DR SALLY SANTEN (Orcid ID : 0000-0002-8327-8002)

Article type : Clinical Teacher's Toolbox

Sally A Santen, MD, PhD, Leif Myklebust, MS, Mark Grichanik, PhD, Johmarx Patton, MD, MHI, Clare Cabrera, MS, Nikki L Bibler Zaidi, PhD,

Sally A Santen: Virginia Commonwealth University - School of Medicine, Department of Emergency Medicine
Richmond Virginia 23284-2512
United States
Sally.santen@vcuhealth.org

Leif Myklebust: University of Michigan Medical School

Mark Grichanik: Rush Medical College of Rush University

Johmarx Patton: University of Michigan Medical School

Clare Cabrera: University of Michigan Medical Center

Nikki L Bibler Zaidi: University of Michigan Medical School

Creating a learner performance dashboard for programmatic assessment

Sally A Santen, MD, PhD, Leif Myklebust, MS, Mark Grichanik, PhD, Johmarx Patton, MD, MHI, Clare Cabrera, MS, Nikki L Bibler Zaidi, PhD,

Editor's note

Programmatic assessment is a process that is gaining traction in medical and health professional education. It is an approach that aligns with competency-based education. Routine and longitudinal data are collected about learners' competence and their progress is regularly reviewed and analysed. Typically a variety of assessment methods are used with the overall aim of allowing both learners and educators to gain an understanding of the extent of learning and to maximise feedback for optimal educational impact (assessment for learning). Eventually the process may be used for high stakes high-stakes decisions at the end of a programme or phase of study (assessment of learning). In this Toolbox the authors describe how they have developed a method of visually representing medical student attainment during programmatic assessment on a dashboard. This development required input from educators, faculty and information technology (IT) staff. The graphs obtained help students and staff to see how a student is performing over time. The authors share feedback on the approach from learners and faculty, and provide recommendations for readers who may wish to develop something similar.

Medical education is moving towards competency-based medical education (CBME), which must be supported by frequent assessment. The foremost challenge of implementing CBME is developing rigorous assessments with validity evidence that demonstrate students' progress towards competency.[1,2]

Programmatic Assessment in Competency-Based Medical Education

In CBME, assessment data can serve multiple purposes, including the assessment *of* learning (summative competency or grading judgements) and the assessment *for* learning (formative feedback for students).[3] These involve the intentional utilization of varied assessment methods

resulting in multiple assessments, purposefully chosen for their alignment with institutional competencies.[3] Together, these assessment data provide a clearer picture of competent performance.

As programmatic assessment incorporates multiple data points over time, aggregating and displaying these data to provide clear communication for the purpose of feedback on learning and inform decision-making becomes critical. The purpose of this teaching toolbox is to assist medical and other health professional educators in designing data visualization for reporting assessment data within a CBME program. The lessons learned along the journey outlined here will enable educators and administrators to plan, design, and implement a data visualization tool —such as a dashboard— to abstract, summarize, or categorize assessment data.[4]

CBME assessment can create tension among stakeholders, especially students, who hold traditional expectations of norm-referenced grades.[1,2] For example, in the traditional pass/ fail grading system of the pre-clerkship phase, students' performance is generally compared to their peers and cut points for passing are commonly determined by class means and standard deviations —this system does not represent CBME assessment. Similarly, there is dissonance between grading and CBME assessment in the clerkships [2] Therefore, it is important to recognize this tension and affirm the purpose of the assessment framework.

Assessment Reporting

In programmatic assessment, CBME assessment data must be presented clearly and succinctly to provide meaning for students, faculty, and those responsible for making judgements about competency and progression.[4-6] One challenge involves communicating performance metrics from both summative and formative assessments. Given the importance of displaying and reporting assessment data to stakeholders in a meaningful way, medical schools are partnering with experts in technology and learning analytics to design methods for displaying students' assessments.[7] Data visualization involves the creation and study of data that has been abstracted, summarized, or categorized in some schematic form.[7,8] It combines data from various assessments to organize and visualize data for interpretation and is

employed to effectively provide feedback to students and assist faculty in making competency-based decisions.[4,6] The data visualization can provide efficient feedback, yet Hauer and colleagues found that students' use of this information to guide learning is variable—some students engaged in feedback to improve performance, others did not.[4,9] While clear communication of data via assessment dashboards is essential to provide feedback to various stakeholders, it should also mirror the CBME assessment schema.

Methods of Assessment Visualization

At University of Michigan Medical School, we were particularly interested in how data visualization could help create meaning for stakeholders. Specifically, students, faculty and administrators were engaged in the design and prototyping. To support CBME assessment, we worked with our internal education software development team to build an assessment dashboard that could display and monitor each student's performance by pulling from existing data files. We used a Java platform. This required significant resources—including four software developers and two business analysts over eight months, as well as biweekly to monthly meetings with a faculty advisory team. A critical success factor in the development of the visualizations was the partnership between faculty and IT staff. When IT staff develop visualizations independent of faculty, the prototypes may not be ideal for the intended purpose. When faculty develop prototypes in isolation, feasibility for IT implementation can be a limiting factor.

Historically, starting with the first-year medical students, we provided a tabled summary of medical knowledge assessment scores. This tabular display included each student's individual quiz and examination scores as well as class performance metrics (i.e. class mean and standard deviation; Table 1) and was the standard method of displaying scores for both learners and faculty.

<Table 1>

Our tabular reporting of scores made it difficult to look at discrete metrics and understand how the student was performing longitudinally. Therefore, we asked our education software development team to create a graphical representation of each student's performance across

the first year of medical school (Figure 1A). This graph provided a clear picture of performance trends for each student, which was much more meaningful than the previous table of scores. In the example (Figure 1A), one can see that the student was below the class mean on the majority of the assessments. The first version of these graphs was norm-referenced, as it was anchored on the class mean and standard deviation.

<Figure 1A>

As we looked at the graph in Figure 1A, we appreciated how it told the story of the student's performance; but we also realized that this graph—centered around the mean of the class—did not reflect our goal towards CBME assessment.[1] In our competency committee, we wanted to shift from norm-referenced to more criterion-based assessment data for decision-making. So, we went back to the software development team and addressed the question—how could we promote criterion and competency-based decisions using data visualization techniques?

Commented [SS1]: Pull out quote

The education software development team created a criterion-based visualization for each student's performance (Figure 1B). The color coding of each band derived its meaning from a stoplight—i.e. red, yellow, and green. The criterion in the pre-clerkship, basic science courses was established with a cut point of 75%. A score below 75% was in the red zone; this signaled that the student failed a course and was not competent in the domain. While 75% was considered passing, it signaled that a student was barely performing at criterion levels. We knew from historical data and analyses that students with 2 year preclinical cumulative scores below 83% were at increased risk of failing the USMLE Step 1 examination, so the band from 75-83% was yellow to indicate this risk and marginal competence. These indicators helped us to identify, communicate, and work with students to improve their performance. Scores ranging 83-91% were coded green, indicating that students were solidly competent. We were concerned that a competency-based system might focus only on lowest performers (the floor); therefore, we also noted high performance, or "excellence" based on students achieving scores above 91%.

We deliberated whether to include any norm-referenced data in our criterion graph.[1] If we believed in a purely competency-based system, it would not matter how the other students

performed and there would be no need to display the mean. However, we ultimately decided to include the class mean on the graph (triangles) because we felt that including norm-referenced data provided meaningful knowledge to the student about performance (Figure 1C). This decision was grounded in historical data; the yellow zone was associated with an increased risk for failing Step 1. Kruger and Dunning have shown that in general,[10,11] people believe they are above average even when their performance is below average. This self-assessment flaw is seen in students as well, so providing class statistics addressed this limitation.

Our data visualization has multiple purposes including providing formative feedback to students, a platform for coaching and counseling students, and informing decision-making. Students review their performance data with coaches or advisors to identify areas of strength and further growth and goal setting.[9,11,12]

Stakeholder Feedback

We found it important to have faculty and assessment expertise in the development process. Our team was comprised of software developers, as well as experts in educational assessment and theory and faculty in core content areas. Together, we explored multiple iterative visualizations. Once the final version was created, it was shared with the educational leadership faculty and the competency committee who appreciated the clarity. In contrast, the students' questioned the decision to include four categories instead of just three (i.e. red, yellow, and green only). Students said they were "trying to figure out how the cutoffs were determined since students will likely want to know" While the marginal (yellow) cut-off score was based in historical data used for advising, it was not predictive of poor performance. We had less evidence to support the "excellent" category cut-off, but we felt it was important to note excellence. In the process of implementation, we recognized the importance of communicating goals to our students.

In our clerkship assessment dashboard, we used a purely competency-based display of data without norm-referenced metrics (Figure 2). As our journey towards CBME assessment progresses, we continue to explore the challenges surrounding standard setting for cut-scores to indicate competency in medical knowledge. Further, we hope to be able to remove all norm-

based references. Yet, this will require a culture shift towards a growth mindset and away from competition and rank-ordering of students.

Recommendations for CBME assessment data visualization

During this journey, the CBME assessment framework helped to guide the development and visualization of student scores and create meaning. Boscardin and colleagues'[7] twelve tips to promote successful development of a learner performance dashboard provides excellent guidance. There were lessons learned during the development of this assessment dashboard (Table 2).

In summary, when designing a reporting data visualization to abstract, summarize or categorize assessment data for a CBME assessment program, it is important to pay attention to alignment with the construct of competency and engage stakeholders in the process.

Glossary

Competency-based medical education (CBME): CBME is an approach to educating trainees in a framework focused on the explicit developmental progression of health care professionals to meet the needs of those they serve with a focus on performance outcomes.

Competency-based medical education assessment (CBME assessment): In CBME a key aspect is the rigorous measurement (assessment) of the performance (competency).

Norm-referenced grades: Graded assessments are based on the mean scores of the class cohort as compared to criterion, where there is a set cut-off.

Programmatic assessment: Programmatic assessment is an assessment program with the intent to optimize learning and decision-making functions through purposefully chosen assessments aligned with the curriculum outcomes.

United States Medical Licensing Examination (USMLE): USMLE is a 3 part examination that measures a student's ability to apply knowledge, concepts, and principles of health and disease that leads to medical licensure.

Validity and validity evidence: The **validity** of a measurement tool is considered to be the degree to which the tool measures what it claims to measure. Validity evidence is the process, data, analysis and theory that supports the interpretation of the measurement tool (such as test scores) for proposed uses.

References:

1. Pereira AG, Woods M, Olson APJ, van den Hoogenhof S, Duffy BL, Englander R. Criterion-Based Assessment in a Norm-Based World: How Can We Move Past Grades? Acad Med. 2018:93:513-515.
2. Hauer Keza, Lucey CR. Core Clerkship Grading: The Illusion of Objectivity. Acad Med. 2018 Aug 14. [Epub ahead of print]
3. Van der Vleuten CP, Schuwirth LW, Driessen EW, et al. A model for programmatic assessment fit for purpose. *Med Teach.* 2012;34:205-14.
4. Monrad SU, Mangrulkar RS, Woolliscroft JO, Daniel MM, Hartley SE, Highet A, Vijayakumar N, Santen SA. Competency committees in undergraduate medical education: Approaching tensions through a polarity framework. *Acad Med* (in press)
5. Friedman KA, Raimo J, Spielmann K, Chaudhry S. Resident dashboards: helping your clinical competency committee visualize trainees' key performance indicators, *Med Ed Online*, 2016;21:1, 29838.
6. Spickard A III, Ahmed T, Lomis K, Johnson K, Miller B. Changing medical school IT to support medical education transformation. *Teach Learn Med*. 2016;28:80–7.
7. Boscardin C, Fergus KB Hellevig B, Hauer KE. Twelve tips to promote successful development of a learner performance dashboard within a medical education program, *Med Teach,* 2018; 40:8, 855-861.
8. Rost, LS. What questions to ask when creating charts: The attempt of a data vis workflow. *Uncharted*. Available here: https://blog.datawrapper.de/better-charts/ accessed November 5, 2017.
9. Hauer KE, Iverson N, Quach A, Yuan P, Kaner S, Boscardin C. Fostering medical students' lifelong learning skills with a dashboard, coaching and learning planning. *Persp Med Ed.* 2018;7 (5);311-7.
10. Kruger J, Dunning D. Unskilled and Unaware of It: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J Pers Soc Psychol*. 1999; 77:1121-34.

11. Sargeant J, Eva KW, Armson H, et al. Features of assessment learners use to make informed self-assessments of clinical performance. Med Educ. 2011;45:636–47.
12. Vance G, Williamson A, Frearson R, O'Connor N, Davison J, Steele C, Burford B. Evaluation of an established learning portfolio. Clin Teach. 2013;10:21-6

Scientific Trunk Course Cumulations

Download CSV

Date	Course	Cumulative Score	Class Mean	Standard Deviation	Grade
08/29/2016	POP500 Genetics Component Cum 2016-2017	76.12	86.91	5.97	S
08/29/2016	POP500 MDM Component Cum 2016-2017	87.05	87.51	8.4	S
08/29/2016	POP500 Pathology Component Cum 2016-2017	76.52	91.38	6.76	S
10/03/2016	CEL500 Course Cum 2016-2017	73.24	90.21	5.41	F
11/07/2016	CAR504 Course Cum 2016-2017	85.24	89.21	5.47	S
11/14/2016	REN506 Course Cum 2016-2017	83.53	87.40	5.04	S
12/19/2016	MUS513 Course Cum 2016-2017	83.21	93.48	4.13	S
01/23/2017	GAS508 Course Cum 2016-2017	83.70	88.59	4.60	S
02/13/2017	END510 Course Cum 2016-2017	83.39	93.39	5.81	S
02/27/2017	IMM501 Course Cum 2016-2017	89.55	86.21	5.27	S
03/27/2017	CNS509 Course Cum 2016-2017	75.33	86.82	5.85	S
05/01/2017	HPT512 Course Cum 2016-2017	81.23	94.88	5.23	S
05/08/2017	INF500 Course Cum 2016-2017	83	85.55	4.87	S
05/29/2017	CLN500/501 Course Cum 2016-2017	92.89	86.81	3.77	S
10/30/2017	PSY614 Course Cumulation 2016-2017	83.22	87.06	4.21	S
11/25/2017	NEU609 Course Cumulation 2016-2017	88.89	86.59	5.35	S
12/18/2017	DER612 Course Cumulation 2016-2017	82.28	89.71	3.29	S
01/15/2017	HEM603 Course Cumulative 2016-2017	94.17	94.43	6.23	S