD†

ABSTRACT  Objective: The purpose of the study was to investigate the relationship between authenticity (how close to actual practice) of preclerkship instructional formats and preclerkship and clerkship outcome measures. A secondary purpose was to investigate the effect of student’s small-group assignment during preclerkship period on instructional formats and preclerkship and clerkship outcome measurements. Methods: A prospective, randomized, crossover study was carried out with preclerkship students taking a clinical reasoning course. Students were randomized to small groups and exposed to three formats of differing instructional authenticity across three subject areas. Three student cohorts were taught using one instructional format per subject area with each cohort receiving a different instructional format for each of the three areas. Outcome measures at the end of the preclerkship year and the clerkship year were selected to determine the effect of each instructional format. Hierarchical linear modeling was performed to assess impact of format on outcomes and to assess potential group effect on outcomes. Results: Increasingly authentic instructional formats did not significantly improve clinical reasoning performance. Small-group assignment did not significantly contribute to the outcomes providing evidence that teacher training was successful. Conclusions: Increasing authenticity of instructional formats does not appear to significantly improve clinical reasoning performance. Faculty can be successfully trained to teach consistently in a clinical reasoning course. Medical educators should balance increasing authenticity with factors such as cognitive load and learner experience as well as engaging in frame-of-reference training to minimize group effects with designing new instructional formats.

BACKGROUND

Expertise in clinical reasoning is essential for a physician’s practice. It entails arriving at a diagnosis and deciding on an action plan that also adjusts to a patient’s circumstances and preferences. One could argue that clinical reasoning is a critical part of nearly everything a physician does in practice.1,2 Despite its being essential to practice, questions about how one develops expertise in clinical reasoning, which is not a directly observable behavior, remain unanswered. The literature suggests that expertise in clinical reasoning requires both a relevant domain of content knowledge and a wealth of experience with a variety of patient presentations in the relevant domain.3,4 For example, recognizing an atypical presentation of congestive heart failure likely requires both knowledge of congestive heart failure and related conditions, as well as experience with diagnosing this disorder so that the nuances of the atypical presentation can be appropriately recognized and interpreted. Indeed, it is the interconnectedness of this relevant knowledge and rich experience that enables expertise. One needs both deep understanding of disease processes and presentations, as well as a tolerance for uncertainty and the ability to exploit experience to make decisions.3,5

Instructional formats may impact both the development of knowledge and experience. For example, initial exposure to authentic presentations of a disease could help build experience.2 As such, an important curriculum design question is the extent to which instructional formats with high (versus low) authenticity differentially impact the development of expertise in clinical reasoning. By authenticity, we mean how close the instructional format is to actual clinical practice. For example, a paper case of a presentation has a high authenticity because in the latter the learner can see and interact with a patient as in an actual clinical encounter but in the former the learner does not see a patient or participate in the relational aspects of a clinical encounter.

The literature would suggest at least two contrasting views of how instructional formats might impact the development of clinical reasoning. Theories of emotional engagement would argue that as instructional formats are made more authentic (i.e., like actual practice), learning is improved since the trainee may be more motivated to learn material, which is perceived as relevant and important. The resulting increase in cognitive effort, attention to the task, and use of learning strategies may improve performance.6 Alternatively, cognitive load theory, which refers to limitations in one’s ability to process information in working memory, would argue that humans can only hold a limited number of information “chunks” in the short-term memory.
(or “the working memory”) at a given time. According to cognitive load theory, this load into is divided into intrinsic (relating to the complexity of the problem to be solved) and extraneous load (everything else that is taken into account by the learner but has no direct relationship to the problem to be solved). From this standpoint, depending upon the learner’s baseline knowledge and the specifics of the learning task, increasing authenticity of instruction could lead to additional verbal and nonverbal information that overwhelm the learner (extraneous load), thereby negatively impacting learning.

We have previously reported on the impact of three different instructional formats on clinical reasoning performance in a 2-year investigation of a single class year (i.e., a single cohort) at our institution. In those studies, we found that increasing the authenticity of second-year instructional format (formats in order from low to high authenticity were paper case, DVD encounter, and SP encounter) did not significantly improve end-of-second-year outcomes or end-of-third-year outcomes, with the exception of one positive effect: students who subsequently reported seeing more patients during the clerkship year (i.e., the most authentic instructional format of all) for each of three studied topics performed better on the corresponding performance outcome. Based on this result, we argued that seeing actual patients represented the most authentic instructional format and, therefore, may improve learning.

Several questions still remained after our initial studies. First, insufficient sample size may have resulted in diminished power in our investigation as we followed a 1-year cohort (i.e., a single class year). Second, there may have been a group effect that confounded our instructional format intervention; i.e., being in a particular small group may have differentially impacted trainee performance because of classmate and instructor effects specific to the different groups. The inability to model these potential effects was a limitation of our prior work, despite our attempts to carefully control the teaching content in each session. In particular, each teacher underwent approximately 1 hour of faculty development in preparation to teach these sessions and was given carefully crafted teaching points, as well as the content for each group. Nonetheless, it is possible group differences still existed, such as SPs who unintentionally portrayed the cases differently (i.e., conveyed emotion or left out minor facts with their presentation). Thus, the objectives of this current study were twofold: (1) to determine the impact of three commonly used instructional formats on educational outcomes in both the second and third year of medical school using a larger sample than in prior study and (2) to determine if there was a group effect which may have confounded our results. To address these objectives, we used hierarchical linear modeling (HLM), an analytic approach markedly different from our previous work. Because we provided each instructor with frame-of-reference training and because the content presented in each group was largely identical, we predicted that our HLM analysis would reveal no group effects.

METHODS

Study Setting

The Uniformed Services University of the Health Sciences (USU) is the United States’ only federal medical school. During the time of the study, the school has 2 years of preclinical education, followed by 2 years of clerkship education. The Introduction to Clinical Reasoning (ICR) Course ran through the entire second year and exposes students to a series of common symptoms, laboratory findings, and syndromes in medicine.

Within the normal context of the ICR course (http://www.usuhs.mil/icr/index.shtml), we selected three ICR small-group topics: abdominal pain, anemia, and polyuria. The standard educational format for the ICR course is to present subject content using paper cases, followed by small group discussion led by a faculty facilitator.

We conducted a 2-year, prospective, randomized, crossover study to determine the effect of educational format on second-year medical students’ performance in ICR as well as third year clerkship outcomes. This article explicitly addresses the potential impact of the group on reasoning outcomes. Further, this study used a larger group (followed two class years of medical school or two cohorts of students through their second and third year of medical school) than our previous study, which followed a 1-year cohort through their second and third years of medical school. All students enrolled in the ICR course at USU were eligible and invited to participate, and the participants were randomly assigned to small groups using a random number generator. The course directors for ICR, in addition to the individual small-group preceptors, were blinded to student participation in the study. Our final sample size was 233 students in 21 groups.

Educational Formats

We chose three instructional formats to present the three selected subject areas/clinical problems: a paper case, a prerecorded video (DVD) presentation of a doctor and patient conveying the same content, and a live recreation of the case content with a SP. Within each of the three subject areas (anemia, abdominal pain, and polyuria), there were three cases presented per subject area over the 90-minute small-group session (i.e., each case had a 30-minute discussion). Each small group was exposed to identical content for each clinical subject area, but was assigned to a different instructional format for the three cases presented within a subject area (paper, DVD, or SP; Fig. 1). Additional details regarding educational formats are described elsewhere.

To help ensure consistency in instruction and to potentially avoid a group effect, all small-group facilitators were oriented to the study, course goals, and small-group session expectations, but not to study outcome measurements. Group facilitators were provided identical discussion points to cover after the presentation of each case within the three topics, regardless of the instructional format used for that case.
Measurements

Preclerkship Outcomes
We collected both graded outcome measures within the normal context of the ICR course (an end-of-second-year objective structured clinical examination (OSCE), and an essay/short answer examination), in addition to a non-graded outcome (video quiz) that was added for this study.

Clerkship Outcomes
We investigated students’ learning outcomes measured by National Board of Medical Examiners (NBME) subject exam score in internal medicine (IM), exam points on the medicine clerkship, clinical points on the medicine clerkship, and OSCE score on abdominal pain at the end of their third year of medical school. During the last week of

FIGURE 1. Study Design.
the internal medicine clerkship, students took the NBME subject examination in medicine and two locally developed examinations. Examinations were summed into a measure called exam points. Teachers working with the students recommended grades and reported the number of clinics spent with the student. Teacher-recommended grades were weighted according to the number of clinics the teacher spent with the student and summarized into a measure called clinical points. The abdominal pain OSCE station contained a case that was identical to one of the study’s second-year cases on abdominal pain.

Explanatory Variables

The explanatory variables were divided into person level (level 1) and group level (level 2) to facilitate the HLM analysis. At the person level, we included gender, pretest score, medical school preclinical grade point average (GPA), and self-reported total number of patients seen, which came from our Internet-based cweblog student–patient tracking system on the IM clerkship (see description below). These variables were all related to individual students. Students took a pretest on the first day of the clerkship, which has been shown to be a good proxy measure of preclerkship ability as it correlated well with United States Medical Licensing Examinations (USMLE) Step 1 in a prior study \(r = 0.66\) and disattenuated \(r = 0.85\)\(^9\). Medical school preclinical GPA was calculated using course grades from the first 2 years. The GPA is a weighted average created by multiplying each course grade by the number of contact hours for the given course, summing the weighted grades across courses, and then dividing the sum by the total number of contact hours. The resulting averages were converted to a common 4-point scale (range: 0.0–4.0). Students recorded patient encounters in an electronic patient logbook (cweblog). Previous research on the reliability and validity of cweblog data suggests that these self-report logs have adequate sensitivity of core problem reporting (60%, SD = 21.5%), and excellent specificity (95%, SD = 4%) and percent agreement (87%, SD = 7%)\(^10\).

For the outcome measure of OSCE score on abdominal pain, we replaced total number of patients with self-reported number of patients seen only on abdominal pain cases because the outcome variable was subject specific.

At group level (level 2), we used instructional method as an explanatory variable because the students within one group received the same instruction. It should be noted that we only estimated the effect of instructional method when any of the effects of person-level explanatory variables on the outcomes showed significant variation at group level. In other words, if the effects of person-level explanatory variables did not vary significantly across groups, it implied no significant variation across groups, and thus modeling instructional method effect would be meaningless. This study was approved by the USU Institutional Review Board.

Statistical Analysis

We utilized HLM for the current study for several important reasons. First, misestimated standard errors occur with multilevel data when one fails to take into account the dependence among individual responses within the same group; in other words, a student’s learning and performance within a small group is likely influenced by other students (and/or the teacher) in the group. In our case, this dependence may arise because students within the same group shared the experience of receiving the same instruction. HLM resolves this problem by incorporating into the statistical model a unique random effect for each group unit. The variability in these random effects is taken into account in estimating standard errors. In essence, HLM allows us to model the fact that observations (i.e., in this case, our subjects) are not independent.

Second, heterogeneity of regression could be a problem since we do not know whether the relationships between person-level predictors (pretest score, GPA, gender, etc.) and outcomes varied across groups. HLM enables us to estimate a separate set of regression coefficients for each group unit and then to model variation among the groups in their sets of coefficients as multivariate outcomes to be explained by group factors.

For each outcome variable, we started with random-coefficient model\(^11\), which provides estimates for the unconditional parameter variability in the random intercepts and slopes. As described above, if any of the person-level explanatory variables showed significant variation at group level, we added the group-level explanatory variable (instructional method) into the model to see whether this would help explain the variation across groups. In addition, we estimated variance explained by level-1 explanatory variables for each outcome variable by comparing level-1 effects estimated by one-way analysis of variance model\(^11\) and random-coefficient model.

RESULTS

From the level-1 results, pretest score and GPA had significant influences on exam points (Table I). None of the slopes of level-1 explanatory variables showed significant variation across groups. By adding the four explanatory variables, the level-1 variances reduced from 43.98 (estimated from one-way
The results of the outcome variable clinical points were similar with those of exam points (Table II). Level-1 results indicated that pretest score and GPA had significant impact on clinical points, whereas level-2 results showed nonsignificant variation of the slopes across groups. By adding the four explanatory variables, the level-1 variances reduced from 53.85 to 32.91, i.e., by 39%. That is, they accounted for 39% of the student-level variance in the outcome of clinical points.

Pretest score and GPA were again significantly associated with the outcome NBME subject examination score (Table III). Surprisingly, number of patients had a significant negative effect on NBME score, though the effect size was quite small. Level-2 results showed nonsignificant variation of the slopes across groups. Adding the four explanatory variables reduced the level-1 variances from 555.11 to 393.67, i.e., 29%, which means that these explanatory variables accounted for 29% of the student-level variance in the outcome of NBME score.

GPA and gender (male students received higher scores) were significantly associated with OSCE score on abdominal pain at the student level (Table IV). Furthermore, their slopes varied significantly across groups. In other words, the strength of association between GPA and gender with the OSCE score was different from group to group. We added the instruction method variable (paper, standard patient, or DVD) as a grouplevel explanatory variable. However, it did not help to explain the variance (level-1 variance changed to 394.95), which indicated that instruction method did not matter here. Adding the four explanatory variables reduced the level-1 variances from 555.11 to 393.67, i.e., 29%, which means that these explanatory variables accounted for 29% of the student-level variance in the outcome of OSCE score on abdominal pain.

**TABLE IV.** OSCE Score on Abdominal Pain

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>SD</th>
<th>t Ratio</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Abdominal Pain Patients $\gamma_{1}$</td>
<td>-0.10</td>
<td>0.10</td>
<td>-1.00</td>
<td>0.33</td>
</tr>
<tr>
<td>Pretest $\gamma_{2}$</td>
<td>0.06</td>
<td>0.25</td>
<td>0.24</td>
<td>0.81</td>
</tr>
<tr>
<td>Gender $\gamma_{3}$</td>
<td>-10.77</td>
<td>5.27</td>
<td>-2.04</td>
<td>0.05</td>
</tr>
<tr>
<td>GPA $\gamma_{4}$</td>
<td>9.22</td>
<td>4.12</td>
<td>2.24</td>
<td>0.04</td>
</tr>
<tr>
<td>Random Effect Variance Component</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Patients $\mu_{1}$</td>
<td>0.06</td>
<td>0.24</td>
<td>0.40</td>
<td>0.06</td>
</tr>
<tr>
<td>Pretest $\mu_{2}$</td>
<td>0.40</td>
<td>0.06</td>
<td>357.11</td>
<td>0.01</td>
</tr>
<tr>
<td>Gender $\mu_{3}$</td>
<td>115.72</td>
<td>0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPA $\mu_{4}$</td>
<td>393.67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level-1 Effect</td>
<td>393.67</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**DISCUSSION**

We hypothesized that students’ second-year, small-group instructional format would impact performance at the end of the third-year clerkship. Our hypothesis was not substantiated, in that the students’ second-year instructional method did not impact performance at the end of the third-year clerkship in our prospective, randomized, crossover trial. We also hypothesized that there would be no group effect because of instructional method given our attention to frame-of-reference training and our carefully constructed instructional content. This hypothesis was substantiated with the finding of nonsignificant variances at the group level.

We found several other notable findings using the HLM approach. Pretest score and GPA were significant predictors of IM clerkship clinical points, exam points, and NBME score. This is consistent with prior studies. Further, GPA and gender were significant predictors of OSCE score on abdominal pain. The gender effect for OSCE exams has been previously reported.

We compared the proportion of variance that can be explained by the studied variables using HLM and ordinary linear regression. If we used ordinary linear regression methods, for example, the $R^2$ was 0.04 for exam points and 0.18 for clinical points, which indicated that 40% of variance of exam points and 18% of variance of clinical points could be explained by the same set of variables. In comparison, the HLM techniques allowed us to gain a more accurate picture by excluding “noise” brought by the variance at group level. In doing so, the explanatory variables were able to account for more variance: 46% of variance of exam points and 24% of
variances of clinical points. This increased variance shows the added value (and added information gained) from using HLM over ordinary linear regression.

There are several limitations in the present study. First, small sample size, especially at the group level, was a limitation. If we had more groups, the estimation of standard errors for the explanatory variables would be more robust, and the relationship between the explanatory variables and the dependent variables would likely to be stronger. Second, the randomized design of the present study minimized variation across groups at the outset. In other words, HLM would be an even more useful method than ordinary linear regression in a study context where randomization of participants was difficult to realize, which is very common in medical educational research. It is important to realize that with the use of exams as a study outcome, students’ other preparation for these summative events, such as question and answer books or studying in groups, may outweigh and effectively “wash out” any effects of instructional format.

In conclusion, we found evidence that faculty development efforts can and do work in this setting as shown by the absence of a group effect with the associated instructional method. Future studies are needed to replicate these findings, ideally with large sample size and more explanatory variables at group level. No group effect occurred despite having teachers with varying levels of experience, and different specialties serve as instructors in this course. We also did not find a linear relationship between increasing authenticity of instruction and performance. We suspect that this may be due to the need to balance instructional format with factors such as cognitive load, learner experience, and knowledge about the content area. Our prior work supports this conclusion.2,8 Finally, we believe that HLM can be a useful approach for determining the impact of learner nesting effects (i.e., learner being instructed in a small group) on outcomes.

REFERENCES

Relationship Between OSCE Scores and Other Typical Medical School Performance Indicators: A 5-Year Cohort Study

Ting Dong, PhD; LTC Aaron Saguil, MC USA; CDR Anthony R. Artino Jr., MSC USN; COL William R. Gilliland, MC USA (Ret.); Donna M. Waechter, PhD; CAPT Joseph Lopreaito, MC USN (Ret.); Amy Flanagan, MFA; Steven J. Durning, MD, PhD

ABSTRACT  Background: Objective Structured Clinical Examinations (OSCEs) are used at the majority of U.S. medical schools. Given the high resource demands with constructing and administering OSCEs, understanding how OSCEs relate to typical performance measures in medical school could help educators more effectively design curricula and evaluation to optimize student instruction and assessment. Purpose: To investigate the correlation between second-year and third-year OSCE scores, as well as the associations between OSCE scores and several other typical measures of students’ medical school performance. Methods: We tracked the performance of a 5-year cohort (classes of 2007–2011). We studied the univariate correlations among OSCE scores, U.S. Medical Licensing Examination (USMLE) scores, and medical school grade point average. We also examined whether OSCE scores explained additional variance in the USMLE Step 2 Clinical Knowledge score beyond that explained by the Step 1 score. Results: The second- and third-year OSCE scores were weakly correlated. Neither second- nor third-year OSCE score was strongly correlated with USMLE scores or medical school grade point average. Conclusion: Our findings suggest that OSCEs capture a viewpoint that is different from typical assessment measures that largely reflect multiple choice questions; these results also support tenets of situated cognition theory.

INTRODUCTION
The Objective Structured Clinical Examination (OSCE) is used at the vast majority of U.S. medical schools to assess student performance. It features standardized, scripted patient presentations and encounters that are either directly monitored by an instructor in the room or indirectly monitored by an evaluator synchronously or asynchronously watching from outside the room. The OSCE aims to evaluate students’ clinical skill, interpersonal communication skill, and medical knowledge. It also affords students the opportunity to receive timely feedback on patient communication and care in a more authentic clinical situation than that provided by the more frequently used multiple choice question (MCQ) test.

A contemporary educational theory, situated learning, proposes that knowledge, thinking, and learning are contextual and experiential. With respect to medical education, performance is the result of interactions between multiple factors (patient, trainee, and encounter setting) and these factors often change as the encounter transpires. Given the dynamic and complex nature of social interactions in OSCEs, situated learning theory would argue that OSCEs provide a unique vantage point from which to adjudicate performance beyond typical medical school measures such as medical school grade point average (GPA) or MCQ tests, which are devoid of these interactions.

Ascertaining if OSCEs provide a unique assessment lens is important given the resources (time, monetary, and human) devoted to constructing and administering these examinations. Currently, relatively little research has been conducted to investigate the relationship between medical students’ performance on OSCEs and more traditional medical school performance indicators, such as USMLE Step examinations and medical school GPA. Further, the few published studies that examined this relationship reported inconsistent findings, which may be due, in part, to inadequate sample sizes. For example, Simon et al reported moderate correlations between second-year medical students’ OSCE scores and USMLE Step 1 scores and Step 2 scores, respectively. In contrast, another study reported students’ OSCE scores only had weak correlations with both Step 1 and Step 2 scores.

In the present study, we examined the relationship between students’ OSCE scores and several important indices of medical school performance. We expected small to moderate correlations between OSCE scores, USMLE Clinical Knowledge (CK) scores, and medical school GPA. We also hypothesized that second-year and third-year OSCE scores would moderately correlate with each other and would also explain additional variance in the USMLE Step 2 CK score beyond that explained by the Step 1 score as Step 2 CK is a more clinically oriented examination than Step 1. In other words, consistent with situated learning theory, we hypothesized that OSCEs would measure something unique that is not captured in these other measurements and expected the associations between the OSCEs and the other measurements to be small to moderate at best.

METHODS

Study Participants
As the United States’ only federal medical school, the Uniformed Services University (USU) matriculates approximately 170 medical students annually. At the time of this study, USU offered a traditional 4-year curriculum: 2 years of basic science courses followed by 2 years of clinical rotations. Upon matriculation, students are commissioned as
Relationship Between OSCE Scores and Other Performance Indicators

### Study Variables

Our analysis included six variables: comprehensive OSCE scores from the second and third year of medical school, cumulative second- and fourth-year GPA, and USMLE Step 1 and Step 2 CK scores.

Our second-year OSCE featured the following stations: chest pain, thirst, weakness, fatigue, oral presentation, and geriatric assessment. Our third-year OSCE featured the following stations: abdominal pain, fatigue, foot pain, loss of memory, and chronic cough. Each OSCE included varying diagnoses over the study period, and different diagnoses were tested in the second- and third-year OSCEs. According to a previously conducted, unpublished generalizability study at our institution, second-year OSCE stations demonstrated a moderate generalizability coefficient ($r = 0.52$) with 40.8% of the overall variance explained by student ability. Pearson’s correlations between these six, second-year OSCE stations and Internal Medicine clerkship clinical points ($r = 0.30, p < 0.01$), examination points ($r = 0.23, p < 0.01$), and total points ($r = 0.30, p < 0.01$) were small to moderate.

Students completed the basic science-oriented USMLE Step 1 after their first 2 years of medical school and the more clinically oriented Step 2 CK examination during their fourth year of medical school. First attempt scores for each examination were used for this analysis. Second-year cumulative GPA reflects performance on basic science material and fourth-year cumulative GPA reflects performance across all 4 years of medical school.

### Statistical Analysis

We performed univariate correlation and multiple regression analyses. The correlation analysis reported Pearson correlations between the variables. The purpose of the multiple regression analysis was to examine the incremental impact of second-year and third-year OSCE scores on USMLE Step 2 CK score beyond that predicted by the USMLE Step 1 score.

### RESULTS

The means and standard deviations of the variables are shown in Table I. The average OSCE score of the third year was slightly lower (4 points) than that of the second year. The variances of both OSCEs were about the same. The correlation between second-year and third-year OSCE scores ($r = 0.29, p < 0.01$) was small. The second-year OSCE score had weak correlations with second-year cumulative GPA ($r = 0.26, p < 0.01$), USMLE Step 1 score ($r = 0.11, p < 0.01$), and Step 2 CK score ($r = 0.14, p < 0.01$). The third-year OSCE score had no significant correlation with Step 1 score ($r = 0.06, p > 0.05$) but had weak correlations with Step 2 CK score ($r = 0.14, p < 0.01$), second-year cumulative GPA ($r = 0.22, p < 0.01$), and fourth-year cumulative GPA ($r = 0.26, p < 0.01$).

The results of multiple regression analysis (Table II) showed that the additional USMLE Step 2 CK score variance accounted for by the second- and third-year OSCE scores beyond that explained by the Step 1 score was minimal ($R^2$ change = 0.01).

### DISCUSSION

Our study investigated the relationship between students’ performance on OSCEs, which are scripted, contextual encounters featuring standardized patients, and several traditional medical school performance indicators. There were two major findings. First, the correlation between students’ performance on the second-year OSCE and third-year OSCE was small. This indicates that students who did well in the second-year OSCE did not necessarily do well in the
third-year OSCE and vice versa. This could be as a result of imperfect reliability and validity of the examinations. On the other hand, from the perspective of situated learning theory, this small correlation can be explained by the fact that second-year and third-year OSCE performance is the culmination of different situations (such as different patient conditions, chief complaints, associated medical conditions, and acuity of illness presentation), as well as different encounter factors (such as different performance goals and objectives for each station). The weak correlations between OSCE examinations are also consistent with the finding of context specificity, that there are factors other than the clinical facts of the patient’s condition, that impact performance.\(^5,6,10\) Second, OSCE scores were weakly correlated with USMLE Step 1 and Step 2 CK scores as well as medical school GPA. Although the Step 2 CK score is more clinically based than Step 1, our study findings suggest that a surrogate for clinical experience (OSCEs) provides little additional variance beyond Step 1. The results are consistent with our theoretical predictions as both Step 1 and 2 do not contain the dynamic and rich social interactions featured in OSCE examinations. This indicates that the OSCE measures something different from what the USMLE measures, such as interpersonal communication skills, realtime patient care skills, and the ability to apply CK at the patient’s bedside. Third, USMLE examinations were highly correlated with each other and with GPA. These examinations primarily assess knowledge and are devoid of rich contextual information. These strong correlations are also consistent with prior publications.\(^11\) Likewise, the strong correlations between GPA and USMLE Step 1 and Step 2 CK are also consistent with prior studies.\(^11,12\)

Our study had several limitations. First, the data came from a single institution and our findings may not be generalizable to other medical schools. Although USU is similar, in many ways, to other U.S. medical schools, there may be important student differences that limit the external validity of our findings. For example, the average age of USU students tends to be older and the proportion of females lower than that found in other medical schools. In addition, at the time of the study, the university offered a traditional 4-year program (2 years of basic science and 2 years of clerkship rotations), whereas many other medical schools are using or are transitioning to a more integrated curriculum. Second, this was a retrospective study. We can only infer associations instead of causations from the results. Third, we included only a limited number of OSCEs in the study. Fourth, our OSCEs have good but not optimal reliability data. However, to the best of our knowledge, this study was the largest and most complete one correlating OSCE score with other measures of medical school performance in terms of sample size and years of medical school included.

In conclusion, OSCE scores at different stages of medical school education were weakly correlated with each other and with other standard indicators of medical school performance. These findings are consistent with situated learning theory. As a result, OSCEs appear to provide a lens into medical student performance that is unique when compared to standard medical student assessment measures. Although the OSCE features a patient encounter in a specific context, the USMLE Step 1 and Step 2 CK are MCQ tests devoid of social interactions. We would expect stronger correlations between OSCE performance and the USMLE Step 2 Clinical Skills examination. Correlating OSCE performance with USMLE Step 2 Clinical Skills and other clinical performance measures (such as mini-clinical evaluation exercise) could provide educators with additional guidance regarding what accounts for the variance in these measurements as well as additional validity evidence for these examinations. This will be our future study direction.

**REFERENCES**

The Association Between Specialty Match and Third-Year Clerkship Performance

LTC Aaron Saguil, MC USA*; LCDR Erin K. Balog, MC USN†; Matthew N. Goldenberg, MD‡; Ting Dong, PhD§; CDR Anthony R. Artino Jr., MSC USN§; Col Christopher M. Zahn, USAF MC‖; Lt Col Jessica T. Servey, USAF MC*; Lt Col E. Matthew Ritter, USAF MC¶; Col David R. Welling, USAF MC (Ret.)‖; LCDR Laura B. Ramsay, MC USN‖; ENS Gerald Ming, MSC USN*; Steven J. Durning, MD PhD**

ABSTRACT  Purpose: The United States is experiencing an accelerating physician shortage, especially within primary care. Medical educators are actively seeking ways to predict student specialty match and workforce requirements. Previous studies investigating specialty match have focused on factors known at the time of matriculation. This study examined whether third-year clerkship performance could be used to predict specialty match later in medical school. Method: The authors evaluated the clerkship performance of 802 students graduating from the Uniformed Services University between 2007 and 2011. They examined the relationship of students’ clerkship grades and National Board of Medical Examiners’ clinical subject examination scores to specialty match. In addition, the authors combined student performance in family medicine, internal medicine, and pediatrics to create composite variables and assessed their associations with the match. Results: Among 802 students, 339 (42.4%) students matched to primary care specialties. There was a positive association between higher family medicine (Odds ratio [OR] 1.65, 95% confidence interval [CI] 1.05, 2.59), general surgery (OR 1.91, 95% CI 1.22, 2.99), internal medicine (OR 2.17, 95% CI 1.35, 3.49), and pediatrics (OR 2.59, 95% CI 1.52, 4.43) clerkship grades and students matching into family medicine, general surgery, internal medicine, and pediatrics, respectively. Only family medicine showed a weak correlation between higher National Board of Medical Examiners’ scores and specialty match. Conclusions: Higher clerkship performance in four of six Uniformed Services University third-year clerkships is associated with matching into the corresponding specialty. Clerkship performance provides a potential tool for educators in counseling students and predicting future specialty match.

INTRODUCTION

The United States is in the midst of an accelerating physician workforce shortage that is compounded by a maldistribution of physicians across specialties and the urban–rural divide.1-9 The active physician workforce, recently estimated at 765,000,10 falls short of current needs and is not expected to grow at a pace sufficient to meet future demands.1 By 2025, population growth and aging will mean an additional 51 million people and 24 million seniors needing health care.10 In addition, 32 million more Americans are anticipated to access the system as a result of health care reform.11 All of these factors are expected to strain increasingly limited physician resources.

Among current and future specialty care needs, the projected shortfall in primary care is among the most dramatic. By 2025, the deficit in primary care physicians is expected to increase to 46,000, accounting for 37% of the total predicted physician shortage.1,12 Undoubtedly, it is in the United States’ interest to shape the future workforce to meet current and projected needs. Efforts are underway to increase the total number of physicians, to improve their health care efficiency, and to address specialty and geographic disparities in distribution.9,13,14

With respect to specialty disparities, much work has been done to explore factors predicting medical student specialty choice and, in particular, to determine which factors predict a match into primary care, which we refer to as family medicine, internal medicine, and pediatrics.10 In general, studies have found that medical student specialty preference is associated with demographic factors, student attributes, and medical school characteristics (Fig. 1).

Demographic factors predicting accession to a primary care specialty include being a woman3,15-19 and having a rural background.3,16,17,20 In addition, married status,3 older age upon entry to medical school,20 lower parental education and socioeconomic attainment,16,20 and having done volunteer
work in a developing nation have been found to predict family medicine specialty choice.

Student attributes predicting primary care specialty choice include social commitment, altruism, valuing patient relationships, a favorable view toward primary care patients, and an interest in providing care for the disadvantaged. A desire for a varied scope of practice and low income expectations are additional factors predicting family medicine choice.

Medical school factors associated with primary care specialty choice include attendance at a public school, attendance at a school receiving Title VII funding, required rotations in primary care, the presence of primary care role models and interest groups, and exposures to medicine away from the academic center (such as with international, rural, or inner-city experiences). The relationship of financial indebtedness to specialty choice is complex, and studies have yielded conflicting results.

To date, only a few studies with contradictory results have looked at how academic performance in medical school is associated with specialty choice. In general, the literature reveals that students tend to choose specialties in which they receive higher evaluations. One study found that an honors grade in internal medicine was associated with choosing internal medicine, another found that higher grades in all clerkships (except family medicine and pediatrics) were associated with specialty choice, and yet another found that there was a correlation between pediatrics grades and pediatrics specialty match.

Our study aims to determine the extent to which performance in each of the third-year clerkships can predict specialty choice among a large, 5-year cohort of medical students. We hypothesized that high performance in a clerkship (as measured by clerkship grade and National Board of Medical Examiners’ clinical science subject examination score), relative to performance in other clerkships, would correlate with matching to that clerkship’s specialty. We also hypothesized that high performance in the primary care (family medicine, internal medicine, and pediatrics) clerkships, relative to performance in other clerkships, would correlate with matching into primary care. Demonstrating either association would provide medical educators with additional data to use in counseling students and in predicting future specialty match.

Finally, we hypothesized that female gender, predictive of matching into primary care in other studies, would also correlate with primary care specialty match.

**METHOD**

We conducted this study at the Uniformed Services University of the Health Sciences (USU), a Liaison Committee on Medical Education-accredited institution dedicated to educating students and officers to serve as physicians in the United States Military and Public Health Service. USU students are not required to pay for tuition, fees, equipment, or books, and they collect active duty officer pay and a housing stipend in return for an active duty service obligation. We collected clerkship performance data, including grades and NBME.
Association Between Specialty Match and Third-Year Clerkship Performance

We first used counts and percentages for categorical variables and means and standard deviations (SD) for continuous variables to describe demographic characteristics, school performance measures, and the residency specialty match for our student cohort. We chose specialty match over specialty choice since match is more objectively reported through the military Joint Graduate Medical Education Selection Board. Second, we used Pearson correlations, $r$, to assess the bivariate associations between clerkship grades and specialty match and between NBME scores and specialty match, respectively, for each of our school’s six required third-year clerkships (family medicine, general surgery, internal medicine, obstetrics and gynecology, pediatrics, and psychiatry). Third, we used multivariate logistic regression and odds ratios (ORs) to assess the associations between clerkship grade and specialty match, while controlling for gender and total third-year grade point average (GPA). Additionally, we used multivariate logistic regression and ORs to assess the associations between NBME scores and specialty match, while controlling for gender and composite NBME clinical subject examination performance (an average of the scores across all six clerkships). Controlling for third-year GPA and composite NBME performance was intended to distinguish students who did relatively better in a given clerkship when compared to their overall performance during the clerkship year. Finally, we averaged the clerkship grades and NBME scores from family medicine, internal medicine, and pediatrics to create the composite variables “primary care clerkship grade” and “primary care NBME score,” respectively. Both of these variables were tested for association with a created variable called “composite primary care specialty match,” which combined all students matching into family medicine, internal medicine, and pediatrics.

We analyzed the data using SPSS version 19 (IBM, New York, New York), and set statistical significance at $p < 0.05$.

RESULTS

Specialty match information was available for all 802 students who graduated in the classes of 2007 through 2011. Table I shows the student gender breakdown, match results, and average third-year, primary care, and individual clerkship performance measures. In the studied classes, 26.9% were female, the mean age at matriculation was 24.5 years, and the mean third-year GPA was 3.18 (SD 0.44). Mean grades for the six clerkships ranged from 2.89 (SD 0.72) to 3.37 (SD 0.58) on a 4-point scale, and the mean NBME clinical subject examination scaled scores ranged from 69.97 (SD = 10.83) to 74.57 (SD = 6.84). In the fourth year, 339 (42.4%) students matched into primary care specialties.

Table II displays the unadjusted and adjusted associations between clerkship grade, NBME scores, and specialty match. When unadjusted, only pediatric clerkship grade was associated with pediatric specialty match (correlation coefficient 0.11, 95% confidence interval [CI] 0.04, 0.18). When adjusted for gender and third-year GPA, there was a positive, statistically significant association between the family medicine (Odds ratio [OR] 1.65, 95% CI 1.05, 2.59), general surgery (OR 1.91, 95% CI 1.22, 2.99), internal medicine (OR 2.17, 95% CI 1.35, 3.49), and pediatrics (OR 2.59, 95% CI 1.52, 4.43) clerkship grade and students matching into family medicine, general surgery, internal medicine, and pediatrics, respectively. In addition, there appeared to be a positive trend between the psychiatry (OR 2.21, 95% CI 0.92, 5.30) and composite primary care (OR 1.95, 95% CI 0.96, 3.96) clerkship grade and matching into psychiatry and a primary care specialty, respectively.

<table>
<thead>
<tr>
<th>TABLE I. Demographic Characteristics, Clerkship Performance Measures, and Specialty Match Results Among 802 Medical Students Graduating From the USU, 2007–2011</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student Characteristics</strong></td>
</tr>
<tr>
<td>Female, No. (%)</td>
</tr>
<tr>
<td>Age at Matriculation, Mean (SD)</td>
</tr>
<tr>
<td>Third-Year GPA, Mean (SD)</td>
</tr>
<tr>
<td><strong>Clerkship and Specialty Characteristics</strong></td>
</tr>
<tr>
<td>Clerkship/Specialty</td>
</tr>
<tr>
<td>Family Medicine</td>
</tr>
<tr>
<td>General Surgery</td>
</tr>
<tr>
<td>Internal Medicine</td>
</tr>
<tr>
<td>Obstetrics/Gynecology</td>
</tr>
<tr>
<td>Pediatrics</td>
</tr>
<tr>
<td>Psychiatry</td>
</tr>
<tr>
<td>Primary Care</td>
</tr>
</tbody>
</table>

$^*NRMP %$ represents the percentage of U.S. seniors matching into the same specialty through the National Residency Matching Program from 2007 to 2011. General surgery, internal medicine, and pediatrics numbers include categorical and preliminary matches.29
When unadjusted, NBME score was not associated with specialty match. Once adjusted for gender and third-year GPA, only the family medicine NBME score and family medicine matched were weakly associated (OR 1.07, 95% CI 1.01, 1.13).

In the model adjusted for clerkship grade and third-year GPA, female gender was significantly associated with matching into pediatrics (OR 2.85, 95% CI 1.73, 4.71), psychiatry (OR 2.27, 95% CI 1.08, 4.77), and primary care (OR 1.81, 95% CI 1.28, 2.55). In addition, there appeared to be a trend between female gender and matching in obstetrics and gynecology (OR 2.93, 95% CI 0.93, 9.19).

**DISCUSSION**

At USU, achieving a higher grade in a required third-year clerkship was positively associated with matching into that clerkship’s specialty. This association was significant across all clerkships except psychiatry and obstetrics and gynecology. The consistency of this finding across four of the six required clerkships (and the appearance of a trend in a fifth, psychiatry) provides evidence to support our hypothesis that performance is a predictor of specialty match. This result broadens and strengthens those of previous studies.31,32

Contrary to our second hypothesis, NBME clinical subject examination scores (accounting for approximately 20% of a clerkship’s grade at our institution) did not appear to be associated with specialty match. Only in family medicine were NBME scores and match weakly associated. Given the moderate to strong associations between clerkship grade and specialty match, the lack of a correlation involving NBME performance indicates that a significant portion of a clerkship grade’s predictive power lies in other components. This capacity may reside in clinical performance evaluations, faculty-constructed examinations (written, oral, or practical), and/or the results of other assignments and experiences. Finally, as we predicted, female gender was associated with matching into primary care; being a woman was also associated with matching into pediatrics as well as psychiatry.

The question of why clerkship grade might predict specialty match is worth exploring. In contrast to demographic factors, student attributes, and medical school characteristics (all of which are set by the time of matriculation), clerkship performance is not determined until the third year. At that time, students receive their first concentrated exposure to clinical medicine and embark on an intense period of professional acculturation, experiencing medicine as practiced across different specialties. Situated learning theory and the principle of legitimate peripheral participation suggest that the clerkship is a powerful moment in professional identity formation.30 Students have increased responsibilities in patient care (“legitimate”), are working with faculty and residents in the field (“peripheral”), and are engaged in meaningful experiences (“participation”).31,32 A higher grade in a clerkship, once controlled for overall academic ability (as reflected by third-year GPA), may indicate a more successful adaptation to a specialty’s culture and community of practice. As a result, clerkship performance may provide an additional, interim measure that can inform advisors, faculty, and residency program directors when discussing career options with students.

Our study has several limitations. First, because this was a retrospective, correlational study, causal statements cannot be made. As opposed to reflecting a causal link between performance and specialty match, our data might as easily be explained by supposing that students desiring a given specialty try harder and thus do better in the respective clerkship, or that advisors steer students to specialties in which they performed well, creating a self-fulfilling prophecy. Alternatively, students may intentionally delay a clerkship in their desired specialty until the end of the academic year, improving their chances of earning a better grade as a result of their participation in, and completion of, other specialty clerkships first. Although data maintained by the USU Office of Student Affairs indicates that over 90% of students successfully match into their chosen specialty, a prospective study exploring the specialty preferences of students before matriculation, through medical school, and into the specialty selection board—when analyzed in conjunction with performance and residency selection—may help untangle the potential causal relationships between preference, performance, and matching into specialty.

### TABLE II. Associations Between Clerkship Grade, NBME Clinical Science Subject Examination Scaled Scores (NBME Scores), and Specialty Match Among 802 Medical Students Graduating From the USU, 2007–2011

<table>
<thead>
<tr>
<th>Clerkship Grade</th>
<th>NBME Scores</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unadjusted Pearson’s r, 95% CI, N</td>
<td>Adjusted* OR, 95% CI</td>
<td>Unadjusted Pearson’s r, 95% CI, N</td>
</tr>
<tr>
<td>Family Medicine</td>
<td>0.01, (−0.06, 0.08), 788</td>
<td>1.65, (1.05, 2.59)*</td>
</tr>
<tr>
<td>General Surgery</td>
<td>0.06, (−0.01, 0.13), 792*</td>
<td>1.91, (1.22, 2.99)*</td>
</tr>
<tr>
<td>Internal Medicine</td>
<td>0.03, (−0.05, 0.10), 698</td>
<td>2.17, (1.35, 3.49)*</td>
</tr>
<tr>
<td>Obstetrics/</td>
<td>0.02, (−0.07, 0.11), 492</td>
<td>1.04, (0.27, 3.99)</td>
</tr>
<tr>
<td>Gynecology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pediatrics</td>
<td>0.11, (0.04, 0.18), 798*</td>
<td>2.59, (1.52, 4.43)*</td>
</tr>
<tr>
<td>Psychiatry</td>
<td>0.03, (−0.04, 0.10), 794</td>
<td>2.21, (0.92, 5.30)**</td>
</tr>
<tr>
<td>Primary Care</td>
<td>−0.04, (−0.11, 0.03), 687</td>
<td>1.95, (0.96, 3.96)**</td>
</tr>
</tbody>
</table>

*p < 0.05, **p < 0.10. *Adjusted for gender, third-year GPA. *Adjusted for gender, and a composite of all clerkship NBME scores. **Adjusted for clerkship grade, third-year GPA. 1Primary care includes a composite of grades or NBME scores from family medicine, internal medicine, and pediatrics.
choice, and match. It would be highly instructive to see if students wishing to match in a given specialty chose another specialty as a result of a higher grade in a different clerkship or were dissuaded from their original specialty as a result of a poor grade in the associated clerkship.

Second, our study was conducted at a single institution with a unique place among American medical schools. All students at USU are officers in the military or the Public Health Service and are paid a monthly salary and housing allowance. Upon completion of medical school, all of our students enter the military graduate medical education match. As a result, our findings may not be generalizable. Indeed, a recent comparison of our graduating classes with U.S. seniors participating in the National Residency Matching Program indicated that more of our students match into family medicine and that substantially fewer match into internal medicine.\(^{33}\) In addition, 26.9% of our graduates were female, compared to the approximately 50% seen at other medical schools.

Although the external validity of our study may be limited, some aspects of the USU experience offer unique advantages from the perspective of study methodology. Since students take on significantly less or no debt to finance their medical education, we were able to control for debt as a determinant of specialty match. Furthermore, though military physicians garner different monetary bonuses by specialty, the divergence in pay between generalists and subspecialists is far less drastic than in civilian medical practice. When coupled with the fact that our students are obligated to 7 years of military medical practice after completing a military sponsored residency (during which all are paid at the same rate), the financial incentive to choose a traditionally higher paying specialty is diminished.

Our unique situation also allows us to partially control for the effects of practice location as a determinant of specialty match. As opposed to the civilian system, USU students have fewer programs to which they can apply, and their subsequent practice locations are generally different from the places at which they were educated. Although these advantages in finance and practice location may be helpful in isolating the effect of clerkship performance on specialty match, a multi-institutional study with a variety of medical schools would be required to assess the generalizability of our results.

A third limitation to our study was that we did not have the data to determine which students matching to internal medicine and pediatrics remained generalists after completing their residencies. The military offers a full range of subspecialty fellowship options, and it may be that a number of these students chose to subspecialize. Although this would not lessen the predictive value of internal medicine and pediatric clerkship performance as a marker for specialty match, it would render it less usable as a marker for eventual primary care practice.

Finally, we did not find a positive association between the psychiatry or obstetrics and gynecology clerkships and their respective matches. The lack of a significant association may be as a result of the fact that 2 years’ worth of data from our student cohort was unavailable for the obstetrics and gynecology clerkship because of staffing turnover. As a result, we only had 23 students matching into obstetrics and gynecology on which to perform our analyses, a number potentially insufficient to detect an association. We encountered a similar difficulty with respect to psychiatry; although we had more complete data, only 37 students matched into the specialty. In contrast, 88 students matched into pediatrics, and over 100 students matched into family medicine, general surgery, and internal medicine. As we accumulate more data in years to come, we hope to further explore potential relationships between clerkship grade and specialty match.

The strengths of this study include looking specifically at clerkship performance and specialty match across multiple specialties with a large cohort of our five most recent graduating classes, and having results not confounded by trainee debt. Our finding that higher grades in required clerkships are positively associated with matching into a clerkship’s respective specialty is consistent with a recent, large-scale study which found that students receiving honors grades in internal medicine were more likely to choose internal medicine.\(^{21}\) Our results were only partially consistent with a 1994 single institution study reporting that higher grades in all clerkships (except family medicine and pediatrics) were associated with matching into the respective specialties.\(^{27}\)

The implications of our findings depend on the potential direction of causality. If clerkship grades help determine specialty match, then undergraduate medical educators and advisors should continue to encourage students to try their very hardest in their preferred specialty’s clerkship. Clerkship directors might wish to devote more rotation time, weight, and effort to graded activities based on clinical performance and participation since our study suggests that these clerkship components—as opposed to NBME performance—are associated with specialty match. Academic departments may consider advocating for longer clerkships both to allow for more specialty acculturation through legitimate peripheral participation and to provide an opportunity to identify excelling students for mentoring; emerging models of longitudinal clerkships may facilitate this change. Academic departments may then be able to provide program directors with information on excelling individuals, allowing directors to focus limited resources on these students, contacting them, advising and mentoring them, and encouraging them to interview and rank their program.

On the other hand, if specialty preference largely determines clerkship grade, the implication is that students tend to do well in the clerkship of the specialty they are actively considering or have already chosen. If this is the case, then undergraduate medical educators and advisors may use the clerkship grade to support and encourage a student’s choice. Among clerkship directors and academic departments, the grade may still serve as a signal to identify which students would benefit from longitudinal mentoring by a member of their field and which students to put in touch with specialty program directors. Directors might still
use grades to focus their efforts in discussing program choice with potential applicants.

Notwithstanding the limitations of the present study, the findings provide evidence that required clerkship grades are associated with specialty match. This association gives medical educators another piece of information with which to counsel students and an additional data point with which to predict future specialty match.

REFERENCES

Reflections on LTCOS Projects During Medical School

Dario M. Torre, MD, MPH, PhD

This series of articles about student learning during medical school shows an innovative, theory-driven, inquisitive approach that furthers our knowledge of current educational issues while providing opportunities for further research. All the manuscripts rest on a wide array of theoretical frameworks, spanning theories of self-regulated learning and self-efficacy\(^1\) to motivational theory and social cognitive orientations. The importance of conceptualizing and conducting research on the platform of established theory in order to generate ideas for new theories to inform practice is an effort clearly evident in all manuscripts in this section. Such an approach allows medical education researchers to become familiar with existing theories and their applications, while broadening and expanding their inquiries and curiosity.\(^2\)

The topic of clinical reasoning, crucial to the growth and development of medical students and often neglected in medical school curricula, is examined from novel perspectives that involve instructional formats, emotions, and motivational aspects. The inquiry on developing a validated tool to measure self-efficacy among medical students is an excellent example of linking theory to practice while exploring and inquiring about a concept that is crucial for the development of medical students as learners and as individuals.

The authors and contributors truly raise the bar and delve into ill-structured problems, posing new questions while avoiding repetitive research that hinders our ability to think, reflect, and define a meaningful research path.

P. Freire describes the concept of unfinishedness of the human being, as the awareness that makes learners eternal seekers of knowledge and education.\(^3\) I believe our call is to be ourselves in a perpetual lifelong movement of search and inquiry. The authors of the articles in this section, as reflective thinkers and researchers utilizing all the skills in their educational arsenal, truly seem to move in such direction.

REFERENCES


Drexel University School of Medicine, Philadelphia, PA.
ABSTRACT  Purpose: Each year military medical students participate in a separate, military match culminating with the Joint Services Graduate Medical Education Selection Board (JSGMESB). Prior studies have explored postinterview communication that occurs during the National Resident Matching Program (NRMP), but not during the JSGMESB. We examined the frequency and nature of communication during the JSGMESB and compared it with the NRMP. Methods: Cross-sectional survey study of senior students conducted at Uniformed Services University of the Health Sciences (USU) and seven civilian U.S. medical schools during March to May 2010. Respondents answered questions regarding communication with residency programs during the match. Results: Significantly fewer USU respondents communicated with programs compared with the civilian cohort (54.1% vs. 86.4%, \( p < 0.01 \)). Specific inquiries regarding rank order were more commonly experienced by USU respondents compared with civilians (17.5% vs. 4.8%, \( p = 0.02 \)). USU respondents found postinterview communication both helpful (41.3%) and stressful (41.3%). 11.1% of USU respondents indicated that they moved a program higher on their final rank lists because of further communication with these programs. Conclusions: Postinterview communication during the JSGMESB process is less common and less stressful than that reported in the NRMP. USU respondents are more likely to be asked directly about their rank list and occasionally do change their lists. Uniform guidance mirroring the NRMP’s dealing with direct inquiries about rank lists could potentially improve the process.

INTRODUCTION

Each year, fourth-year medical students across the country participate in the National Residency Match coordinated by the National Resident Matching Program (NRMP). The vast majority of the 18,000 fourth-year medical students participating are civilian; less than 1% are military.¹

Military medical students are typically either recipients of the Health Professions Scholarship Program (HPSP) or are students at Uniformed Services University of the Health Sciences (USU), the nation’s only federal medical school. HPSP and USU fourth-year students participate in an independent military match process culminating in the Joint Services Graduate Medical Education Selection Board (JSGMESB) and only rarely in the NRMP (i.e., if selected for training in civilian status by the JSGMESB).

There are a number of notable differences between the JSGMESB and the NRMP. First, the JSGMESB process assigns military medical students to residency positions exclusively at Department of Defense (DoD) residences. Residents at these DoD programs are active duty military officers, and thus civilian students typically do not apply through the JSGMESB as they would have to agree to an active duty service obligation as military officers. Conversely, HPSP students sometimes apply through the NRMP in addition to or in conjunction with the JSGMESB if they meet one of three criteria: (1) they are applying to and receive permission from the JSGMESB to pursue a residency slot in a specialty that is not supported by a DoD residency, (2) they are granted their preference to train in a non-DoD residency, or (3) they fail to match for a residency through the JSGMESB.

Communication (i.e., phone calls or e-mails between programs and applicants) taking place after the interview, but before the match, is known to occur regularly during the NRMP process.² It is likely that military programs and applicants also communicate after interviews, but before the selection board, during the JSGMESB process. One important difference between the NRMP and the JSGMESB is that although the former has specific written guidance addressing
communication between programs and applicants after inter-
views, the latter does not appear to. In other words, although
military program directors receive guidance from service-
specific Graduate Medical Education (GME) leadership (i.e.,
U.S. Air Force, Army, and Navy) during the military match
process, no specific, universal mandate similar to that in the
NRMP appears to exist. Instead, each service communicates
service-specific guidance to DoD program directors regard-
ing rules of engagement for the JSGMESB process. In
addition to reviewing these rules of engagement, we also
spoke with several current program directors representing
each service and they further confirmed that no such written
guidelines on communications exist (although, a few pro-
gram directors indicated that it was understood by them that
they were not to ask applicants directly about rank lists).

However, despite the NRMP’s specific guidance regulat-
ing postinterview communication, studies have found that
these rules are often “bent” or even broken by both programs
and applicants during the NRMP. We recently described
the types of postinterview communication that take place
between civilian programs and applicants during the NRMP.
The nature and types of communication between programs
and applicants after the interview, but before the selection
board that take place during the JSGMESB have not previ-
ously been described.

In this article, we sought to (1) determine the frequency
and nature of communications between USU fourth-year stu-
dents and programs during the JSGMESB and (2) define the
impact of communications between USU students and pro-
grams on (a) students’ perceptions of the match and (b) how
they ranked programs. Finally, we compared these findings in
the USU student population with those observed in a cohort
of civilian fourth-year medical students going through a sep-
ate, parallel NRMP.

METHODS

USU students were included as part of a cross-sectional study
of all senior medical students after the 2010 match using a
convenience sample of 8 U.S. medical schools (USU; Har-
vard Medical School; University of California, San Francisco;
University of Chicago Pritzker School of Medicine;
University of Florida; University of Pittsburgh; Warren
Alpert Medical School of Brown University; and Wright
State University Boonshoft School of Medicine). We chose
to survey USU fourth-year students in lieu of all military
medical students (i.e., HPSP students) because of easier
access and a single Institutional Review Board approval pro-
cess. Results from the 7 civilian schools participating in the
NRMP match have been described previously. The present
study was approved by the Institutional Review Boards at all
participating sites.

The survey and our methodology are described in detail
elsewhere. Briefly, we developed a survey to assess post-
interview communication between applicants and programs
during the match. The survey questions were developed fol-
lowing a review of the literature and discussions among
faculty at participating schools. The survey was pilot tested
locally first by faculty and then by residents at two institu-
tions before administration.

The survey (Appendix) consisted of questions about
demographics, academic record, application characteristics,
and communication between residency programs and appli-
cants. Self-reported demographics included gender and
under-represented minority status. Items also included spe-
cialty choice, number of residency applications submitted,
number of residency interview offers, number of residency
interviews attended, and matched program’s rank list posi-
tion. Specialties were grouped into 12 categories.

Fifteen of the questions addressed communication be-
tween residency programs and applicants. Students were
asked whether they were contacted by a program after
interviewing, by whom, and whether programs asked where
they would rank the program. Participants also were asked
whether and how often programs explicitly stated any of the
following: that a student would “fit well” into the program, be
“ranked highly,” or be “ranked to match.” We also asked
participants whether they altered their rank list based on
communications from programs and about statements they
made to programs (e.g., telling a single or multiple programs
that it/they would be ranked first or ranked highly). Par-
ticipants were asked if others had recommended notifying their
top program of their choice and, if so, who made that rec-
mendation. The survey concluded with questions about the
match outcome and overall experience. We also asked par-
ticipants if, when they ranked a program first, they felt a pro-
gram explicitly or implicitly told them that they would match
even though they did not eventually match at that program.
Finally, participants rated by Likert scale whether they found
postinterview communications stressful. For the pur-
purposes of data analysis, communications from programs to applicants
were noted to be “specific” if the program explicitly told the
applicant that they would “match at their program if they
wanted to” (i.e., “ranked to match”). Communications were
considered “nonspecific” if the students received feedback
that they would either “fit well” into the program or be
“ranked highly.”

Site investigators distributed an anonymous online survey
after the NRMP match day inviting participation (Survey
Monkey; Palo Alto, CA). Students received up to 3 follow-up
e-mails. No compensation was provided and participation
was voluntary.

Descriptive statistics and correlations were performed
using SPSS version 16 (Chicago, Illinois). Comparisons
between USU and civilian schools were made using a non-
parametric test (Mann–Whitney U) given the skewed distri-
bution of data.

RESULTS

The survey response rate for USU students was 63 of 159
(39.6%). The overall survey response rate across the seven
civilians (42.9% vs. 78%, *p* < 0.05). Eight USU respondents were female (12.7%), 6 (9.5%) identified themselves as under-represented minorities, 10 (15.9%) obtained an additional advanced degree while in medical school, 7 (11.1%) were members of AOA, and 25 (39.7%) received honors in the specialty clerkship to which they applied.

USU respondents applied to family medicine (16; 25.4%), general surgery (9; 14.3%), combined surgical subspecialties (7; 11.1%), emergency medicine (6; 9.5%), radiology (6; 9.5%), pediatrics (5; 7.5%), and internal medicine (4; 6.3%). The specialties USU students applied to differed from those applied to by their civilian counterparts (54.1% vs. 86.4%, *p* < 0.01). Programs compared with the 10.3 programs of their civilian counterparts (5.9 vs. 24.1, *p* < 0.01) and interviewed at a mean of 2.9 programs compared with the 10.3 programs of their civilian counterparts (*p* < 0.01). USU respondents matched at their top choice 61.9% of the time, a similar rate to that observed at the national average for 2010, which was 52.7%.1,7

Although just over half of our USU respondents were contacted by residency programs after they interviewed, significantly fewer were contacted when compared with their civilian counterparts (54.1% vs. 86.4%, *p* < 0.01). Military program directors contacted USU applicants most commonly, but they were less likely to do so compared with their civilian counterparts (42.9% vs. 78%, *p* < 0.01). USU respondents were also commonly contacted by residents (25.5% e-mail, 11.1% phone call). USU respondents were infrequently e-mailed or called by associate program directors (3.2% e-mail, 3.2% phone call), faculty interviewers (4.8% e-mail, 3.2% phone call), or alumni from the military student’s medical school currently at the program (4.8% e-mail, 3.2% phone call). When USU respondents did report that they received unsolicited communication, typically between one and four programs contacted them (Table II).

When programs did contact USU respondents, specific inquiries by the programs regarding rank order were more commonly reported by USU respondents compared with that reported by civilian student counterparts (17.5% vs. 4.8%, *p* = 0.02). Programs made a number of positive statements regarding where students would rank, although these statements occurred less commonly when compared to the civilian cohort. These statements included “that you would fit well into their program” (39.7% vs. 76.2% civilian, *p* < 0.01), “that you will be highly ranked at their program” (27% vs. 52.8% civilian, *p* < 0.01), and “that you would match at their program if you wanted to” (20.6% vs. 34.6% civilian *p* < 0.01).

Information regarding rank order flowed in both directions. Students reported receiving advice from mentors, colleagues, and/or peers to inform programs where that program ranked on his or her list. About a quarter (23.8%) of USU respondents reported being advised to tell their top choice that that program was first on their individual rank list. Similar to the civilian respondents, the majority of USU respondents (45/63, 71.4% vs. 62.9% civilian, *p* = 0.37) told their top choice program that they would rank them first, whereas about a third of respondents (23/63, 36.5% vs. 59.9% civilian, *p* = 0.21) informed their top choice they would rank them highly. Although no respondents endorsed intentionally misleading programs by informing multiple programs that each would be ranked first, 23 respondents (36.5%) did tell more than one program that they would be ranked highly.

Communication between programs and USU respondents appeared to affect subsequent student ranking of programs. In 7 (11.1%) cases, USU respondents indicated that they either

<table>
<thead>
<tr>
<th>TABLE I. Residency Specialty Breakdown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specialty</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Anesthesiology</td>
</tr>
<tr>
<td>Emergency Medicine</td>
</tr>
<tr>
<td>Family Medicine</td>
</tr>
<tr>
<td>Internal Medicine</td>
</tr>
<tr>
<td>Neurology, Pathology, and Rehabilitation</td>
</tr>
<tr>
<td>Obstetrics/Gynecology</td>
</tr>
<tr>
<td>Pediatrics</td>
</tr>
<tr>
<td>Psychiatry</td>
</tr>
<tr>
<td>Radiology</td>
</tr>
<tr>
<td>General Surgery</td>
</tr>
<tr>
<td>Surgical Specialty</td>
</tr>
<tr>
<td>Dermatology, Ophthalmology, Radiation Oncology</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE II. Frequency and Types of Communication for Each Respondent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program Representative</td>
</tr>
<tr>
<td>------------------------</td>
</tr>
<tr>
<td>Program Director</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Resident</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Associate Program Director</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Faculty Interviewer</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Alumnus</td>
</tr>
</tbody>
</table>
moved a program higher or lower on their final rank lists based on communication with programs. Altering the final rank list because of postinterview communication was uncommon among military students and their civilian counterparts (11.1% vs. 23.4%, \( p = 0.33 \)).

Amongst USU students who did not match at the program desired (16 respondents), 4 respondents (25.0%) recalled that a program told them that they would match there if they wanted, but despite ranking it above other programs, the students did not match there. Twelve respondents (75.0%) stated that although they were not told that they would match at the program, they felt that they would match there based on the feedback received.

Similar to their civilian counterparts (47.7%), 41.3% of USU respondents agreed or strongly agreed that communication with programs after interviews was helpful to them in selecting a residency position. Despite this response, a notable number of civilian (65.2%) and USU respondents (41.3%) found the decision to communicate with programs and the issues involved with the communication to be stressful (although USU respondents were less likely to report stress, \( p < 0.01 \)).

**DISCUSSION**

Communication between USU fourth-year applicants and programs participating within the JSGMESB process occurs regularly. This communication is important and allows programs and applicants to resume dialogue started earlier in the application process and facilitates the flow of information amongst participants. In general, most types of postinterview communication between applicants and program representatives are not prohibited by either the JSGMESB or the NRMP. The NRMP, in regards to postinterview communication, states:

“Both applicants and programs may express their interest in each other; however, they shall not solicit verbal or written statements implying a commitment.”

The military match lacks similar uniform, specific guidance across participating services regarding this point.

As mentioned earlier, the JSGMESB process differs in many ways from the NRMP. First, the JSGMESB is reserved for applicants who will serve on active duty during their residency training. Also, although the NRMP oversees the match between thousands of applicants and hundreds of institutions, the scale of the JSGMESB is much smaller, on the order of hundreds of applicants and scores of institutions. This difference in scale results in significantly fewer program options within a given specialty for USU applicants.

During the NRMP, programs and applicants each submit lists blinded to the other’s contents. The JSGMESB process differs slightly among the three participating services, making direct comparison to the NRMP more nuanced. Although all programs participating in the JSGMESB follow the same core rules, each service implements these rules differently. In practical terms, the processes range from a slightly modified Electronic Residency Application Service-based match approximating the NRMP to a paper-based application process with a formal selection board meeting in person at the JSGMESB. All programs participating in the JSGMESB follow a uniform interviewing process that includes use of standardized program director interview form.\(^8\)

With these differences between the military match and NRMP in mind, it is interesting to note that communication occurred less often in almost all regards for the military cohort when compared with the civilian cohort. This difference observed may be partly because of differences in the number of programs civilian respondents applied to when compared with USU respondents, and thus civilian respondents simply had more opportunities for programs to contact them. The content of communication reported by civilian respondents and USU respondents also differed significantly. Although civilian respondents were more likely than USU respondents to receive nonspecific, positive statements regarding their ranking by programs, USU applicants were more likely to be asked directly about their rank list order.

Although communication between USU fourth-year applicants and programs was noted to increase stress in some respondents, this effect was reported significantly less often in the military than in civilian cohort. Similarly to civilian respondents, 40% of USU respondents also reported that the ability to communicate with programs was beneficial to them. Students may find the ability to freely communicate with programs beneficial to them for a number of reasons, some of which may be military unique. As noted previously, there are often far fewer options within the JSGMESB for a specialty when compared with the NRMP. Although their civilian counterparts might be drawn to a specific geographic region to pursue a specialty, USU respondents and other military applicants have a limited number of options in distinctly varied geographic locations. It follows that the ability to communicate freely with representatives from prospective programs may be valuable for applicants in order to better assess regions of the country to which they are considering a move. Also, in contrast to civilian programs operating through the NRMP, individual military programs are not at risk to underfill (i.e., even when applicants to a particular specialty are low, the programs typically share in the shortfall). These differences in the competitive nature of military programs along with fewer military programs to apply to could contribute to a more favorable application and interviewing climate for students and also may explain why students reported less stress associated with postinterview communication.

At the same time, the military medical corps is ultimately a small, close-knit group of physicians, and applicants are well aware of this fact. This knowledge may increase pressure on military applicants when approached by a representative of a program they know they may work closely with in the future, not just as a trainee, but potentially as a colleague. The military rank structure may also contribute to student...
stress as program directors tend to be high-ranking medical corps officers. Although academic rank exists at civilian academic teaching centers, in the military, where physicians wear rank on their uniforms, differences in rank are made much more explicit. These military-unique factors should be considered when discussing postinterview communication.

With all of this in mind, it is notable that a small number of USU respondents (7; 11%) reported that they changed their rank lists because of postinterview communication (“positive postinterview feedback”), moving some programs higher on their lists. The proportion of respondents who changed their lists was similar to that reported in the civilian cohort. The specific reason in each applicant’s case that caused him or her to move a program or programs higher on his or her rank list was not explored in our survey. Whether or not programs directly asking students about their rank lists caused the rank list to change was also not addressed by our survey. Further, our questionnaire did not ask who queried students about their rank lists (e.g., program leadership vs. residents at the program).

There were some important limitations to our current study. First, our survey response rate was lower than desired and thus our findings could have suffered from nonresponse bias. This may be partially explained by the potentially sensitive nature of the survey questions and the resulting concerns of respondents that they might be bringing scrutiny to GME programs. However, despite the low response rate, there were no statistically significant differences between USU responders and nonresponders on several important measures, which provides evidence to mitigate this limitation. Second, the questions were developed primarily for the NRMP, not the military match; thus, potentially important differences among military services could not be explored further. Another limitation with our survey design was that it did not quantify the number of times applicants and individual programs interacted. This quantification might have allowed for exploration into the amount of communication between applicants and programs and other outcomes of interest (e.g., stress). Further, because the survey was released after NRMP match day, USU applicants had already found out the results of the JSGMESB 4 months earlier, possibly introducing more recall bias compared with civilian applicants. Also, although we searched diligently for a JSGMESB regulation similar to the NRMP regulation dealing with asking directly about rank order preference, we did not find one, it does not rule out the possibility that one already exists. Finally, although we asked students to comment on whether or not they found the decision to communicate with programs stressful, we do not know why students found it stressful or what aspects of the process were stressful.

Areas to consider for future study include adding military students who attend institutions other than USU, assessing military medical students’ overall satisfaction with the military match, and exploring differences in student perceptions by service.

CONCLUSION

Communication after interviews between USU respondents and programs is relatively common and offers value to both programs and applicants. Postinterview communication during the military match occurs less frequently compared with civilian counterparts and is reported to be less stressful and more helpful. Despite USU respondents being more likely to be asked directly about their rank order preference, USU respondents are not more likely to change their match lists than their civilian counterparts and report less stress associated with communicating with programs. Uniform guidance, similar to those of the NRMP, for participants in the JSGMESB, regarding asking directly about rank order preference, would be unlikely to negatively influence communication and might make programs and applicants more comfortable with the process. Further, educating all parties involved in postinterview communication, including residents who may or may not be aware of existing regulations, could potentially improve the process as well. In the end, regardless of whether or not the JSGMESB process is modified in regards to postinterview communication, the effect on this already well-performing system is likely to be modest.

APPENDIX

1. What is your gender □ Male □ Female
2. Do you classify yourself as an underrepresented minority?
   ○ Yes
   ○ No
3. Did you obtain any of the following advanced degrees during your time in medical school? (e.g., MD/PhD/MPH)
   ○ Yes
   ○ No
4. Are you a member of your school’s AOA chapter? (Note, please answer no if you were designated as AOA eligible rather than as a member)
   ○ Yes
   ○ No
   ○ N/A; My school does not have an AOA chapter
5. In which third of the class will you graduate?
   ○ Top 1/3
   ○ Middle 1/3
   ○ Bottom 1/3
6. What grade did you ultimately receive during your required clerkship in the specialty to which you applied? Please choose the most applicable answer.
   □ A/Honors □ C/Pass □ F/Fail
   □ B/High Pass □ D/Low Pass
7. Did you couples match? □ Yes □ No
8. On your interview days, did any faculty member question you regarding a significant other?
   □ Yes
   □ No
   If so, approximately how many times did you receive this question? [drop down list 0 to 50]

9. To what primary field did you apply (i.e., not prelim or transitional):
   □ Internal Medicine
   □ Internal Medicine prelim
   □ Internal Medicine/Pediatrics
   □ Emergency Medicine
   □ Family Practice
   □ Neurology
   □ Pediatrics
   □ Psychiatry
   □ Ob/gyn
   □ Anesthesiology
   □ Surgery
   □ Surgery prelim
   □ Surgical Subspecialty [plastics, ENT, urology, Neurosurgery orthopedics, etc]
   □ Other [radiology/ophthalmology/dermatology/pathology/rad onc/physical med. and rehabilitation]

10. How many programs did you apply to in your main specialty? [drop down list 0 to 50]

11. How many interview invitations did you receive? [drop down list 0 to 50]

12. At how many programs did you interview? [drop down list 0 to 50]

13. Some residency programs communicate with applicants after the interview. How many programs did you receive the following unsolicited (e.g., not a response to a message from you personalized communication (e.g., addressed specifically to you) from one of the following individuals?

14. Among the programs that contacted you, how many programs explicitly stated the following?

<table>
<thead>
<tr>
<th>Statement</th>
<th>Number of programs who made the statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>That you would fit well into their program</td>
<td>Drop down list for each row (number 0–50):</td>
</tr>
<tr>
<td>That you will be ranked highly at their program</td>
<td></td>
</tr>
<tr>
<td>That you would match at their program if you wanted to</td>
<td></td>
</tr>
</tbody>
</table>

15. Among the programs that contacted you, how many asked where you would be ranking them? [drop down list 0 to 50]

16. Thinking back on how you composed your match list, how many programs did you move up on your list because you received encouraging postinterview feedback from that program? [drop down list 0 to 50]

17. Some applicants expect to match at a particular program and are disappointed when they do not match there. Did any of the following happen to you?
   ○ I was explicitly told by a program that I would match there if I wanted to. I ranked it above others, but I did not match there.
   ○ Although I was not explicitly told that I would match at the program, I felt that I would match there based on feedback I received. I ranked it above others, but I did not match there.
   ○ Neither apply.

18. Some residency programs pay special attention to applicants from their affiliated medical school. Did any of the following happen to you?
   ○ I was explicitly told by my own medical school’s affiliated residency program that I could match there if I wanted to.
   ○ Although I was not explicitly told that I would match at my own medical school’s affiliated residency program, I expected to based on encouraging postinterview feedback from the program.
   ○ Neither apply.

19. Applicants often indicate to programs that they will rank them first on their match list. Did you feel pressured to tell one or more programs that you would rank them first?
   ○ Yes
   ○ No
   → if yes, go to 20
   → IF no, go to 21

20. From whom did you feel pressure to notify your number one program? (check all that apply) ALPHABETIZE THESE CHOICES
   ○ Fellow applicants
Postinterview Communication Between Military Residency Applicants and Programs

○ Web-based resources (e.g., student-doctor.net)
○ Advisor in the specialty
○ Other advisor
○ Program director from your medical school
○ Program director from other program
○ Alumni
○ Parent
○ Significant other
○ Residents in the specialty
○ OTHER:

21. Please select which of the following statements apply to your ultimate communications with programs about their position on your rank list.
○ I told my top choice program that I would rank them “first”
○ I told my top choice program that I would rank them “highly”
○ I told more than one program that I would rank them “first”
○ I told more than one program that I would rank them “highly”
○ I did not tell any programs where I would rank them on my match list.

22. At what position on your rank list did you rank the program at which you ultimately matched?
Drop down list from 1 to 50 and “I did not procure a residency spot during the regular match.”
Please rate your level of agreement or disagreement with the following statements.

23. Overall communication with program directors, faculty, and/or residents after the interview was helpful to me in selecting a residency position.
   a. Strongly agree
   b. Agree
   c. Neutral
   d. Disagree
   e. Strongly disagree
   f. N/A (I did not communicate with programs after my interviews and before the match.)

24. The decision whether and/or how to communicate with residency programs after my interviews made the match process more stressful for me.
   a. Strongly agree
   b. Agree
   c. Neutral
   d. Disagree
   e. Strongly disagree
   f. N/A (I did not communicate with programs after my interviews and before the match.)

Thank you for participating in this survey. If you have additional comments about the postinterview Match process, please let us know in the box below.

Acknowledgment

Thank you for completing this survey!

Acknowledgment

This study was funded by the Department of Medicine at the University of Chicago.

References

Leadership Success and the Uniformed Services University: Perspectives of Flag Officer Alumni

Ting Dong, PhD; Steven J. Durning, MD, PhD; COL William R. Gilliland, MC USA (Ret.); LTC Kent J. DeZee, MC USA; Donna M. Waechter, PhD; COL John E. McManigle, USAF MC (Ret.); David F. Cruess, PhD; Sharon K. Willis; CDR Anthony R. Artino Jr., MSC USN

ABSTRACT The Uniformed Services University of the Health Sciences (USU) houses the nation’s only federal medical school, the F. Edward Hebert School of Medicine. A key aspect of the curriculum at USU is leadership education as graduates go on to serve the Department of Defense through a variety of senior positions in the military. We surveyed a specific group of USU graduates who have achieved the rank of General or Admiral (“flag officers”) to enhance our understanding of successful leadership for military physicians and to gain an understanding of how USU might shape its curriculum in the future. Methods: We sent an Internet-based survey to 13 flag officer graduates. The first section of the survey contained items from the multifactor leadership questionnaire-6S, a questionnaire with evidence of reliability and validity for evaluating leadership styles. The second section of the survey contained open-ended questions addressing key characteristics of an effective leader in the Military Health System, experiences that prepared them for leadership, USU’s role in leadership positions, and advice for USU for better educating future leaders. The second section of the survey was coded using the constant comparative method. Results: Eight flag officers (63%) responded to the survey. They all scored highly on transformational leadership style. Qualitative themes reached saturation for each open-ended question. The flag officers identified characteristics consistent with published literature from other fields regarding effective leadership. They endorsed USU’s role in achieving their leadership positions and suggested areas for improvement. Conclusions: Characteristics of effective leadership (transformational leadership style) identified by the flag officers surveyed in this study are consistent with the literature from other fields. These finding have important implications for leadership education at USU and potentially other institutions. The results also provide additional data to support the notion that USU is meeting its societal obligation to educate future leaders in military medicine.

INTRODUCTION The Uniformed Services University of the Health Sciences (USU) was established by the 92nd Congress in 1972. Its founding was largely the result of a 25-year effort of Congressman F. Edward Hebert (D-LA).1 USU was established to provide a comprehensive education in medicine to men and women who demonstrate potential for and commitment to careers as medical corps officers in the Army, Navy, Air Force, or Public Health Service. Several years later, to recognize Congressman Hebert’s efforts, legislation officially designated the USU School of Medicine, our nation’s only federal medical school, as the F. Edward Hebert School of Medicine.

USU admitted its charter class of 32 students in 1976. We currently matriculate about 170 medical students each year. USU has awarded Doctor of Medicine degrees to more than 4,200 graduates who have gone on to serve in the Army, Navy, Air Force, and Public Health Service. The medical school’s motto, and societal charge, is “caring for those in harm’s way,” which includes caring for active duty members, military retirees, and their dependents. Our graduates practice in military treatment facilities throughout the United States and around the world, as well as in combat, humanitarian assistance missions, natural disasters, and among a multitude of cultures.

The current overarching learning objectives for the medical school are organized into the following six areas, which are consistent with the current Accreditation Council of Graduate Medical Education core competencies: (1) medical and population health knowledge, (2) interpersonal and communication skills, (3) patient care, (4) practice-based learning and improvement, (5) professionalism and officerhood, and (6) systems-based practice. Thus, an important component of the educational program has always involved the development of physicians who understand principles of leadership and teamwork and who develop critical thinking skills. Previous studies2 have identified several important leadership roles that our graduates assume in their careers to meet USU’s societal charge. These prior studies, however, did not attempt to ascertain factors or elements of leadership success from the standpoint of the graduates. In the present study, we sought to take a closer look at USU’s goal of educating future leaders in military medicine. We explored this through the perspective of a group of flag (General or Admiral) officers who graduated from USU, as reaching the military rank of flag officer is a recognized pinnacle of military leadership.

Over the past 20 years, transformational leadership theories have been some of the most frequently researched leadership theories.3,4 Transformational leadership depicts leader behaviors that transform and inspire followers to perform beyond expectations while transcending self-interest for the good of organization.7 Transformational leadership differs...
from more traditional leadership models, such as transactional leadership, which are largely based on the exchange of rewards contingent on performance. Several researchers have argued that transformational and transactional leaders use different schemas or mental frameworks to interpret events and subsequently result in different leadership behaviors and actions. A series of meta-analyses found that transformational leadership was positively associated with leadership effectiveness and important institutional outcomes across many different types of organizations, situations, and levels of analysis. Thus, transformational leadership is believed by many leadership scholars to be an optimal leadership approach.

In the current study, we were interested in answering several research questions:

1. Are flag officers who graduated from USU transformational leaders?
2. Opinions of these flag officers
   (a) What are the characteristics of an effective leader?
   (b) What was USU’s role in preparing them for leadership?
   (c) Which experiences helped prepare them for leadership (USU or other)?
   (d) How can USU shape its curriculum in the future?

METHODS

Participants

Since the foundation of USU in 1972, 13 graduates have achieved the rank of flag officers (O-7 or higher). These 13 flag officers were asked to complete an online survey with the assistance of the USU Alumni Affairs office. The survey was a flag-officer-specific part of the Alumni survey sent to all USU graduates.

Instrument

Flag officers were sent a link to the online survey (Appendix). At the beginning of the survey, we briefly introduced the Long-Term Career Outcome Study research team and informed them that this year marks the 40th anniversary of the founding of USU. We also stated that we were specifically interested in learning about their leadership styles and the ways in which USU education might have prepared them for their current (or former) role as a leader in medicine.

The survey had two sections. The first section included items from the multifactor leadership questionnaire-6S (MLQ-6S), which has been reported to be a reliable and valid leadership questionnaire and widely used in leadership research. The MLQ-6S aims to evaluate leadership preferences, especially transformational leadership style. It consists of 21 five-point, Likert-scale questions designed to assess seven components (or factors) related to leadership—idealized influence, inspirational motivation, intellectual stimulation, individualized consideration, contingent reward, management-by-exception, and laissez-faire. These first four factors measure degree of transformational leadership. Idealized influence indicates whether leaders hold subordinates’ trust, maintain their faith and respect, show dedication to them, appeal to their hopes and dreams, and act as their role model. Inspirational motivation measures the degree to which leaders provide a vision, use appropriate symbols and images to help others on their work, and try to make others feel their work is significant. Intellectual stimulation shows the degree to which the leaders encourage others to be creative in looking at old problem in new ways, creates an environment that is tolerant of seemingly extreme positions, and nurtures people to question their own values and beliefs and those of the organization. Individualized consideration indicates the degree to which leaders show interest in others’ well-being, assign projects individually, and pay attention to those who seem less involved in the group.

The next two factors measure degree of transactional leadership. Contingent reward indicates the degree to which leaders tell others what to do in order to be rewarded, emphasize what they expect from them, and recognize their accomplishments. Management-by-exception is a leadership style that can be described as simply telling others the job requirements and being content with standard performance. The last factor, laissez-faire measures degree of passive/avoidance leadership. This leadership style is characterized as requiring little of others, being content to “let things ride,” and letting others do their own things.

Items scores on the MLQ-6S can range from 1 (extremely negative on the statement) to 5 (extremely positive on the statement), and there are three items measuring each factor. Therefore, 15 is the highest possible score one can get on a particular factor, and 3 is the lowest possible score. Scores in the range of 3 to 7 are considered low on a particular factor, scores in the range of 8 to 11 are considered medium, and scores in the range of 12 to 15 are considered high.

In the second section of the questionnaire, we asked the flag officers to answer four open-ended questions, which addressed our second set of research questions. These four questions were derived from the literature as well as discussions with Long-Term Career Outcome Study team members and other educational leaders at USU: (1) What do you think are the key characteristics needed to be an effective leader in medicine? (2) What experiences helped prepare you for leadership? Please be as specific as you can. (3) Do you think your USU education prepared you for your leadership positions? Why or why not? (4) What specific advice would you give USU’s educational leadership for how they might better educate future leaders?

Data Analysis

We calculated each flag officer’s score on the seven factors of the MLQ-6S. We analyzed the open-ended questions adopting the constant comparative method to identify underlying themes. The constant comparative method is a form of grounded theory and represents a process in which themes
are identified in the data with no a priori hypotheses. New themes are compared with previous themes and data. Consistent with this approach, we used an open coding method and sought to identify themes that relate to the phenomena of interest, in this case leadership. The themes were created and reviewed by two study investigators (S.J.D. and T.D.). The USU Institutional Review Board approved this study.

RESULTS

Eight out of 13 (62%) flag officer graduates completed the survey. The flag officer scores on the seven factors of the MLQ-6S are shown in Table I. In general, all flag officers scored highly on the four factors measuring transformational leadership style (idealized influence, inspirational motivation, intellectual stimulation, individualized consideration; Table I). They scored moderate on transactional leadership style (contingent reward, management-by-exception) and low on passive/avoidant leadership style (laissez-faire).

For the open-ended questions, we achieved saturation as well as complete agreement, following discussion of themes and coding differences, for the themes from the four leadership items. Saturation implies that if additional responses were obtained, new themes would not be expected to emerge. We list the identified themes for each question and give examples of quotes from the flag officers (bullet points) in Table II.

Our flag officers also provided a number of suggestions for improvement. These suggestions included emphasizing leadership as a “core competency.” For example, one respondent commented:

“I think that the services expect that USU Grads will be leaders wherever they go so the school should have this as a core competency of the education.”

Others suggested that we be clear about expectations of leadership and emphasizing leadership role models and mentorship. One respondent noted:

“Having former USU students come back and talk about their military careers and experiences really drives home that it is so much more than just being a good doctor and being facile with the science of medicine. The art of medicine and the art of military service are equally important.”

Others commented on USU need to maintain our focus and mission, which includes leadership.

“I strongly recommend that you not lose focus on the mission of the university. USU is special.”

Finally, several flag officers suggested that USU consider additional training and research opportunities, such as completion of a master’s degree. One respondent noted:

“Encouraging students to focus on research as a way to solve problems in medicine will lead to leadership learning. I would consider offering some students a 5th year to spend balancing clinical rotations and research or even a Master’s Degree.”

### Table I. Flag Officers’ Scores on the Seven Factors of Multifactor Leadership Questionnaire

<table>
<thead>
<tr>
<th>Flag Officer</th>
<th>Idealized Influence</th>
<th>Inspirational Motivation</th>
<th>Intellectual Stimulation</th>
<th>Individualized Consideration</th>
<th>Contingent Reward</th>
<th>Management-by-Exception</th>
<th>Laissez-Faire</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>13</td>
<td>11</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>B</td>
<td>12</td>
<td>13</td>
<td>12</td>
<td>13</td>
<td>13</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>C</td>
<td>14</td>
<td>13</td>
<td>15</td>
<td>13</td>
<td>10</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>D</td>
<td>13</td>
<td>13</td>
<td>12</td>
<td>15</td>
<td>14</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>E</td>
<td>14</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td>F</td>
<td>13</td>
<td>12</td>
<td>12</td>
<td>15</td>
<td>14</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>G</td>
<td>13</td>
<td>12</td>
<td>12</td>
<td>9</td>
<td>14</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>H</td>
<td>12</td>
<td>15</td>
<td>14</td>
<td>9</td>
<td>10</td>
<td>13</td>
<td>7</td>
</tr>
</tbody>
</table>

Score range on a particular factor: High = 12–15; Moderate = 8–11; Low = 3–7.
Question 1: What do you think are the key characteristics needed to be an effective leader in medicine and/or public health?

Theme 1: Effective communicator
- Excellent public presentation skills
- Excellent communication skills
- Ability to convey vision for the group

Theme 2: Empathy
- Genuine concern for others
- Respect and concern for others, especially subordinates

Theme 3: Integrity
- Irreproachable integrity and honesty
- Sense of honor and ethical decisions
- Humility and trustworthiness

Theme 4: Ambition and energy
- Willingness to put oneself “out there” and take the heat
- Energy to make change happen and make it worth doing

Theme 5: Expertise
- Technical expertise in his/her respective executive/management field
- Competency in your own specialty

Theme 6: Know and motivate subordinates
- Genuine interest in people, how they work
- Motivate others to make the vision a reality; understanding team members while expecting the best from them

Question 2: What experiences helped prepare you for leadership?
Please be as specific as you can.

Theme 1: USU-based experience
- Medical History Course—learning that other things are important beside the grind of medicine and the importance of understanding what history brings to the present.
- Rugby team—learning how to depend on each other, how to win, how to lose and get up, how to have fun, collaboration, and relationship building.

Theme 2: Other prior experience (that may or may not be the result of USU’s education)
- Every leadership opportunity prepared me for future opportunities. Some of what I learned was trial and error, some was observing, some was the result of mentoring, some was from various leadership training/courses, and some was from reading and self-learning. Actually serving in leadership positions, however, was by far the most effective preparation for other leadership jobs.
- There is no substitute for the opportunity to serve others when in Command. I have found each opportunity unique and also a learning experience.
- Clinical medicine practice. Teaching and being responsible for students and house staff caring for patients. Working through difficult problems.

Question 3: Do you think your USU education prepared you for your leadership positions? Why or why not?
All eight respondents answered yes to this question. We identified two themes as below.

Theme 1: USU education built upon prior experience
- I came from an academy so most of this was well instilled before USU but this grew at the school.
- I think my college and ROTC background shaped my foundation on officer ship and USU, built on this foundation, preparing me to become a medical officer.

Theme 2: Leadership preparation directly from USU
- USU set high standards for conduct and performance, helped establish a foundation of service to others rather than self, made me acutely aware of my role as a Military Physician, an Officer with the inherent responsibilities of such, as well as a Physician, the leader of the health care team, expected to work through problems on my own and in groups.

Question 4: What specific advice would you give USU’s educational leadership for how they might better educate future leaders?

Theme 1: Emphasize leadership as a “core competency”
- I think that the services expect that USU Grads will be leaders wherever they go. So the school should have this as a core competency of the education. That should include some core curriculum along with bringing medical and other leaders in to talk to the students.
- Students listen when someone they respect talks to them about leadership.

Theme 2: Be clear about expectations for leadership
- Be clear about what is expected of these young officers in terms of personal character development, behavior, and performance.
- If they fail, pick them up and challenge them again. If they choose not to meet expectations, help them find a career that better suits them.

Theme 3: Emphasize leadership role models and mentorship
- High emphasis on leadership role models with aggressive mentorship while in the school as well as stressing their roles as mentors toward their junior officers.
- More interaction with line unit and line unit commanders during school rotations (not just classes, but actual assignments). Place intense emphasis on the difference between “practicing medicine in the military” vs. practicing military medicine!
- Exposing students to leaders who are active in their fields through special seminars and causal encounters can go a very long way. They learn, even in small doses, what a leader looks like and what makes him or her tick.
- Bring senior alumni back to mentor the young officers.

Theme 4: Maintain our focus and unique mission
- USU is special. Don’t lose that.
- Don’t allow the students forget that the only reason they get to become doctors at USU is to serve the warfighters in their respective service as competent medical officers.
- Embrace the military traditions. Don’t let the students convince you otherwise.

Theme 5: Additional training opportunities
- Encouraging students to focus on research as a way to solve problems in medicine will lead to leadership learning. I would consider offering some students a 5th year to spend balancing clinical rotations and research or even a master’s degree.
- I do think that USU students would benefit from gaining managerial experience if an MBA were offered in conjunction to an MD degree.
- I think the number of leadership opportunities could and should be increased. Bushmaster was (a good experience), but even there, more focus on leadership skills would be appropriate.

*The bullets under each theme are examples of quotes given from respondents.
perspectives of flag officer alumni

Another officer commented:

“I think the number of leadership opportunities could and should be increased.”

Many respondents commented on several USU-related experiences that they believed helped prepare them for their leadership positions. Some of these experiences were not surprising, such as the Combat Casualty Care Course, Bushmaster (our 3-day military simulation exercise), clinical rotations in our military hospitals, mentors, and the medical history course. These courses and rotations emphasize progressive responsibility in real-world educational settings, where there is also an emphasis on providing feedback to students to improve their performance. For Combat Casualty Care Course and Bushmaster, leadership principles are explicitly taught and evaluated. Some experiences, however, were also a little surprising, such as the following items:

Rugby team—learning how to depend on each other, how to win, how to lose and get up, how to have fun, collaboration, relationship building.

We also included a category of “other experiences” because it was not clear if the respondent was speaking about USU-specific experiences or not. Many of the comments in this category related to mentoring, and we have included responses from three different flag officers below:

“Mentoring—see how the ‘big boys’ think.”

“Mentors: Cannot overemphasize the criticality of mentors. Even today I call them to pulse them about issues.”

“Of significant importance to my success later in my career has been the personalized mentoring I received from a more senior Officer.”

Several comments in the other experiences category also related to changing assignments. One flag officer noted:

“I’m convinced the key to my success has been my alternating assignments in both clinical and operational medicine throughout my career. To remain solely in one facet of medicine will ultimately limit your ability to lead an extremely complex, multifaceted organization at its highest levels.”

DISCUSSION

A survey of USU flag officers provides additional evidence to support the notion that USU is meeting its goal to educate future leaders in military medicine. This is consistent with prior work.2 USU flag officers scored highly on transformational leadership style (Table I), a finding that is consistent with work from other fields.5 This result provides additional evidence to support the credibility of the opinions offered by the respondents on the open-ended questions.

The reported characteristics of an effective leader in our system are consistent with those reported in other contexts.5 Further, all these characteristics—effective communication, empathy, integrity, ambition and energy, expertise, as well as knowing and motivating subordinates—are important components of the USU curriculum. For example, in Operations Kerkeshner and Bushmaster, where students conduct field exercises simulating battle situations, they are given the unique opportunities to practice all these leadership components with direct observation by superiors. Other important leadership threads are covered in our military medicine courses, such as military contingency medicine and introduction to medical planning, as well as other components of our military unique curriculum that occur throughout all four years of medical school at USU.11

USU physicians who have achieved the rank of flag officer reported that leadership should be considered a core competency in medical school and graduate medical education. We concur with this suggestion, especially given the increasing complexity of health care in the United States and abroad, as well as the increasingly scarce resources for providing this care. Educating physician leaders is clearly one strategy to help ensure the effective delivery of health care on a regional and national basis.

Based on the findings reported here, we believe that medical schools wishing to emphasize leadership may want to consider the following educational initiatives. First, focus on activities that foster our identified leadership components, which are consistent with transformational leadership—effective communication, empathy, integrity, ambition and energy, expertise, as well as knowing and motivating subordinates. For example, in the preclinical period, seminars on transformational leadership followed by small group activities could be instituted. In the clinical period, emphasize the importance of teamwork in caring for patients, focusing on our identified components, with the physician serving as a leader. Indeed, we believe these findings support the movement of U.S. medical schools to expose students to early clinical experiences where they can practice and subsequently demonstrate mastery of these components. Second, we believe that perhaps the best model for achieving mastery of these components is the longitudinal clerkship, which enables a student to work with the same group of faculty who can provide guidance and feedback on these components of leadership in the context of clinical practice. A third recommendation is the idea of placing increased emphasis on mentoring outside of the “formal curriculum.” The role of mentors was an important theme identified by our flag officers, and medical educators could be proactive in establishing mentoring relationships early in the curriculum which are periodically reinforced throughout medical school. A fourth suggestion is to establish leadership as a core competency in the medical school and graduate medical education in the curriculum. Currently, leadership is not explicitly identified in the Accreditation Council of Graduate Medical Education competencies, and adding leadership as a competency with transformational leadership principles as a domain of competence to be addressed could improve
leadership education. Doing so would require that curriculum planners have a “thread” of leadership education and experiences throughout their time in medical school. Finally, advocate leadership as an important component of residency (registrar) training.

The present study had several limitations. Flag officers were surveyed because they represent the pinnacle of leadership in our system. However, with this approach, reporting bias could play an important role. The specific examples cited and rich qualitative responses, however, suggest that this may not be the case. Second, other established leaders at ranks other than flag officer could have differing opinions. Third, we sampled a small number of participants and the response rate for participants was suboptimal. The fact that we achieved saturation with each response, however, suggests that we would not have identified additional themes if more individuals were included. Fourth, we only queried USU graduates. Having gone through our curriculum, these individuals are arguably the most credible sources for USU’s role in their leadership position as well as suggestions for improvement. Some could argue this also makes them the most biased. As such, we believe that triangulating this data with non-USU graduates at the flag or other senior officer levels could provide another important perspective. Fifth, we had flag officers conduct a self-assessment using survey methodology. Given that our flag officers are distributed throughout the United States, this was the most feasible approach to addressing our research questions.

In conclusion, our findings provide additional data to support the claim that USU is meeting its societal obligation in terms of training future leaders in military medicine. Flag officers represent a distinct group of leaders, and the MLQ results, with high scores for transformational leadership, further support their expertise in leadership. The characteristics of effective leadership identified by these flag officers are consistent with the existing literature from other fields. Integrating the tenets of transformational leadership into the curricula could facilitate the development of physician leaders. Finally, participants enthusiastically endorsed USU’s role in their leadership skill, and they also suggested a number of means that USU could consider improving our curriculum in the future.

APPENDIX: LEADERSHIP SURVEY TO FLAG OFFICERS

Multifactor Leadership Questionnaire-6S

Instructions: This questionnaire provides a description of your leadership style. Twenty-one descriptive statements are listed below. Judge how frequently each statement fits you. The word others may mean your followers, clients, or group members.

Key: 5 = frequently, if not always; 4 = fairly often; 3 = sometimes; 2 = once in while; 1 = not at all.

1. I make others feel good to be around me.
2. I express with a few simple words what we could and should do.
3. I enable others to think about old problems in new ways.
4. I help others develop themselves.
5. I tell others what to do if they want to be rewarded for their work.
6. I am satisfied when others meet agreed-upon standards.
7. I am content to let others continue working in the same way as always.
8. Others have complete faith in me.
9. I provide appealing images about what we can do.
10. I provide others with new ways of looking at puzzling things.
11. I let others know how I think they are doing.
12. I provide recognition/rewards when others reach their goals.
13. As long as things are working, I do not try to change anything.
14. Whatever others want to do is OK with me.
15. Others are proud to be associated with me.
16. I help others find meaning in their work.
17. I get others to rethink ideas that they had never questioned before.
18. I give personal attention to others who seem rejected.
19. I call attention to what others can get for what they accomplish.
20. I tell others the standards they have to know to carry out their work.
21. I ask no more of others than what is absolutely essential.

Open-ended questions:
1. What do you think are the key characteristics needed to be an effective leader in medicine and/or public health?
2. What experiences helped prepare you for leadership? Please be as specific as you can.
3. Do you think your USU education prepared you for your leadership positions? Why or why not?
4. What specific advice would you give USU’s educational leadership for how they might better educate future leaders?

REFERENCES
Where are They Now? USU School of Medicine Graduates After Their Military Obligation is Complete

LTC Kent J. DeZee, MC USA*; Steven J. Durning, MD, PhD†; Ting Dong, PhD‡; CDR Anthony R. Artino Jr., MSC USN‡; COL William R. Gilliland, MC USA (Ret.)*; Donna M. Waechter, PhD†; Col John E. McManigle, USAF MC (Ret.)‡; LTC Aaron Saguil, MC USA§; David F. Cruess, PhD†; John R. Boulet, PhD∥

ABSTRACT The Uniformed Services University’s (USU) F. Edward Hébert School of Medicine was chartered in 1972, with the goal of providing high-quality physicians for the Uniformed Services. In exchange for their education, USU graduates incur an active duty service obligation, after which they may choose to stay on active duty or transition to civilian practice. The purpose of this study is to describe the practice characteristics of USU graduates after this obligation has been completed in order to determine the societal benefits during this phase of their careers. To accomplish this purpose, we performed a retrospective cohort study of the first 20 years of USU graduates (1980–1999). We used the American Medical Association Physician Masterfile to determine the graduates’ current practice location and characteristics, as well as their board certification status. Of these 2,760 graduates, nearly all (91%) were involved in active clinical practice in over 100 self-declared specialties, the vast majority (89%) practiced in locations other than the immediate vicinity of the medical school (i.e., Maryland and the District of Columbia), and most still worked for the federal government (71%). Finally, USU graduates in full-time clinical practice had a board certification rate of 93%, which was better than the average of all other graduates of U.S. Medical Schools (88%) in the same time period. Thus, it seems USU is attaining its goal of producing high-quality physicians who continue to benefit the nation after their service obligation has been completed, with many still in federal service.

INTRODUCTION

The Uniformed Services University (USU) was chartered 40 years ago when President Nixon signed the legislation on September 21, 1972.¹ The major push for the new school came from Congressman F. Edward Hébert, who for years advocated for a “West Point for doctors,” with the goal to train physicians for service in the military and public health service.² The first class of 32 medical students matriculated in 1976 on the Bethesda, Maryland campus and graduated in 1980. Class size gradually increased, as faculty and physical space became available, to the current size of approximately 170 per year. Students matriculating to USU may come from the civilian U.S. population or be members of the U.S. military. Upon entering USU, students are commissioned as junior military officers (Second Lieutenant for the Army and Air Force, Ensign for the Navy and Public Health Service). Each USU class is comprised of slightly more than 1/3 Army students, followed by nearly equal numbers of Navy and Air Force students, along with a few (2%) Public Health Service students.

The USU program in undergraduate medical education is tuition free. Medical students receive a salary and comprehensive health benefits for themselves and their family members and are provided a stipend for housing. In return, students incur a 7-year active duty service obligation to the U.S. Government, which is served after medical school when not in further training (e.g., residency or fellowships). For example, a student matriculating to USU in 1976 and selecting pediatrics (students may apply for any specialty in the military match) would graduate from medical school in 1980, complete his/her 3-year residency in 1983, then serve his/her active duty obligation from 1983 to 1990. At this point (assuming the physician has no other sources of obligation, such as a military academy or Reserve Officers’ Training Corps scholarship for their college or university years), the physician may choose to stay on active duty or separate from active military service. Those separating immediately after the 7-year commitment incur a 6-year individual ready reserve (IRR) commitment. After 10 years of active service, there is no IRR requirement. While in the IRR, the physician is free to practice medicine as a civilian.

From a medical perspective, transitioning from military to civilian practice is seamless as USU is an accredited medical school and all military physicians must maintain an active license to practice medicine in the Military Health System or Public Health Service from the same state licensing boards as...
USU Graduates After Service Obligation Completion

civilian physicians. Therefore, other than the requirement to obtain a new medical license if the physician intends to practice in a state where he/she is currently unlicensed, there are no new administrative, regulatory, or training requirements for USU physicians to transition to civilian medical practice.

Based on our school’s experience, nearly all USU graduates will choose to continue to use their medical skills once they have completed their service obligations. Thus, society will realize additional benefits after USU graduates’ Uniformed Service is over, improving the “return on investment.” The goal of this report is to describe the practice patterns of USU graduates after their initial obligation is complete, with particular attention to continued government service, geographic distribution of graduates, and the quality of their training (as measured by board certification rates).

METHODS

Study Design and Population

This study is a retrospective cohort of all medical school graduates of USU in Bethesda, Maryland. As the goal was to describe graduates after their initial 7-year active duty service obligation was completed, we only included the first 20 years of graduates, from 1980 (the initial class) until the class of 1999. We did not consider more recent graduates as we wished to allow at least 12 years from graduation to permit graduate medical education training and obligatory active duty service time to be completed (e.g., 5 years of general surgery residency followed by 7 years of service). Naturally, this definition will inevitably include some individuals who have not completed their initial obligatory service (e.g., a neurosurgeon graduating in 1999), but it was not possible to reliably determine which of those more recent graduates had not yet completed their obligation. Hence, we elected to select a cohort that would include a few of these individuals in the study, rather than choosing one with an earlier graduation date as an earlier date would exclude many more graduates who have left military service and would be eligible for this study. In order to provide a reference standard, we defined a comparison cohort as all other graduates of U.S. medical schools in the same time period.

Data Sources

We used the 2011 American Medical Association (AMA) Physician Masterfile,2 which was queried for all graduates of U.S. medical schools from 1980 to 1999. The following data fields were used (1) type of practice (direct patient care, administration, medical teaching, medical research, resident, nonpatient care, retired, semiretired, temporarily not in practice, not active for other reasons, unclassified); (2) major professional activity (office-based practice, full-time hospital staff, administration, medical teaching, resident, research, locum tenens, inactive, other); (3) present employment (solo practice, two-physician practice, locum tenens, group practice, health maintenance organization, medical school, nongovernmental hospital, city/county/state hospital, hospital, military government hospital, Public Health Service, Veterans Affairs, other federal agency, other nonpatient care, unclassified); (4) primary self-designated specialty; (5) medical school; (6) year of graduation; (7) board certification (the version of the Masterfile we used included board certification data from the American Board of Medical Specialties, which is compiled and maintained in a proprietary and copyrighted database by the American Board of Medical Specialties); and (8) state of practice. We defined the type of practice by combining the first two items (type of practice and major professional activity) as there was a great deal of overlap. We used the present employment item and the primary self-designated specialty as the physician’s employment characteristic and specialty, respectively. As current board certification is linked to the quality of the physician,3,4 we used this metric to define physician quality but restricted the cohort to only those having a major clinical activity of office-based practice or full-time hospital staff, as those who were not in full-time practice in this manner may have intentionally let their board certification lapse.

Statistical Analysis

We compared board certification rates between the USU cohort and the other U.S. medical school graduates by using a χ² test; we set statistical significance at p < 0.05. The USU Institutional Review Board approved this study.

RESULTS

The AMA Physician Masterfile identified 2,760 USU graduates from 1980 to 1999 (inclusive), which is 99% of the 2,793 graduates during this time period as per USU’s records. Practice locations were available for all but 3 graduates. USU graduates were located in all 50 states, the District of Columbia, and Guam. Six states individually had more than 5% of the graduates and together had 52% of all graduates: California (n = 320, 12%), Texas (n = 284, 10%), Maryland (n = 268, 10%), Virginia (n = 235, 9%), Florida (n = 163, 6%), and Washington (n = 151, 5%). From another perspective, the vast majority of the graduates (n = 2,443, 89%) had practice locations in states or territories other than those in the immediate vicinity of USU (i.e., Maryland and the District of Columbia).

Self-designated specialty was available for all but 55 graduates. The graduates identified with 118 different specialties. The most popular was family medicine (n = 441, 16%), followed by anesthesia (n = 167, 6%) and emergency medicine (n = 157, 6%). The top 15 specialties are listed in Table I.

Practice type was available for 2,680 graduates, the vast majority of whom were in direct patient care (n = 2,442, 91%), followed by administration (n = 120, 4%), and retired/semiretired/temporarily out of practice (n = 43, 2%). Other practice types included medical teaching (n = 34, 1%), residency (n = 23, 1%), and research (n = 18, 1%). Of those in
TABLE I. Top 15 Self-Identified Specialties for Graduates of the USU School of Medicine, Graduating Classes 1980–1999

<table>
<thead>
<tr>
<th>Specialty*</th>
<th>n, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family Medicine</td>
<td>441, 16</td>
</tr>
<tr>
<td>Anesthesiology</td>
<td>167, 6</td>
</tr>
<tr>
<td>Emergency Medicine</td>
<td>157, 6</td>
</tr>
<tr>
<td>Diagnostic Radiology</td>
<td>149, 6</td>
</tr>
<tr>
<td>Orthopedic Surgery</td>
<td>143, 5</td>
</tr>
<tr>
<td>Pediatrics</td>
<td>132, 5</td>
</tr>
<tr>
<td>Obstetrics and Gynecology</td>
<td>129, 5</td>
</tr>
<tr>
<td>General Surgery</td>
<td>114, 4</td>
</tr>
<tr>
<td>Ophthalmology</td>
<td>92, 3</td>
</tr>
<tr>
<td>Psychiatry</td>
<td>85, 3</td>
</tr>
<tr>
<td>Internal Medicine</td>
<td>84, 3</td>
</tr>
<tr>
<td>General Preventive Medicine</td>
<td>78, 3</td>
</tr>
<tr>
<td>Dermatology</td>
<td>69, 3</td>
</tr>
<tr>
<td>Otolaryngology</td>
<td>58, 2</td>
</tr>
<tr>
<td>Anatomic/Clinical Pathology</td>
<td>48, 2</td>
</tr>
</tbody>
</table>

*Specialty listings do not include subspecialties, e.g., adult cardiology for internal medicine, pediatric gastroenterology for pediatrics, etc. Only the top 15 specialties are listed as the remaining 28% of graduates listed one of more than 100 other specialties.

the direct patient care (n = 2,442), most were full-time hospital staff (n = 1,512, 62%), followed by office-based practice (n = 927, 38%), and very few were locum tenens (n = 3, 0.1%).

The federal government remained the primary employer for most of the graduates. Of the 2,760 in the cohort, 2,286 (83%) had available data for the primary employer. Of these, 1,614 (71%) worked for the federal government in an Army (n = 612), Navy (n = 534), Air Force (n = 366), Public Health Service (n = 71), Veteran’s Administration (n = 14), or other federal facility (n = 17). The next largest group (n = 605, 26%) consisted of nongovernmental employers, including group practice (n = 452), solo practice (n = 79), nongovernmental hospitals (n = 57), two-physician practices (n = 10), locum tenens (n = 4), and health maintenance organization (n = 3). Another 48 graduates (3%) worked for a city or state medical facility. Finally, 19 (1%) worked for other organizations.

The quality of USU graduates, as measured by board certification rates, was excellent. During the study time period (1980–1999 graduates), there were 2,439 USU graduates and 418,452 other U.S. graduates with their major clinical activity of office-based practice or full-time hospital staff. Of these, 2,269 (93%) USU graduates were board certified compared to 366,255 (88%) of the other U.S. graduates, which was statistically significant (p < 0.0001, \( \chi^2 \)).

DISCUSSION

Our study revealed that nearly all USU graduates remain in active practice after their service obligations are complete. Further, nearly three-quarters of them continue to serve the public via local, state, or federal government employment. The graduates are not simply returning to the location of their school, however, as 89% of graduates did not practice near USU and all 50 states were represented. Last, but perhaps most importantly, USU graduates have better board certification rates compared to the average among graduates of other U.S. medical schools.

The vision of Congressman Hébert was to provide a new source for high-quality physicians to provide care for the nation’s military. The results from this study strongly support that USU is meeting this goal. USU graduates are high-quality physicians as measured by board certification rates, which is a well-supported measure for physician quality. Further, the majority of USU graduates stay in federal service, either on active duty or as civilians, after their initial service obligation is complete. Thus, the federal government keeps high-quality physicians within the system, which is also beneficial as the start-up training needs would be minimal since the physicians would already be familiar with the specific aspects of military medicine, such as policies for medical retention, illness absence forms, etc. The graduates also provide stability to the workforce through “institutional memory.” Since these graduates have transitioned to civilian practice, they would be able to remain at a location longer than most active duty military physicians, who frequently (approximately every 3 years) change duty locations, and thus must spend time and energy to learn the local system.

Since the school is federally funded, the entire nation should benefit from the investment, not just the areas around the school. Our study indicates that this is indeed the case as all 50 states have USU graduates and 89% of graduates do not practice in the immediate area around the school. Thus, it appears that USU graduates’ eventual practice location is influenced by factors other than the state of their undergraduate medical education. Previous work in this area, though dated, suggests that 40% of physicians practice in the same state as their medical school, with higher rates for public schools (48%) than private schools (33%). These authors suggest that state of birth, site of graduate medical education, climate, reputation of medical school, and physician density also influence practice location. USU students come from essentially all areas of the United States, thus graduates might be returning to their state of birth (or, alternatively, to their state of residency immediately before USU matriculation). Location of graduate medical education is also likely to be an important factor, as there are only two locations in the immediate vicinity of the school (Walter Reed National Military Medical Center in Bethesda, Maryland, and Fort Belvoir Community Hospital in Fort Belvoir, Virginia) for graduate medical education since all military USU graduates must perform their residency in the military system. The vast number of graduates will train in military training programs elsewhere, such as Hawaii, Washington, California, Texas, Georgia, Florida, Ohio, and Virginia. Finally, the location of graduates’ final assignment while on active duty may also influence their ultimate practice location.

Our study had several limitations. Unfortunately, we could not precisely define the cohort to exclude graduates who still had service obligations after 12 years; so, our estimate of...
continued government service is likely slightly overestimated. Second, the AMA Physician Masterfile, although robust, did have some missing data, so nonresponse or misclassification bias is possible. The board certification rate is assumed to be fairly reliable, as we have previously found board certification rates of 99% for the first 10 years of USU graduates. That said, the current report’s board certification rate is likely lower than the previous study (93% vs. 99%) as the previous study used self-reported certification and had an incomplete response rate.

In conclusion, USU School of Medicine graduates are high-quality physicians who continue to serve the country in multiple ways long after their initial obligation is complete. From this perspective, the vision of Congressman Hébert has clearly been realized.

REFERENCES
Using Functional Neuroimaging Combined With a Think-Aloud Protocol to Explore Clinical Reasoning Expertise in Internal Medicine

Steven J. Durning, MD, PhD*; John Graner, PhD*; CDR Anthony R. Artino Jr., MC USN*; Louis N. Pangaro, MD*; Thomas Beckman, MD†; Eric Holmboe, MD‡; Terrance Oakes, PhD*; COL Michael Roy, MC USA*; Gerard Riedy, MD, PhD§; CPT Vincent Capaldi, MC USA§; CPT Robert Walter, MC USA§; Cees van der Vleuten, PhD∥; Lambert Schuwirth, MD, PhD∥

ABSTRACT Background: Clinical reasoning is essential to medical practice, but because it entails internal mental processes, it is difficult to assess. Functional magnetic resonance imaging (fMRI) and think-aloud protocols may improve understanding of clinical reasoning as these methods can more directly assess these processes. The objective of our study was to use a combination of fMRI and think-aloud procedures to examine fMRI correlates of a leading theoretical model in clinical reasoning based on experimental findings to date: analytic (i.e., actively comparing and contrasting diagnostic entities) and nonanalytic (i.e., pattern recognition) reasoning. We hypothesized that there would be functional neuroimaging differences between analytic and nonanalytic reasoning theory. Methods: 17 board-certified experts in internal medicine answered and reflected on validated U.S. Medical Licensing Exam and American Board of Internal Medicine multiple-choice questions (easy and difficult) during an fMRI scan. This procedure was followed by completion of a formal think-aloud procedure. Results: fMRI findings provide some support for the presence of analytic and nonanalytic reasoning systems. Statistically significant activation of prefrontal cortex distinguished answering incorrectly versus correctly (p < 0.01), whereas activation of precuneus and midtemporal gyrus distinguished not guessing from guessing (p < 0.01). Conclusions: We found limited fMRI evidence to support analytic and nonanalytic reasoning theory, as our results indicate functional differences with correct vs. incorrect answers and guessing vs. not guessing. However, our findings did not suggest one consistent fMRI activation pattern of internal medicine expertise. This model of employing fMRI correlates offers opportunities to enhance our understanding of theory, as well as improve our teaching and assessment of clinical reasoning, a key outcome of medical education.

INTRODUCTION Expertise in clinical reasoning improves patient care and involves establishing a diagnosis and an individual care plan; it is an essential outcome of medical education. Clinical reasoning encompasses the sum of the thinking and decision-making processes within clinical practice. However, investigating clinical reasoning involves making inferences about internal mental processes. Despite being at the core of clinical expertise, the difficulty in assessing clinical reasoning greatly challenges our understanding.

Until recently, researchers have largely avoided investigating the functional activation of complex psychological constructs such as cognition. However, functional magnetic resonance imaging (fMRI) enables the assessment of regional brain activity by detecting changes in oxygenation reflecting perfusion. Regions that use more oxygen to process stimuli (an “artifact” of thought) evoke a rapid feedback loop that dilates blood vessels to more markedly increase oxygenation to such areas. These changes can be reliably detected by fMRI (blood oxygen level-dependent signal). Pairing fMRI with a think-aloud procedure, whereby participants are asked to vocalize whatever comes to mind as they work through a task such as solving a clinical case, could facilitate identification of links between cognition and its neural substrate, providing insight into the cognitive processes underlying clinical reasoning.

Clinical reasoning, like many outcomes in medical education, is difficult to measure. One current gold standard for physician performance is multiple-choice exams. Indeed, to become a practicing physician, one must pass a series of multiple-choice question (MCQ) examinations.

A primary goal of the Long-Term Career Outcome Study is tracking the performance (outcomes) of our graduates. Given the limitations of MCQ examinations, we therefore investigated fMRI neuroimaging patterns of our graduates and other graduates who practice in the Department of Defense to enhance our understanding of what these MCQ examinations are measuring. Such understanding could ultimately help us to establish the limits of these MCQ examinations and plan new outcome measures that could be used in our and other systems.

Educational Theory Results of empirical studies allow for the development and refinement of educational theory, and indeed, researchers

*Uniformed Services University of the Health Sciences, 4301 Jones Bridge Road, Bethesda, MD 20814.
†Mayo Clinic, 200 First Street SW, Rochester, MN 55905.
‡American Board of Internal Medicine, 510 Walnut Street, Suite 1700, Philadelphia, PA 19106.
§Walter Reed Army Medical Center, 6900 Georgia Avenue, NW # 77, Washington, DC 20307.
∥Maastricht University, Minderbroedersberg 4-6, 6211 LK Maastricht, The Netherlands.
have called for more studies assessing educational theory with neuroimaging.\textsuperscript{5–7} Current leading educational theories, based on decades of research, propose that physicians use two primary systems for clinical reasoning in patient care\textsuperscript{8}; a rapid, automatic, and unconsciously controlled nonanalytic system and a slow, rule-based, and consciously controlled analytic system.\textsuperscript{3,8}

Nonanalytic reasoning (or “pattern recognition”) is believed to be used predominantly by experts, whose knowledge is presumably more highly organized in memory.\textsuperscript{9–11} Conversely, with analytic reasoning, physicians consciously select diagnoses and explicitly seek additional data to confirm or refute these diagnoses. It is believed that physicians rely on analytic reasoning when they are unsure of the diagnosis (i.e., guessing or when there is poor problem representation) or when selecting between multiple potential diagnoses. Although experts probably use both systems, many believe that nonanalytic reasoning predominates.\textsuperscript{4,9–11}

**fMRI and Reasoning**

Until recently, fMRI had been used primarily to study simple on/off stimuli such as pictures and words, whereas more complex mental processes, such as problem solving, have been avoided for several reasons.\textsuperscript{4} One notable reason addressed in this article is that a road map—such as educational theory—is necessary to interpret the multifaceted neural responses.\textsuperscript{4}

Simple task fMRI studies\textsuperscript{12} have shown both increased cortical tissue activity devoted to tasks after training (“specialized” to the task) and functional reorganization,\textsuperscript{3} representing restructuring of brain activity or activation of different brain areas during progressive stages of learning. This suggests two relevant themes: (1) learning appears to be localized and highly specialized and (2) learning can produce decreased activation reflecting improved data processing efficiency and also increased activation reflecting greater cortical tissue devoted to the task.\textsuperscript{13–15} Therefore, decreased activation may represent nonanalytic reasoning, whereas increased cortical activation may represent analytic reasoning.

Studies from other disciplines suggest that working memory processing may occur in multiple brain regions\textsuperscript{16} and that there may be both slow and fast systems. The slow system is often referred to as the “control network” or “domain general network,”\textsuperscript{17} and may equate with high effort (analytic) reasoning.\textsuperscript{12,17} This system is believed to involve portions of the frontal, cingulate, and parietal cortices. The fast system is believed to overlap these areas, presumably because it requires support from the control network before practice (and learning) occurs.\textsuperscript{15} The fast system is thought to involve the prefrontal, anterior cingulate, and parietal-temporal regions.\textsuperscript{18} The caudate and precuneus have also been implicated in learning and memory.\textsuperscript{19}

Educational theory, substantiated by empirical findings, has improved our understanding of clinical reasoning. Given that clinical reasoning is not directly observable, there is a need to employ new methods to enhance our understanding of this process, which is paramount to patient care. New insights to inform leading theory could also shape future investigation of clinical reasoning. We therefore paired fMRI with a think-aloud protocol to assess expert performance on validated measures of clinical reasoning. We used established educational theory to assist with neuroimaging interpretation. Specifically, we presented board-certified internists with vignette-based MCQs in an fMRI scanner to examine neuroimaging correlates of analytic and nonanalytic reasoning and expertise in medicine.

We also explored potential differences in brain activity with think-aloud data consistent with guessing vs. nonguessing as well as deep versus superficial problem representation. This was done as guessing and superficial problem representation would be expected to correlate with analytic reasoning. Deep problem representation is associated with expertise in multiple fields outside of medicine.\textsuperscript{3} We used each of these independent analyses (correct, deep) because studies suggest that they are markers of expertise.\textsuperscript{3,8}

Further, we contrasted three phases in the fMRI scanner: reading, answering, and reflecting. We sought to capture the cognitive process of arriving at a diagnosis. Reading, answering, and reflecting are hypothesized to involve different cognitive processes and therefore may involve different functional neuroactivation. Reading, for example, can involve different levels of comprehension, and data processing and understanding can vary from literal meaning to deep inference, which is different from the act of answering. Reflecting can involve metacognition (awareness of one’s reasoning process).\textsuperscript{3}

Our objective in this study was to examine fMRI correlates of analytic and nonanalytic reasoning theory, with the hypothesis that such neuroimaging differences exist between analytic and nonanalytic reasoning and can be observed by fMRI; given prefrontal cortex involvement in executive function, we expected to observe differences in this area. We contrasted performance with answering correctly vs. incorrectly, guessing vs. nonguessing, and deep vs. superficial problem representation. Our outcomes measures included: performance on MCQs, think-aloud protocol analysis (guessing and problem representation), and fMRI patterns of activation. We compared answering and reflecting phases for each participant, using the reading phase as the baseline, looking for differences in the activation patterns in areas of the brain (activation patterns involve changes in oxygenation which is detected by fMRI and reflects changes in regional brain perfusion). The fMRI outcome was a significant difference in the activation of an area of the brain between these phases (reading and reflecting). Although we focused on the prefrontal cortex for reasons stated above, all areas of the brain were included in the analysis.

**METHODS**

**Participants**

Board-certified internal medicine physicians (general internists and internal medicine subspecialists) affiliated with the
Uniformed Services University of the Health Sciences (USU) participated in the study following informed consent. We included this group as they all have passed the American Board of Internal Medicine (ABIM) certifying examination (an indicator of expertise); they served as our expert group. Exclusion criteria included the presence of surgical metal devices or other pieces of metal as this could lead to medical complications because of the magnet in the fMRI scanner, an inability to complete an fMRI because of anxiety or claustrophobia as this would prevent us from having participants complete the task of answering all 32 questions and would also be expected to bias thought processes during the task, taking calcium channel blockers as this impacts cerebral blood flow of participants and fMRI looks specifically at blood flow changes, and being in practice as an attending physician >20 years because fMRI findings can vary with age. Further, work suggests that expertise wanes overtime (beyond the 5th decade) and we were trying to capture the “true age. Further, work suggests that expertise wanes overtime (beyond the 5th decade) and we were trying to capture the “true age of experts.”3 Our study was approved by the Institutional Review Boards at USU and Walter Reed Army Medical Center.

**Measurements**

**Multiple-Choice Questions**

We used MCQs from the ABIM and National Board of Medical Examiners (NBME) as well-validated measures for expertise in diagnostic reasoning.20 These certification and licensing questions have been tested on thousands of subjects and have been modified, as needed, to optimize validity. We selected questions from the fields of cardiology and rheumatology. Higher-order MCQs were explicitly selected as these require physicians to integrate and synthesize data. Participants answered a total of 32 questions based on previous statistical analysis: 8 easy and 8 hard NBME items, 8 easy and 8 hard ABIM items (Maintenance of Certification MCQ). We selected questions that fit on a single computer screen, and the questions contained words only. The items selected also had favorable discrimination indices when previously administered. The opportunity for participants to push hand-held buttons for answer options “A” to “E” also made the MCQs ideal for use in an fMRI scanner, which precludes speaking as jaw motion impairs image interpretation.

**Think-Aloud Protocol**

Participants completed a formal think-aloud procedure for all items immediately following completion of the items in the scanner. Participants received formal training on the think-aloud procedure before entering the scanner.321 Specifically, participants were asked to describe in words their process of answering the questions (to “think aloud”) while reviewing each item just answered in the fMRI scanner. Participants were handed a printout of items with “A” to “E” answer options; however, the actual responses provided by individuals while in the scanner were not included. Each think-aloud procedure was performed by a trained investigator and audio recorded. Guessing was determined to be present when the participant voiced phrases during the think aloud such as “I have no idea” or “I am guessing.” Guessing was predicted to involve analytic reasoning as answer options would have to be compared and contrasted.

**fMRI Measurement**

Subjects were scanned on a 3T 750 MRI scanner (General Electric, Milwaukee, Wisconsin) with a 32-channel head coil. Details of fMRI measurement and fMRI analyses are included in the Appendix. An fMRI task presentation of the 32 questions was created using E-Prime software (Psychology Software Tools, Sharpsburg, Pennsylvania) and displayed via a goggle system (Nordic NeuroLab, Milwaukee, Wisconsin) while each participant was in the fMRI scanner. The questions were presented in a random order for each subject over the course of four fMRI acquisition runs, with 8 questions per run. The exact length of each run varied, depending on the amount of time each subject took to progress through the reading and answering phases of each question. The mean run length (± SD) was 392 ± 62 seconds. Subjects received pretraining to acquaint them with aspects of answering MCQs in the scanner, including the method of question presentation and the correct use of the response buttons. Participants were also trained in the think-aloud procedure,320 for which a modified version was used in the reflection phase for items in the scanner; that is, participants were asked to think aloud (reflect) without speaking during the reflection phase. For each MCQ, participants were given up to 60 seconds to read the question and each question was projected in three phases. In the first phase, the question stem appeared (reading phase), ending with “what is the most likely diagnosis?” or a related diagnostic question, but not displaying answer options “A” to “E.” Each participant could push any button during the reading phase to move on to the answer options (the second or answering phase). Subjects were then given 7 seconds to choose an answer option using the finger response buttons. The final phase (reflection phase) was then entered, where participants were instructed to think, without speaking, about how they arrived at the diagnosis (“how did you establish the diagnosis for this item”). This phase lasted 14 seconds. The time for each phase was chosen based on typical time for answering MCQs on high-stakes examinations and literature suggesting estimated time needed to answer and think aloud.

**Data Analysis**

The Appendix provides a detailed description of the data analysis. Of the 18 subjects, 1 subject’s fMRI data set was excluded from further analysis because of motion artifact while in the scanner.

Initial activation analysis was performed using a general linear model approach concatenating the four data sets for each subject. Hemodynamic response estimates were modeled for two of the three question phases (answer, reflect). As we sought to capture the construct of reasoning, we compared
phases (answering, reflecting) to elicit task-specific findings. Thus, the “reading” time periods were treated as a baseline in our analyses (activations reported in this work are therefore relative to the reading period).

Each subject’s response to each of the 32 questions was categorized based on three characteristics: correct vs. incorrect, guess vs. nonguess, and deep vs. superficial; the latter two categories were arrived at by analyzing audio-recorded transcripts of the postscan think-aloud procedure. For all group analyses, age and gender were included as covariates.

For determining deep versus superficial problem representation and guessing versus not guessing, three coders independently reviewed a sample (n = 4) of the audio-recorded think-aloud protocols of participants and then discussed results. Agreement on categorization was discussed after reviewing these transcripts. Two coders then reviewed all the transcripts in their entirety, which were then coded and analyzed independently. A $\kappa$ statistic was calculated to assess agreement between coders.

RESULTS
A total of 17 board-certified internists were included in our analysis. The mean number of correct responses was 18.5/32 questions (range = 15–25). There were 15 men and 2 women. Mean age was 39.5 ± 7 (range = 32–51 years).

All participants were board certified in internal medicine, and all but one (a cardiology fellow) were board certified in their respective subspecialty. Significant fMRI activation locations are shown in Table I.

Inter-rater agreement for the think-aloud transcripts was considered acceptable. For guess versus nonguess phases, $\kappa$ was 0.96, and for deep versus superficial problem representation, $\kappa$ was 0.72.

Examining neuroimaging activation locations for correct versus incorrect answers revealed one statistically significant difference—greater activation of the prefrontal cortex with answering incorrectly versus correctly (corrected $p < 0.01$; Fig. 1). Although there were no other significant differences in brain activation when making comparisons across the group for correct versus incorrect answers, a nonsignificant cluster in the right temporal gyrus was observed. There were no activation differences in the reflection phase with correct versus incorrect responses. Guessing was established using the formal think-aloud data. Two brain areas were significantly less activated (corrected $p < 0.01$) with guessing versus not guessing during the answering phase: bilateral precuneus and left midtemporal gyrus (Fig. 2). There were no significant differences (corrected $p < 0.01$) in guessing versus not guessing in the analysis of the reflection phase.

DISCUSSION
Our study builds on the fMRI and educational theory literature, and our findings provide some validity evidence because they replicate a number of findings from fMRI research in other academic realms. These findings provide preliminary data regarding functional neuroimaging processes involved in answering clinical reasoning MCQ examinations, an important outcome in medical education.

| TABLE I. fMRI Activation Patterns With Each Comparison or Contrast; Locations Where the Contrast Led to Significant (and 1 Nonsignificant) Results (Comparing Reading and Answering) and Three-Dimensional Co-ordinates Where Contrasts Were Seen. |
|---|---|---|---|---|
| Contrast | Location | Montreal Neurological Institute Coordinates | Corrected p |
| Incorrect vs. Correct | Medial Prefrontal Cortex | $-2$ | $-47$ | $-4$ | $<0.01$ |
| Right Superior Temporal Gyrus | $46$ | $9$ | $-21$ | $>0.1$ |
| Guess vs. Nonguess | Medial FrONTAL Gyrus | $-1$ | $1$ | $47$ | $<0.01$ |
| Precuneus | $1$ | $50$ | $49$ | $<0.01$ |
| Left Midtemporal Gyrus | $57$ | $55$ | $7$ | $<0.01$ |

Contrast = comparison; incorrect vs. correct and guess vs. nonguess obtained from think-aloud analysis.
validity to our findings of the use of the prefrontal cortex in internal medicine expertise.

**Precuneus**

Structures in the basal ganglia have recently been implicated as playing an important role in learning and memory. Recent fMRI studies connect the precuneus with self-referential goal-directed actions as well as memory retrieval. These functions are consistent with our finding of increased activation when not guessing on an item and with reflection on answered items. These patterns of activation would not be expected by control network theory, and the precuneus may represent a focus of expertise (i.e., nonanalytic reasoning or both analytic and nonanalytic reasoning as studies suggest experts can use both systems).

**Temporal Lobe**

We observed accentuated activation of the medial temporal lobe when not guessing; like the precuneus, this area may represent a focus for clinical reasoning expertise. Several studies have shown that damage to the medial temporal lobe is associated with an impaired ability to learn and remember new information (i.e., anterograde amnesia). This finding has been replicated in other studies, and the parietal-temporal area has been implicated in reasoning about meaningful content. This result suggests that participants were activating learning areas when not guessing and perhaps were giving up on the item with guessing, resulting in less medial temporal lobe activation.

Several of our analyses revealed no statistically significant differences. We did not identify functional neuroimaging differences for think-aloud categorization of deep versus superficial problem representation. Unlike what could be anticipated from “control network” or “domain general network” theory, our findings suggest that expertise in internal medicine may not have a uniform pattern of fMRI activation. There may be several reasons for this.

Our study has several limitations. First, our study may have had insufficient power to detect fMRI differences. Second, we may have failed to find fMRI differences because of range restriction from sampling only experts, though our performance means and SDs for items suggest that this did not play a large role. Third, participant performance may have been affected by answering items in the scanner. A fourth limitation is the lack of a formal rest or inactive period to compare with the task phases; as we sought to capture the construct of reasoning, we believe that comparing phases (reading, writing, and reflecting) would result in more meaningful, task-specific findings, understanding that there can be features of answering or reflecting while being engaged in the process of reading an item. Further, “resting” state images are not necessarily resting/neutral and may not be reproducible within a subject. Fifth, we used a single group (board-certified internists). Comparing expert
responses with novices (i.e., senior medical students or interns), for example, may have provided a more robust range of functional neuroimaging responses to differentiate findings. Finally, even though performance means and SDs for items were small, there may be no association between reasoning and brain activity with the group analyses used (i.e., brain activation patterns may be highly variable within a given individual).

To our knowledge, this is the first study to utilize fMRI to investigate clinical reasoning expertise, and it yielded several notable findings that are relevant for future research in medical education. First, we did find limited fMRI evidence to support analytic and nonanalytic reasoning theory as we found functional differences with correct vs. incorrect answers and guessing vs. not guessing. Ultimately, findings from our small-scale study suggest there is no one consistent fMRI activation pattern of internal medicine expertise. Obtaining measurable functional neuroimaging differences may require larger sample sizes or contrasting groups at different levels of training, such as novices versus experts.

Our findings have several implications for medical education theory and research. First, we found limited functional neuroimaging evidence to support dual processing, but the distinction between analytic and nonanalytic processes proved to be more complicated than originally thought. Alternatively, expert physicians may use multiple methods while solving a given problem. Work has previously suggested that analytic processing, by its nature of being effortful, would be slow; this may not always be the case—there may be times that analytic processing is a bit faster and so well-practiced that it requires minimal effort, effort that the participant may not even consciously perceive. That is, analytic and nonanalytic theories may oversimplify an expert’s cognitive processing. Second, working memory for clinical reasoning may be diffuse and given the areas of activation, likely multimodal in experts. This work supports cognitive load theory’s notion of multimodal processing and the potential educational benefits of teaching with multiple modes (i.e., auditory and visual). Third, clinical reasoning in experts may require larger sample sizes or contrasting groups at different levels of training, such as novices versus experts.

Our findings would support the need for additional work in functional neuroimaging such as identifying patterns of activation that could potentially assist with teaching and remediating clinical reasoning and integrating existing theories on clinical reasoning and expertise. For example, comparing experts’ findings with individuals who have failed to meet board certification or recertification may reveal important functional neuroimaging differences. Unfortunately, current behavioral measures such as answering MCQs or completing post-encounter forms do not appear to be optimal indicators of proficiency with clinical reasoning. Our findings support emerging fMRI literature, indicating that fMRI may help educators with assessing complex cognitive processes, such as diagnostic reasoning. Ultimately, we believe the measurement approach of combining neuroimaging and think-aloud data with more traditional assessment measures offers great potential for advancing our understanding of diagnostic reasoning.

**APPENDIX : fMRI MEASUREMENT AND ANALYSIS METHODOLOGY**

**fMRI Measurement**

Acquisitions were performed using an echo-planar imaging sequence of 40 contiguous sagittal slices per brain volume (TR = 2000 ms, TE = 25 ms, flip angle = 60°, slice thickness = 4.0 mm). In-plane resolution was 3.75 × 3.75 mm (64 × 64 voxels).

During the imaging session, a high-resolution T1-weighted image was acquired for anatomical reference (three-dimensional GRE; TR = 6.6 ms, TE = 2.5 ms, flip angle = 12 degrees). This image consisted of 312 sagittal slices with a slice thickness of 0.6 mm and an in-plane resolution of 0.468 × 0.468 mm (512 × 512 voxels).

**fMRI Analysis**

All fMRI data were processed using the AFNI software package. Image preprocessing included: removal of the first three volumes (6 seconds) from each 4D time series, slice-time correction, motion correction, registration to the T1 anatomical image, smoothing with an 8 mm full-width-half-max Gaussian kernel, and conversion of voxel values to percent-change-from-mean rather than absolute intensity.

Neuroimaging activation analysis was performed using a general linear model (GLM) approach concatenating the four data sets for each subject. Hemodynamic response estimates were modeled for the answering and reflecting question phases. The “answer” times varied from question to question (depending on how quickly the participant answered) and were modeled with a gamma-variate function with variable duration and variable relative amplitude (amplitude variation was based on duration variation). The “reflection” time was constant at 14 seconds and was modeled with a nonvariable gamma-variate. The GLM analysis was used to determine the significance of these model time-courses within each voxel, using the “reading” phase as a “baseline” (thus, activations reported in this work are relative to the reading period). Time courses associated with preprocessing motion correction parameters were also included in the GLM regression analysis to regress out any further image intensity changes due to motion.

Six group-wide contrasts were run on a voxel-wise basis to determine if these distinctions correlated with differences in the “answer” or “reflection” response (contrasts were run between “answer, correct” and “answer, incorrect,” “reflection, correct” and “reflection, incorrect,” “answer, deep” and “answer, superficial”). These contrasts were established to determine if there were any brain regions that showed different activation responses based on response categorizations.
These contrasts were carried out using a linear mixed-effects (LME) model approach (using AFNI’s 3dLME.R program). Specifically, using the GLM method described above, initial voxel-wise statistical parametric maps (SPM) of BOLD signal intensity changes (betas) relative to the reading period baseline were calculated for the answering and reflecting periods for the MCQs answered correctly and the MCQs answered incorrectly (for each subject, this produced four SPMs: “Correct Answer,” “Incorrect Answer,” “Correct Reflection,” and “Incorrect Reflection”). Similar sets of SPMs were calculated based on whether or not the subject guessed on a MCQ or not (producing SPMs for “Guess Answer,” “NonGuess Answer,” “Guess Reflection,” and “NonGuess Reflection”). Finally, sets of SPMs were calculated based on whether the subject’s problem representation for a given MCQ was determined to be deep or superficial (producing SPMs for “Deep Answer,” “Superficial Answer,” “Deep Reflection,” and “Superficial Reflection”). A group-wide LME model analysis was then carried out for each MCQ categorization method (correct vs. incorrect, guess vs. nonguess, deep vs. superficial). In each analysis, the contrasts of interest were “Correct Answer vs. Incorrect Answer” and “Correct Reflection vs. Incorrect Reflection”). In all LME analyses, subject age and gender were included as regressors of no interest. Results of the group-wide LME model analyses were corrected for multiple comparisons using family-wise error correction (using AFNI’s LME model analyses were corrected for multiple comparisons using family-wise error correction (using AFNI’s LME model analyses were corrected for multiple comparisons using family-wise error correction (using AFNI’s LME model). These contrasts were carried out using a linear mixed-effects (LME) model approach (using AFNI’s 3dLME.R program). Specifically, using the GLM method described above, initial voxel-wise statistical parametric maps (SPM) of BOLD signal intensity changes (betas) relative to the reading period baseline were calculated for the answering and reflecting periods for the MCQs answered correctly and the MCQs answered incorrectly (for each subject, this produced four SPMs: “Correct Answer,” “Incorrect Answer,” “Correct Reflection,” and “Incorrect Reflection”). Similar sets of SPMs were calculated based on whether or not the subject guessed on a MCQ or not (producing SPMs for “Guess Answer,” “NonGuess Answer,” “Guess Reflection,” and “NonGuess Reflection”). Finally, sets of SPMs were calculated based on whether the subject’s problem representation for a given MCQ was determined to be deep or superficial (producing SPMs for “Deep Answer,” “Superficial Answer,” “Deep Reflection,” and “Superficial Reflection”). A group-wide LME model analysis was then carried out for each MCQ categorization method (correct vs. incorrect, guess vs. nonguess, deep vs. superficial). In each analysis, the contrasts of interest were “Correct Answer vs. Incorrect Answer” and “Correct Reflection vs. Incorrect Reflection”). In all LME analyses, subject age and gender were included as regressors of no interest. Results of the group-wide LME model analyses were corrected for multiple comparisons using family-wise error correction (using AFNI’s 3dClustSim). This produced corrected p values (α values) for each cluster in the LME results based on cluster size.

REFERENCES

Secrets of Success in Medical Training

Cees van der Vleuten, PhD

In front of you is a theme issue of “Military Medicine” on the accomplishments of the F. Edward Hébert School of Medicine, Uniformed Services University of the Health Sciences (USU), in reaching their mission to train high-quality physicians for the Uniformed Services. It reports on a number of studies carried out within a larger Long-Term Career Outcome Study (LTCOS) done by a very successful team of investigators. I have had the honor to serve as a guest editor of this issue and was asked to reflect on these studies in this editorial. I have a long history in medical education as an educationalist and as a researcher. I took the liberty to review the outcomes of the studies and relate them to what seem to be more universal patterns of findings, a sort of a bigger picture, at least as I see them—a bigger picture that may bring us hopefully closer to the secrets of success in medical training. I will use the before, during, and after training framework as is used to structure the research in LTCOS.

BEFORE TRAINING
What strikes me, as many times before, in the admission research is the relative weak predictive power of single test formats. The studies indicated that the bulk of variables studied accounted for little variance. It is therefore no surprise that school performance of alternate candidates on the admission list was rather similar to that of initially selected candidates. Does this mean that admission is not an important process? I think not. However, these findings strengthen me in my view that single data points of measurement always have their limitations. Past performance predicts future performance, but in order to make meaningful predictions we need many data points. That is why GPA is predictive for school performance, at least to a certain extent. We need more longitudinal information over a longer period of time in order to make more solid predictions. This seems not only to be true in admission research but also in assessment research. We find similar patterns of predictive indicators. We may continue to look for the magical bullet, but I am afraid that any one-off measurement will always have those limitations.

DURING TRAINING
In a traditional approach to curriculum delivery, we simply deliver a set of relevant courses. The integration of those courses, and the use of the information from those courses when appropriate, is basically left to the learner. By offering the blocks, we assume the learner is able to build a house. In my “bigger picture” this assumption can be heavily challenged. I think that “transfer” of knowledge to the use of knowledge is a huge step that it does not happen automatically and that a medical training program should provide deliberate support to make that transfer. Major transitions in the curriculum such as the one from school to the workplace are indeed huge transitions. Thinking about strategies to support the learner in making steps of transfer is an important way forward. I was therefore not surprised by the success of implementing “meta themes” in the USU curriculum and the value of these for reflection and self-directed learning. They are vital strategies for the learner to see the bigger picture. Recent research shows that mentoring support may further facilitate these strategies. In my view, feedback and reflection is a dialogue between people and very effective for developing excellence in medical training.

The purposeful strategies to support the learner may also be beneficial for other lessons that were learned from the research done by the LTCOS team. The first is that the “one size fits all” approach to curriculum delivery does not fit the needs of the individual learner. Learners vary in background, in learning styles, in motivation, and in aspirations. Catering the curriculum to the needs of the individual learner is a secret of success. The LTCOS research showed that standard performance measures could serve as diagnostic instruments; a system of learner dialogue and learner support may do the rest. Interestingly also, as was found in this research and in many other instances, effective teaching and support may not only come from faculty staff but from peers as well. We could benefit from this insight much more. The second reason why learner support might be a beneficial strategy consists of the (recurrent) imposing data on well-being of students. In the LTCOS studies, a high prevalence of non-well-being was found (burnout = 49%, depressive symptoms = 38%, and low mental quality of life = 34% for fourth year medical students). This is impressive and not new. Apparently, we need to better address the emotional side of medical training as well, and some form of learner support might have further utility.

A final token of the transfer issue was the research in which positive findings in relation to authenticity of instruction were reported. A logical educational strategy to narrow the gap between learning and applying is to contextualize learning with authentic professional tasks. To mimic professional reality in an educational setting and to allow the learner to experience and practice with these tasks, we
directly address the ease of transfer. To me, this is yet another secret of success.

**AFTER TRAINING**

The bigger lesson of the after education research is a simple but equally important one: training matters! The research provides some compelling evidence that the USU training programs train high-quality physicians, in some instances performing better than graduates from other schools, showing committed lifelong service to the country, many of them within the federal system and others successfully outside. It shows that being committed to excellence and getting that across to learners is a solid basis for lifelong high-quality medical service.

There is one more nuance that I would like to make to the observation that training matters. In recent years, competency frameworks have been proposed in various parts of the world. In the United States, these are the Accreditation Council for Graduate Medical Education competency outcomes, but there are several others in different regions of the world. They all have lots in common; they call attention to the more generic skills of being an effective physician. Traditionally, we emphasize the cognitive part of medical training, but these more generic skills such as professionalism or communication (just to name two) are equally important. Research shows that when things go wrong in clinical practice, very typically these kinds of skills are involved. There is also research showing the opposite; these generic skills are very determinant for how successful we are in the labor market. The LTCOS research showed that the USU curriculum was effective in fostering transformational leadership. This resonates with other research showing that when training really addresses these generic competencies, they tend to be effective in developing these competencies. By the way, there is evidence that it does this without loss in the knowledge base of graduates. A clear win-win situation it seems to me, and a clear hint to another secret of success for medical training.

A very final secret of success is evidenced by the collection of education research articles in this issue of “Military Medicine” in itself: research in medical education matters! As I said in the beginning, I have enjoyed a long career in medical education. I have witnessed personally how medical training has professionalized substantially in a number of decades. Integration of curricula, early patient practice, simulation centers, active learning methods, structured clerkships, quality assurance, modern assessment strategies, e-learning, staff development, and curriculum governance are all areas where we have made massive progression. In part, this is because of education research. We could not imagine clinical practice without (bio)medical research, why could we imagine education practice without impetus from education research? Practice fuels research, research fuels practice. The same holds true in education. This theme issue is a perfect illustration of this final secret of success.
The Long-Term Career Outcome Study (LTCOS): What Have We Learned From 40 Years of Military Medical Education and Where Should We Go?

Steven J. Durning, MD, PhD; CDR Anthony R. Artino Jr., MSC USN; Ting Dong, PhD; David F. Cruess, PhD; COL William R. Gilliland, MC USA (Ret.); LTC Kent J. DeZee, MC USA; LTC Aaron Saguil, MC USA; Donna M. Waechter, PhD; Col John E. McManigle, USAF MC (Ret.)

ABSTRACT The work of the Long-Term Career Outcome Study (LTCOS), F. Edward Hébert School of Medicine, Uniformed Services University of the Health Sciences (USU) has been a multidisciplinary effort spanning more than 5 years. Borrowing from the established program evaluation and quality assurance literature, the LTCOS team has organized its evaluation and research efforts into three phases: before medical school, during medical school, and after medical school. The purpose of this commentary is to summarize the research articles presented in this special issue and to answer two fundamental questions: (1) what has been learned from LTCOS research conducted to date, and (2) where should the LTCOS team take its evaluation and research efforts in the future? Answers to these questions are relevant to USU, and they also can inform other medical education institutions and policy makers. What is more, answers to these questions will help to ensure USU meets its societal obligation to provide the highest quality health care to military members, their families, and society at large.

INTRODUCTION
On our University’s 40th Anniversary, the Long-Term Career Outcome Study (LTCOS) investigators believe it is important to reflect on the many institutional efforts over the past 4 decades, and we are delighted to have the opportunity to report to Military Medicine readers on a variety of LTCOS-related research projects. In this article, we will discuss our answers to two cardinal questions: (1) what have we learned, including how this work can help others within and outside our institution, and (2) where should we take our research in the future? We believe answers to these questions are relevant to our institution, to other institutions with an interest in medical school outcomes, and to policy makers.

What Have We Learned?
The work of the LTCOS has spanned more than 5 years now. Indeed, it has been both a privilege and a pleasure to conduct this work on behalf of our Dean, F. Edward Hébert School of Medicine. Through his vision and unwavering support, as well as that of the members of the LTCOS team, our emerging findings are helping our institution and others. We believe—and LTCOS is “living proof”—that any institution seeking to explore the performance of their students and graduates needs vision, leadership, support, interdepartmental collaboration and expertise (both MD- and PhD-trained faculty), and, most importantly, personnel dedicated to the success of the project.

As the name LTCOS suggests, we are committed to evaluate long-term outcomes, but to do so requires assessing short- and intermediate-term outcomes since these help to inform longer-term measures and practice. The LTCOS team has also been interested in collaboration from early on in our history. Accordingly, there has been an ongoing emphasis on establishing multidisciplinary collaborations and conducting multi-institutional investigations. We have discussed some of these efforts in a previous Military Medicine article and in our introductory editorial at the beginning of this special issue.

Early in our work, we sought to establish a framework to further define our efforts. Borrowing from the established program evaluation and quality assurance literature, we compartmentalized our efforts into three phases—the time before entering medical school (prematriculation), the time during medical school, and the time after medical school (postgraduation). We have intentionally followed this three-phase division in the current special issue.

The below “lessons learned” or “take-home points” are listed using this framework—before, during, and after. Emphasis is placed on the articles in this edition; however, the reader is also encouraged to read a sample of other LTCOS-related papers—both single institution work and multi-institutional collaborations. In addition, we encourage readers to review three Uniformed Services University (USU) articles that have recently been accepted for presentation at the Research in Medical Education (RIME) meeting and for publication in Academic Medicine, the journal of the American Association of Medical Colleges (AAMC). Of the 15 articles accepted for the RIME meeting in 2012, the three USU articles represent one-half of the six articles accepted from U.S. institutions.

BEFORE MEDICAL SCHOOL
Determining what factors are associated with future performance in medical school and beyond is of particular interest,

F. Edward Hébert School of Medicine, Uniformed Services University of the Health Sciences, 4301 Jones Bridge Road, Bethesda, MD 20814.
at USU. Our medical students go on to become Department of Defense residents and faculty physicians for at least 7 years after completing their residency (and sometimes fellowship) education. Lessons learned from our institution in the before phase address the work of the USU admissions committee and feature results that may be of general interest to other admission committees.

So, what have we learned? From the Gilliland et al 28 article, we found that members of our admission committee use the entire admission package when completing their summary comments and making their final recommendation for each applicant. The themes generated had good reliability across raters and saturation was achieved. However, it appears from our thematic analysis that certain components of the admission package are used less often, and the reasons for this are not known. As a result, our school may consider revising the admission committee member (tertiary reviewer) applicant rating form to reflect our identified themes and improve our use of the spectrum of data available from the application package. In a second study, 29 we also assessed the predictive validity of these themes. In this article, the themes, especially when characterized by negative comments, do seem to predict subsequent performance in medical school, albeit weakly.

Recently, we have also explored two specific components of the admission package: self-reported research and clinical experience and their associations with subsequent medical school performance. Our data suggest that self-reported research experience has a weak, positive association with preclinical performance. 30 This finding is in contrast to our results from a study of self-reported clinical experience. 10 In particular, self-reported clinical experience had a negative association with subsequent medical school performance (cumulative grade point average at the end of medical school). Interestingly, self-reported clinical experience was unrelated to superior performance in internship, the very place where one might expect to find an effect. Taken together, if presented with two otherwise identical candidates, our data suggest that our admissions committee may do well to choose the one with self-reported research experience over the candidate with self-reported clinical experience. Lessons learned from these two articles include the need to measure actual clinical and research experience (i.e., as ascertained by interviews) in an effort to further explore the influence of these experiences on medical school performance. If one uses self-reports alone (available from the AAMC), our findings raise some legitimate concerns about the usefulness of such reports.

We also conducted an investigation of our alternate list candidates. 31 Our findings suggest that alternate list candidates who eventually matriculate to USU perform on par with regularly accepted candidates. Our data also suggest that as long as the applicant pool remains robust, additional matriculants from the alternate list could be used to expand class size. Based on our investigation, these students would likely be able to meet current performance criteria. We are currently exploring whether this finding translates into the arena of graduate medical education (GME).

Implications from these before studies prompt us to reconsider how admission processes occur at USU. Although some effects were significant and practically meaningful, the bulk of the variables we studied ultimately accounted for little variance beyond traditional admissions measures like grade point average and Medical College Admission Test. Perhaps, some of the AAMC application categories should be revised or abolished? The same scrutiny needs to be given to other contemporary measures for admissions, such as the Multiple Mini Interview, to determine if these methods are predictive (both statistically and clinically) of better performance in medical school and beyond. To date, studies of the Multiple Mini Interview have yielded promising results. 32

**DURING MEDICAL SCHOOL**

Finding from the during phase of medical school can also help inform our policies and practices, as well as those of other institutions. Such understanding can be particularly helpful to inform initiatives such as curriculum reform. Lessons learned include the finding that a preclinical course fostering reflection does appear to impact performance, albeit weakly, throughout medical school. 33 Indeed, our work suggests that mapping “meta-themes” of the curriculum (such as reflection and self-regulated learning) and ensuring that these topics are introduced, reinforced, and practiced in medical school is important. We also found that the authenticity of instruction (how close instructional methods are to actual clinical practice; e.g., a standardized patient presentation would be considered more authentic than a paper case of the same condition) in a preclinical course does not appear to improve clinical reasoning performance at the end of the second or third year of medical school. 34 From both a practical and theoretical standpoint, we believe this finding may indicate that medical school curricula should not be designed using a “one size fits all” approach. Indeed, these prospective, randomized studies suggest that we should consider customizing curricula based on students’ needs and that these needs can be determined, in large part, by standard performance measures. Furthermore, these studies suggest that faculty can be successfully trained to teach clinical cases as our analysis method (hierarchical linear modeling) allowed us to determine that the specialty of the faculty member did not impact performance either in the course or in the internal medicine clerkship the following year. 34 Thus, based on these findings, it seems that, if done properly, faculty development can produce identical educational outcomes without respect to the specific experiences of the teacher. Moreover, this work suggests that residents or even senior students can serve as teachers, provided that teaching materials are standardized and teachers are properly oriented. Indeed, recent work suggests that peer teaching benefits both the teacher and the learner. 35 as schools continue to revise their curricula, they...
may consider more formally integrating very junior faculty and even residents or senior students in the education of other students. Finally, the results from this 2-year prospective study indicate that scarce resources used for developing expensive and trendy instructional materials (DVDs and standardized patients for preclinical students in an introductory reasoning course) could be diverted to other needs in the preclinical setting if the development of superior clinical reasoning performance is the overall goal.\textsuperscript{13,14,34}

In the Saguil et al article,\textsuperscript{36} we have learned that high performance during a particular clerkship (when compared to a student’s overall performance) appears to be related to pursuing that particular career path. What is not known is if high performance during the clerkships leads to career selection or if having a desired career profession leads to high performance in that clerkship (i.e., in which direction does the causal arrow point)?\textsuperscript{36} Alternatively, this result may be a combination of these two causal mechanisms, which likely interact in reciprocal fashion. Future work using advanced research designs and analyses (e.g., longitudinal data collection and structural equation modeling, respectively), might allow us to untangle these phenomena. Such work could also lead to revisions of how student performance is viewed on the clerkships.

We are also learning that performance on National Board of Medical Examiners (NBME) subject exams does appear to be correlated with performance on United States Medical Licensing Examinations (USMLE); the more NBME examinations included, the more variance in USMLE for which we can account.\textsuperscript{37} It also appears that primary care NBME examinations may be particularly important as these explained the majority of the variance in our study when entered first into our regression model. This information is helpful for schools like ours that are undergoing curriculum reform and are wrestling with questions of exam timing. Another “action item” from this work refers to how institutions should respond to students who exhibit failing performance on NBME examination(s). Our work suggests that these examinations can serve as an “early warning” for suboptimal performance on USMLE, and that early intervention following NBME failures should be considered.

In our Objective Structured Clinical Examination (OSCE) article,\textsuperscript{38} we explored the association of clinical performance in subsequent iterations as measured by sequential OSCE examinations. This work is important as all medical students are required to take the USMLE Step 2 Clinical Skills Examination. Our findings suggest that performance on one OSCE does not predict future OSCE performance, and we provide a theoretical perspective for why this may be the case. This is akin to the vexing issue of why physician performance can vary when seeing two patients with the same condition. In other words, the specifics of the clinical situation—patient, physician (or trainee), and environment—interact in dynamic ways to determine the outcome (clinical performance). An implication of this work is that specific training in these patient, environments, and physician factors could be provided in medical school to help reduce the cognitive load during patient care encounters. Such an approach could potentially improve performance. This is consistent with our prior work on the topic, which has also been conducted at USU.\textsuperscript{39–41}

An important question in GME (residency) is the extent to which process measures are associated with outcomes. Indeed, the Accreditation Council on Graduate Medical Education is now requiring program directors to make this link. What are we learning about this link at the undergraduate (medical school) level? We investigated the association between a variety of clerkship process measures (i.e., number of patients and medical problems seen) and ultimate performance on the internal medicine clerkship.\textsuperscript{42} Similar to our research on self-reported research experience,\textsuperscript{31} we found that a self-report system for documenting patient problems and diagnoses is weakly associated with subsequent performance at the end of the clerkship. There are many possible reasons for this weak association, and additional exploration may help both medical schools and GME programs (including ours) develop better systems for tracking and monitoring learner process and outcome measures.

There has been renewed attention in medical education to “non-cognitive” factors and how these factors may impact medical student and resident performance. This renewed interest is consistent with emerging findings from contemporary education research.\textsuperscript{43} We have previously reported on the importance of well-being for our USU trainees as part of a multi-institutional collaborative.\textsuperscript{17,19,20,23} We are also learning that student self-efficacy, or the belief in one’s capability to successfully execute the behaviors needed for attaining a desired performance, is an important predictor of student success. This is consistent with other LTCOS-related work addressing motivation and emotion as predictors of performance in a clinical reasoning course at our institution.\textsuperscript{15,16,44} Indeed, it appears that our institution (and possibly others) may need to attend to these less-traditional measures and their potential impact on future performance.\textsuperscript{15,16,44} We continue to gather evidence for reliability and validity of a tool for measuring self-efficacy in our USU students across medical school.\textsuperscript{45} This tool could have applicability in GME (residency) and future practice. Additionally, coupling self-efficacy (a self-perception) with measures of actual performance could provide powerful feedback for students and practicing physicians. By helping trainees learn to properly “calibrate” their self-efficacy (or task-specific confidence) with their actual competence, we may be able to considerably improve their future performance as more accurate “self-assessment” may be a critical component to improvement and life-long learning. Indeed, findings from contemporary medical education research suggest that improved self-assessment is indeed possible, particularly when assessed on a more situation-specific level (as opposed to globally).\textsuperscript{46} Findings from our work are adding to this
emerging research on the importance of accurate calibration for the development of essential life-long learning habits.

Broader implications from our during medical school investigations are several fold. We have demonstrated the importance of theory as a means of informing medical education practice. Several studies utilized well-established educational theories, and theoretical predictions have informed our findings and led to further theory development and improved practice. A second important outcome of this work is the realization that multi-institutional research is required to improve the generalizability of our work. Third, alternate list applicants, who subsequently matriculate, appear to perform as well as “regular” applicants, which is consistent with emerging work in the area of admissions. And finally, further attempts to relate process measures, such as number of patients seen, with performance outcomes could make important contributions to designing effective and efficient curricular experiences in medical school and beyond.

AFTER MEDICAL SCHOOL
As our name suggests, assessing longer-term outcomes in medical education is of paramount importance. In this vein, we are gathering important data to help us revise our curriculum and ultimately enhance the performance of USU graduates. Our graduates are practicing in virtually all specialties and in all 50 states (not just the states where they were trained), and our graduates appear, on average, to have higher board certification rates than their U.S. civilian counterparts. Our results also suggest that USU and its affiliated residency training programs are providing high-quality education to physicians who become board certified and go on to serve our country’s active duty members, their dependents, and retirees.

Our graduates enter the military-specific match—a unique, longitudinal, and “trackable” match process. Our initial investigation into this match process suggests that we have fewer communications between our residency directors and potential interns (i.e., after interviewing with a program but before official match results received) than they do in the civilian match. This difference in communication occurs despite not having explicit rules and regulations against such communications, which our civilian counterparts have for such communications. Potential implications include establishing more explicit criteria for communications, which could result in even less such communications.

Leadership is a core tenet of our medical school. As such, we actively surveyed the opinions of some of our most senior leaders, our flag (or general) officers. Notably, our findings suggest that our flag officer graduates are transformational leaders, which is consistent with high-performing leaders from other fields. These officers appear to attribute part of their leadership success to their USU education, a finding that provides additional evidence to support the university’s role in training future leaders. Implications of this study include focusing our leadership curriculum on learning the tenets of transformational leadership. Our general officers also suggested some key areas to improve leadership education at USU, including additional leadership experiences, mentoring opportunities, and master’s degree programs. Based, in part, on these results, we are currently expanding offerings in all three of these areas.

Some of the difficulties with establishing longer term outcomes may be due, in part, to the precision of the measurements we employ. This argument is particularly relevant to skills that are not directly observable, such as clinical reasoning. Clinical reasoning is often considered a “black box” in medicine; it entails practically everything a physician does in practice, yet it is not directly observable. Thus, we are exploring novel assessment instruments for assessing clinical performance in our graduates, including the use of advanced technologies to assess clinical reasoning (S.J. Durning, J. Graner, A.R. Artino, L.N. Pangaro, T. Beckman, E. Holmboe, T. Oakes, M. Roy, G. Riedy, V. Capaldi, R. Walter, C. van der Vleuten, L. Schuwirth, unpublished data). Functional magnetic resonance imaging offers promise for helping us test theoretical predictions and provide additional evidence for the use of typical performance measurements. These technologies offer unique opportunities to better understand physician performance, which is of particular importance to our graduates who are often asked to practice medicine in austere environments.

Broader lessons learned from articles in the after phase of education include the extensive success of the USU curriculum in terms of training tomorrow’s leaders and serving our nation while in uniform (and beyond). Our data suggest our graduates display diversity in practice locations and specialties while highlighting quality of performance as reflected by board certification rates. Nonetheless, there is room for improvement and our work has identified a number of areas to target, as well as several areas for future research, which we outline below.

Where Should LTCOS Focus Its Future Research?
We have identified a number of potential future directions to help further inform the aforementioned work. Here, we list several future directions, in no particular order, with the understanding that this list is simply a starting point for discussion. Specifically, future LTCOS research will target the following:

1. Enhance multi-institutional work. Such work can boost the generalizability of our findings and our ability to work with other experts in the field. We have conducted a number of multi-institutional studies but are always looking to expand our collaboration with other credentialing bodies and civilian institutions. We also plan to pursue extramural grant funding to help support the efforts of such collaborations. We will also continue to publish in top-tier medical education journals and present at international conferences, such as the upcoming RIME conference in 2012 (these articles are published in Academic Medicine).
Expand on the use of theory to inform our selection, curriculum, and evaluation efforts. We will continue to use contemporary educational theories to frame our empirical work and help us improve the medical school education at USU. A number of previous studies have explored educational theory. 12–16,34,40,41,44,45

Move beyond the use of traditional performance measurements. These traditional measurements have led to great progress in medical education. Future exploration will include studying performance evaluation (i.e., OSCEs, miniclinical evaluation exercises), as well as other novel instruments such as functional magnetic resonance imaging, facial feature analysis, ecological momentary assessments, and eye tracking techniques.

Expand upon the breadth and depth of medical school outcome measurements. We are currently using educational theory to assist with defining and measuring “life-long learning” across the continuum of medical education. We are also collecting an end-of-third-year program director evaluation form and asking postgraduate year (PGY)-1 and PGY-3 graduates to self-assess USU’s role in their preparation for current assignments. We believe this is particularly important since we are the only “closed” medical education system. USU is one of the few medical education institutions capable of performing studies that span medical school and practice. Such studies are not feasible for many civilian institutions where medical students graduate to attend one of hundreds of residency training programs to one of thousands of practice environments.

Expand our “study lab” to be more inclusive of GME and continuing medical education. Novel assessments and contemporary educational theories should be applied to all new PGY-1’s at Department of Defense GME programs, not just USU graduates. The results from these investigations could be used to improve the generalizability of our findings.

Explore additional ways to educate our medical students in leadership. We hope to begin a Masters of Health Professions Education degree program with a focus on leadership, in addition to the already successful Masters in Health Administration and Policy program. These programs are currently not offered to medical students; however, components of these programs could be incorporated into the medical school curriculum. Additionally, efforts could support having more USU junior faculty and residents and medical students participate in selected studies. Articles in this special issue include medical students as authors. 31,36,37

Maintain a focus on social accountability. We will continue to obtain outcomes that demonstrate the ways in which USU graduates are serving our nation and our taxpayers. To date, we have some evidence that USU and our graduates are meeting the societal charge to train competent physicians. 40

Conduct more prospective studies, which will allow us to draw more definitive conclusions regarding prediction, causation, and intervention effectiveness.

In summary, from the inception of the LTCOS to this current edition, we have tried to conduct research that will inform USU’s educational processes. Our future work will continue to expand on our previous findings in novel, practical, and, we hope, interesting ways. Ultimately, the LTCOS team will continue to strive to improve the education of our students, which we believe will translate into improved care of their future patients. We believe it is our societal obligation to ensure our patriots, their family members, and society at large all receive the highest quality care, care that they have earned and deserve.

REFERENCES

LTCOS INVESTIGATORS

Steven J. Durning, MD, PhD (Co-Principal Investigator)
Professor of Medicine and Pathology

Anthony R. Artino Jr., PhD (Co-Principal Investigator)
Associate Professor of Preventive Medicine and Biometrics and Medicine

David F. Cruess, PhD
Professor and Vice Chair, Preventive Medicine and Biometrics

Kent J. DeZee, MD, MPH
Associate Professor of Medicine

Ting Dong, PhD
Research Associate

William R. Gilliland, MD, MHPE
Professor of Medicine and Assistant Dean for Medical Curriculum

John E. McManigle, MD
Vice Dean, School of Medicine

Aaron Saguil, MD, MPH
Assistant Professor of Family Medicine

Donna M. Waechter, PhD
Associate Dean of Medical Education

ACKNOWLEDGEMENTS

The LTCOS team would like to thank Dr. Ting Dong for her tireless and exemplary work on this supplement. The issue truly would not have been possible without her efforts. We are also extremely grateful for the outstanding work of our Guest Editors and Peer Reviewers. LTCOS work would also not be possible without the support of the Dean’s Education and Endowment Fund.
“Learning to Care for Those in Harm’s Way”