How we developed and piloted an electronic key features examination for the internal medicine clerkship based on a US national curriculum

KIRK A. BRONANDER1, VALERIE J. LANG2, L. JAMES NIXON3, HEATHER E. HARRELL4, REGINA KOVACH5, SUSAN HINGLE5 & NORMAN BERMAN6

1University of Nevada School of Medicine, USA, 2University of Rochester School of Medicine, USA, 3University of Minnesota College of Medicine, USA, 4University of Florida College of Medicine, USA, 5Southern Illinois University School of Medicine, USA, 6Geisel School of Medicine at Dartmouth, USA

Abstract

Background: Key features examinations (KFEs) have been used to assess clinical decision making in medical education, yet there are no reports of an online KFE-based on a national curriculum for the internal medicine clerkship.

What we did: The authors developed and piloted an electronic KFE based on the US Clerkship Directors in Internal Medicine core curriculum. Teams, with expert oversight and peer review, developed key features (KFs) and cases.

Evaluation: The exam was pilot tested at eight medical schools with 162 third and fourth year medical students, of whom 96 (59.3%) responded to a survey. While most students reported that the exam was more difficult than a multiple choice question exam, 61 (83.3%) students agreed that it reflected problems seen in clinical practice and 51 (69.9%) students reported that it more accurately assessed the ability to make clinical decisions.

Conclusions: The development of an electronic KFs exam is a time-intensive process. A team approach offers built-in peer review and accountability. Students, although not familiar with this format in the US, recognized it as authentically assessing clinical decision-making for problems commonly seen in the clerkship.

Introduction

Medical educators are continually looking for improved methods of medical student assessment. In the United States, 93% of internal medicine clerkship directors rely on the National Board of Medical Examiners (NBME) subject exam in Internal Medicine to assess core knowledge (Kelly et al. 2012). However, 33% also use locally developed exams. The majority developed these exams to assess content that reflects their clerkship curricula, but local exams often have limited validity evidence to support their use (Kelly et al. 2012). Furthermore, many clerkship directors indicated clinical decision making was an area that was underrepresented on the NBME subject exam and this contributed to their decision to create a locally developed exam (Kelly et al. 2012).

What we did

In response to these concerns, we developed an online key features examination (KFE) assessing clinical decision-making for the internal medicine clerkship based on the Clerkship Directors in Internal Medicine (CDIM) core curriculum objectives. While KFE have been developed for clerkship students before (Hatala & Norman 2002; Fischer et al. 2005), this is the first KFE developed using a national curriculum. Following development, we conducted a pilot test to determine feasibility and acceptability of the exam.

Practice points

- A key feature is a critical step that is necessary for the successful resolution of a problem.
- Key features examinations have been used to assess clinical decision making, but are less familiar to medical educators in the US.
- Successful development of a key features exam is a time-intensive and team-based process.
- Students reported this key features exam resembled clinical practice.

Background of KFEs

A key feature (KF) is defined as a critical step in the successful resolution of a clinical problem (Page et al. 1995). A key features exam (KFE), by focusing on these critical steps, can reliably assess reasoning skills in a particular area using as few as 2–3 items per case vignette (Norman et al. 2006). The KFE...
format is flexible and allows sequential pieces of clinical information to be provided between questions (Farmer & Page 2005). This allows a student to follow a case longitudinally from initial work-up and diagnosis to final management. It also allows for more than one correct answer per question, which is more representative of actual clinical practice (Farmer & Page 2005).

There is strong evidence linking KFE scores and actual practice. The Medical Council of Canada Qualifying Exam (CQE) has used a KFs approach since 1992 (Pague & Bordage 1995). Scores on the KFs component of the CQE are more predictive of patient adherence to antihypertensive regimens (Tamblyn et al. 2010) and complaints to medical regulatory authorities than other portions of the CQE (Tamblyn et al. 2007).

Developmental process

Development group

Nine experienced undergraduate internal medicine educators from a diverse group of LCME-accredited medical schools were recruited to be members of the KFs development group (KFDG). An expert in KFE development guided the group throughout the process.

Exam blueprint

The KFs cases were drawn from content in the Simulated Internal Medicine Patient Learning Experience (SIMPLE) virtual patient instructional program, which was designed to address all CDIM core curriculum training problems (an established US national curriculum). Drawing from the 36 SIMPLE virtual patient cases, 71 KF case vignettes were developed (3 case vignettes from one SIMPLE case; 2 case vignettes from 33 SIMPLE cases; and 1 case vignette each from two SIMPLE cases). Each case vignette contained 2–3 KFs. In total, 67% of the cases were in an outpatient setting (including offices and emergency departments) and 34% in an inpatient setting. Fifty-nine percent of the KFs focused on diagnosis and 41% focused on management.

KF identification

The KFDG convened for a two-day workshop on how to identify the KFs. The KFDG worked in teams of two or three to develop the KFs, providing feedback, accountability and peer review throughout the process. In addition, an expert in KF development provided immediate, early feedback to each team on their KF drafts. Any disagreements regarding the content of the KFs were resolved by consensus among the full KFDG. After the initial workshop, teams worked online using a file-sharing program to develop case vignettes, questions and answer options. Final components were written at a second two-day workshop with final editing and revision online. Each case was peer reviewed by at least two members of the group and the expert consultant.

Case vignette development

Case vignettes included patient demographics (name, age and gender), the setting (inpatient, emergency room and outpatient) and any other clinical information needed to resolve the problem. The language in the vignettes was written in lay terms to decrease the likelihood of cuing and increase reliability (Eva et al. 2010). For example, instead of “he had tonic–clonic muscle activity” the case vignette stated, “he twitched and jerked for a minute or two”.

Question development

Questions followed the vignettes and consisted of two sentences. The first sentence asked the student to make a diagnostic or therapeutic choice (or choices). Some examples were as follows: “What is your leading diagnosis at this time”? or “what actions will you take at this time”? The second sentence clarified the limit of choices, such as: “Select only one”; “you may select up to five”; or “select as many as applicable”. Usually one question corresponded to one KF, however, occasionally a single question contained multiple KFs. Answer options consisted of a “short-menu format” (Page et al. 1995) of 15–25 different options arranged in alphabetical order. The answer options included the correct answers, plausible but incorrect options and rarely a dangerous or “poison” option.

Scoring key development

The final step was to create a scoring key for each KF. Each KF was scored separately and was worth one point. A single question could include two KFs and would be worth two points in that instance. Some KFs contained multiple components and students could achieve partial credit. For example, if the KF was “administer intravenous benzodiazepines, thiamine and glucose” to a patient with alcohol withdrawal seizures, the student would receive 0.33 points for each correct management option they selected. Students who selected a dangerous option (poison option) or selected more than the maximum number of options allowed lost all credit for that KF. Appendix 1 (available as Supplemental Material) shows an example of two KFs, a case vignette, a question, answer options and a scoring key.

The exam was administered on-line through a custom-designed web interface. Students accessed the test by signing into a secure website; once in the testing interface, they were given instructions and an example of a KF case before the start of the exam. Questions were delivered sequentially, and the electronic format prevented students from changing their answers to previous questions.

Pilot study

Methods

After receiving institutional review board exemptions, eight medical schools including public and private schools of varying size and geographic location participated in the pilot study. The 71 cases were randomly combined into several exams consisting of nine cases, each based on an estimation that this exam would take approximately one hour. When students had excessive time remaining after the first two exam administrations, the remaining exams were lengthened to
12 cases. This study was conducted at the beginning of the academic year. Most students were rotating in their first clerkship of the third year, and at one site, the exam was administered to 13 students in the fourth year sub-internship to obtain feedback from more clinically experienced students. Students at each site took the same version of the exam.

A survey assessing students’ acceptance of the exam was developed and reviewed by the exam developers (Appendix 2, available as Supplemental Material). There were 11 questions with answer options on a five-point Likert scale, one question with a binomial (yes/no) response and two free-text questions. In free-text questions, students were asked what they liked most about the KFs exam and what they would most like to change about the exam. The survey was accessed on-line at the end of the exam. Participation was voluntary. Participants were entered into a drawing for an iPad. All data were de-identified prior to analysis.

Analyses

Time spent on each exam case was recorded automatically by the exam software. Survey responses were analyzed and divided into categories of strongly agree/agree, neutral, and disagree/strongly disagree. Free text comments were coded and reviewed independently by two authors (H. E. H. and V. J. L.). Themes were generated using an iterative process.

Results

The exam was administered to 162 students at eight medical schools. A total of 96 students responded to the survey; students from three of the sites chose not to participate (Table 1). Eighty-three (86.4%) participants were in their third year and 13 (13.5%) were in their fourth year. All students were in an internal medicine clerkship or sub-internship. For many students, the exam took place during their first clerkship (mean completed clerkships 0.86; range 0–6).

Students spent a mean of 4.6 min on each case. The nine-case exams took a mean of 35.7 min to complete. The 12-case exams took a mean of 45.3 min to complete. Fifty-eight (79.4%) students reported that they had just enough time to complete the exam, 12 (16.4%) reported that they had too much time and 5 (6.8%) stated that they had too little time. Compared to a standard multiple-choice exam, 49 (68.1%) students found the KFs exam more difficult, 3 (4.2%) found it easier and 20 (27.8%) were neutral about the relative difficulty. Forty-three (59.7%) students reported that they could not find their preferred answer option (Table 2).

Qualitative analysis

Acceptability

There were 50 comments from 44 respondents to the question asking what they liked most about the exam. There were 49 comments from 43 respondents to the question asking what they would most like to change about the exam. Five responded that they would not change anything about the exam, and one responded that he or she did not like anything. Four comments were too vague to code and one did not address the exam. The remaining 88 comments were coded into six themes: format, authenticity, feedback, technical, cognitive level and clarity of instructions. Representative comments are included below.

Format

There were 28 comments about the exam format, 13 of which were positive and 15 of which were negative. Positive comments addressed the intuitive nature of the format, with the option to choose multiple correct answers when making a decision. Others appreciated the conciseness of the vignettes and the straightforward questions.

“Short, bread-and-butter, without attempts to be tricky”.

Other students expressed a desire for more information in the vignettes. Some did not like questions that limited the number of correct answers, while others did not like questions that allowed them to select as many options as they thought were appropriate.

“I prefer multiple choice exams with one right answer or an oral exam where my ability to make decisions is not limited by the computer program and I can explain my reasoning”.

Authenticity

There were 18 positive comments and 7 negative comments about authenticity. Students recognized a connection between the assessment cases and real patient cases they had encountered, describing the key features cases as “realistic” and “clinically oriented”.

“Many of the cases correlated almost perfectly with what I’ve experienced in Medicine”.

However, some students were uncomfortable with the variability they had observed in how different clinicians make decisions with similar real cases.

---

**Table 1.** Participants in the pilot exam and survey.

| Medical School | Students completing exam (n) | Students completing the survey (n) |
|----------------|-------------------------------|-----------------------------------|
| A              | 3                             | 3 (100%)                          |
| B              | 66                            | 33 (50%)                          |
| C              | 19                            | 18 (94.7%)                        |
| D              | 29                            | 29 (100%)                         |
| E              | 13                            | 13 (100%)                         |
| F              | 23                            | 0 (0%)                            |
| G              | 5                             | 0 (0%)                            |
| H              | 4                             | 0 (0%)                            |
| Total          | 162                           | 96 (59.3%)                        |

---

809
“Some of the questions I felt were a matter of clinical judgment and different physicians would handle these situations differently”.

Feedback

There were three positive and eight negative comments about feedback. While some students appreciated receiving their scores for each assessment case, they desired even more specific information about the answers, which had not been reported in order to keep the exams secure.

“I liked knowing roughly how I did after each case”. “It is frustrating not to be able to review the correct answers, and its usefulness as a learning tool was limited by this”.

Technical

There were four positive and nine negative comments about technical aspects of the exam. Some appreciated the computer-based delivery format.

“On-line, easy to navigate”.

However, some respondents had slow Internet connections or issues with seeing the laboratory values and electrocardiograms clearly. Because some of the cases progressed serially, responses could not be changed after advancing to the next question, which some students did not like.

“I did not like that I could not change my answer after clicking submit”.

Cognitive level

There were eight comments about the cognitive level assessed by the exam, all of which were positive. One described the exam as “challenging”, and another stated that it required “not a strict recall of facts”.

“The format was more integrative than other multiple choice exams”.

Clarity of instructions

Three respondents had negative comments about the clarity of instructions, all of which addressed how to select answer options. For questions that allowed students to select as many tests as they deemed appropriate, one student was concerned that being allowed to over-order tests might lead to selection of a “poison” option (a test which was not indicated and could harm the patient, thus forfeiting all points for that question).

“I think that a statement that encourages not using all the choices you’re given would be helpful. Several times I only needed 3 tests, but since I could choose 6, I did and ended up with 0%”.

Next steps

Authenticity emerged as a positive theme in the pilot study. We were encouraged that the majority of students found that the exam problems resembled clinical practice and accurately reflected internal medicine clerkship content and their ability to make clinical decisions. The next step is to gather validity

Table 2. Internal medicine student attitudes about the feasibility and acceptability of a pilot on-line key features examination.

| Agree/strongly agree n (%) | Neutral n (%) | Disagree/strongly disagree n (%) |
|---------------------------|---------------|---------------------------------|
| Exam problems resemble problems from clinical practice | 61 (83.6%) | 8 (11.0%) | 4 (5.5%) |
| Understood the instructions for each of the questions | 59 (80.8%) | 10 (13.7%) | 4 (5.5%) |
| Exam accurately reflected medicine clerkship content | 45 (61.6%) | 16 (21.9%) | 12 (16.4%) |
| Exam accurately assessed ability to apply what was learned in the medicine clerkship | 41 (56.2%) | 23 (31.5%) | 9 (12.3%) |
| Exam more accurately assessed ability to make clinical decisions than a standard multiple choice exam | 51 (69.9%) | 14 (19.2%) | 8 (11.0%) |
| Software ran smoothly with no technical problems | 62 (86.1%) | 5 (6.9%) | 5 (6.9%) |
| Exam schedule ran without problems | 69 (95.8%) | 1 (1.4%) | 2 (2.8%) |
| Exam classroom was comfortable and reasonably free of distraction | 65 (90.3%) | 2 (2.8%) | 5 (6.9%) |
| Prefer to take this type of exam on-line rather than on paper | 50 (69.4%) | 17 (23.6%) | 5 (6.9%) |
| The amount of time I was given to complete this exam was... | Too much time | 12 (16.4%) | Just enough time | 58 (79.4%) | Too little time | 5 (6.8%) |
| Compared to a standard MCQ exam, how would you rate the format of this exam? | Moderately more/much more difficult | 49 (68.1%) | Neutral | 20 (27.8%) | Moderately less/much less difficult | 3 (4.2%) |
| There were questions for which I was not able to find the answer I was looking for | Yes | 43 (59.7%) | No |
evidence for the KFE and modify items that did not perform well. This work is in progress. In addition, periodic, updating, standard-setting and comparison among other testing modalities will be critical to its future success. In our pilot study, students reported that they prefer exams that allow them to go back to previously answered questions and change their answers. This preference has been reported previously by Fischer et al. (2005). However, preventing backward navigation to completed questions was an important advantage in developing items that assess clinical decision-making as a case unfolded.

Some students reported having difficulty finding their preferred answer in the list of options. One potential solution to this problem is to have students enter their preferred option as a free text. However, this requires manual scoring of the examination, the cost of which could be prohibitive. The short-menu of options were peer-reviewed among the members of the KFDG, but it is possible that additional outside peer review would result in a more expansive number of options.

We found that administration of a high-stakes examination via the internet required significant resources, including maintenance of a server and software, managing complicated security challenges to maintain the integrity of the exam and access to a testing room with computers. However, the web-based format allowed us to administer the test to multiple sites easily and automate the marking of the exam.

Although students lacked familiarity with the KF format, the use of KFEs in medical education may increase, given their test characteristics and acceptance as an assessment of clinical decision-making (Hrynczak et al. 2014). Therefore, US students may gain more exposure to the format over time. Future research to understand the optimal way to introduce the KFE format to learners is important.

Conclusion
Developing an electronic KFE based on a national curriculum was a challenging and time-intensive process. Although US student participants were not familiar with the KF format, they overwhelmingly agreed the KFE problems resembled clinical practice. Students also reported the exam more accurately assessed the ability to make clinical decisions compared to a standard multiple choice exam.

Notes on contributors
KIRK A. BRONANDER, MD, FACP, is Clerkship Director and Associate Professor of Medicine at the University of Nevada School of Medicine. Dr. Bronander is an internal medicine physician and hospitalist. His academic interest is in using various simulation modalities to teach students and residents.

VALERIE J. LANG, MD, is Clerkship Director and Associate Professor in the Hospital Medicine Division at the University of Rochester School of Medicine and Dentistry. Her academic interests center on the teaching and assessment of clinical reasoning and high value care decision-making.

L. JAMES NIXON, MD, MHPE, is Vice Chair for Education and Professor in the Department of Medicine at the University of Minnesota Medical School. He also directs the Essentials of Clinical Medicine Course and is interested in teaching and assessing clinical reasoning.

HEATHER E. HARRELL, MD, FACP, is Clerkship Director and Associate Professor of Medicine at the University of Florida College of Medicine. Dr. Harrell is an internal medicine physician. Her academic interest is in using portfolios to promote reflection and self-directed learning in medical students.

REGINA KOVACH, MD, is the Director of the Year Three Curriculum and Chair of the Student Progress Committee at Southern Illinois University School of Medicine. She is interested in the assessment of clinical reasoning and professionalism, and the educational applications of technology.

SUSAN HINGLE, MD, FACP, is Interim Department Chair and Clerkship Director at the Southern Illinois University School of Medicine. Dr. Hingle is a Professor of Medicine and internal medicine specialist. Her academic interests include how to teach and assess interpersonal and communications skills, systems-based practice and practice-based learning.

NORMAN BERMAN, MD, is Professor of Pediatrics at Geisel School of Medicine at Dartmouth and Section Chief of pediatric cardiology at Children’s Hospital at Dartmouth. He is Executive Medical Director at MedU and works actively developing and researching virtual patients.

Acknowledgments
The authors would like to acknowledge the contribution of Dr. Georges Bordage who was an invaluable resource and consultant for the key features development group (KFDG); Amy Holthauser, MD, for her leadership of the KFDG; and Debra Stottlemeyer, MD, and Janet Jokela, MD, who were also members of the KFDG.

Declaration of interest: Kirk A. Bronander, MD, FACP, is an associate editor for SIMPLE virtual patient cases; Valerie J. Lang, MD, is an editor-in-chief of SIMPLE virtual patient cases; L. James Nixon, MD, MHPE, is an editor-in-chief of SIMPLE virtual patient cases; Heather E. Harrell, MD, FACP – no declarations of interest; Regina Kovach, MD, is an associate editor for SIMPLE virtual patient cases; Susan Hingle, MD, FACP – no declarations of interest; and Norman Berman, MD, is executive medical director at MedU, publisher of SIMPLE virtual cases.

References
Eva KW, Wood TJ, Reddie J, Touchie C, Bordage G. 2010. How clinical features are presented matters to weaker diagnosticians. Med Educ 44:775–789.

Farmer EA, Page G. 2005. A practical guide to assessing clinical decision-making skills using the key features approach. Med Educ 39:1188–1194.

Fischer MR, Kopp V, Holzer M, Ruderich F, Jungier J. 2005. A modified electronic key features examination for undergraduate medical students: Validation threats and opportunities. Med Teach 27(5):450–455.

Hatala R, Norman GR. 2002. Adapting a key features examination for a clinical clerkship. Med Educ 36:160–165.

Hrynczak P, Takahashi SG, Nayer M. 2014. Key-feature questions for assessment of clinical reasoning: A literature review. Med Educ 48:870–885.

Kelly WF, Papp KK, Torre D, Hemmer PA. 2012. How and why internal medicine clerkship directors use locally developed, faculty-written examinations: Results of a national survey. Acad Med 87:924–930.

Norman G, Bordage G, Page G, Keane D. 2006. How specific is case specificity? Med Educ 40:618–623.

Page G, Bordage G, Allen T. 1995. Developing key-feature problems and examinations to assess clinical decision-making skills. Acad Med 70(3):194–201.
Supplementary material available online.