Validity Evidence for a Key Features Examination in the Internal Medicine Clerkship

BY

VALERIE J. LANG

B.S., Allegheny College, 1992

M.D., SUNY Syracuse College of Medicine, 1997

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Master of Health Professions Education
in the Graduate College of the
University of Illinois at Chicago, 2015

Chicago, Illinois

Defense Committee:

Georges Bordage, Chair and Advisor

Rachel Yudkowsky

Yoon Soo Park
Dedication

To my husband Alex, the doctors’ doctor, whose encouragement and self-proclaimed status as my biggest fan mean more than he will ever know; and to my son Charlie, whose curiosity and humor inspire me to look at the world through the eyes of a child. It is my hope that this project provides one small step toward advancing medical educators’ abilities to assess and, more importantly, to improve the decisions of the future physicians who will take care of you.
Acknowledgements

The following collaborators merit sincere gratitude for contributing their time and talent to this project: Amy Holthouser, MD, Kirk Bronander, MD, Heather Harrell, MD, Regina Kovach, MD, Norm Berman, MD, Susan Hingle, MD, and MedU for conceptualization, development, and validation of the exam; L. James Nixon, MD, MHPE, Janet Jokela, MD and Debra Stottlemyer, MD for exam development, T. Robert Vu, MD, Debra Leizman, MD, and Clifford Packer, MD for data gathering and critique; Saad Alvi, MD, for assistance with data gathering; Sandra Monteiro, PhD for statistical analyses; Audra Bucklin for her assistance with managing the project; Yoon Soo Park, PhD and Rachel Yudkowsky, MD, MHPE for their time and council, and Georges Bordage, MD, PhD, for his guidance in the exam development, validation, and interpretation of results.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>SECTION</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>METHODS</td>
<td>7</td>
</tr>
<tr>
<td>RESULTS</td>
<td>10</td>
</tr>
<tr>
<td>CONTENT</td>
<td>11</td>
</tr>
<tr>
<td>RESPONSE PROCESS</td>
<td>12</td>
</tr>
<tr>
<td>INTERNAL STRUCTURE</td>
<td>13</td>
</tr>
<tr>
<td>RELATIONSHIP TO OTHER VARIABLES</td>
<td>15</td>
</tr>
<tr>
<td>CONSEQUENCES</td>
<td>19</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>21</td>
</tr>
<tr>
<td>CONTENT</td>
<td>22</td>
</tr>
<tr>
<td>RESPONSE PROCESS</td>
<td>21</td>
</tr>
<tr>
<td>INTERNAL STRUCTURE</td>
<td>22</td>
</tr>
<tr>
<td>RELATIONSHIP TO OTHER VARIABLES</td>
<td>24</td>
</tr>
<tr>
<td>CONSEQUENCES</td>
<td>25</td>
</tr>
<tr>
<td>CONCLUSIONS</td>
<td>26</td>
</tr>
<tr>
<td>CITED LITERATURE</td>
<td>27</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>31</td>
</tr>
<tr>
<td>VITA</td>
<td>35</td>
</tr>
</tbody>
</table>
LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I NUMBER (PERCENTAGE) OF PARTICIPANTS FROM 8 MEDICAL SCHOOLS</td>
<td>10</td>
</tr>
<tr>
<td>II NUMBER (PERCENTAGE) OF CASES ACCORDING TO LOCATION AND ORGAN SYSTEM, NUMBER (PERCENTAGE) OF KFS ACCORDING TO DECISION FOCUS, AND NUMBER (PERCENTAGE) OF QUESTIONS PER EXAM FORM</td>
<td>11</td>
</tr>
<tr>
<td>III NUMBER (PERCENTAGE) OF STUDENTS AND RESPONSES TO SURVEY REGARDING THEIR EXPERIENCES WITH THE ONLINE KEY FEATURES EXAMINATION FOR THE INTERNAL MEDICINE CLERKSHIP</td>
<td>12</td>
</tr>
<tr>
<td>IV MEAN SCORE FOR THE FOUR FORMS OF A 15-CASE KFE AND NBME-SE IN THE INTERNAL MEDICINE CLERKSHIP</td>
<td>13</td>
</tr>
<tr>
<td>V INTERNAL STRUCTURE: NUMBER OF CASES ACCORDING TO 4 LEVELS OF DISCRIMINATION</td>
<td>14</td>
</tr>
<tr>
<td>VI INTERNAL STRUCTURE: PROJECTED RELIABILITY WITH REMOVAL OF LEAST DISCRIMINATING CASE</td>
<td>15</td>
</tr>
<tr>
<td>VII INTERNAL STRUCTURE: GENERALIZABILITY ANALYSIS</td>
<td>16</td>
</tr>
<tr>
<td>VIII PRIOR CLINICAL EXPERIENCE OF STUDENTS TAKING 4 DIFFERENT FORMS OF THE KFE</td>
<td>16</td>
</tr>
<tr>
<td>IX CORRELATION COEFFICIENTS BETWEEN PERFORMANCE ON THE KFE, THE NBME-SE, AND CLINICAL EXPERIENCE</td>
<td>17</td>
</tr>
<tr>
<td>X CONSEQUENCES</td>
<td>20</td>
</tr>
<tr>
<td>FIGURE</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>1 Example of a clinical situation and key features</td>
<td>1</td>
</tr>
<tr>
<td>2 Example of a key features vignette, question stems, answer options, and scoring</td>
<td></td>
</tr>
<tr>
<td>key</td>
<td>2</td>
</tr>
</tbody>
</table>
### LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDIM</td>
<td>Clerkship Directors in Internal Medicine</td>
</tr>
<tr>
<td>DERM</td>
<td>Dermatology</td>
</tr>
<tr>
<td>DI</td>
<td>Discrimination Index</td>
</tr>
<tr>
<td>ENDO</td>
<td>Endocrinology</td>
</tr>
<tr>
<td>GU</td>
<td>Genitorurinary</td>
</tr>
<tr>
<td>KF</td>
<td>Key Feature</td>
</tr>
<tr>
<td>KFE</td>
<td>Key Features Exam</td>
</tr>
<tr>
<td>MCC</td>
<td>Medical Council of Canada</td>
</tr>
<tr>
<td>NBME</td>
<td>National Board of Medical Examiners</td>
</tr>
<tr>
<td>NS</td>
<td>Not Significant</td>
</tr>
<tr>
<td>OSCE</td>
<td>Objective Structured Clinical Examination</td>
</tr>
<tr>
<td>PMP</td>
<td>Patient Management Problem</td>
</tr>
<tr>
<td>QE</td>
<td>Qualifying Examination</td>
</tr>
<tr>
<td>RHEUM</td>
<td>Rheumatology</td>
</tr>
<tr>
<td>SD</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>SGIM</td>
<td>Society of General Internal Medicine</td>
</tr>
<tr>
<td>SIMPLE</td>
<td>Simulated Internal Medicine Patient Learning Experience</td>
</tr>
</tbody>
</table>
SUMMARY

Key features exams (KFEs) are used to assess clinical decision making. Four forms of an online KFE were developed to assess third-year medical students’ decision-making abilities during internal medicine clerkships in the U.S. We used Messick’s unified assessment framework to gather validity evidence regarding content, response process, internal structure, relationship to other variables, and consequences.

At 8 U.S. medical schools, 759 students during their internal medicine clerkship were given 75 minutes to complete one of four 15-case KFE forms. They also completed a survey regarding their experiences. Item analyses and generalizability studies were performed. KFE scores were compared with prior clinical experience and scores on the National Board of Medical Examiners Subject Examination (NBME-SE).

A total of 515 (67.9%) students consented to participate. Mean form scores ranged from 54.5% to 60.3% (SD 8.4-9.6). Phi-coefficient ranged from 0.36 to 0.52; adding 5 cases to the most reliable form would increase the Phi-coefficient to 0.59. Removing the least discriminating case from the two most reliable forms would increase the alpha coefficient to 0.58 and 0.57. The main source of variance came from the interaction of students (nested in schools) and cases. Correlation between KFE and NBME-SE scores ranged from 0.24 to 0.47 (p<0.01).

Three of the KFE forms had comparable reliability to other KFEs reported in the literature, despite requiring less testing time. This study provides the first validity evidence for KFEs for assessing clinical decision making in U.S. internal medicine clerkships.
Introduction

A core goal of the internal medicine clerkship is for students to learn to make diagnostic and management decisions about common internal medicine problems.\(^{1-3}\) One means of assessing clinical decision making in medicine is the "key features" approach\(^ {4}\) that focuses exclusively on the critical steps in solving clinical problems. One operational way of defining these critical steps is to focus on where learners find it most difficult or are most likely to make errors.\(^ {5}\) Key features exams (KFEs) were originally developed in the 80’s to replace the longer Patient Management Problems (PMPs)\(^ {6}\) that assessed all aspects of a case, from chief complaint to follow-up, rather than focusing on the critical decision points. The shorter KF cases allowed examiners to sample a greater number of cases for a given period of testing time, thereby increasing the reliability of the exam scores.\(^ {7}\)

To develop a KFE case, first the critical steps (key features) are identified for the resolution of the problem or situation (Figure 1).

---

**Clinical problem:** Gallstone pancreatitis

Given a patient with risk factors for gallstones and clinical presentation consistent with pancreatitis, the 3\(^ {rd}\)-year medical student will:

- **KF1.** Consider pancreatitis as the most likely diagnosis;
- **KF2.** Order an ultrasound (to look for gallstones); and
- **KF3.** Administer volume resuscitation and analgesia.

**Figure 1. Example of a clinical situation and key features.** (Note: This information is not seen by the student).
Given the KFs, a case vignette is written followed by a set of questions that test the KFs (and only the KFs). The information in the vignettes is described using lay language as often as possible because the use of lay language better discriminates among examinees.(8) Norman et al also showed that test score reliability is maximized by testing 2 to 3 key features per case; testing one KF provides insufficient information about the examinee's ability and more than three KFs do not improve reliability and is a poor use of testing time.(9) The vignette is then followed by 2 to 3 questions, typically testing one KF per question, although one question may test more than one key feature.(9) Answering the questions can require either selecting options from a short menu or providing a free-text response. Examinees receive points for selecting or writing the correct options and may fail to garner points for errors of omission (e.g. not ordering an essential diagnostic test), and may lose points for errors of commission (e.g. ordering a potentially harmful diagnostic test) or over-ordering. See Figure 2 for an example of a KF vignette, question stems, answer options, and scoring key.

---

A 43-year-old female presents with abdominal pain and nausea for the past two days. When it first began, the pain was more on her right side, and it came and went. For the past day, the pain has been more in her middle upper abdomen, and it is now present all the time and more intense. Her last bowel movement was yesterday and was normal. She did not vomit. In the past she has had plantar fasciitis for which she takes ibuprofen 400mg twice a week, and high cholesterol for which she takes a statin every now and then. She does not smoke and drinks 1 glass wine/week.

VS: Temp 38.1, pulse 110, BP 130/80, RR 16, SaO2 99% RA.

Her conjunctivae are clear. Her abdomen is obese, soft, and mildly distended. Palpation of her middle upper abdomen causes the most pain. She tenses her abdominal muscles but is able to relax them
when distracted. When you press deeply in her right upper abdomen, she is able to take a full breath in.

There are very few bowel sounds.

PART-1

Q1. What is your leading diagnosis at this time? Select only one.

A. Acute coronary syndrome
B. Ascending cholangitis
C. Cholecystitis
D. Clostridium difficile colitis
E. Gastroenteritis (bacterial)
F. Gastroenteritis (viral)
G. Hepatitis due to statin
H. Hepatitis (viral)
I. Ischemic colitis
J. Mesenteric ischemia
K. Pancreatitis
L. Peptic ulcer disease
M. Pericarditis
N. Somatization
O. Splenic infarction

Scoring

KF1. Consider pancreatitis as the most likely diagnosis
   1 point
   K. Pancreatitis
   0 point
   More than 1 option selected
Q2. Lab test results are:

<table>
<thead>
<tr>
<th>Test</th>
<th>Results</th>
<th>Normal</th>
<th>Normal (SI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WBC</td>
<td>18.0</td>
<td>4.5-11</td>
<td>4.5-11</td>
</tr>
<tr>
<td>Hgb</td>
<td>13.0</td>
<td>11.7-16.1</td>
<td>117-161</td>
</tr>
<tr>
<td>Hct</td>
<td>39%</td>
<td>35-47</td>
<td>0.35-0.47</td>
</tr>
<tr>
<td>Lipase</td>
<td>2100</td>
<td>0-160</td>
<td>0-160</td>
</tr>
<tr>
<td>Amylase</td>
<td>400</td>
<td>28-100</td>
<td>0.48-1.7</td>
</tr>
<tr>
<td>AST</td>
<td>67</td>
<td>&lt;35</td>
<td>&lt;0.53</td>
</tr>
<tr>
<td>ALT</td>
<td>80</td>
<td>&lt;45</td>
<td>&lt;0.58</td>
</tr>
<tr>
<td>T-Bil</td>
<td>2.1</td>
<td>0-2</td>
<td>0-34</td>
</tr>
<tr>
<td>Alk Phos</td>
<td>143</td>
<td>39-117</td>
<td>0.65-1.95</td>
</tr>
<tr>
<td>Sodium</td>
<td>141</td>
<td>136-145</td>
<td>136-145</td>
</tr>
<tr>
<td>BUN</td>
<td>28</td>
<td>6-20</td>
<td>2.1-7.1</td>
</tr>
<tr>
<td>Creat</td>
<td>1.0</td>
<td>0.6-1.3</td>
<td>53-115</td>
</tr>
<tr>
<td>Urine β-hCG</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
</tr>
</tbody>
</table>
What action(s) will you take at this time? You may select up to 4. Select T if no action is needed at this time.

A. 0.9% NaCl 2L bolus IV
B. Abdominal free air series
C. Acetaminophen PO
D. Aspirin 325mg PO
E. Ciprofloxacin and metronidazole PO
F. Contact precautions
G. CT scan abdomen with IV and PO contrast
H. D5 0.45% NaCl IV at 100 cc/h
I. Discontinue statin
J. Fecal occult blood test
K. Lipid profile
L. Lorazepam IV
M. Magnetic resonance cholangiopancreatography (MRCP)
N. Meperidine IV
O. Morphine IV
P. Pantoprazole bolus IV followed by continuous drip
Q. Piperacillin-tazobactam IV
R. Stool *C difficile* toxin assay
S. Ultrasound abdomen, right upper quadrant
T. No action needed at this time.

**Scoring**

KF2. Order an ultrasound to look for gallstones
   1 point U. Ultrasound abdomen, right upper quadrant
   0 point T. No action needed at this time or more than 4 options selected.

KF3. Administer volume resuscitation and analgesia
   0.5 point A. 0.9% NaCl 2L bolus IV
   0.5 point O. Morphine IV
   0 point T. No action needed at this time or more than 4 options selected.

**Figure 2.** Example of a KF vignette, question stems, answer options, and scoring key.
KFEs have been part of the Medical Council of Canada (MCC) Qualifying Exam (QE) since 1992.(7) The MCC QE is administered at the end of medical school and includes an applied knowledge part (comprised of multiple choice questions), a clinical decision-making part (the KFE), and a clinical performance part (an objective structured clinical exam (OSCE)). Tamblyn et al have shown that the KFE and the communication score on the OSCE best predict future complaints in practice,(10) and KF scores most strongly predict patients’ adherence to antihypertensive regimens in practice.(11) KFEs have also been developed for internal medicine clerkships in Canada(12) and Germany,(13) medical school core curriculum progress tests in the Netherlands,(14) self-assessment in subspecialty surgery in the U.S.,(15) and fellowship in the Royal Australian College of General Practitioners.(16) However, no KFE has yet been validated for the assessment of U.S. medical students.

The great majority (93%) of U.S. clerkships use the National Board of Medical Examiners Subject Exam (NBME-SE), a multiple choice exam, to assess students’ applied knowledge of internal medicine.(17) In addition, 33% use locally developed exams to complement this exam, most of which have gathered little or no validity evidence for the locally developed exams.(17)

An online KFE was developed to assess the clinical decision-making abilities of U.S. students during their internal medicine clerkship.(18) The purpose of this study was to gather validity evidence for the KFE. Using Messick’s unified assessment framework, evidence was gathered from all five categories of construct validity, that is, content, response process, internal structure, relationship to other variables, and consequences.(19, 20)
Methods

Exam development and pilot testing are described in detail elsewhere. (18) In brief, internal medicine faculty members were trained by an experienced medical educator (GB) to develop KF cases. The test blueprint for the KFE came from the Simulated Internal Medicine Patient Learning Experience (SIMPLE) program that contains a set of 36 instructional virtual patient cases, (21) itself based on the Clerkship Directors in Internal Medicine (CDIM) Core Curriculum for the Medicine Clerkship. (2) KFs were defined for clinical problems from each of the 36 instructional cases by two exam developers (experienced internists); when there was disagreement, the KFs were discussed by the development group until consensus was reached. Overall about two KF-cases (range 1 to 3) were developed to assess the critical clinical decisions for each SIMPLE instructional case, for a total of 71 KF-cases. Each KF-case assessed a mean of 2.6 KFs (range 1 to 5). Students selected their answers from a preset list of options; there were no free-text answers. Each case was peer reviewed by an additional 2-3 test developers.

Scoring keys were developed in accordance with the MCC guidelines. (22) For each KF, students could receive full, partial, or no credit. The unit of measurement for KFEs is the case score. Thus, KF scores within each case were averaged (weighted evenly), regardless of the number of KFs or questions within a case. Final total test scores were calculated by averaging case scores (weighted evenly) across the whole test. After a pilot test with 162 students, (18) 60 of the 71 cases were selected and modified as necessary and four 15-case exam forms (A-D) were created according to blueprint specifications, that is, balanced according to organ systems, location of care (outpatient - inpatient), and decision focus (diagnosis - management).

The study was conducted from February, 2012 to January, 2013. Schools were solicited during a national meeting of CDIM and from among the test developers’ institutions. Nine schools initially enrolled in the study, including 7 public and 2 private institutions with a range of class sizes (24-220) and
spanning various regions of the United States (West, Midwest, Southeast, Northeast). Ethics approval was obtained from the institutional review boards from all study sites and the University of Illinois at Chicago.

The four KFE forms were administered to students at or near the end of their third-year internal medicine clerkship. Each group of students in a clerkship block took the same exam form (because of online logistical limitations that only allowed one form to be used at a time), and the four KFE forms were rotated at each study site throughout the year. Students were provided an online introductory tutorial regarding the KF’s approach and format. Students had 75 minutes to complete the 15-case exam on line in a proctored classroom; this was a non-speeded test. Upon completion of the KFE, students responded to an online survey regarding the clarity of the instructions and questions, technical issues, test difficulty, use of the KFE for clerkship assessment, and whether using the KFE for a portion of their grade would impact their study habits (Appendix A). In accordance with the study protocol, completion of the survey constituted consent by the students to participate in the study (Appendix B).

Students completed the NBME-SE in Internal Medicine, which was also administered at the end of the clerkship, either on the same day, or within the week. The NBME-SE is a 100-item multiple-choice exam assessing application of knowledge in internal medicine. Students were allotted 150 minutes to complete the NBME-SE.

Each form of the KFE was created based on the same blueprint but contained a different set of 15 cases. Each exam form (A, B, C and D) was analyzed separately. Descriptive statistics were performed for each exam form. Internal-consistency reliability (Cronbach’s alpha) was calculated for each exam form using case scores. Case difficulty was calculated. Discrimination was determined using an item-total correlation coefficient, with each case treated as an “item” and the case removed from the total to which it was compared. KFE test scores were used to conduct a generalizability study to determine the
variance contribution of students, medical schools and cases. Because of the design of this study, variance due to students is nested within schools (S:Sc), and the interaction of students and cases (case specificity) included the nesting factor (CxS:Sc), which also serves as the error term in this design. Finally, correlation analyses were conducted to examine the relationship between the KFE scores (percent correct) and four other measures, that is, the NBME-SE scores, number of other clerkships completed, and number of inpatient and outpatient training weeks in the internal medicine clerkship. Because decision making requires applied knowledge, it was hypothesized that scores on the KFE would correlate moderately with scores on the NBME-SE.
Results

One of the 9 medical schools was not included in the study because students did not have access to the survey and could not be consented. All four forms of the KFEs were administered at six schools; three forms were administered at one school and two forms were administered at one school due to variations in the timing of their clinical rotations. The data from these eight schools were included in the analyses. A total of 759 students took the KFE; 515 (67.9%) completed the survey, that is, consented to participate, of which 501 (97.3%) had data from both the KFE and NBME-SE (Table I).

<table>
<thead>
<tr>
<th>Site</th>
<th>Eligible n</th>
<th>Consented n (%)</th>
<th>Eligible n</th>
<th>Consented n (%)</th>
<th>Eligible n</th>
<th>Consented n (%)</th>
<th>Eligible n</th>
<th>Consented n (%)</th>
<th>Eligible n</th>
<th>Consented n (%)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>7 (70.0)</td>
<td>10</td>
<td>0 (0)</td>
<td>16</td>
<td>9 (56.3)</td>
<td>15</td>
<td>15 (100)</td>
<td>51</td>
<td>31 (61.4)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>21</td>
<td>16 (76.2)</td>
<td>23</td>
<td>17 (73.9)</td>
<td>20</td>
<td>19 (95.0)</td>
<td>22</td>
<td>20 (90.9)</td>
<td>86</td>
<td>72 (83.7)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>55</td>
<td>36 (65.5)</td>
<td>57</td>
<td>57 (58.8)</td>
<td>47</td>
<td>32 (68.1)</td>
<td>92</td>
<td>53 (57.6)</td>
<td>291</td>
<td>178 (51.2)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>55</td>
<td>41 (74.5)</td>
<td>26</td>
<td>20 (76.9)</td>
<td>21</td>
<td>21 (100)</td>
<td>32</td>
<td>26 (81.3)</td>
<td>134</td>
<td>108 (80.5)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>3 (75.0)</td>
<td>6</td>
<td>5 (83.3)</td>
<td>7</td>
<td>7 (100)</td>
<td>0</td>
<td>0 (0)</td>
<td>17</td>
<td>15 (88.2)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>19</td>
<td>18 (94.7)</td>
<td>33</td>
<td>21 (63.6)</td>
<td>34</td>
<td>28 (82.4)</td>
<td>15</td>
<td>13 (86.7)</td>
<td>101</td>
<td>80 (79.2)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>20</td>
<td>5 (25.0)</td>
<td>6</td>
<td>1 (16.7)</td>
<td>14</td>
<td>13 (92.9)</td>
<td>18</td>
<td>6 (33.3)</td>
<td>58</td>
<td>25 (43.1)</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0 (0)</td>
<td>0</td>
<td>0 (0)</td>
<td>14</td>
<td>3 (21.4)</td>
<td>7</td>
<td>3 (42.9)</td>
<td>21</td>
<td>6 (28.6)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>184</td>
<td>126 (70.0)</td>
<td>201</td>
<td>121 (60.2)</td>
<td>173</td>
<td>132 (76.3)</td>
<td>201</td>
<td>136 (67.7)</td>
<td>759</td>
<td>515 (67.9)</td>
<td></td>
</tr>
</tbody>
</table>

Results are presented according to Messick’s five categories of construct validity evidence.
Content

Care was given to create exam forms that were representative of the blueprint characteristics, that is, proportionate to the two locations (inpatient and outpatient), eight organ systems, and two decision foci (diagnosis and management); see Table II for details. The same care was given to train the test developers to define KFSs and write KF cases and questions, both led by an experienced KFE test developer.

<table>
<thead>
<tr>
<th>Location*</th>
<th>Form A n (%)</th>
<th>Form B n (%)</th>
<th>Form C n (%)</th>
<th>Form D n (%)</th>
<th>Mean n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inpatient</td>
<td>5 (33.3)</td>
<td>4 (26.7)</td>
<td>4 (26.7)</td>
<td>5 (33.3)</td>
<td>4.5 (30.0)</td>
</tr>
<tr>
<td>Outpatient</td>
<td>10 (66.7)</td>
<td>11 (73.3)</td>
<td>11 (73.3)</td>
<td>10 (66.7)</td>
<td>10.5 (70.0)</td>
</tr>
<tr>
<td>Organ system</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>2 (13.3)</td>
<td>2 (13.3)</td>
<td>2 (13.3)</td>
<td>2 (13.3)</td>
<td>2 (13.3)</td>
</tr>
<tr>
<td>Endo/Rheum/Derm</td>
<td>2 (13.3)</td>
<td>2 (13.3)</td>
<td>2 (13.3)</td>
<td>2 (13.3)</td>
<td>2 (13.3)</td>
</tr>
<tr>
<td>Gastroenterology</td>
<td>3 (20.0)</td>
<td>2 (13.3)</td>
<td>2 (13.3)</td>
<td>2 (13.3)</td>
<td>2.3 (15.0)</td>
</tr>
<tr>
<td>Hematology/Oncology</td>
<td>2 (13.3)</td>
<td>2 (13.3)</td>
<td>1 (6.7)</td>
<td>1 (6.7)</td>
<td>1 (6.7)</td>
</tr>
<tr>
<td>Infectious Disease</td>
<td>1 (6.7)</td>
<td>1 (6.7)</td>
<td>1 (6.7)</td>
<td>1 (6.7)</td>
<td>1 (6.7)</td>
</tr>
<tr>
<td>Neurology/Psychiatry</td>
<td>1 (6.7)</td>
<td>0 (0)</td>
<td>1 (6.7)</td>
<td>1 (6.7)</td>
<td>0.8 (5.0)</td>
</tr>
<tr>
<td>Pulmonary</td>
<td>1 (6.7)</td>
<td>2 (13.3)</td>
<td>1 (6.7)</td>
<td>1 (6.7)</td>
<td>1.3 (8.4)</td>
</tr>
<tr>
<td>Renal/GU/Fluid/</td>
<td>2 (13.3)</td>
<td>2 (13.3)</td>
<td>3 (20.0)</td>
<td>3 (20.0)</td>
<td>2.5 (16.7)</td>
</tr>
<tr>
<td>Electrolytes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Key features (total per exam)</td>
<td>40</td>
<td>39</td>
<td>39</td>
<td>37</td>
<td>38.8</td>
</tr>
<tr>
<td>Decision Focus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnosis</td>
<td>23 (57.5)</td>
<td>25 (64.1)</td>
<td>23 (62.1)</td>
<td>21 (56.8)</td>
<td>23.0 (59.4)</td>
</tr>
<tr>
<td>Management</td>
<td>17 (42.5)</td>
<td>14 (35.9)</td>
<td>16 (43.2)</td>
<td>16 (43.2)</td>
<td>15.8 (40.6)</td>
</tr>
<tr>
<td>Questions</td>
<td>33</td>
<td>35</td>
<td>33</td>
<td>31</td>
<td>33.0</td>
</tr>
<tr>
<td>Questions per case (mean)</td>
<td>2.2</td>
<td>2.3</td>
<td>2.2</td>
<td>2.1</td>
<td>2.2</td>
</tr>
<tr>
<td>Key features per case (mean)</td>
<td>2.7</td>
<td>2.6</td>
<td>2.6</td>
<td>2.5</td>
<td>2.6</td>
</tr>
</tbody>
</table>

* "Outpatient" location included both the emergency department and clinics.
Response Process

Of the 515 students who responded to the survey, 463 (89.9%) indicated that 75 minutes was sufficient to complete the exam. The students who wanted more time requested a mean of 23.9 (range of means 12.0-31.5) additional minutes. See Table III for details. The majority of students reported having no issues with the online technology (76.5%), clarity of instructions (91.8%), and questions (87.5%). When technical issues were reported (in decreasing order), 29 (26%) were related to uncertainty over whether the exam was finished, 27 (24%) were related to uncertainty about navigating between cases and locking in answers, 27 (24%) were related to logging in, 19 (17%) were related to slow internet or local computer issues, 7 (6%) were related to content within the exam itself, and 4 (3%) were unrelated to the exam.

---

**TABLE III.** NUMBER (PERCENTAGE) OF STUDENTS AND RESPONSES TO SURVEY REGARDING THEIR EXPERIENCES WITH THE ONLINE KEY FEATURES EXAMINATION FOR THE INTERNAL MEDICINE CLERKSHIP.

<table>
<thead>
<tr>
<th></th>
<th>Form A</th>
<th>Form B</th>
<th>Form C</th>
<th>Form D</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students (n)</td>
<td>126</td>
<td>121</td>
<td>132</td>
<td>136</td>
<td>515</td>
</tr>
<tr>
<td>Sufficient time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>104 (82.5)</td>
<td>115 (95.0)</td>
<td>122 (92.4)</td>
<td>122 (89.7)</td>
<td>463 (89.9)</td>
</tr>
<tr>
<td>No</td>
<td>22 (17.5)</td>
<td>6 (5.0)</td>
<td>10 (7.6)</td>
<td>14 (10.3)</td>
<td>52 (10.1)</td>
</tr>
<tr>
<td>Added time needed (mean minutes)</td>
<td>31.5</td>
<td>12.0</td>
<td>15.3</td>
<td>24.3</td>
<td>23.9</td>
</tr>
<tr>
<td>Exam process issues: students responding “yes” n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technical</td>
<td>23 (18.3)</td>
<td>38 (31.4)</td>
<td>19 (14.4)</td>
<td>41 (30.1)</td>
<td>121 (23.5)</td>
</tr>
<tr>
<td>Clarity of instructions</td>
<td>9 (7.1)</td>
<td>12 (9.9)</td>
<td>13 (9.8)</td>
<td>8 (5.9)</td>
<td>42 (8.2)</td>
</tr>
<tr>
<td>Clarity of questions</td>
<td>13 (10.4)</td>
<td>26 (21.5)</td>
<td>9 (6.8)</td>
<td>14 (10.3)</td>
<td>59 (11.5)</td>
</tr>
</tbody>
</table>
Scores for each exam form were reviewed manually for accuracy. The review revealed that, for the questions in which students were instructed to select as many answer options as they deemed appropriate, the limit on the scoring regarding the allowable maximum number of options was not enforced. This scoring error was corrected, and scores were recalculated. Subsequent review indicated that the limits were correctly enforced.

*Internal Structure*

The mean test score from the four KFE forms was 58.4 (range: 54.5% to 60.3% (SD 8.4%-9.6%)). The mean score for the NBME-SE was 78.5 (range: 76.1% to 79.6% (SD 7.7%-14.0%)). The Cronbach’s alpha for the 15-case KFE forms varied from 0.44 to 0.53 (Table IV).

| TABLE IV. MEAN SCORE FOR THE FOUR FORMS OF A 15-CASE KFE AND NBME-SE IN THE INTERNAL MEDICINE CLERKSHIP. Only data from students with available NBME scores were included. |
|--------------------------------------------------|-----------------|-----------------|-----------------|-----------------|
| Students (n)                                      | 125             | 109             | 132             | 135             |
| KFE Score                                         |                 |                 |                 |                 |
| Mean (SD)                                         | 54.5 (9.6)      | 59.0 (8.7)      | 60.3 (8.4)      | 59.6 (9.5)      |
| Low                                               | 20.2            | 41.9            | 43.0            | 30.9            |
| High                                              | 75.5            | 83.0            | 82.6            | 85.6            |
| Reliability (Cronbach’s alpha)                    | 0.53            | 0.51            | 0.44            | 0.46            |
| NBME exam score (mean, SD)                        | 76.1 (14.0)     | 77.4 (8.0)      | 79.6 (10.6)     | 79.6 (7.7)      |

Fifty-nine of 60 (98.3%) cases had a positive discrimination index, and 32 of the 60 (53.3%) cases had a discrimination index greater than 0.20 (Table V). The mean discrimination index for all the cases was 0.27.
TABLE V. NUMBER OF CASES ACCORDING TO 4 LEVELS OF DISCRIMINATION.

<table>
<thead>
<tr>
<th>Discrimination Index</th>
<th>Form A n (%)</th>
<th>Form B n (%)</th>
<th>Form C n (%)</th>
<th>Form D n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;0</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (7)</td>
</tr>
<tr>
<td>0.0-0.19</td>
<td>7 (47)</td>
<td>6 (40)</td>
<td>9 (60)</td>
<td>5 (33)</td>
</tr>
<tr>
<td>0.20-0.30</td>
<td>5 (33)</td>
<td>8 (53)</td>
<td>4 (27)</td>
<td>8 (53)</td>
</tr>
<tr>
<td>&gt;0.30</td>
<td>3 (20)</td>
<td>1 (7)</td>
<td>2 (13)</td>
<td>1 (7)</td>
</tr>
<tr>
<td>Total &gt;0.20</td>
<td>8 (53)</td>
<td>9 (67)</td>
<td>6 (40)</td>
<td>9 (67)</td>
</tr>
</tbody>
</table>

By removing the one least discriminating case from each exam form, the Cronbach's alpha increased from 0.44-0.53 to 0.48-0.58 (Table VI).

TABLE VI. PROJECTED RELIABILITY WITH REMOVAL OF LEAST DISCRIMINATING CASES.

<table>
<thead>
<tr>
<th>Case</th>
<th>Form A</th>
<th>Form B</th>
<th>Form C</th>
<th>Form D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DI</td>
<td>Alpha if Case Removed</td>
<td>DI</td>
<td>Alpha if Case Removed</td>
</tr>
<tr>
<td>1</td>
<td>.40</td>
<td>.484</td>
<td>.43</td>
<td>.483</td>
</tr>
<tr>
<td>2</td>
<td>.39</td>
<td>.481</td>
<td>.30</td>
<td>.517</td>
</tr>
<tr>
<td>3</td>
<td>.39</td>
<td>.486</td>
<td>.28</td>
<td>.526</td>
</tr>
<tr>
<td>4</td>
<td>.27</td>
<td>.515</td>
<td>.28</td>
<td>.526</td>
</tr>
<tr>
<td>6</td>
<td>.24</td>
<td>.526</td>
<td>.24</td>
<td>.532</td>
</tr>
<tr>
<td>7</td>
<td>.23</td>
<td>.522</td>
<td>.24</td>
<td>.541</td>
</tr>
<tr>
<td>8</td>
<td>.20</td>
<td>.526</td>
<td>.23</td>
<td>.533</td>
</tr>
<tr>
<td>9</td>
<td>.17</td>
<td>.533</td>
<td>.23</td>
<td>.534</td>
</tr>
<tr>
<td>10</td>
<td>.16</td>
<td>.533</td>
<td>.17</td>
<td>.544</td>
</tr>
<tr>
<td>11</td>
<td>.16</td>
<td>.534</td>
<td>.13</td>
<td>.552</td>
</tr>
<tr>
<td>12</td>
<td>.12</td>
<td>.539</td>
<td>.10</td>
<td>.554</td>
</tr>
<tr>
<td>13</td>
<td>.12</td>
<td>.544</td>
<td>.10</td>
<td>.556</td>
</tr>
<tr>
<td>14</td>
<td>.04</td>
<td>.561</td>
<td>.07</td>
<td>.560</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>.584</td>
<td>.05</td>
<td>.572</td>
</tr>
</tbody>
</table>
Using generalizability analysis, the absolute reliability coefficient varied from 0.36 to 0.52 (Table VII).

Adding 5 cases to the most reliable form (A) would increase the G-coefficient to 0.59. The majority of variance was attributed to cases (16%) and students nested within schools (5%). The smallest proportion of variance came from schools (1%).

**TABLE VII. INTERNAL STRUCTURE: GENERALIZABILITY ANALYSIS.** S = student; Sc = medical school; C = Case.

<table>
<thead>
<tr>
<th>Component</th>
<th>Form A</th>
<th>Form B</th>
<th>Form C</th>
<th>Form D</th>
<th>Mean</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sc</td>
<td>0.0002</td>
<td>0.0008</td>
<td>0.0012</td>
<td>0.0003</td>
<td>0.000635</td>
<td>1</td>
</tr>
<tr>
<td>S:Sc</td>
<td>0.00354</td>
<td>0.0034</td>
<td>0.0022</td>
<td>0.0046</td>
<td>0.0034275</td>
<td>5</td>
</tr>
<tr>
<td>C</td>
<td>0.0087</td>
<td>0.0170</td>
<td>0.0116</td>
<td>0.0110</td>
<td>0.01209</td>
<td>16</td>
</tr>
<tr>
<td>C x Sc</td>
<td>0.001</td>
<td>0.0013</td>
<td>0.0017</td>
<td>0.0006</td>
<td>0.0011425</td>
<td>2</td>
</tr>
<tr>
<td>C x S:Sc</td>
<td>0.0500</td>
<td>0.0555</td>
<td>0.0589</td>
<td>0.0679</td>
<td>0.05792</td>
<td>77</td>
</tr>
<tr>
<td>Phi coefficient (absolute)</td>
<td>0.46</td>
<td>0.37</td>
<td>0.31</td>
<td>0.46</td>
<td>0.41</td>
<td></td>
</tr>
<tr>
<td>G-coefficient for 15 cases</td>
<td>0.52</td>
<td>0.48</td>
<td>0.36</td>
<td>0.50</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td>No. cases for G to approach 0.6</td>
<td>20 (0.59)</td>
<td>30 (0.65)</td>
<td>40 (0.60)</td>
<td>30 (0.57)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Relationship to Other Variables*

Prior to taking the KFE, the students had completed a mean of 3.2 clerkships (SD 2.3) and spent a mean of 1.3 (SD 1.9) weeks in an outpatient setting and 6.8 weeks (SD 2.1) in an inpatient setting during the internal medicine clerkship (TABLE VIII).
TABLE VIII. PRIOR CLINICAL EXPERIENCE OF STUDENTS TAKING FOUR DIFFERENT FORMS OF THE KFE.

<table>
<thead>
<tr>
<th></th>
<th>Form A</th>
<th>Form B</th>
<th>Form C</th>
<th>Form D</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Clerkships</td>
<td>2.57</td>
<td>2.14</td>
<td>3.11</td>
<td>4.98</td>
</tr>
<tr>
<td>completed (mean)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medicine clerkship</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>length</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inpatient weeks (mean)</td>
<td>6.86</td>
<td>6.64</td>
<td>6.61</td>
<td>6.73</td>
</tr>
<tr>
<td>Outpatient weeks (mean)</td>
<td>1.12</td>
<td>1.4</td>
<td>1.46</td>
<td>1.18</td>
</tr>
</tbody>
</table>

The disattenuated correlation coefficients between scores on the KFE forms and the NBME exam varied from 0.241 to 0.466 (p<0.01) (TABLE IX). The correlation coefficient between the number of clerkships completed and exam scores varied from 0.163 (p=NS) to 0.268 (p<0.01) for the KFE and -0.067 to 0.099 (p=NS for all correlations) for the NBME-SE.
TABLE IX. CORRELATION COEFFICIENTS BETWEEN PERFORMANCE ON THE KFE, THE NBME-SE, AND CLINICAL EXPERIENCE.

Form A

<table>
<thead>
<tr>
<th></th>
<th>NBME Score</th>
<th>Clerkships Completed</th>
<th>Outpatient Weeks</th>
<th>Inpatient Weeks</th>
<th>KFE Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>NBME Score</td>
<td>1</td>
<td>.071</td>
<td>-.246&quot;</td>
<td>.230&quot;</td>
<td>.241&quot;</td>
</tr>
<tr>
<td>Clerkships Completed</td>
<td>1</td>
<td>-.251&quot;</td>
<td>.380&quot;</td>
<td>.189&quot;</td>
<td></td>
</tr>
<tr>
<td>Outpatient Weeks</td>
<td>1</td>
<td></td>
<td>-.698&quot;</td>
<td>-.002</td>
<td></td>
</tr>
<tr>
<td>Inpatient Weeks</td>
<td>1</td>
<td></td>
<td>.024</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p< 0.05 level (2-tailed). ** p< 0.01 level (2-tailed).

Form B

<table>
<thead>
<tr>
<th></th>
<th>NBME Score</th>
<th>Clerkships Completed</th>
<th>Outpatient Weeks</th>
<th>Inpatient Weeks</th>
<th>KFE Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>NBME Score</td>
<td>1</td>
<td>.099</td>
<td>-.053</td>
<td>.060</td>
<td>.417&quot;</td>
</tr>
<tr>
<td>Clerkships Completed</td>
<td>1</td>
<td>.069</td>
<td>-.084</td>
<td>.288&quot;</td>
<td></td>
</tr>
<tr>
<td>Outpatient Weeks</td>
<td>1</td>
<td></td>
<td>-.724&quot;</td>
<td>.088</td>
<td></td>
</tr>
<tr>
<td>Inpatient Weeks</td>
<td>1</td>
<td></td>
<td>.079</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| KFE Score | 1 |
### Form C

<table>
<thead>
<tr>
<th></th>
<th>NBME Score</th>
<th>Clerkships Completed</th>
<th>Outpatient Weeks</th>
<th>Inpatient Weeks</th>
<th>KFE Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>NBME Score</td>
<td>1</td>
<td>.007</td>
<td>-.088</td>
<td>.018</td>
<td>.370**</td>
</tr>
<tr>
<td>Clerkships Completed</td>
<td>1</td>
<td>.163</td>
<td>-.204*</td>
<td>.163</td>
<td></td>
</tr>
<tr>
<td>Outpatient Weeks</td>
<td>1</td>
<td>-.347**</td>
<td>-.153</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inpatient Weeks</td>
<td>1</td>
<td>.237**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KFE Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

### Form D

<table>
<thead>
<tr>
<th></th>
<th>NBME Score</th>
<th>Clerkships Completed</th>
<th>Outpatient Weeks</th>
<th>Inpatient Weeks</th>
<th>KFE Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>NBME Score</td>
<td>1</td>
<td>-.067</td>
<td>.088</td>
<td>-.100</td>
<td>.466**</td>
</tr>
<tr>
<td>Clerkships Completed</td>
<td>1</td>
<td>-.179*</td>
<td>.162</td>
<td>.204*</td>
<td></td>
</tr>
<tr>
<td>Outpatient Weeks</td>
<td>1</td>
<td>-.757**</td>
<td>-.006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inpatient Weeks</td>
<td>1</td>
<td>.013</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KFE Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
Consequences

On the survey, 387 of the 515 (75.1%) students indicated that the difficulty of the KFE was just right compared to 127 (24.7%) who indicated it was too hard. None reported that it was too easy. When asked about preferences regarding how the KFE scores should be used, 381 (74%) recommended using the exam for formative feedback and 311 (60.4%) recommended that it should not be used as part of the clerkship grade. Of the 214 students who stated the KFE should (25 students, 4.9%) or maybe should (179 students, 35.8%) be used for grading, most often they recommended using it for 5% (86 students, 40.2%) or 10% (81 students, 37.9%) of the final clerkship grade. If the KFE was used for a portion of the clerkship grade, 352 (68.3%) reported that they would change their study habits, 89 (17.3%) would not change their study habits, and 74 (14.4%) were uncertain. Generally students commented that they would focus more on case-based learning such as with their own patients or with SIMPLE cases, though they worried that this would detract from studying for the NBME-SE.
<table>
<thead>
<tr>
<th></th>
<th>Form A</th>
<th>Form B</th>
<th>Form C</th>
<th>Form D</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students (n)</td>
<td>126</td>
<td>121</td>
<td>132</td>
<td>136</td>
<td>515</td>
</tr>
<tr>
<td>Difficulty -n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Too easy</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Just right</td>
<td>75 (61.0)</td>
<td>102 (84.3)</td>
<td>105 (80.8)</td>
<td>105 (77.8)</td>
<td>387 (75.1)</td>
</tr>
<tr>
<td>Too hard</td>
<td>48 (39.0)</td>
<td>19 (15.7)</td>
<td>30 (19.2)</td>
<td>30 (22.2)</td>
<td>127 (24.7)</td>
</tr>
<tr>
<td>Portion of grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5%</td>
<td>20</td>
<td>23</td>
<td>27</td>
<td>16</td>
<td>86</td>
</tr>
<tr>
<td>10%</td>
<td>25</td>
<td>11</td>
<td>18</td>
<td>27</td>
<td>81</td>
</tr>
<tr>
<td>15%</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>18</td>
</tr>
<tr>
<td>20%</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td>25%</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>30%</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>40%</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>&gt;40%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Change study habits if graded -n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>88 (69.8)</td>
<td>81 (66.9)</td>
<td>89 (67.4)</td>
<td>94 (69.1)</td>
<td>352 (68.3)</td>
</tr>
<tr>
<td>No</td>
<td>20 (15.9)</td>
<td>26 (21.5)</td>
<td>22 (16.7)</td>
<td>21 (15.4)</td>
<td>89 (17.3)</td>
</tr>
<tr>
<td>Not sure</td>
<td>18 (14.3)</td>
<td>14 (11.6)</td>
<td>21 (15.9)</td>
<td>21 (15.4)</td>
<td>74 (14.4)</td>
</tr>
<tr>
<td>KFE for formative feedback -n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>97 (77.0)</td>
<td>91 (75.2)</td>
<td>99 (75.0)</td>
<td>94 (69.1)</td>
<td>381 (74.0)</td>
</tr>
<tr>
<td>No</td>
<td>7 (5.6)</td>
<td>9 (7.4)</td>
<td>7 (5.3)</td>
<td>13 (9.6)</td>
<td>36 (7.0)</td>
</tr>
<tr>
<td>Not sure</td>
<td>22 (17.5)</td>
<td>21 (17.4)</td>
<td>26 (19.7)</td>
<td>29 (21.3)</td>
<td>98 (19.0)</td>
</tr>
</tbody>
</table>
Discussion

This is the first study to gather validity evidence for a KFE to assess clinical decision making in students in a U.S. clerkship. The results will be discussed using Messick's five categories of validity evidence.

Content

The process of developing the content for the four KFEs provides support for their validity. The content was drawn from the SIMPLE virtual patient case curriculum, which was, in turn, based on the CDIM Curriculum for the Core Internal Medicine Clerkship. Both curricula were developed based on input from medical student educators nationally and are widely used in U.S. medical schools. The cases and key features were blueprinted to each exam form to ensure a relatively even distribution according to organ system, location of care, and type of clinical decision-making, as illustrated in TABLE I. Basing the exams on national curricula and matching them to the blueprint ensured that the content represents a breadth of problems encountered in internal medicine, and in a variety of contexts. Exam quality was ensured by engaging educators with extensive experience teaching medical students to write the cases, under the guidance of an expert in the key features approach.

Response Process

A minority of students reported technical issues with the online exam. The majority of these were related to slow internet connections or uncertainty about navigation between cases, locking in answers, and whether all the questions were completed. These issues did not interfere with ultimate completion and correct scoring of the exam, but may have created distractions for the students. The instructions and navigation issues have since been addressed. Local computer and internet capacity are essential for students to sit any online exam. The process for scoring KFEs is complex, which increases the likelihood
of errors. We did find an error with the original scoring of a subset of questions. This error was fixed, but this experience highlighted the importance of monitoring and ensuring quality control of the scoring key.

Internal Structure

The KFE was a somewhat challenging exam with an overall mean score of 58.4 (compared to 78.5% for the NBME-SE). However only a minority of students (24.7%) reported that the exam was too hard. This discrepancy echoes the difficulty students encounter in self-assessing their performance. Students may have been more optimistic about the difficulty of the exam because the cases seemed familiar to them. In the pilot study students had reported that the cases were authentic and similar to cases encountered during internal medicine clerkships. The lower mean score for the KFE (compared to the MBME-SE) raises the question of whether students are getting sufficient practice and feedback on making clinical decisions during their clerkships. It is also possible that students would have put more effort into answering questions if the exam score were included as part of their grades. Participation rate in the study was moderate at 66%, and it is unclear whether students who chose not to participate performed differently on the exam. However, the mean NBME-SE for participants at one institution was similar for participants (79.3, SD 8.7) and the class as a whole (77.5, SD 8.5), and the mean NBME-SE score for all participants (78.5, SD 10.1) was similar to that of all students nationally (77.6, SD 8.3).

The reliability of the scores for three of the four KFE forms was moderate and comparable to the reliability of other medical student KFEs, despite requiring less testing time (75 minutes).(12, 13) Internal medicine clerkship KFEs in Canada and Germany, respectively, had reliabilities (Cronbach’s alpha) of 0.49 with 15 cases over 120 minutes(12) and 0.65 with 15 cases over 90 minutes.(13) The Canadian KFE included free text responses, which may account for the increased testing time, and required manual scoring. The German KFE included a greater number of KFs (4 per case), and students
selected responses from predetermined menus. Progress tests assessing medical student performance in a core curriculum in the Netherlands had high reliability (0.85-0.87) with 40 cases, though the allotted time was not reported.(14) For the high stakes Medical Council of Canada Qualifying Examination (MCC QE) the KFE has a reliability (G coefficient) of 0.622-0.639 for 35-40 cases over 3.5 hours,(9) and the Royal Australian College of General Practitioners Fellowship KFE has a reliability (Cronbach’s alpha) of 0.76-0.84 with 24 cases over 3 hours.(16) In general, increasing the number of cases for KFEs increases the reliability of the scores for an exam; however, this also adds to testing time.(26) Because the majority of clerkships already use the NBME-SE,(17) which took 2.5 hours at the time of the study and currently takes 3 hours, our KFE was kept brief to avoid over-burdening students and clerkship directors.

Norman et al found optimal reliability with 2-3 questions per KF case in the MCC QE; the reliability (G coefficient) was 0.579 with 20 cases with 2 KFs each and 0.579 for 15 cases with 3 KFs each.(9) Adding 5 cases to our highest performing KFE form in our study would increase the reliability to 0.59, which would extend testing time by 25 minutes. However, removing the least discriminating case from the same exam form would result in nearly the same increase on reliability (to 0.58), without increasing testing time.

Using generalizability analyses, the medical school factor was a negligible contributor to the variance (1%); students (nested in schools) contributed five times more to the variance (5%). Cases also contributed substantially to the variance (16%). The case effect may also be a reflection of the idiosyncratic nature of the clerkship experiences of the students.(27) Most of the variance was attributable to the interaction between students (nested in schools) and cases. Given the design of this study and logistical limitations, this source of variance reflects the error variance and is confounded with case specificity (S x C) usually associated with this source of variance. Future studies should better isolate this source of variance either by adding KFs to the existing design [(S:Sc) x (KF:C)] with case
specificity represented by the (S x C: Sc) term, akin to the 2006 Norman study,(9) or by removing the
school facet and adding KFs [S x (KF:C)] with case specificity represented by the (S x C) term.

The reliability and error variance varied among the four exam forms, likely due to the variable quality of
the cases in the different forms, with form A with the highest reliability and greatest number of cases
with high discrimination indices, and form C with the lowest reliability and discrimination indices. On
review post-study, some of the cases had minor technical issues (e.g., it was not clear to which
component of troponin the normal values were referred to) or the definition of the KFs for a clinical
problem for which a multitude of diverse diseases could account (e.g., fever of unknown origin).

It is possible that the reliability of the scores from the KFE forms in this study will increase over time
with ongoing reviews of the cases and item analyses. Reliability of the Royal Australian College of
General Practitioners Fellowship KFE improved with each year of administration, which the authors
attributed to review of the cases after implementation and increasing experience of the test
developers.(16) In the 15-case exam forms in this study, the impact of removing a single case with low
discrimination is significant, increasing the reliability from 0.52 and 0.51 to 0.57 and 0.58, respectively,
for the highest performing forms (Table VI). However, removing a case may also decrease alignment
with the exam blueprint.

Relationship to Other Variables

The positive, but moderate, correlation between scores on the KFE and NBME-SE (0.241 to 0.466)
suggests that the two exams assess some overlapping constructs. A recent review of the literature on
KFEs identified moderate correlations between KFEs and exams designed to assess more general
knowledge (correlation coefficient 0.44-0.52).(12, 13, 23) The NBME-SE is a rigorously developed single
best-answer multiple choice exam. This format lends itself well to assessing applied knowledge and, to
an extent, clinical reasoning. The KFE was designed to assess clinical decision making, which represents
the outcome of clinical reasoning. Thus there may be different and complementary roles for the KFE and the NBME subject exam within a program of assessment.

The relationship between KFE scores and prior clinical experience was mixed. The weak positive correlation between number of clerkships completed and KFE scores on 3 of the 4 KFE forms may reflect the benefit of additional experience making clinical decisions, but this impact is minor with 4.8% of the variance accounted for by prior experience. In a single institution study, Reteguiz et al found a small but significant increase in NBME internal medicine SE scores (absolute difference 3.7 points, p=.018) if students had previously completed other core clerkships (psychiatry, obstetrics-gynecology, or surgery).(28) We did not find a correlation between number of clerkships completed and NBME-SE scores in our study (Table IX). A national study comparing curricula at most U.S. medical schools to NBME-SE scores is currently underway.

Consequences

Students’ preference to use the KFE format for formative rather than summative purposes are understandable, as students may be reluctant to add an exam format specifically targeting decision making ability that would count as part of the grade. However, students’ preferences are not necessarily predictive of the utility of an exam and increased familiarity with the format may improve acceptance over time.

Conclusions

The KFE adds a new tool to the repertoire for assessment methods in the internal medicine clerkship, focusing specifically on the assessment of clinical decision making. Advantages of the KFE include the focus on clinical decision making for common internal medicine problems, the blueprinting to a national curriculum, the capacity to lock in answers and allow a case to unfold, and practical considerations,
including short testing time and automated scoring. There is no single gold standard for assessing clinical reasoning or decision making because these are complex skills(29) that are influenced by contextual factors such as location, patient factors, members of the health care team, and one’s own physical and emotional state.(30, 31) Standardized written examinations should be considered as part of a program of assessment that includes multiple episodes of assessment over time and in the context of actual patient care.(32, 33) Given its moderate reliability, it is recommended that the KFE in its current form should be used for low or moderate stakes assessment, but should not be used for high stakes assessment of students. It is possible that, with ongoing review of cases and/or lengthening of the exam, it would be suitable for higher stakes assessment in the future.

Future studies should repeat analysis of items with ongoing review and revision, and further address consequences, including standard-setting procedures to recommend pass/fail cut-offs and the impact on students’ study habits with decision-focused assessment.
Cited Literature


APPENDIX A: Survey of students.

1. What is your current year in medical school (excluding time out for research or other activities):
   A. 2
   B. 3
   C. 4

2. How many required clerkships, excluding your current clerkship, have you completed? (numerical)

3. During your current internal medicine clerkship, how many weeks were inpatient? (numerical)

4. During your current internal medicine clerkship, how many weeks were outpatient? (numerical)

5. Did you encounter any technical problems while completing the examination today?
   A. No
   B. Yes; please describe: (free text)

6. Were the instructions clear?
   A. Yes
   B. No; please explain: (free text)

7. Overall (considering the entire set of cases as a whole) rate the level of difficulty of this exam:
   A. Too easy
   B. Just right
   C. Too hard

8. Were you given enough time to complete this exam?
   A. Yes
   B. No; how much more time would you need? (free text)

9. Were the test questions following each case scenario clear?
   A. Yes
B. No; please explain: (free text)

10. The exam today did not count towards your grade. Do you think this type of exam should count as part of your grade in the future?
A. Yes; please explain: (free text)
B. No; please explain: (free text)
C. Maybe; please explain: (free text)

11. If yes, what percentage of your grade should the exam count for?
A. 5%
B. 10%
C. 15%
D. 20%
E. 25%
F. 30%
G. 35%
H. 40%
I. >40%
J. Other (free text)

12. Do you think the results from this exam should be used to give you formative feedback only (to help you learn, but not count as part of the grade)?
A. Yes
B. No
C. Not sure

13. If this exam were required as part of your grade, do you think it would change how you prepare for this exam?
A. No

B. Yes; explain in what way

C. Not sure
APPENDIX B: Information Letter for informed consent.

SCHOOL OF MEDICINE AND DENTISTRY

Department of Medicine

UNIVERSITY OF ROCHESTER MEDICAL CENTER

MEDICINE OF THE HIGHEST ORDER

Information Letter

Title of Study: Validity Evidence for a SIMPLE Key Features Examination to Assess Clinical Decision Making in Medical Students

Principal Investigator: Valerie J. Lang, MD
University of Rochester Department: Medicine

This form describes a research study that Dr. Valerie Lang from the University of Rochester is conducting. The purpose of this research study is to look at how well performance on the key features exam measures clinical-decision-making ability among third year medical students, how much it correlates with prior clinical experience, and how students view this type of exam.

This study involves a survey about your prior clinical training and your opinions about the exam. We estimate that approximately 30 subjects will take part in this study at the University of Rochester. This is a multi-institutional study and is being conducted at medical schools across the country. You will be asked to complete a survey after you finish the SIMPLE exam toward the end of your Medicine Clerkship. This should take about 10 minutes. There is a small chance that some of the questions may make you feel uncomfortable. You don’t have to answer those questions if you don’t want to. In fact you don’t have to answer any question that you choose not to answer. Just skip that question and go on to the next one.

Your participation in this survey is completely voluntary. You are free not to participate or to withdraw at any time, for whatever reason without penalty or loss of benefit to which you are otherwise entitled. Participation will have no impact on your grading or evaluation in the Clerkship. Completion of the SIMPLE key features exam is a required component of the clerkship, and you will receive a report of your scores and the corresponding SIMPLE cases for further study, whether or not you choose to complete the survey.

All the information I receive from you, including your name and any other identifying information, will be strictly confidential and will be kept under lock and key. I will not identify you or use any information that would make it possible for anyone to identify you in any presentation or written reports about this study. Only summarized data will be presented at meetings or in any publications. If you choose to complete the survey, your SIMPLE key features scores and NBME Medicine Subject Examination (aka “shelf exam”) scores will be used in the analysis. The Institute for Innovative Technology in Medical Education (iInTIME), the organization which administers the SIMPLE exam and the survey, will not have access to your shelf exam scores at any time. All identifiable information (e.g. name, email) will be removed prior to analysis.

You should be aware, however, there is a small possibility that responses could be viewed by unauthorized parties (e.g. computer hackers because your responses are being entered and stored on a web server). There are no other expected risks to you for helping me with this study. There are also no expected benefits for you either.

For more information or questions about this research you may call Dr. Valerie Lang (585) 275-4912. If you have questions, concerns or complaints about your rights as a research subject you may contact, anonymously if you wish, Human Subjects Protection Specialist at the University of Rochester Research Subjects Review Board, Box 420315, 265 Crittenden Boulevard, Rochester, NY 14642-8315. Telephone 1-(585) 276-0005. For long-distance you may call toll-free, 1-(877) 449-4441. You may also call these numbers if you cannot reach the research staff or wish to talk to someone else.

RSRB case number: 00040172 Page 1 of 1 Version Date: 12/15/2011
VITA

Personal Information

Work: 601 Elmwood Avenue, Box MED-HMD
       Rochester, NY 14642

       Tel: (585) 273-4881
       Fax: (585) 273-1069

       e-mail: Valerie.Lang@urmc.rochester.edu

Home: 455 Mt. Vernon Avenue
       Rochester, NY 14620

       Tel: (585) 461-2893

Citizenship: United States

       Born: Rochester, NY, USA

Education

2009-present  Masters of Health Professions Education, in progress

       Thesis: "Validity Evidence for a Key Features Examination for the Internal
               Medicine Clerkship"

       University of Illinois at Chicago

       Chicago, IL

1993-1997  Doctor of Medicine

       SUNY Health Science Center College of Medicine
Syracuse, NY

1988-1992  Bachelor of Science

Magna Cum Laude

Major in Biology, Minor in Philosophy

Allegheny College

Meadville, PA

University of Aberdeen

Aberdeen, Scotland

Residency

1997-2000  Categorical Residency in Internal Medicine

University of Rochester Medical Center; Rochester, NY

Licensure

2000-present  New York State License #215167

Professional Certification

2000, 2010  American Board of Internal Medicine
Additional Certifications

2006    Advanced Cardiac Life Support

2003    Patient Safety Certification Course
         University of Rochester School of Medicine and Dentistry
         Department of Community and Preventive Medicine

Faculty Appointments

2011-present    Associate Professor of Medicine
         University of Rochester Medical Center
         Rochester, NY

2003-2011    Assistant Professor of Medicine
         University of Rochester Medical Center
         Rochester, NY

2000-2003    Senior Instructor of Medicine
         University of Rochester Medical Center
         Rochester, NY

Professional Appointments

2000-Present    Hospitalist
Hospital Medicine Division
University of Rochester Medical Center

2002-present  Director, Adult Inpatient Medicine Clerkship

University of Rochester School of Medicine & Dentistry
Rochester, NY

- Supervise training for 104 students per year at four clerkship sites
- Conduct ECG interpretation workshops weekly
- Conduct evidence-based medicine workshop semi-monthly
- Lead grading committee
- Counsel students interested in Internal Medicine residencies and write 30-60 departmental letters per year

2003-present  Director, Medicine Sub-Internship

University of Rochester School of Medicine & Dentistry
Rochester, NY

2008-present  Director, Hospital Medicine Division Faculty Development Program University of Rochester School of Medicine & Dentistry
Rochester, NY

2010-present  Editor-in-Chief, SIMPLE (Simulated Internal Medicine Patient Learning Experience) Virtual Patient Program
Institute for Innovative Technology in Medical Education (MedU)
Lebanon, NH
Professional Society Membership

2003-2011  Society of General Internal Medicine

2002-present  Clerkship Directors in Internal Medicine

2002-present  Upstate New York Chapter of Society of Hospital Medicine
               President and Founder, 2002-2006

2000-present  Society of Hospital Medicine

1997-present  American College of Physicians

Honors and Awards

2013  Special Commendation for Third Year Teaching
      University of Rochester School of Medicine & Dentistry

2012, 2013  Patient and Family-Centered Care Top Performer Award
              University of Rochester Medical Center

2009  Louis N. Pangaro Educational Program Development Award
      Clerkship Directors in Internal Medicine
2009  Fellow, Society of Hospital Medicine

2009, 2010  Top 10% Reviewer, Journal of Hospital Medicine

2007  Research Abstract Award, Clerkship Directors in Internal Medicine

2005  Fellow, American College of Physicians

2004-2006  Jules Cohen Dean’s Teaching Fellow

1996  Alpha Omega Alpha

1992  Phi Beta Kappa

Educational Contributions of Past 10 Years

Recurring Activities

<table>
<thead>
<tr>
<th>Years</th>
<th>Course/ Clerkship</th>
<th>Activity</th>
<th>Topic</th>
<th>Learner Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002-present</td>
<td>Inpatient Medicine Clerkship</td>
<td>Clerkship Director</td>
<td>Inpatient internal medicine</td>
<td>3rd year</td>
</tr>
<tr>
<td>2003-present</td>
<td>Medicine Sub-Internship</td>
<td>Sub-Internship Director</td>
<td>Advanced inpatient internal medicine</td>
<td>4th year</td>
</tr>
<tr>
<td>2001-present</td>
<td>Inpatient Medicine Clerkship</td>
<td>Preceptor Rounds</td>
<td>Clinical reasoning and bedside clinical skills</td>
<td>3rd year</td>
</tr>
<tr>
<td>Date</td>
<td>Role</td>
<td>Activity</td>
<td>Description</td>
<td>Year</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------------------------</td>
<td>------------------------------------</td>
<td>-------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>2002-present</td>
<td>Inpatient Medicine Clerkship</td>
<td>ECG Workshops</td>
<td>Introduction to interpretation of ECGs</td>
<td>3rd</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Application of evidence-based medicine skills in clinical decision-making</td>
<td></td>
</tr>
<tr>
<td>2002-present</td>
<td>3rd Year Comprehensive Assessment</td>
<td>Oral examiner</td>
<td>Direct observation of clinical skills, oral exam of clinical reasoning</td>
<td>3rd</td>
</tr>
<tr>
<td>2005-present</td>
<td>2nd Year Comprehensive Assessment</td>
<td>Oral examiner, video review facilitator</td>
<td>Direct observation of clinical skills, oral exam of clinical reasoning</td>
<td>2nd</td>
</tr>
</tbody>
</table>

**Residents**

<table>
<thead>
<tr>
<th>Date</th>
<th>Role</th>
<th>Activity</th>
<th>Description</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001-present</td>
<td>Internal Medicine Residency</td>
<td>R1 Morning Report</td>
<td>Clinical reasoning</td>
<td>R1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Clinical decision-making, direct observation of clinical skills</td>
<td>R1-R4, MS3, MS4</td>
</tr>
<tr>
<td>2008-2011</td>
<td>Chief Residents Training Program (all fields)</td>
<td>Rising Chiefs Workshop</td>
<td>Feedback skills</td>
<td>R3-R5</td>
</tr>
<tr>
<td>2005-present</td>
<td>Internal Medicine Intern Retreat</td>
<td>Breakout sessions</td>
<td>Work-life balance, errors in medicine, women in medicine</td>
<td>R1</td>
</tr>
</tbody>
</table>

**Additional Activities**

1. **Medical Students**

**2009-10**

- Faculty Mentor: Stress, Adaptation, and Transition Clinical Practice Essay; 3 hours
- Facilitator; Mindfulness Series: Suffering; August 13, 2009; 1.5 hours
2008-9

- Faculty Mentor: Stress, Adaptation, and Transition Clinical Practice Essay; 4 hours
- Professional Development Group Facilitator, Successful Interning Course; 2 hours
- Professional Development Group Facilitator, 1st Year Student Orientation; 2 hours
- Facilitator, Introduction to Clinical Medicine Art and Observation PBL; 2 hours

2007-8

- Tutor, 2nd Year Case Seminar Clinical Practice Essays; 4 hour

2006-7

- Master Clinician Rounds Preceptor; Ambulatory Clerkship, Jan-March, 2007; 10 hours
- Tutor, 2nd Year Case Seminar Clinical Practice Essays; 4 hours

2005-6

- Group Leader, “Art of Observation” project, 1st year orientation; 2 hours
- Tutor, 2nd Year Case Seminar Clinical Practice Essays; 4 hours
- Preceptor, 1st year orientation professionalism seminars; 10 hours

2004-5

- Group Leader, “Art of Observation” project, 1st year orientation; 2 hours
- Director and Preceptor, Medical Student Hospitalist Elective; 80 hours
- Preceptor, 1st year orientation professionalism seminars; 10 hours
- Preceptor, Introduction to Clinical Medicine Physical Exam (1st year students); 12 hours
- Standard Setter, National Board of Medical Examiners Medicine Subject Exam

2. Residents

2014-15

- Resident Poster Day Judge
2010-11

- Reviewer, Intern Survival Guide “Alcohol Management” section

2009-10

- Master Teacher Rounds; 6 hours
- “Suffering”; Mindfulness series; 1 hour
- “Alcohol Withdrawal”; Resident Noon Conference; 1 hour

2008-9

- Master Teacher Rounds; 6 hours
- “Teaching Medical Students”; R2 Skills Day; 1 hour
- “Balance”; Mindfulness series; 1 hour
- “Time”; Mindfulness series; 1 hour
- “Interpersonal Mindfulness: Awareness in Action”; communication skills for residents; 1 hour

2007-8

- “Burnout”; Mindfulness Series; 1 hour
- “How Doctors Think”; Resident Noon Conference; 1 hour
- “There are Medical Students on My Team”; Becoming an R2 Day; 1 hour

2006-7

- “What to do when you have a medical student on your team”: Becoming an R2 Day; 1 hour
- Research Mentor for Anne Marie Mattingly, M.D.; U. of Rochester Internal Medicine Residency Program, 2006-08

2004-5

- “What to do when you have a medical student on your team”, Becoming an R2 Day; 1 hour
3. Continuing Professional Education

2014-15

- “DOCS-REx: Direct Observation of Clinical Skills and Reasoning Exercise”. University of Rochester Highland Hospital Preceptor Faculty Development.
- “DOCS-REx: Direct Observation of Clinical Skills and Reasoning Exercise”. Rochester General Hospital Department of Medicine Faculty.

2013-14

- “Teaching High Value Care”. University of Rochester Faculty Development Colloquium.
- “Direct Observation of Clinical Skills and Reasoning Exercise”. University of Rochester Hospital Medicine Division.
- “Choosing Wisely: Hospital Medicine”. University of Rochester Department of Medicine Grand Rounds.
- “Direct Observation: Part 2”. University of Rochester Hospital Medicine Division.
- “The Power of Watching: Direct Observation at the Bedside.” University of Rochester Hospital Medicine Division.
- “Direct Observation of Clinical Skills”. Unity Health Department of Medicine Faculty.
- “Providing Effective Feedback to Residents and Students.” Unity Health Department of Medicine Faculty.
- “Medical Student Education and Hospitalists: Q & A.” University of Rochester Hospital Medicine Division.
- “The Medical Student Experience Today: The Spectrum of Simulation in the Double Helix Curriculum.” University of Rochester Meliora Weekend

2012-13

- “Innovations in the Medicine Clerkship.” University of Rochester Combined Instruction Committee.
- “The Gift of Feedback in the Clinical Setting.” Rochester General Hospital; Hospital Medicine Division Faculty Development; 1 hour.
- “Direct Observation of Clinical Skills.” Rochester General Hospital; Hospital Medicine Division Faculty Development; 1 hour.
• “Medical Education Scholarship”. University of Rochester; Hospital Medicine Division Faculty Development; 1 hour.
• “Virtual Patients: In Class, On-Line, and On Rounds.” University of Rochester Instruction Committee Retreat.
• “Direct Observation of Clinical Care.” University of Rochester; Department of Family Medicine; 1 hour.
• “Introduction to High Value Care.” University of Rochester General Medicine and Hospital Medicine Faculty Development; 1 hour.

2011-12

• “Assessing Clinical Decision Making”. University of Rochester Faculty Development; 1 hour.
• “How to Teach Diagnostic Reasoning Using Likelihood Ratios”. University of Rochester Hospital Medicine and General Medicine Faculty Development; 1 hour

2010-11

• “Validation of a Key Features Examination to Assess Clinical Decision-Making in Medicine Clerkship Students”; University of Rochester Medical Education Research Interest Group; 1 hour
• “Feedback and the Feed-forward Interview”; University of Rochester Medical Education Research Interest Group Journal Club; 1 hour
• “Teaching Scripts”; University of Rochester Faculty Development Colloquium; 2 hours
• “Effective Feedback”; University of Rochester Faculty Development Colloquium; 2 hours
• Development of position paper on expanding the University of Rochester’s on-line CME program: “Using Evidence-based Education Practices to improve Physicians’ Practices”
• “Feedback”: University of Rochester Dean’s Teaching Fellows; 2 hours

2009-10

• “Aligning Compensation with Education: EVUs”; GMD/HMD Journal Club; 1 hour
• “Introduction to Teaching Scripts and the UTI Teaching Script”; HMD Faculty Development Workshop; 1 hour
• “Feedback I”; GMD/HMD Faculty Development Workshop; 1 hour
• “Feedback II: Difficult Feedback Scenarios”; GMD/HMD Faculty Development Workshop; 1 hour
• “Feedback Refresher”; University of Rochester Physical Medicine and Rehabilitation Faculty Development Program; invited workshop; 2 hours

• “Effective Feedback”; University of Rochester Faculty Development Colloquium; 2 hours

2008-9

• “Medical Students on the Inpatient Service”; Hospital Medicine Faculty Orientation; 1 hour

• “Effective Feedback”; University of Rochester Dean’s Teaching Fellowship Program; invited workshop; 2 hours

• “Providing Feedback that Works” University of Rochester Faculty Development Program; 2 hours

• “Providing Effective Feedback”; University of Rochester Faculty Development Colloquium; 2 hours

2007-8

• “Understanding Meta-Analyses; Low-Molecular Weight Heparin vs. Unfractionated Heparin for VTE Prophylaxis”; HMD/GMD Journal Club; 1 hour

• “Medical Students on the Inpatient Service”; Hospital Medicine Faculty Orientation; 1 hour

• “Direct Observation of Clinical Skills and Providing Effective Feedback”; invited workshop; Mary Imogene Bassett Hospital Faculty Development Program; 8 hours

• “Providing Effective Feedback”; University of Rochester Physical Medicine and Rehabilitation Faculty; invited workshop; 2 hours

• “Providing Effective Feedback”; University of Rochester Faculty Development Colloquium; invited workshop; 2 hours

• “Making the Most of Feedback”; University of Rochester Physical Medicine and Rehabilitation Residents; invited workshop; 2 hours

2006-7

• “Tolvaptan for Treatment of Hyponatremia”; HMD/GMD Journal Club; 1 hour

• “Urethral Catheters”; HMD/GMD Journal Club; 1 hour

• “Providing Effective Feedback”; University of Rochester Faculty Development Colloquium; 2 hours

2005-6
• “Errors in Clinical Reasoning”; HMD/GMD Journal Club; 1 hour
• Panelist, University of Rochester Department of Medicine Grand Rounds Case Discussion, 1 hour
• “Direct Observation of Clinical Skills”; HMD/GMD Faculty Development Workshop; 1 hour
• “Direct Observation of Clinical Skills”; Unity Hospital Faculty Development Workshop; 1 hour

2004-5

• University of Rochester Department of Medicine Grand Rounds: “Update in Hospital Medicine”; 0.5 hours
• “Prevention of Contrast-Induced Nephropathy with Sodium Bicarbonate”, GMU Journal Club; 1 hour

4. Interprofessional Education

2013-14

• “Choosing Wisely in Hospital Medicine”. University of Rochester Hospital Medicine Advanced Practice Provider conference; 1 hour.
• Interprofessional Patient Safety conference facilitator; University of Rochester School of Medicine & Dentistry and School of Nursing; 1 hour

2012-13

• “Effective Feedback”; University of Rochester Nurse Practitioner Faculty Development
• “Virtual Patients in APP Clinical Education”; University of Rochester; Hospital Medicine Advanced Practice Providers.

2011-12

• “Direct Observation and Effective Feedback on Clinical Skills”; New York Chiropractic College Clinical Faculty Development Conference.

2009-10

• “Pulsus Paradoxus”; U. of Rochester Advanced Practice Providers; 1 hour
• “Effective Feedback”; University of Rochester School of Nursing Advanced Practice and Accelerated Nursing Programs Faculty Retreat; invited workshop; 2 hours

Committee Assignments

National

2014-2015 President
Clerkship Directors in Internal Medicine

2013-present Board of Directors
Alliance for Academic Internal Medicine

2014-present Senior Communications Council
MedU

2014-present Women in Academic Internal Medicine Forum Working Group
Alliance for Academic Internal Medicine

2013-present Vice Chair, High Value Care Working Group
Alliance for Academic Internal Medicine

2013-present High Value Care Advisory Board
American Board of Internal Medicine- American Colleges of Physicians- Alliance for Academic Internal Medicine
2013-present  MedU Research Committee

2013-present  Program Planning Committee

  Clerkship Directors in Internal Medicine
   •  Co-Director, CDIM-APDIM Pre-course on Medical Education Research (2014-15)

2013-present  Millenium Conference Task Force on Value-Added Care

2013  Working Group on Alliance for Academic Internal Medicine Committees

2011-2013  Councilor

  Clerkship Directors in Internal Medicine

2012-2013  Leader, CDIM-APDIM Task Force on Department of Medicine Letters

2012-2013  Educators’ Toolbox Working Group

  MedU

2009-2011  Chair, SIMPLE Committee

  Clerkship Directors in Internal Medicine

2008-2009  USMLE Step 3 Utilization of Resources Test Material Development Committee

2008-2009  Vice Chair, SIMPLE Committee

  Clerkship Directors in Internal Medicine
2005-2008  Education Committee
            Society of Hospital Medicine

2005-2007  Educational Innovations Committee
            Clerkship Directors in Internal Medicine
            Vice Chair, 2007

2005-2006  Education Committee Liaison to Geriatrics Resource Room
            Society of Hospital Medicine

2005-2006  Task Force on Industry Support of Local Chapters
            Society of Hospital Medicine

Regional

2002-2006  Northeast Regional Council
            Society of Hospital Medicine

University of Rochester: University-Wide

2010-2011  Medical School Liaison to Learning Outcomes Steering Committee

University of Rochester: School of Medicine & Dentistry
2002-present  Third and Fourth Year Instruction Committee
  • Chair, 2006-7; 2010-11
  • Vice Chair, 2005-6; 2009-10; 2014-15

2014-2016  Dean’s Teaching Fellowship Advisory Committee

2014-2016  Continuing Medical Education Advisory Committee

2014-2016  Disease Process & Therapeutics and Bed-to-Bench Advisory Committee

2014-2016  Curriculum Steering Committee
  (& 2009-2011)

2014-2015  LCME Reaccreditation Subcommittee
  • Review standards for competencies and curricular objectives, design, and content and write report

2011-present  Medical Education Research Interest Group

2011-2013  Comprehensive Assessment Design Team

2011-2012  Continuing Medical Education Advisory Committee
  Department of Medicine Representative

2008-2009  Committee on Redesign of Year 2 Case Seminars

2006-2013  Medical School Admissions Committee
  Committee Member, 2006- 2009
Interviewer, 2009-2013

2006-2007  Committee on the Educational Program for the MD Degree
           Third Year Redesign subcommittee member

2005-2007  Electrolyte Replacement Protocol Development Committee

2005      Faculty Development Task Force

2003-2009  Strong Memorial Hospital Clinical Council

2003-2005  Continuity of Care Committee

2004-2005  Strong Memorial Hospital Provision of Care Task Force

2003-2004  Alpha Omega Alpha Selection Committee

2003      Root Cause Analysis Committee

2002-2003  Alcohol Withdrawal Protocol Development

2002-2003  Inpatient Steering Committee

2001-2006  Internal Medicine Residency Selection Committee
Research Activities

Grants

  
  Leader of challenge grant to develop a hospitalist faculty development program for educating medical students about delirium and falls in geriatric inpatients.

- **Jules Cohen, MD Dean’s Teaching Fellowship (2004-2006)**
  
  Competitive fellowship for advanced training in medical education and completion of project: Development of virtual patient cases to replace lecture series in the Medicine

- **Dean of Faculty Development: Medical Education Grant (2010).** PI: V. Lang
  
  Developed and evaluated a program to collaboratively develop teaching scripts for hospitalist faculty development. Grant also funded tuition for masters-level Leadership in Health Professions Education course.

- **American Board of Internal Medicine Foundation Grant; Josiah Macy, Jr. Foundation President’s Grant; Alliance for Academic Internal Medicine Innovations Grant.** PI: V Lang, CD Smith, N Berman. (2013-14)
  
  On-Line High Value Care Curriculum Development: A collaborative project of Med-U, the American College of Physicians, and the Alliance for Academic Internal Medicine. We developed an interactive, case-based national on-line curriculum in high value care for medical students. To evaluate the impact of the curriculum, the project also involved developing and validating two tools: an attitudinal scale and a clinical decision-making exam. A 1-year multi-institutional curriculum evaluation study launched July, 2014.

Research: Other

- Development and validation of an internal medicine clerkship key features examination as a companion to SIMPLE virtual patient cases; 2011-present.

- Randomized, controlled trial of student reactions to varying context in on-line virtual patient cases

Presentations at National or International Professional Meetings


4. Lang VJ. Nixon LJ. Kogan J. Kovach R. Fall LH. Berman N. It’s SIMPLE, and It’s Here! Workshop for clerkship directors to incorporate web-based cases into the curriculum. Academic Internal Medicine Week, 2008.


Reviewer/Editorial Assignments in Professional Journals or National Organizations


2004-2005  Reviewer, Medical Knowledge Self-Assessment Program (MKSAP) 3 for Students

2005  Reviewer, Society of General Internal Medicine Annual Meeting Education Research Abstracts

2007  Reviewer, Medical Knowledge Self-Assessment Program (MKSAP) 4 for Students

2008-2011  Assistant Editor, *Journal of Hospital Medicine*

2011-present  Reviewer, *Academic Medicine*

2012- present  Reviewer, *Journal of Hospital Medicine*

2014  Reviewer, American College of Physicians National Abstract Competition
Other Editorial Assignments

Editor-in-Chief, SIMPLE (Simulated Internal Medicine Patient Learning Experience) Virtual Patient Program; www.med-U.org; 2009- present

- 36-case, peer-reviewed, interactive virtual patient program
  - Comprehensive coverage of national clerkship curriculum
  - Over 782,000 case sessions completed at 118 medical schools (as of 2014)
- Oversee 8-member editorial board, case updates, and supplemental projects
  - Led collaboration with American College of Physicians to integrate MKSAP for students self-assessment questions into cases
  - Piloted manuscript-style case review process

Med-U Editorial Board

- Oversee pedagogical approach and development of collaborative programs with Editors-in-Chief of five interactive computer-assisted learning programs, including internal medicine (SIMPLE), pediatrics (CLIPP), family medicine (fmCASES), radiology (CORE), and surgery (WISE-MD)

Guest Editor, Hospital Medicine Clinics, 2013. 2(4):e481-e612.

Curriculum Development

1. Simulated Internal Medicine Patient Learning Experience (SIMPLE) Project Development Group

- Interactive on-line case-based curriculum for Medicine Clerkship
- Developed case outlines to comprehensively cover national learning objectives for the internal medicine clerkship
- Trained 31 clerkship directors from around the country to author 36 virtual patient case modules
- Edited nine virtual patient case modules
- Authored two virtual patient case modules
• Monitored peer-review process
• Update curriculum annually
• Utilized at 118 medical schools; over 782,000 cases completed

   • Interactive on-line case-based curriculum for undergraduate education
     • Collaborated with the American College of Physicians, Alliance for Academic Internal Medicine, and MedU
     • Led adaptation of learning objectives and developed module outlines for 6 modules in high value care
     • Recruited and trained authors and peer reviewers
     • Authored module: “High Value Care and the Importance of Clinical Reasoning”

3. High Value Care Elective
   • 1-week curriculum in high value care and multimedia instructional design for medical students in the medical education pathway

4. Teaching Script Curriculum for Hospitalists
   • 1-year program for hospitalist faculty development

5. Medical Student Hospitalist Elective
   • 2-week elective in hospital medicine and quality improvement

6. Dx: Diagnostic Excellence (in development)
   • MedU Project Co-Leader in collaboration between MedU and the Society to Improve Diagnosis in Medicine.
   • Online interactive curriculum to teach medical students about cognitive errors in diagnosis and approaches to mitigating errors.
Publications

Original Research


Books, Monographs, Chapters, and Review Articles


Virtual Patient Case Modules


Letters, Editorials, Short Articles, and Other Contributions

1. Lang VJ, Rudmann A. Hospitalist Practice Profile. The Hospitalist. 2002;6(2).


Abstracts


7. O’Connor AB, Lang VJ, Lambert DR, Rudmann A, Robbins B, Bordley DR. Impact of non-teaching services on internal medicine resident inpatient case distribution. Association of Program Directors in Internal Medicine; Spring, 2007 Meeting. Winner, Research Abstract Award.


