



Progress Testing of Basic Science Application During a Clinical
Competency-Based Curriculum Pilot

R. Brent Stansfield, PhD., Larry D. Gruppen, Chris Ricketts, Ph.D.,
Rajesh Mangrulkar, MD.

The 14th Ottawa Conference, Miami, FL
May 17th, 2010

Abstract

Background: In a clinical competency-based curriculum, assessment of students' basic science knowledge is important. During a feasibility pilot of such a curriculum, six students acquired basic science knowledge about three clinical domains (renal failure, trauma, hyperglycemia) from clinical encounters with real patients as well as traditional learning resources.

Summary of Work: Multiple choice items used for progress tests at Peninsula Medical School were assembled into three 50 item tests, administered online at 2-week intervals. Items used clinical scenarios to test knowledge in the 3 domains plus a control domain (infectious disease). Complete item and answers were provided after each test as formative feedback.

Summary of Results: Tests were difficult (mean topic scores from 32% to 80%). Repeated items were easier (80%) than novel items (58%). Scores on items that specifically addressed learning objectives improved, though not statistically significantly.

Conclusions: Brief progress tests are feasible assessments of student learning, though the reliability is a concern in this pilot.

Take home messages: Short progress tests may not be the most sensitive way to evaluate basic science knowledge in a clinical competence curriculum. Longer or more focused tests given at longer intervals may be more sensitive to knowledge application abilities acquired during training.

This presentation describes a curricular innovation program (“ENCORE”) which was piloted at the University of Michigan in the summer of 2009. Dr. Mangrulkar was the project director, Drs. Stansfield and Gruppen provided assessment and evaluation guidance, and Dr. Ricketts of Peninsula Medical College was a key consultant and provided most of the multiple choice items used in the progress test.

The work presented here is part of a collaboration of many people who spent many hours, weeks, years preparing, documenting, and implementing a rather difficult and resource-intensive curriculum. These are just a few of their names.

This talk and the images used are released under a Creative Commons 2.0 license (attribution and share-alike). A pdf of this talk, slides, and photographer credits is available for download on the web.

Contributors

- Rajesh Mangrulkar
- Larry Gruppen
- Chris Ricketts
- Ken Pituch
- Paul Gauger
- Marilyn Roubidoux
- Joseph Fantone
- Tamara Stein
- David Stern



<http://deepblue.lib.umich.edu/>

We ran a pilot of a highly student-centered experimental curriculum with many clinically-based assessments. We tried adapting the progress test to be 1) brief, 2) taken at short intervals, and 3) mostly formative.

Progress Testing for the ENCORE Pilot

Contents

1	ENCORE: A student-centered, experimental curriculum	2
2	A Progress Test made Fast, Frequent, Formative	3
3	Results	6
3.1	Students learn from feedback	6
3.2	Topic differences remained consistent	7
3.3	Students were frustrated	8
4	Conclusions	8

1 ENCORE: A student-centered, experimental curriculum

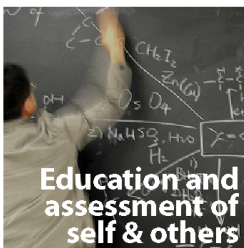
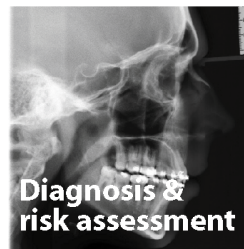
ENCORE: A Clinical competency-based curriculum



ENCORE was conceived as a curriculum in which students had control over their learning and mastery of the practice of medicine. The vision was to define clear learning outcomes and objectives, to give students access to and guidance with a variety of learning resources, and to allow them to be formatively assessed at their discretion—all in the service of helping them achieve the given objectives.

We started by defining the broad areas of competence on which we would focus this curriculum: The Michigan 9.

Outcomes: The Michigan 9



Basic science knowledge was intentionally not defined as a competency *per se*. We made the philosophical decision early in development that knowledge is a means to an end rather than an end itself, and so should not be walled off from these other competencies.

Learning objectives were constructed for specific clinical problem domains. We developed a full set of objectives for 3 domains for the pilot of the ENCORE program: Hyperglycemia, Renal Failure, and Trauma. A typical learning objective looked like this:

Clinical learning objectives

Hyperglycemia 3.08-1: “Determine if changes need to be made to the anti-diabetic treatment regimens. Propose and justify specific recommendations to your attending. Review and implement them with your patient.”

Renal failure had 33 objectives (21 objectives with 1 to 4 sub-objectives each). Hyperglycemia had 32 (17 objectives with 1 to 3 sub-objectives each). Trauma had 53 (23 objectives with 1 to 7 sub-objectives each). All objectives were primarily clinically or orally demonstrable; verbs like “know” or “understand” were explicitly banished from the objective-writers’ vocabularies, replaceable with physical action verbs like “demonstrate” or “identify.”

2 A Progress Test made Fast, Frequent, Formative

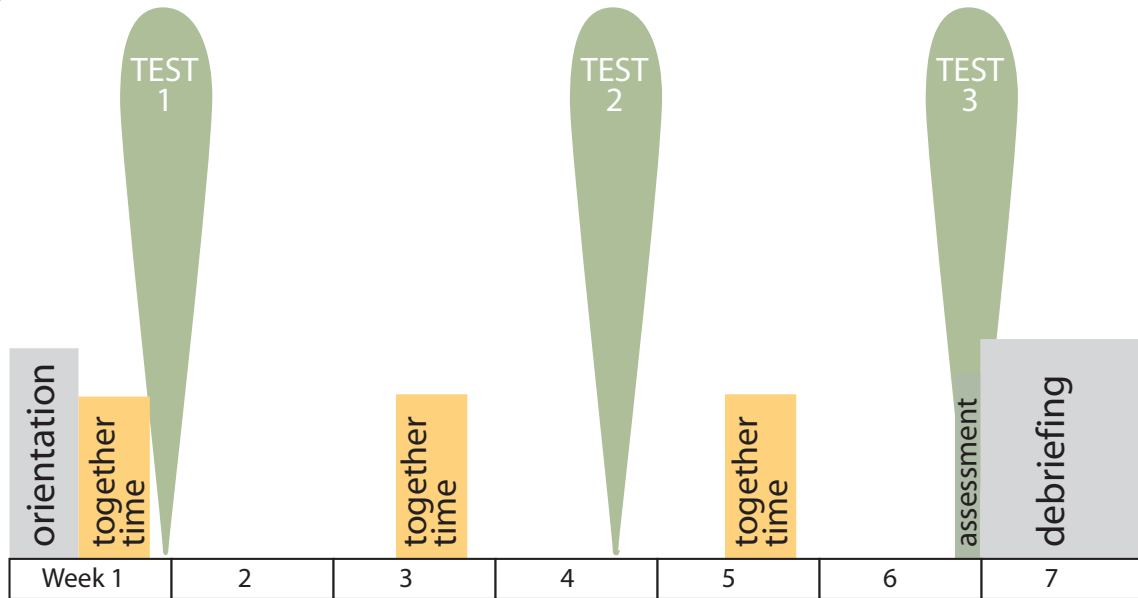
We wanted to co-opt the progress test from its summative role to a formative tool. We wanted it to act as a gauge which students could check to chart their performance on the knowledge aspects of clinical performance. We envisioned a process that was fast (lasting an hour or so), frequent (such that students could test themselves often), and formative (with a lot of explicit feedback).

Fast, Frequent, and Formative



When planning the pilot of the ENCORE program in 2009, we felt it necessary to assign testing times. Ideally, students would volunteer to take the progress tests when they wanted to know how their knowledge application skills were developing. But assigned test times allowed us to gather controlled pre-, mid- and post-pilot scores and to make sure all students took the same test at the same time.

Progress Tests



Three tests: *pre*, *mid*, and *post*

Students were given items describing clinical cases and asking for diagnosis or treatment decisions; a small number of items asked scenario-free questions about the meaning of certain tests or prevalence of certain procedures. A few items were Americanized (word-choice, place names changed from places in England to places in Michigan). Items were selected somewhat randomly, avoiding duplication of content within each test, but not directly tied to specific learning objectives, though all items were arguably relevant at least one objective.

There were several reasons to not tie items to objectives directly: first, is the large number of objectives (118 total). Second, the types of questions that make good multiple choice items are much more specific than a learning objective (a whole test could be constructed from just a handful of objectives). These two together mean that any test must be a small sample of the knowledge domains in question. Third, we expected non-objective-based learning to happen; collateral benefit. Fourth, the items donated by Peninsula saved our team an enormous amount of time and work—writing new items and validating them with students for difficulty level—for which we sacrificed control over item content.

Fifth and perhaps most importantly, the test was to be a formative exercise: items targeting knowledge not central to the learning objectives were at once probing the students' familiarity with the domain in general and familiarizing students with other aspects of the clinical domain.

Each of the three tests consisted of 50 items: 14 per clinical domain and 8 from a clinical domain not included in the curriculum ("infectious disease"). This was included as a control topic for experimental comparison. The second test contained items repeated from the first (in the diagram, each column represents a test and bar represents an item with repeated items connected between tests by thin lines). The trauma curriculum team deemed many of the donated trauma items irrelevant and so new ones had to be written; an unexpected personal tragedy from the lead item writer delayed this effort so the second test contained mostly repeated trauma items.

Progress Tests

Renal failure



Hyperglycemia



Trauma



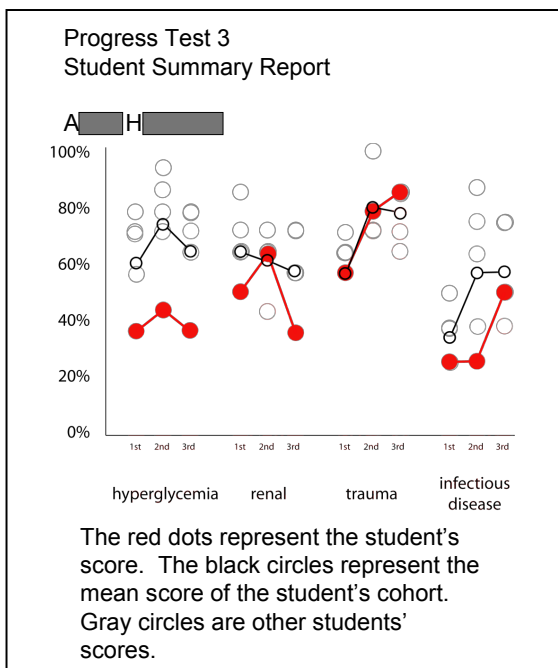
Infectious disease



Repeating items to probe learning from feedback

Tests were given online and students could take the test from any computer. For each test, students had a two-day window in which to log in, from which time they had 90 minutes to complete the test. Reference material was permitted though the time limit likely prohibited any such “just-in-time” research. For two days after the test window had closed, students had complete access to all items, their given answer, the correct answer, and a justification for the correct answer. Students were on their honor not to copy or distribute the items. Summary statistics were given to each student after each test showing his or her mean performance relative to the other students in the pilot:

Feedback

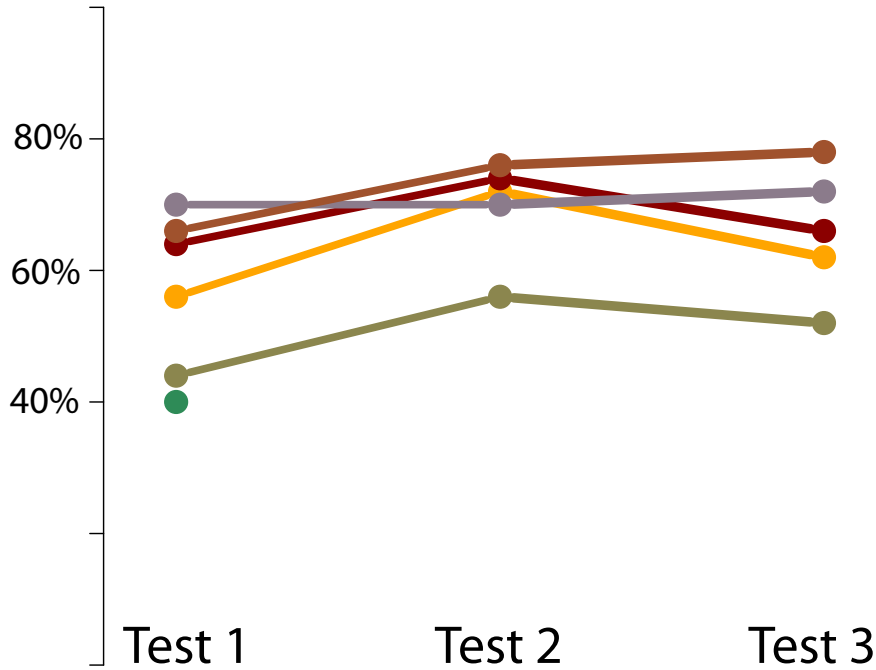


3 Results

3.1 Students learn from feedback

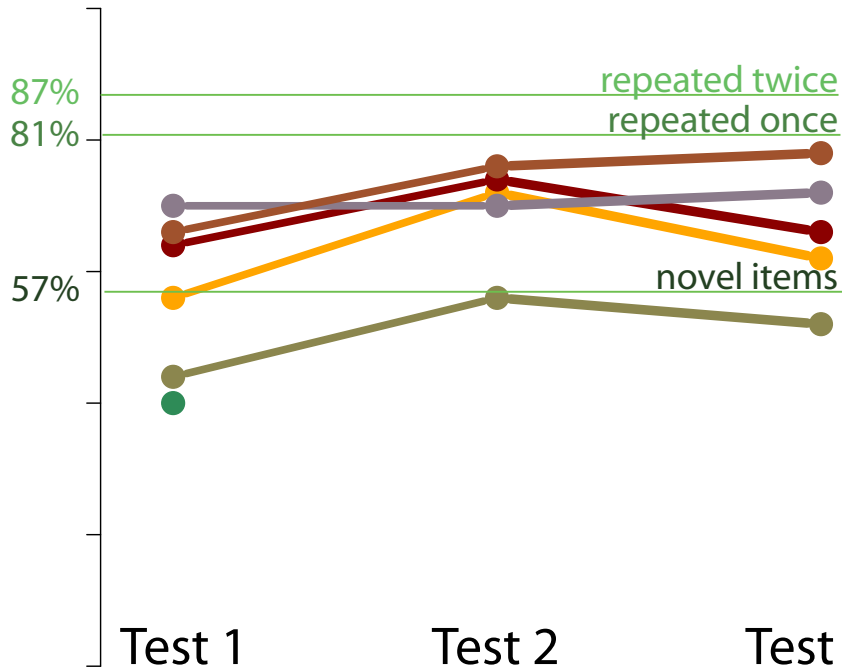
Student performance improved somewhat from test to test, with the largest improvement occurring between the 1st and 2nd test:

Students' scores improved slightly



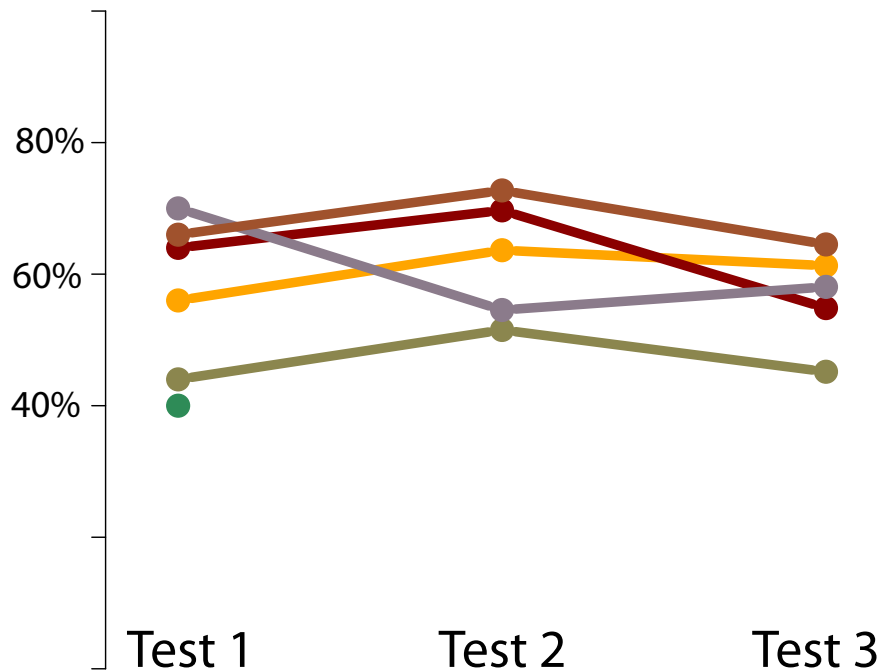
However, this improvement was driven almost entirely by the tendency of students to get repeated items correct; repeated items were answered correctly 85% of the time—much higher than the average scores.

Students' scores improved slightly



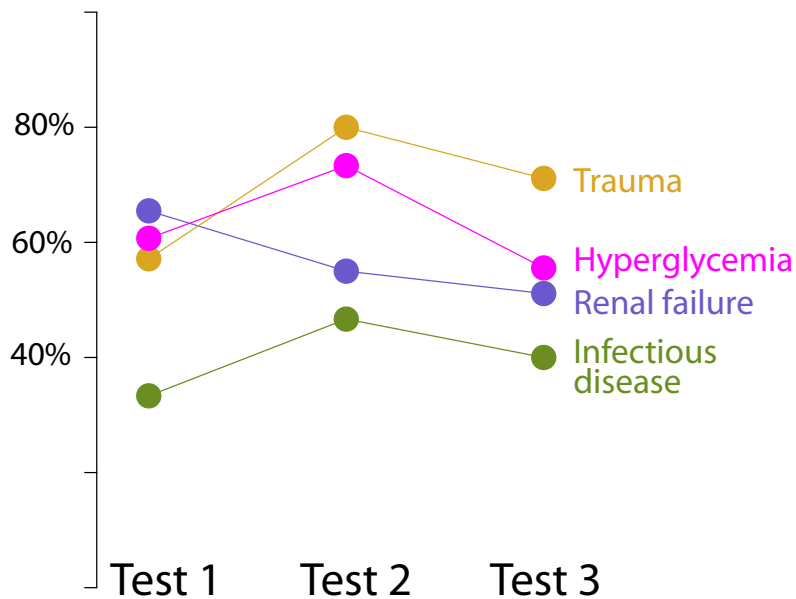
This indicates that students were remembering item content from the transparent feedback. This is unsurprising, of course, and renders the test a poor summative tool. Scores using only non-repeated items showed no student improvement:

Removing repeated items: no score improvement



3.2 Topic differences remained consistent

Target topic performance consistently above control



Means of student topic subscores; novel items only

3.3 Students were frustrated

Feedback

“At first I was counting on the progress test to tell me what to learn. After the second progress test I realized it didn’t help me at all.”

Feedback

“Sometimes I felt like the questions did not apply to what I was learning. They were more detailed than the learning objectives.”

4 Conclusions

Conclusions

- Fast, Frequent, and Formative is Feasible
- More harmony between learning objectives and items
- Focused tests (maybe 50 items per topic)
- Best frequency? Probably > 2 weeks

Thank You

Photo credits: All photos found on www.flickr.com released under a Creative Commons license 2.0.



- by ernstl
<http://www.flickr.com/photos/ernstl/sets/1142809/>



- by cesarastudillo
<http://www.flickr.com/photos/cesarastudillo/with/3959444257/>



- by Roomic Cube
<http://www.flickr.com/photos/roomiccube/3441640970/>



- by Lee Nachtigal



- by Paul Watson



- by Heliøs
<http://www.flickr.com/photos/helios89/1373343153/>



- by Colon Rose
<http://www.flickr.com/photos/73416633@N00/286417542/>



- by alvi2047
<http://www.flickr.com/photos/alvi2047/3483848619/>



- by sparktography
<http://www.flickr.com/photos/sparktography/389824828/>



- by Leo Reynolds
<http://www.flickr.com/photos/lwr/65212080/>