Above average, below expectations: shortfalls in using class averages to inform the education of medical students

Fernando Villaseñor BSc*

Max Rady College of Medicine, University of Manitoba
727 McDermot Avenue, Winnipeg, R3E 3P5

Abstract

In this paper, I argue that the release of cohort-specific class averages through the Examplify™ scoring reports do not adequately encourage development of the Scholar role within the CanMEDS framework. Specifically, I argue that the release of cohort-specific class averages in their current format encourages students to focus on more individual metrics, thus dissuading students from participating in more collaborative efforts to generate collective improvements in practice.

Keywords: CanMEDS, scholar role, medical school, class average, academic performance

"CanMEDS is, at its heart, an initiative to improve patient care by enhancing physician training." [1]

This paper aims to examine the role that cohort-specific class-average exam marks have in informing medical students of their performance. For this piece, when I refer to class averages, I am referring specifically to the use of cohort-specific class averages that are provided to medical students during their pre-clerkship years in the Max Rady College of Medicine at the University of Manitoba. Exam reports, providing these averages, are generated by the online test-taking software known as Examplify™, created by Examsoft Inc.

As a foundation for my discussion of class-average grades, I will be making specific use of the CanMEDS role of Scholar. A physician's role as a scholar is defined as demonstrating "a lifelong commitment to excellence in practice through continuous learning and by teaching others, evaluating evidence, and contributing to scholarship." [1] Within this role, I wish to place a particular emphasis on competencies 1.2 and 1.3, which are as follows:

> 1.2 Identify opportunities for learning and improvement by regularly reflecting on and assessing their performance using various internal and external data sources
>
> 1.3 Engage in collaborative learning to continuously improve personal practice and contribute to collective improvements in practice [1]

Following every computer examination written by pre-clerkship medical students, students receive a score report from Examplify™. The score report details the student's raw score, the class-average score, and the report lists the breakdown of the course material by unit and learning objectives. The report is further broken down into Session, Unit, and Objective, and each section provides colour-coded qualitative feedback based upon the student's performance relative to the class average. "Doing Well" (green) indicates that the student has scored significantly above class average, "Needs Review" (yellow) suggests that the student has neither scored considerably above or below average, and "Needs Improvement" (red) indicates that the student has scored significantly below the class average.

Even before opening the official exam report on the Examplify™ website, students can view their raw exam score. The student's exam score is presented in green, yellow, or red, immediately informing the student of their exam performance in comparison to the performance of their peers. In addition to the objective feedback students receive regarding personal performance, this colour-coded system provides students with additional information informing them of their relative academic standing in the class. This value-laden feedback can help to guide students' efforts.

While I will argue that the provision of class averages can be beneficial in addressing CanMEDS Scholar Competency 1.2, it may be undesirable in adequately encouraging development in competencies 1.2 and 1.3.

---

*correspondence to: villasef@myumanitoba.ca

How a class average may encourage development in Competency 1.2

Anecdotally, medical students tend to learn new academic strategies during the pre-clerkship years of medical school as they adapt to various new demands, including a busy coursework schedule, a large number of examinations, high expectations for examination scores, and an overarching need to efficiently learn and apply complex information. Before entering medical school, many medical students will have previously received near-top marks in their respective undergraduate course programs. By releasing a cohort-specific class-average grade, medical schools provide their students with valuable information about how their results compare to the results of others in their new peer group. In addition to providing an initial comparison of how a student's efforts compare to those of their cohort, the large number of exams students write during pre-clerkship allows students to compare their performance to their peers over time. By continually evaluating their performance outcomes relative to those of their peers, students can gauge effectiveness of their study strategies, and modify their strategies as indicated.

Degree of difficulty can vary significantly from exam-to-exam. With a goal of optimally representing the course material, exam questions at the Max Rady College of Medicine are created with intent to widely canvas the course objectives. Once created, questions are subjected to a secondary review by an outside assessment team. Despite this standardized process, some questions are still inevitably more difficult than others. Each course is led by a course director who is in charge of organizing the course and ultimately selecting the questions that will appear on examinations. Due to the variability in teaching styles, and what one course director versus another may deem an appropriate exam question, exam difficulty and class performance can vary significantly throughout the year. In personal experience, during my first year of pre-clerkship, I have observed class-average scores range from 71%-84% on modular exams. Providing a class-average grade for each exam allows students to assess their performance relative to that of the class.

A high-achieving class can motivate individual students to study harder or more effectively. If the class-average score on an exam is high, students are aware that their peers have effectively learned the material. Desiring to compare favourably to their peers, students may choose to increase their study efforts, consequently increasing their knowledge base.

Measuring progress within a cohort — such as the Max Rady College of Medicine Class of 2022 —eliminates many confounding variables that may otherwise be present if measures from multiple cohorts were to be combined into a single measure (e.g., combining grades from the classes of 2022, 2021, 2020, 2019 etc.). Confounding variables may present as follows:

1) Different timing of holidays and notable events relative to the class exam schedule (e.g., one year, a class social event may occur before a final exam, whereas the next year the event may take place afterwards).

2) A stressful event, impacting the entire class, may occur in one year and not occur in the next (e.g., injury or death of a student in the class).

3) A modular course changes its course leader from one year to the next, which could impact the presentation of course materials and the types of questions selected to appear on examinations.

Presumably, holding all else equal, events such as these will produce observable differences in performance across cohorts.

How feedback relative to class averages might be undesirable in fostering development of Competency 1.2

Even though the class average is generally used as a barometer for individual performance, the class average is a relative measure. Examplify^{TM} score evaluations can change from one year to the next. Depending on the distribution of the normal curve for a class' exam scores, the exact same raw score on an identical exam can result in different qualitative feedback from year to year. It is theoretically possible that one year a raw score of 80% receives feedback stating the student "needs improvement" in many sections, while the next year the same examinations score is considered "doing well." Even if the questions appearing on the examinations were the same, each student receives feedback that is relative to the performance of the class.[1]

CanMEDS Scholar Competency 1.2 states that students should be engaged in the continuous enhancement of their learning using various internal and external data sources. However, with the current exam software, self-reflection is limited by unstandardized and relative data. If a class average is lower than that of a previous year, then the relative nature of this scoring system could provide lower-achieving students with positive feedback about their performance. Without including the data of the past year's performance, students may lack additional data in reference to which they would be motivated to improve their academic achievement.

How feedback relative to a class average may be undesirable in promoting Competency 1.3

Despite of some of the shortfalls associated with providing student feedback relative to the class average,

---

[1] This outcome depends on the exact parameters which Examplify^{TM} uses to differentiate above-average exam scores from below-average exam scores. This is an exaggerated, and statistically improbable, example.

this feedback can nonetheless be a significant motivator for individual student improvement. Class averages aid students by providing an objective data source (albeit within a relative framework) to continually monitor and identify the effectiveness of their learning strategies. Providing a class average can also be a useful tool to regularly inform students about how their classmates are performing and can encourage students to strive to meet the learning standards set by their peers.

Indeed, class averages can promote significant student development in line with Competency 1.2. However, where Competency 1.2 focuses on individual student development, Competency 1.3 places a focus on student engagement in a collaborative learning environment. For many students, competing against classmates in their undergraduate classes has become *de rigueur* to attain admission into medical school among increasingly competitive pools of applicants. In preclerkship, students are primarily evaluated on their exam performance. The way evaluations are structured, students are incentivized to maximize their own individual efforts. Consequently, little focus is placed on working towards collaborative efforts that can contribute to collective improvements in medical education.

Consider the earlier hypothetical example where the same raw score on an exam could generate two different qualitative exam reports. Assuming that each class wrote the identical exam, differing qualitative feedback could affect an individual's preference for joining one class versus the other. With the option of joining either: the class where receiving 80% on an exam would result in positive feedback, or joining the class where 80% would result in less favourable feedback, I personally suspect that many students would prefer to be placed in the class where they are told they have "done well" by receiving a mark of 80%. Rather than embrace membership in the high-achieving class, I believe some students may be unwilling to sacrifice a relatively high-achieving performance and ranking among their peers that could be achieved for a given level of effort. If this unwillingness were to exist, it would speak to a certain psychological mechanism running counter to the spirit of the CanMEDS Scholarship competencies, and would act in opposition to the goal of having individual learners working to contribute to collective improvements in practice.

Concluding Remarks

Providing a class average to individual students for feedback can have numerous benefits in helping the individual student assess their performance relative to their colleagues. By assessing their Examplify^{TM} exam reports, students can identify areas in which they are underachieving relative to their classmates. Moving forward, students can then take steps to address these shortcomings by reflecting on previous approaches and implementing new methods. However, I believe that by informing students of their qualitative standing relative to the raw class average we may be dissuading collaborative efforts. This qualitative feedback may serve as a mechanism which promotes individual efforts above collaborative efforts. Consequently, I believe that this form of evaluative feedback is not promoting development in Competency 1.3 and is falling short of the current CanMEDS Scholar framework.

References

[1] Jason Frank, Linda Snell, and Jonathan Sherbino. CanMEDS 2015 Physician Competency Framework. Ottawa: Royal College of Physicians and Surgeons of Canada. *CanMEDS 2015 Physician Competency Framework. Ottawa: Royal College of Physicians and Surgeons of Canada*, pages 1–30, 2015.