

# Aggregated student confidence estimates support continuous quality improvements in a competencies-oriented curriculum

Frank Joseph Papa,<sup>1</sup> Jerry H Alexander

**To cite:** Papa FJ, Alexander JH. Aggregated student confidence estimates support continuous quality improvements in a competencies-oriented curriculum. *BMJ Open Quality* 2019;8:e000398. doi:10.1136/bmjopen-2018-000398

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2018-000398>).

Received 4 April 2018

Revised 5 December 2018

Accepted 2 January 2019

## ABSTRACT

**Introduction** Competencies oriented medical curricula are intended to support the development of those specific tasks likely to improve patient care outcomes. In 2005, our institution developed curricular objectives and instructional activities intended to enable our students to competently perform four specific clinical tasks (diagnose, treat, manage and explain phenomena) for each of approximately 100 common and/or important patient presentations (eg, dyspnoea). However, competencies oriented curricula must also develop outcome metrics aligned with their objectives and instructional activities in order to launch a continuous quality improvement (CQI) programme. This investigation describes how a novel course evaluation methodology produced presentation and task-focused outcome metrics sufficient to support CQIs in our competencies-oriented curriculum.

**Methods** Literature suggests that aggregated, group opinions are much more reliable than individual opinions in a variety of settings, including education. In 2010, we launched a course evaluation methodology using aggregated student self-assessments of their confidence in performing the four tasks trained to in each presentation-focused instructional activity. These aggregated estimates were transformed into a variety of graphic and tabular reports which faculty used to identify, and then remediate, those specific instructional activities associated with suboptimal presentation and task-focused confidence metrics.

**Results** With academic year 2010–2011 serving as a baseline and academic year 2015–2016 as an endpoint, analysis of variance revealed a sustained and statistically significant gain in student confidence across this 6-year study period ( $p < 0.001$ ).

**Discussion** This investigation demonstrated that aggregated, presentation and task-specific confidence estimates enabled faculty to pursue and attain CQIs in a competencies-oriented curriculum. Suggestions for new approaches to confidence-related research are offered.

## INTRODUCTION

Since the early 2000s, medical education has been undergoing a transformational, competencies-oriented curricular reformation.<sup>1–3</sup> One hallmark of such curricula is the creation of objectives and instructional activities designed to develop those specific clinical

capabilities likely to improve patient care outcomes.<sup>4,5</sup> In 2005, curriculum planners at the University of North Texas Health Science Center introduced their vision of a competencies-oriented curriculum in the form of patient presentation-focused instructional ‘modules’ (eg, acute chest pain, dyspnoea, abnormal vaginal bleeding) within year-two system courses.

Each module is designed to support students in competently performing four core, clinically relevant physician tasks, that is, how to diagnose, treat and manage the common and important disease aetiologies causing the patient presentation at hand, and use the biomedical sciences to explain the clinical phenomena associated with each of the diseases addressed in each module. Students attain these four task-specific competencies via an instructional methodology which provides numerous case-based application opportunities and feedback regarding their performance.

The rationale underlying the construction of these modules was derived from decades of learning sciences literature which can largely be summarised via two competencies-oriented, instructional design principles: (1) competence is much more heavily predicated on presentation-focused and task-specific knowledge rather than generalisable problem solving skills<sup>6–8</sup> and (2) the development of competencies is expedited via multiple presentation and task-focused application opportunities and feedback.<sup>9,10</sup> We refer to this instructional methodology as a Presentation-focused, Task-specific, Application-oriented Learning Module (PTALM).

However, the realisation of a fully developed, competencies-oriented curriculum requires more than highly specified performance objectives and instructional activities. It also requires the creation of outcome metrics aligned with the curriculum’s



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

Medical Education, University of North Texas Health Science Center, Fort Worth, Texas, USA

## Correspondence to

Dr Frank Joseph Papa;  
[frank.papa@unthsc.edu](mailto:frank.papa@unthsc.edu)

objectives and instructional activities.<sup>11–13</sup> Such alignment makes it possible to create a robust continuous quality improvement (CQI) programme whereby outcome metrics provide a reliable estimate of the degree to which the curriculum's objectives and instructional activities enable students to competently perform those specific tasks likely to lead to improved patient care outcomes.<sup>14,15</sup>

Our movement towards presentation and task-focused objectives and instructional activities caused us to realise the limitations associated with our previous curriculum's outcome metrics. Specifically, our year two course examinations were designed to assess student performance via a single, overall grade for each course. Further, our previously employed course evaluations produced a single estimate of the students' overall satisfaction for each course. Similarly, licensure board scores reflected a student's overall performance on the examination, along with an overall grade for each discipline and system assessed.

Simply put, these broadly defined outcome metrics (course grades, course satisfaction evaluations and licensure scores) could not be meaningfully aligned with our PTALMs' highly granular, presentation and task-specific objectives and associated instructional activities. The inability to align our highly granular objectives and instructional activities with equally granular outcome metrics would significantly impede our ability to launch and sustain a CQI programme.

Rationale underlying the use of highly specified course evaluation-based outcomes metrics: We quickly recognised that significant logistical impediments would need to be overcome before our own faculty or the licensing boards could create examinations containing the number of test items needed to produce objective and reliable assessment metrics reflecting the students' patient presentation-focused and task-specific capabilities. We subsequently concluded that the initial development of our CQI programme would need to be predicated on outcome metrics produced by our course evaluations.

A suggestion was made to consider the use of student self-assessments, reflecting their confidence in performing each task associated with each PTALM, as the foundation of a new, highly granular approach to course evaluations-based outcomes metrics. However, previous research made clear that novice-derived self-assessments of confidence were unreliable at the level of the individual student. Fortunately, a new line of research suggested that the aggregation of individual opinions, such as self-assessments from novices, can produce a reliable evaluation of the issue at hand.<sup>16,17</sup>

Study purpose: In 2010, we launched a course evaluation methodology which used individual student self-assessments of confidence to produce an aggregated estimate of the classes' confidence in performing each task associated with any given PTALM. We called these aggregated presentation-focused, task-specific estimates 'confidence indices' (CI). The purpose of this study was to determine if these aggregated student CI metrics

could support CQIs in our new competencies-oriented curriculum.

Study overview: In the Methods section of this report, we describe the design and implementation of our CI-based course evaluation methodology. In the Results section, we provide examples of how this new metric provided faculty with formative feedback reflecting the students' aggregated confidence in their ability to perform each of the four core physician tasks associated with each PTALM. We also provide evidence demonstrating that, over time, this three step CQI process sustained measurable improvements in CIs associated with our year-two preclinical training programme. In the Discussion section, we elaborate on faculty acceptance of the CI metric and suggest opportunities for further research.

## METHODS

### Context

Our year two medical training program represents a hybrid curriculum consisting of: 1) Systems-based courses with each course augmented with 2) approximately 6–10 PTALMs. At the end of each year-two System course, students are asked to self-assess their level of confidence in performing each of the four core physician tasks (diagnose, treat, manage and explain) associated with each PTALM offered in the course. Students express their confidence using the qualitative scale of Very Confident, Somewhat Confident, Uncertain, Somewhat Not Confident and Not at all Confident. A schematic portrayal of the confidence data collection instrument for our year-two Cardiovascular System course and the six PTALMs offered during that course is provided (figure 1). The same data collection template is used for all year-two System courses.

### Measures

To calculate an individual student's CI for a given task, the Likert scale associated with each response is multiplied by 20. Thus, a 'Very Confident' response of 5 becomes a CI of 100 and a 'Not at all Confident' response of 1 becomes a CI of 20. We then average all students' responses for any given task to create a task-specific CI for each of the

PTALM Name	Physician Task																			
	Diagnose					Treat					Manage					Explain				
	V	S	U	S	N	V	S	U	S	N	V	S	U	S	N	V	S	U	S	N
Acute Chest Pain	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
Heart Failure	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
Heart Murmur	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
Palpitations/Dysrhythmias	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
Shock	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
Syncopal	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O

**Figure 1** Data collection scheme for students' end-of-course self-assessment of their confidence in performing each task (Diagnose, Treat, Manage and Explain) for each PTALM in a given course. This example portrays the scheme used to collect student Confidence Indices for each of the six PTALMs associated with the year-two Cardiovascular System. PTALM, Presentation-focused, Task-specific, Application-oriented Learning Module.

four tasks associated with each PTALM. We posited that this transformation of confidence estimates from a 1 to 5 scale to CI values ranging from 20 to 100 would make it easier for faculty to interpret a scale similar to that used to grade performance on an examination. For example, CIs hovering around a 70 would represent a marginally acceptable score, while CIs below this suggested that the associated instructional activity was suboptimal. Multiple confidence estimates are derived from the students' self-assessments:

1. Task CI values for each of the four tasks associated with each PTALM.
2. A PTALM CI by averaging the four Task CIs associated with each PTALM.
3. A System CI for each year-two System course, by averaging the CIs for all PTALMs comprising that course.
4. An Annual CI representing the average of all System CIs for that year.

### Intervention

These four CI metrics are used to produce a variety of tables and graphs designed to facilitate our three step CQI programme. First, faculty use these tables and graphics to readily identify those Task-specific instructional activities, PTALMs and System courses where instruction was suboptimal. Second, faculty draw on their understanding of our two competencies oriented curricular design principles and associated faculty development initiatives to remediate those instructional activities identified as suboptimal. Third, the CI metrics associated with all Tasks, PTALMs and System courses offered the following year are reviewed to determine the effectiveness of both the previous year's remediation efforts, and whether, year-after-year, their remediation efforts are producing gradually improving, gradually declining or unchanged CI metrics. These tables/graphics are further described in the Results section of this report.

It should be noted that no individual student's CIs are identified or reported to faculty, per the requirements of our IRB approved research protocol. SPSS Statistics, V.23 (IBM, Armonk, New York, USA) was used to produce the statistical analyses reported in the Results section.

### RESULTS

Self-assessments of confidence were received from 1040 students across the 6years of this study. On average, this represents a participation rate of 77% of each class. It should be noted that student participation in course evaluations is expected, but not mandated.

Using the Cardiovascular System course as an example, one of our 2015–2016 reports to faculty includes both Task CI and PTALM CI values (table 1). When distributed, all CI values are colour coded with shades of green or red creating a gradient of colours making it easy to distinguish the relative strengths and weaknesses of each task-focused instructional activity within each presentation-focused PTALM. Note that the six Cardiovascular PTALMs are not listed alphabetically, but rather rank ordered in terms of their respective Presentation CI values (located in the last column of table 1). This reporting format makes it is easy for faculty to determine that two PTALMs (Palpitations/Dysrhythmias and Heart Murmur) received markedly lower CIs (72 and 76, respectively) than the four other PTALMs (Heart Failure, Acute Chest Pain, Shock and Syncope).

Also note that table 1 enables faculty to rapidly review each of the four task-specific CI estimates associated with each PTALM and thereby identify which particular task-focused instructional activities were most in need of remediation. This task-specific level review makes clear that the instructional activities associated with all four tasks trained to in both Palpitations/Dysrhythmias and Heart Murmur received essentially the same low CI estimates. This finding suggested that improvements along the entirety of these two PTALMs needed to be pursued.

Another faculty report presents a longitudinal, summative perspective of the Cardiovascular course over the 6years of this study (table 2). This longitudinal format, with both tabular and graphical elements, makes it easy for faculty to determine which PTALMs improved over time, remained relatively stagnant or declined. The longitudinal perspective represented in this report reveals a consistent upward trend in the CI metrics associated with the PTALMs comprising the Cardiovascular course over the 6years of this study.

**Table 1** Task-specific and presentation-specific confidence indices cardiovascular system PTALMs, 2015-2016

PTALM name	Physician task				PTALM CI
	Diagnose	Explain	Treat	Manage	
Heart Failure	86	86	84	82	84
Acute Chest Pain	86	86	82	81	84
Shock	83	84	82	79	82
Syncope	82	83	79	78	80
Heart Murmur	77	78	75	76	76
Palpitations/Dysrhythmias	73	73	70	70	72

PTALM CI is an average of the CI values of the four physician tasks.

PTALM, Presentation-focused, Task-specific, Application-oriented Learning Module.



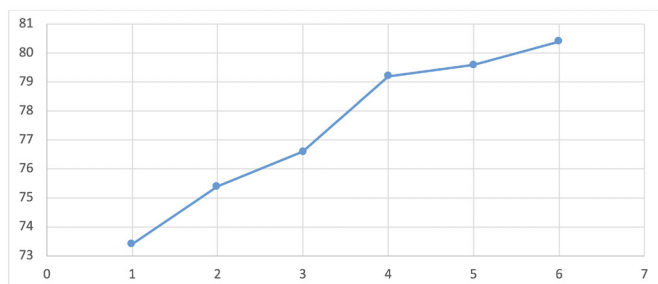
**Table 2** Presentation-specific confidence indices (CI), cardiovascular system PTALMs, annual trend, 2010–2011 through 2015–2016

PTALM name	Year						Annual trend
	1	2	3	4	5	6	
Heart failure	70	72	76	80	78	84	
Acute chest pain	70	73	76	78	75	84	
Shock	49	60	77	79	74	82	
Syncope	56	69	75	75	72	80	
Heart murmur	66	71	66	72	73	76	
Palpitations/dysrhythmias	62	68	66	71	71	72	

PTALM, Presentation-focused, Task-specific, Application-oriented Learning Module.

For example, the CI for the presentation of Shock showed dramatic improvement, rising from an initial CI of 49, in the Uncertain/Somewhat Not Confident range in academic year 2010–2011, to a CI of 82 in the Very Confident/Somewhat Confident range in 2015–2016. However, while our Palpation/Dysrhythmia and Heart Murmur PTALMs demonstrated a 10-point improvement in their CIs over the 6 years of this study, the most current year's CI for each of these two PTALMs has remained markedly lower than the averaged CI score for other Cardiovascular PTALMs. Thus, longitudinal data may be used to design learning sciences principles-based faculty development for those responsible for poorly ranked PTALMs or, where necessary, to support an administrative decision to reassign faculty whose PTALM(s) have shown marginal performance over time. Similar gains in CI values were also found in the other eight Systems courses offered each year, over the 6 years of this investigation, but are not represented in this report.

Gross, year-over-year gains in confidence are summarised by the Annual CI values gathered throughout this 6-year study (figure 2). While a visual inspection of the trend line reveals continuous improvements, an analysis of variance (ANOVA) was conducted to determine



**Figure 2** This figure portrays year-over-year increases in the annual Confidence Indices from 2010–2011 through 2015–2016. These annual Confidence Indices represent the average across all system-level Confidence Indices within each year.

if this trend represented significant improvements. The ANOVA confirms that the improvements were statistically significant across this 6-year period ( $p < 0.000$ ) (table 3).

Scheffé *a posteriori* comparisons revealed that the CI gains from the benchmark year, 2010–2011, to each subsequent year were statistically significant, with  $p < 0.05$ ,  $< 0.001$ ,  $< 0.001$ ,  $< 0.001$ . Further, in comparing each year's Annual CI to the Annual CI of the year immediately following, it was determined that the CI gain from 2010–2011 to 2011–2012 was statistically significant ( $p < 0.05$ ), as was the CI gain from 2012–2013 to 2013–2014 ( $p < 0.001$ ).

It should be noted that consideration of student self-assessed confidence as a predictor of any individual student's performance level was not originally a part of the study plan. However, in response to numerous faculty inquiries, we performed a 1-year analysis of the relationship (correlation) between our students' individually averaged CI for the nine system courses and both their individually average grade for these courses as well as their performance on the United States Medical Licensing Examination (USMLE) 1. We found a positive and significant correlation between our students' CI and their class grade ( $r = 0.300$ ,  $p = 0.000$ ,  $df = 165$ ), and also between their CI and USMLE 1 score ( $r = 0.202$ ,  $p = 0.018$ ,  $df = 165$ ).

## DISCUSSION

### Review of rationale

For decades, the learning sciences have provided ever mounting evidence demonstrating that competence is presentation and task-specific. This evidence is now playing a significant role in the evolution of competencies-oriented curricula as witnessed by a recent AAMC survey revealing that 'curricula organised around clinical presentations will be the next evolutionary step of curriculum renewal for many (American) medical schools'.<sup>18</sup>

We described how this evidence and two learning sciences-derived principles were used to construct our competencies-oriented curriculum; an approach to training predicated on the establishment of highly specified course objectives and instructional activities delivered via an instructional methodology referred to as PTALMs. Faculty and students were generally pleased with our learning sciences-derived, principles-driven curriculum. Following the implementation and stabilisation of our new curriculum, we sought to establish outcome metrics (in the form of student performance assessments) with a level of granularity equal to our PTALMs' objectives and instructional activities. Unfortunately, the logistical overhead associated with the production of multiple, presentation and task-specific test items exceeded our resources.

However, it was much simpler to create a course evaluation methodology that would align with our PTALM's objectives and instructional activities. We posited that if we were to view our students as physicians-in-training, with the collective capacity to reliably estimate how well any given PTALM enabled them to perform its four core

**Table 3** Analysis of variance, annual confidence indices, 2010–2011 through 2015–2016

Source	Sum of squares	df	Mean square	F	Significance
Between groups	5443.96	5	10888.79	53.72	0.000
Within groups	1 857 387.47	9164	202.68		
Total	1 911 831.43	9169			

physician tasks (ie, used aggregated CI metrics), such metrics could serve as a reliable foundation for launching a meaningful CQI programme.

### Interpretation of findings

Our aggregated confidence metrics were used to construct visually oriented (tabular and graphic) portrayals representing the relative rank ordering of the PTALMs comprising each System course within a given year. These visuals enabled faculty to readily identify suboptimally performing PTALMs and any suboptimal task-specific instructional activities within any given PTALM. Once identified, faculty could readily set out to improve these suboptimally performing instructional activities via the use of our learning sciences derived curricular design principles. These visuals were also used to portray how any given System course's PTALMs were progressing over the several years of this study. Throughout this 6-year study, our use of aggregated student confidence estimates enabled us to pursue and attain incremental improvements in targeted instructional activities as witnessed by the results of our ANOVA and posthoc Scheffé analysis.

### Generalisability

The authors readily concede that presentation and task-focused student performance metrics would represent an optimal means of assessing and documenting the development of highly granular clinical capabilities in a competencies-oriented curriculum. However, the logistical overhead associated with the production of multiple, presentation-focused and task-specific test items is substantial. Thus, highly granular, presentation and task-specific course evaluations, such as produced in this investigation, might be among the only pragmatic, near-term means of initiating CQI programmes capable of sustaining continual improvements in the clinical capabilities of students within competencies-oriented curricula.

In regards to the utility of course evaluation-driven CQI programmes, Kogan and Shea have argued that 'The best, yet most underutilized, reason to evaluate courses is to gather feedback for faculty and where necessary provide remediation to improve their teaching'.<sup>19</sup> Evidence of the validity and utility of course evaluations based on aggregated medical student and resident self-assessments has been previously reported by Peterson *et al*.<sup>20 21</sup> However, unlike Peterson's work, our aggregated confidence-based metrics were used by faculty as ongoing, formative feedback intended to improve the curriculum's instructional effectiveness in preparation for the next class of students. Evidence of the utility of aggregated learner feedback, as

reported by Peterson *et al* and as also described herein, would appear to support Kogan and Shea's position. We therefore suggest that most competencies-oriented curricular initiatives could easily implement and subsequently benefit from the development of course evaluation methodologies designed to produce outcome metrics aligned with their curriculum's objectives and instructional activities.

Faculty acceptance and utilisation of the CI metrics introduced in this evaluation methodology proceeded cautiously. Initially, some faculty expressed concern that we might use our CI metric as a replacement for objective assessments of student performance. Accordingly, it was important to continually make clear that the CI metrics represented a curricular CQI initiative and that it was never our intention to use them as a replacement for objective assessments of student performance. In fact, no individual student confidence levels were ever reported. Thus, any potential utility for these CIs was always based on the aggregation of individual student confidence estimates.

### Limitations

The purpose of this investigation was to determine whether a course evaluation methodology predicated on aggregated presentation and task-focused outcome metrics could support CQIs in our competencies-oriented curriculum. However, some faculty were interested in pursuing the factors contributing to the finding identified at the end of our first year of gathering CI estimates, that is, a statistically significant yet low correlation between our students' CI and both their course grades and subsequent USMLE scores. Unfortunately, the scope of this investigation precluded efforts to use confidence estimates as a predictor of future performance.

We recognise that the literature regarding the utility of self-assessments by individuals remains unsettled. For example, Caputo and Dunning found little relationship between an individual learner's confidence and their performance, as did Dunning and Heath and Morgan and Cleave-Hogg.<sup>22–24</sup> However, Stankov *et al* found confidence to be 'by far the best predictor of both concurrent and future performance.'<sup>25</sup> Are insights to be gained from launching new investigations into the use of individual self-assessments of confidence and performance against presentation and task-specific situations?

Future directions for confidence-based research: We speculate that the level of correlation between an individual's confidence and their performance might be

higher under the following circumstances. First, when the number of test items is sufficient to produce a reliable measure of a subject's performance against each specific presentation and/or task of interest. Second, when subjects receive training sufficient to enable them to consciously, insightfully and reliably reflect on and monitor their confidence in performing against presentation and task-specific situations.

Curiously, there is little evidence that health sciences training programmes support the development of those metacognitive and/or executive skills theorised as responsible for self-reflection/self-monitoring. We also suggest that previous investigators did not fully appreciate the presentation and task-specific nature of competence, and thereby did not create the number of presentation and task-specific test items needed to produce a reliable performance metric. Thus, we posit that future efforts to correlate confidence with performance might be higher (and thereby more useful) if both of these concerns were addressed.

One benefit of research demonstrating the existence of higher confidence/performance correlations could be the reliable identification of poorly calibrated (high confidence/low performing) students or residents. Feedback regarding the specific presentations and/or tasks in which learners performed poorly yet were highly confident, could enable them to gain meaningful insights into the origin of their poor performance. Questions such as how many high confidence/low performing individuals exist in a class of students or residents, and how to support learners in gaining insights into the origins of poor performance, remain largely unaddressed. However, improving the learner's capacity to identify and self-calibrate disparities between confidence and performance, could lead to improvements in diagnostic performance; an area of increasing concern to all working to reduce the number of unnecessary deaths due to diagnostic errors.<sup>26 27</sup>

## Summary

Competencies-oriented curricular initiatives require the careful alignment of course objectives and instructional activities with outcomes metrics such as student performance assessments and course evaluations. Once aligned, CQI programmes can be designed to monitor and provide feedback by which continual curricular improvements might be pursued. In this investigation, the authors demonstrated the utility of a novel course evaluation methodology involving aggregated student confidence estimates as a means of implementing a CQI initiative in pursuit of incremental improvements in a presentation-focused, task-specific, application-oriented curriculum.

We encourage medical education researchers to explore the utility of aggregated estimates of confidence in efforts to continually improve their own highly specified, competencies-oriented curricular reforms. Further, institutions using aggregated confidence-based course evaluation metrics would be well positioned to conduct research

exploring the use of individual student confidence estimates as a means of developing training programmes directed at improving their students' evolving metacognitive skills.

**Contributors** FJP was the primary author of the principles underlying the development of the presentation-focused, task-specific competencies-oriented curriculum and the course evaluation metrics described herein. JHA established the basis for increasingly granular course evaluation metrics and their use as formative feedback for launching curricular CQI programming. JHA was also responsible for handling and analysing the data described herein. Both contributed equally to the initial drafting and multiple revisions of the manuscript.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** None declared.

**Patient consent** Not required.

**Ethics approval** Ethical approval has been granted by University of North Texas Health Science Center, Office for the Protection of Human Subjects, protocol number 2012-024, annually updated.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** No additional data are available.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## REFERENCES

- 1 Simpson JG, Furnace J, Crosby J, *et al*. The Scottish doctor--learning outcomes for the medical undergraduate in Scotland: a foundation for competent and reflective practitioners. *Med Teach* 2002;24:136-43.
- 2 Frank JR, Snell LS, Cate OT, *et al*. Competency-based medical education: theory to practice. *Med Teach* 2010;32:638-45.
- 3 Whitehead CR, Austin Z, Hodges BD. Flower power: the armoured expert in the CanMEDS competency framework? *Adv Health Sci Educ Theory Pract* 2011;16:681-94.
- 4 Swing SR. The ACGME outcome project: retrospective and prospective. *Med Teach* 2007;29:648-54.
- 5 Morcke AM, Dornan T, Eika B. Outcome (competency) based education: an exploration of its origins, theoretical basis, and empirical evidence. *Adv Health Sci Educ Theory Pract* 2013;18:851-63.
- 6 Elstein AS, Shulman LS, Sprafka SA. *Medical Problem Solving: An Analysis of Clinical Reasoning*. Cambridge, MA: Harvard University Press, 1978.
- 7 Norman G, Bordage G, Page G, *et al*. How specific is case specificity? *Med Educ* 2006;40:618-23.
- 8 Papa FJ, Harasym PH, Schumacker RE. Evidence of second-order factor structure in a diagnostic problem space: Implications for medical education. *Structural Equation Modeling: A Multidisciplinary Journal* 1997;4:25-36.
- 9 Ericsson KA, Krampe RT, Tesch-Römer C. The role of deliberate practice in the acquisition of expert performance. *Psychol Rev* 1993;100:363-406.
- 10 Ericsson KA. The search for general abilities and basic capacities: Theoretical implications from the modifiability and complexity of mechanisms mediating expert performance. In: Sternberg RJ, Grigorenko EL, eds. *Perspectives on the Psychology of Abilities, Competencies, and Expertise*. New York, NY US: Cambridge University Press, 2006:93-125.
- 11 Martone A, Sireci SG. Evaluating Alignment Between Curriculum, Assessment, and Instruction. *Rev Educ Res* 2009;79:1332-61.
- 12 Cameron T, Krane K, Moineau JLL, *et al*. Mapping medical education curricula. 2012 <https://members.aamc.org/eweb/upload/100Cameron116.pdf> (Last accessed 2017 November 29).
- 13 Barzansky B, Hunt D, Moineau G, *et al*. Continuous quality improvement in an accreditation system for undergraduate medical education: Benefits and challenges. *Medical Education* 2015;37:1032-8.

- 14 Rogers EM. *Diffusion of Innovations*. 4th ed. London (UK): Free Press, 1995.
- 15 O'Neill NA, Palmer A. Importance-performance analysis: a useful tool for directing continuous quality improvement in higher education. *QAE* 2004;12:39–52.
- 16 Surowiecki J. *The wisdom of crowds*. New York (NY): Doubleday, 2004.
- 17 D'Eon M, Sadownik L, Harrison A, et al. Using self-assessments to detect workshop success. *Am J Eval* 2008;29:92–8.
- 18 Mavis B. ange. 2017 <https://www.aamc.org/download/476938/data/ciic04-1jan2017.pdf> (Accessed 28 Nov 2018).
- 19 Kogan JR, Shea JA. Course evaluation in medical education. *Teach Teach Educ* 2007;23:251–64.
- 20 Peterson LN, Eva KW, Rusticus SA, et al. The readiness for clerkship survey: can self-assessment data be used to evaluate program effectiveness? *Acad Med* 2012;87:1355–60.
- 21 Peterson LN, Rusticus SA, Wilson DA, et al. Readiness for residency: A survey to evaluate undergraduate medical education programs. *Acad Med* 2015;90:S36–S42.
- 22 Caputo D, Dunning D. What you don't know: The role played by errors of omission in imperfect self-assessments. *J Exp Soc Psychol* 2005;41:488–505.
- 23 Dunning D, Heath C, Suls JM. Flawed self-assessment: Implications for health, education, and the workplace. *Psychol Sci Public Interest* 2004;5:69–106.
- 24 Morgan PJ, Cleave-Hogg D. Comparison between medical students' experience, confidence and competence. *Med Educ* 2002;36:534–9.
- 25 Stankov L, Lee J, Luo W, et al. Confidence: A better predictor of academic achievement than self-efficacy. *self-concept and anxiety? Learning and Individual Differences* 2012;22:747–58.
- 26 Balogh EP, Miller BT, Ball JR, eds. *Improving diagnosis in health care*. Washington (DC): National Academies Press, 2015.
- 27 Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med* 2008;121:S2–S23.