

## “I’ll Never Play Professional Football” and Other Fallacies of Self-Assessment

KEVIN W. EVA, PhD; GLENN REGEHR, PhD

*It is generally well accepted in health professional education that self-assessment is a key step in the continuing professional development cycle. While there has been increasing discussion in the community pertaining to whether or not professionals can indeed self-assess accurately, much of this discussion has been clouded by the fact that the term self-assessment has been used in an unfortunate and confusing variety of ways. In this article we will draw distinctions between self-assessment (an ability), self-directed assessment seeking and reflection (pedagogical strategies), and self-monitoring (immediate contextually relevant responses to environmental stimuli) in an attempt to clarify the rhetoric pertaining to each activity and provide some guidance regarding the implications that can be drawn from making these distinctions. We will further explore a source of persistence in the community’s efforts to improve self-assessment despite clear findings from a large body of research that we as humans do not (and, in fact, perhaps cannot) self-assess well by describing what we call a “they not we” phenomenon. Finally, we will use this phenomenon and the distinctions previously described to advocate for a variety of research projects aimed at shedding further light on the complicated relationship between self-assessment and other forms of self-regulating professional development activities.*

**Key Words:** *self-assessment, reflection, continuing education, self-regulation, self-monitoring, self-directed assessment seeking*

While the term *self-assessment* has made its way deep into the lexicon of educators in the health professions, the variety of uses to which it has been put can make it a challenging concept to discuss. *Self-assessment* has been used to describe a variety of thought processes,<sup>1</sup> various pedagogical strategies for individual learners<sup>2</sup> and numerous quality assurance endeavors for institutions and professions as a whole.<sup>3</sup> While each of these contexts may, individually, be a justifiable use of the term *self-assessment*, collectively, couching such very different concepts under a single label can be a significant source of confusion and conflict for educators and theoreticians alike. Of course, terminology can provide useful shorthand for discussions among small groups of individuals who are using terms in an explicitly constrained way. It is important, however, to recognize that when such terms are exported to the lar-

ger community, they run the risk of becoming so all-encompassing as to include everything and, therefore, ultimately define nothing.

Lingard and Haber have stated that “the language we use both makes possible and constrains the thoughts we can have.”<sup>4</sup> This is certainly no less true for our community’s use of the term *self-assessment*. The negative impact of the imprecise use of the term on research in the field, for example, is evidenced by the numerous papers and talks that make claims about self-assessment using outcome measures or educational interventions that bear no relation to using oneself to generate a judgment of one’s ability (typically using an externally derived instrument in a self-administered manner).<sup>5</sup> In an effort to address this imprecision in the literature, the current article was written to clarify some apparent confusions and ambiguities about the concept of self-assessment as it pertains to individuals’ efforts at self-regulation. Its purposes are threefold. First, we will offer some distinctions in terminology that may provide the opportunity for greater precision when considering the concept of self-assessment and its role in professional training and practice. Others may offer different opinions regarding how particular aspects of self-assessment should be described, but at a minimum we hope this section of the article will provide the basis and motivation for these discussions to occur. Second, we will identify and address one potential source of the persistence of our community’s efforts to assess and improve self-assessment, what we will refer to as the “they not we” problem. Finally, we will conclude by

---

*Dr. Eva:* Associate Professor and Associate Chair, Department of Clinical Epidemiology and Biostatistics, Program for Educational Research and Development, McMaster University, Hamilton, Ontario, Canada; *Dr. Regehr:* Richard and Elizabeth Currie Chair in Health Professions Education Research, and Professor, Department of Surgery, University of Toronto, Toronto, Canada, Associate Director, The Wilson Centre, Toronto, Canada.

Correspondence: Kevin Eva, PhD, MDCL 3522, McMaster University, Hamilton, ON, L8N 3Z5, Canada; e-mail: evakw@mcmaster.ca.

© 2008 The Alliance for Continuing Medical Education, the Society for Academic Continuing Medical Education, and the Council on CME, Association for Hospital Medical Education. • Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/chp.150

refining the call for different forms of research that we initiated in an earlier article.<sup>1</sup>

## Distinguishing Self-Assessment From Related Concepts

### *Self-Assessment Versus Self-Directed Assessment Seeking*

It seems generally well accepted in the health professions that self-assessment is a key step in the continuing professional development cycle. That is, the archetype of the self-regulating professional is seen as one who regularly self-identifies areas of professional weakness for the purposes of guiding continuing education activities that will overcome these gaps in practice.<sup>6</sup> In this construction, self-assessment is often (implicitly or otherwise) conceptualized as a personal, unguided reflection on performance for the purposes of generating an individually derived summary of one's own level of knowledge, skill, and understanding in a particular area. For example, this conceptualization would appear to be the only reasonable basis for studies that fit into what Colliver has described as the "guess your grade" model of self-assessment research,<sup>7</sup> the results of which form the core foundation for the recurring conclusion that self-assessment is generally poor.<sup>8</sup>

This "unguided, internally generated" construction of self-assessment stands in stark contrast to the model put forward by Boud, who argued that "the phrase self-assessment should not imply an isolated or individualistic activity; it should commonly involve peers, teachers, and other sources of information."<sup>9</sup> The conceptualization of self-assessment as enunciated in Boud's description would appear to involve a process by which one takes personal responsibility for looking outward, explicitly seeking feedback and information from external sources, then using these externally generated sources of assessment data to direct performance improvements. In this construction, self-assessment is more of a pedagogical strategy than an ability to judge for oneself; it is a habit that one needs to acquire and enact rather than an ability that one needs to master. As a result, we would like to suggest that Boud's construction be labeled not as *self-assessment* ability, but rather as a process of *self-directed assessment seeking*. While the evidence pertaining to the accuracy of self-assessment as an ability is robust and clear—we do not do it well—there appears to be little research that directly tests whether or not the habit of self-directed assessment seeking can be taught in a manner that leads the learner to apply the habit cross-contextually, or whether intentionally engaging in this sort of activity is pedagogically advantageous. It is interesting to note that the phrase *self-assessment* originally made its way into the medical education lexicon by virtue of papers that were promoting self-directed assessments such as self-administered multiple-choice question exams;<sup>10</sup> it is only in recent years that the concept of self-assessment seems to have become more closely associated with the introspec-

tive reflective processes that we typically think of today when striving to help students become good self-assessors for the purpose of promoting appropriate lifelong learning activities. Nonetheless, given that the transition has occurred, we will use the term *self-assessment* to reference the new implicit definition and offer a more specific descriptive term (*self-directed assessment seeking*) to reference the pedagogical activity of looking outward for formative and summative assessments of one's current level of performance.

### *Self-Assessment Versus Reflection*

The apparent mutation of the term *self-assessment* as used in medical education raises another important distinction: that of self-assessment versus reflection. We, along with many others, have argued (and continue to believe) that the evidence reveals not merely that humans are poor at producing self-generated summative assessments of their own performance or ability, but that humans are actually predisposed to being poor at this form of self-assessment. There are cognitive reasons (eg, information neglect and memory biases),<sup>11</sup> sociobiological reasons (it being adaptive to maintain an optimistic outlook),<sup>12</sup> and social reasons (eg, not always receiving adequate feedback from peers and supervisors)<sup>13</sup> that render the task of generating "accurate" summative assessments of one's own level of performance or ability particularly challenging. The conclusion that humans do not self-assess well, however, should in no way imply that reflection on performance is a useless activity. Before elaborating on this statement it is necessary to consider a definition of *reflection* as it, too, is a term that has been rather loosely defined in the literature. Recent reviews have reinforced our perception that the term *reflection* is intended to indicate a conscious and deliberate reinvestment of mental energy aimed at exploring and elaborating one's *understanding* of the problem one has faced (or is facing) rather than aimed simply at trying to solve it.<sup>14</sup> Thus, reflection involves activities such as trying to understand *why* a patient's health state deteriorated in the way it did or *why* a social interaction went particularly well. Exploring these sorts of "why" questions may very well prove to be an effective pedagogical strategy that can lead to better understanding of both the world and the adequacy of one's own personal constructions of it. Certainly the expertise literature would suggest that the tendency to reinvest mental energy in this way is a defining determinant of who achieves true expert status in any given field and who evolves into an "experienced non-expert."<sup>15</sup> Again, however, promoting those sorts of reflective behaviors—aimed at understanding the world better—should not be considered the same as promoting self-assessment as a mechanism for judging personal competence. Asking "why" in an effective manner does not require insight into one's own level of knowledge or abilities, because the more useful answers are likely to be

derived through exploration of other sources of information (eg, first principles or the experiences of others).

### *Self-Assessment Versus Self-Monitoring*

A third important distinction that should be recognized is that of self-assessment as an overall personal judgment of one's ability versus awareness, in the moment, of whether or not the current situation is going well, a process better described as "self-monitoring." In previous papers, we have described the latter activity as "slowing down when one should" and "knowing when to look it up" to point out that in many instances people's self-limiting behaviors are probably driven more by the specific circumstances they find themselves in than by an overall judgment of ability.<sup>16,17</sup> To reiterate an example we have used in the past, we suspect most people are prompted to open a dictionary as a result of encountering a word for which they are uncertain of the meaning rather than out of a broader assessment that their vocabulary could be improved. One may recognize weaknesses in performance, in the moment, but still not have those observations impact upon one's broader self-concept. It is easy (and typically automatic) to find reasons to discount negative performances by saying things like "If only [insert favourite excuse here] hadn't occurred."<sup>18</sup> The end result is that a series of moments in which we are aware of the effectiveness of our immediate performance does not guarantee that those moments will be aggregated to generate an accurate self-assessment overall. It is arguable that the health professional community should predominantly be concerned with identifying contextual factors that influence self-monitoring behaviors in the moment of action rather than worrying about the accuracy of generic and broader self-assessments of ability. A physician's ability to recognize when a particular situation is beyond his boundaries of competence is likely to be a greater determinant of patient safety than whether or not that physician is able to determine which broad-based continuing education activities would most effectively fill his gaps in knowledge/skill.

### *Summary*

In this section, we have contrasted self-assessment (an ability) with self-directed assessment seeking (a pedagogical strategy) as potential mechanisms for determining one's areas of strength and weakness, we have contrasted self-assessment (an ability) with reflection (a pedagogical strategy) as potential mechanisms for improving one's understanding of the world, and we have contrasted self-assessment (an ability) with self-monitoring (an immediate contextually relevant response to environmental stimuli) as potential mechanisms for determining the need to recruit additional resources to facilitate performance in particular situations. These distinctions among self-assessment, self-directed assessment seeking, reflection, and self-monitoring represent only a few of what we suspect are many blendings of concepts through the

ubiquitous and broad use of the term *self-assessment* and its operational instantiations in the medical education literature. In each case described, we would argue that it is the alternate activity, not "self-assessment," that is the critical operative process driving safe and effective professional practice and development. So where does the persistent faith in the ability to self-determine one's personal strengths and weaknesses arise? In the next section we focus on this question and try to understand why the faith in this form of self-assessment persists in light of empirical evidence to the contrary.

### **The Illusion of Personal Accuracy in Self-Assessment**

The basic human need to rely on external assessments when trying to determine one's own ability level seems to strike most people as counterintuitive. The most common response to the findings that self-assessment is poor appears to be bewilderment at how "they" can be so bad, with a concomitant belief that if "we" can just get "them" to self-assess as well as we do, then everything will be okay. This reaction is sufficiently pervasive that, ironically, the majority of people think they are above average in self-assessment ability.<sup>19</sup> In the preceding section we enumerated and described several reasons for this illusion to exist, and an elaboration of some of these ideas is provided by Dunning, Heath, and Suls,<sup>11</sup> but a brief example might further illuminate some fundamental errors in reasoning that perpetuate the myth that "we" are good self-assessors. At a recent reception following a talk on self-assessment, a colleague suggested that his self-assessment was fine—after all, he knew he was never going to be a professional football player, so he clearly knew his limitations. This sort of claim raises three issues that highlight some of the pitfalls regarding thinking about self-assessment.

First, judging the quality of physical skills for which there is an objectively observable outcome is probably importantly different from judging cognitive aptitudes or less objective physical skills in that, unlike cognitive aptitudes, the mental processes required to judge the quality of physical performances are different (often derived from external information conveyed via the senses) from the internal mental processes required to enact the performance.<sup>20</sup> Whether or not the ball goes through the goal posts after one kicks it is, for the most part, indisputable so one can quite easily judge one's performance through the immediate feedback provided by the outcome of the action (though even in objective physical domains many of us would have to admit to having discounted even such blatant feedback with thoughts like "I could do it, if only I tried a little harder" or "If only the wind wasn't blowing the wrong direction"—a process of reinterpretation that Gilovich has labeled the gambler's fallacy<sup>18</sup>).

More importantly, this anecdote of an individual's recognizing that he will never play professional football provides an example of arguing for the accuracy of self-assessment on the basis of a logical fallacy. To elaborate, we might think of

the world as a  $2 \times 2$  table in which there are some activities at which we thrive and others at which we perform poorly, crossed with some activities we think we do well and others we think we do poorly. To argue for accuracy of our own self-assessment by saying that we know we are poor at something that we actually do poorly is to look at only one of the four cells. Similar sentiments apply to arguments in favor of the accuracy of self-assessment founded upon "knowing" that we can do things we actually can do. Regardless of which cell is the more predominant cause of belief in self-assessment it must be recognized that the claim that individuals are poor at self-assessment does not imply that people are never accurate in their judgments of ability. The empirically derived conclusion that humans are generally poor at self-assessment merely indicates that we are inaccurate as often as not and that, as a result, one should not rely on self-assessments to provide valid indications of ability.

Finally, this particular example involves a process of reasoning from extreme examples, another erroneous rhetorical strategy. Extreme examples, by definition, are not representative of more common situations. As a result, ability to identify extreme examples accurately (eg, knowing accurately that you are unable to play professional football or claiming that there are individual differences in the general capacity to self-assess by recollecting a previously encountered narcissist) does not enable extrapolation to more common and general judgments.

It is important to remember that we are all subject to the same cognitive errors that lead to the inflated sense of ability to self-assess, and, as a result, our own confidence in our ability to self-assess is as flawed as everyone else's.<sup>19</sup> Unfortunately, it is this personal, flawed self-confidence in our own self-assessment ability that has led us as educators to perpetuate the myth that the effective self-identification of strengths and weaknesses is even possible. As a result, we educators have not only failed to be part of the solution, we have actually been part of the problem.

### Truths About Self-Assessment

Given the issues expressed in this article, we believe that the most important educational activity related to self-assessment as a generalizable, reflective process of generating an unguided, personal, summative assessment of one's own level of ability or performance should entail helping people overcome their personal belief that they can rely on it.<sup>21</sup> Personal, unguided reflections on practice simply do not provide the information sufficient to guide performance improvements adequately.<sup>8</sup> This has been shown in the context of the continuing education exercises one chooses to enact,<sup>22</sup> in the electives students choose to take,<sup>23</sup> and in the effort one exerts to develop specific skills like suturing techniques.<sup>24</sup> We wish to reemphasize that the inadequacy of self-assessment must be viewed as a "we" problem rather than a "they" problem. The same phenomenon impacts upon all of us. As humans, we all tend to believe that we are less

susceptible to cognitive biases than the average person and, therefore, see these phenomena in others as something we should be able to correct.<sup>19</sup> However, inflated self-assessments are not a result of inexperience, inadequate training, or inadequate practice. Rather, the tendency to be overly optimistic about one's abilities is a fundamental property of the way our brains are wired.<sup>25</sup> We do not store literal records of past events in memory. Rather, we store the gist of past events and reconstruct the memories each time we think about them.<sup>26</sup> In doing so we pay selective attention to the evidence around us and ignore details of past experiences that are crucial to determining the likelihood of success in future situations. Previous situations in which we did not perform well are very memorable but tend to be reinterpreted in a light that enables us to maintain a positive view of ourselves.<sup>17</sup> When we have a poor performance, usually it is easy to find a way to blame external circumstances.<sup>18</sup> We do it quite automatically because it is adaptive to maintain a positive outlook; the self-efficacy literature teaches us that those who expect to do well are more likely to do well than those who expect to do poorly.<sup>12</sup> How teachers, peers, and parents couch the praise that leads to these expectations is important,<sup>27</sup> but the general principle that we are better off being overly optimistic is a robust one. As Gilbert has written in another context, "Evolution deserves the Microsoft Windows Award for installing these mental processes in every one of us without asking permission."<sup>28</sup>

The problem is not that these mental phenomena exist. It is that people do not appreciate that they are active. As a result, self-regulating professionals and educators have not yet institutionalized the need to draw on their community for external feedback rather than prioritizing self-generated personal opinions. As humans, we perceive that we have more insider knowledge about the events and experiences that led us to our current performance or state of ability and, therefore, that we are better judges of ourselves than are those around us. However, such feelings of triangulation are illusory and can be misleading.<sup>26</sup> The Physician Review and Enhancement Program in Ontario, charged with evaluating and assessing family physicians' ongoing competence within the province, have reported that the second best predictor of incompetence is working in isolation.<sup>29</sup> In part, this is probably because the individual working in isolation simply does not have available the external feedback necessary to enable accurate recognition of the limits of knowledge and skill. The evidence is clear and overwhelming: self-assessment is not and will never be a generic skill that one can develop. Its "accuracy" is at best dependent on context and, perhaps more importantly, on the individual's level of competence within that particular context.<sup>20</sup>

In addition, as a community we must explicitly recognize that the term *self-assessment* tends to be facily used to describe a complicated, multifaceted, and multipurpose set of phenomena. It has been so vaguely and variously defined as almost to warrant being thrown out of the health professional educators' lexicon if not for the fact that it provides a useful



shorthand for getting people close to a similar mental space when discussing continuing competence, lifelong learning, and professional self-regulation. Still, we should remember that language creates reality and, as a result, we need to strive for increased precision and care in the words that we use.<sup>4</sup>

These issues have important implications for the direction in which the health professional education community should be focusing its educational activities and research agendas. We believe that research questions that take the form of “How well do various practitioners self-assess?” “How can we improve self-assessment?” or “How can we measure self-assessment skill?” should be considered defunct and removed from the research agenda. There have been hundreds of studies into these questions and the answers are “Poorly,” “You can’t,” and “Don’t bother.” Usually when researchers claim to have improved self-assessment they have fallen prey to one of the fallacies described by Ward et al. (eg, placing undue faith in group-based correlations or mistaking distance of individual guesses from true scores as a sensible measure of self-assessment accuracy).<sup>30</sup> Instead, we see three potential programs of research that parallel the three concepts we have distinguished from self-assessment in the first section of this article.

From the perspective of self-directed assessment seeking, we must recognize that for maintenance of competence efforts to be in any way meaningful, external feedback is essential. With this in mind, we would argue that the predominant concern regarding the professional issues of maintenance of competence and continuing professional development (in Schön’s terms, issues of reflection on practice)<sup>31</sup> should focus upon developing habits of self-directed assessment seeking and upon understanding factors that influence our ability to absorb these external sources of feedback in developing a coherent self-awareness of our strengths and weaknesses. Given the influence attributional judgments have on people’s interpretation of the meaning of external feedback, our goal, as a community going forward, should be to determine how best to enable individuals to develop the habits of seeking external feedback, how to convince them that external assessments are legitimate and should be sought, and as a result, how best to deliver external assessments in a way that maximizes the chance the information provided will be incorporated into behavioral changes. The relevant issue is not whether a person can predict his score on a test, but what he does with the information when he finds out his score. Thus, we should be asking questions like “What forms of external data would help individuals recognize areas that require updating?” “How can we collect and deliver these data in a meaningful form?” “How can we convince people to believe this feedback and incorporate it into their self-concept?” and, more generally, “How can we get people to act on externally derived information?” External feedback may not be a panacea, but it is these questions (and questions like them) that we believe will truly advance and productively refine the health professions’ models of professional maintenance of competence and practice improvement.

From the perspective of self-reflective exercises, we should be trying to understand the role of reflection on practice as a pedagogical strategy for better understanding the world around us. Summative forms of self-assessment need not be accurate, nor be generic skills, to enable exercises in self-reflection to formatively facilitate performance improvements. The focus of these exercises is not to determine that one is great or at least good enough, but rather to determine how one understands the world and how one might increase this understanding to the benefit of future performance. Research addressing questions like “Does engaging in self-reflection result in improved performance” could parallel the emerging literature that reveals the pedagogical benefits of externally derived assessment strategies (eg, multiple-choice tests of knowledge).<sup>32,33</sup> More sophisticated questions could address (a) whether or not sharing one’s self-reflections with peers, a tutor, or a mentor is necessary to elicit full advantage of the activity and (b) whether or not developing the habit of self-reflection in one context tends to transfer readily to maintaining that habit in novel contexts or at variable stages of one’s career. It is these questions (and questions like them) that we believe will help the health professions better understand what activities its training programs should be encouraging its members to undertake and what activities should be valued as evidence of having implemented continuous professional development strategies.

Finally, from the perspective of self-monitoring, the forefront of our discussions about professional self-regulation of practice should address the appropriateness of slowing down and/or help-seeking behavior in the context of interacting with specific patients (again, borrowing from Schön, issues of reflection in practice).<sup>16,17,31</sup> Thus, we should be asking questions like “Do individuals show behavioral indications of slowing down/help seeking when they reach the boundaries of their knowledge/abilities in their moment-to-moment interactions with patients?” “What cues (external or internal) initiate such slowing down processes?” “Does the initiation of these processes impact upon the appropriateness of the care provided?” and “How best can the skills associated with slowing down and help seeking be taught?” It is these questions (and questions like them) that we believe will have the greatest impact on patient safety, as it seems likely that these moment-to-moment self-monitoring processes (or lack thereof) are most likely to impact upon the care received by individual patients relative to any of the other forms of “self-assessment” differentiated in this article.

These three suggested research foci are just guideposts for potential approaches that we believe will ultimately be more fruitful than the standard “guess your grade” approaches to studying “self-assessment” in an undifferentiated manner that have dominated the health professional education literature in recent years. Regardless of which program(s) the community chooses to pursue and which proves to be the most productive, it is time to move beyond the rhetoric that self-assessment as a general, personal, unguided judgment of ability should be taught and developed as a valid

### Lessons for Practice

- To maintain meaning it is necessary to distinguish the act of self-assessment as a self-determined judgment of one's ability from other related, but conceptually distinct activities.
- Various arguments in favor of the accuracy of self-assessment rest upon fallacious rhetorical strategies and ironic misperceptions of one's own abilities to self-assess.
- Further research is needed to better understand the pedagogical value of self-directed assessment seeking and reflection, as well as the cues related to the sufficiency of self-monitoring activities, but questions relating to the accuracy of self-assessment as a generic skill that one can develop should be considered defunct.

basis on which to direct performance improvements because this belief, when all is said and done, is the fundamental fallacy of self-assessment.

### References

1. Eva KW, Regehr G. Self-assessment in the health professions: A reformulation and research agenda. *Acad Med.* 2005;80(suppl 10):S46–S54.
2. Bosma J, Laszakovits D, Hattery RR. Self-assessment for maintenance of certification. *J Am Coll Radiol.* 2007;4:45–52.
3. Liaison Committee on Medical Education. Guide to the institutional self-study for programs of medical education leading to the M.D. degree (for full accreditation surveys scheduled in 2008–2009). Available at: <http://www.lcme.org/selfstudyguide0809.pdf>. Accessed September 6, 2007.
4. Lingard LA, Haber RJ. Teaching and learning communication in medicine: A rhetorical approach. *Acad Med.* 1999;74:507–510.
5. Collins J. Education techniques for lifelong learning: Writing multiple-choice questions for continuing medical education activities and self-assessment modules. *Radiographics.* 2006;26:543–551.
6. Handfield-Jones R. MAINPRO: Self-regulation, professionalism, and continuing professional development. *Can Fam Physician.* 2002;48:856–858.
7. Colliver JA, Verhulst SJ, Barrows HS. Self-assessment in medical practice: A further concern about the conventional research paradigm. *Teach Learn Med.* 2005;17:200–201.
8. Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: A systematic review. *JAMA.* 2006;288:1057–1060.
9. Boud D. Avoiding the traps: Seeking good practice in the use of self assessment and reflection in professional courses. *Soc Work Educ.* 1999;18:121–132.
10. Carmichael HT. Self-assessment tests: The psychiatric knowledge and skills self-assessment program. *JAMA.* 1970;213:1656–1657.
11. Dunning D, Heath C, Suls JM. Flawed self-assessment: Implications for health, education, and the workplace. *Psychol Sci Public Interest.* 2004;5:69–106.
12. Bandura A. *Self-Efficacy: The Exercise of Control.* San Francisco: Freeman; 1997.
13. Ende J. Feedback in clinical medical education. *JAMA.* 1983;250:777–781.
14. Mann K, Gordon J, McLeod A. Reflection and reflective practice in health professions education: A systematic review. *Adv Health Sci Educ.* DOI 10.1007/s10459-007-9090-2.
15. Bereiter C, Scardamalia M. *Surpassing Ourselves: An Inquiry Into the Nature and Implications of Expertise.* La Salle, Ill: Open Court; 1993.
16. Eva KW, Regehr G. Knowing when to look it up: A new conception of self-assessment ability. *Acad Med.* 2007;82(10 Suppl.):S81–S84.
17. Moulton C, Regehr G, Mylopoulos M, MacRae H. Slowing down when you should: A new model of expert judgment. *Acad Med.* 2007;82(10 Suppl.):S109–S116.
18. Gilovich T, Douglas C. Biased evaluations of randomly determined gambling outcomes. *J Exp Soc Psychol.* 1986;22:228–241.
19. Pronin E, Lin DY, Ross L. The bias blind spot: Perceptions of bias in self versus others. *Person Soc Psychol Bull.* 2002;28:369–381.
20. Kruger J, Dunning D. Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J Personality Soc Psychol.* 1999;77:1121–1134.
21. Regehr G, Eva K. Self-assessment, self-direction, and the self-regulating professional. *Clin Orthop Relat Res.* 2006;449:34–38.
22. Sibley JC, Sackett DL, Neufeld V, Gerrard B, Rudnick KV, Fraser W. A randomized trial of continuing medical education. *N Eng J Med.* 1982;306:511–515.
23. Dauphinee WD, Patel VL. Early career choice: An unsuccessful program. *J Med Educ.* 1983;58:695–702.
24. Boehler ML, Rogers DA, Schwind CJ, et al. An investigation of medical student reactions to feedback: A randomized controlled trial. *Med Educ.* 2006;40:746–749.
25. Wilson TD. *Strangers to Ourselves: Discovering the Adaptive Unconscious.* Cambridge, Mass: The Belknap Press of Harvard University Press; 2002.
26. Loftus EF. Our changeable memories: Legal and practical implications. *Nature Reviews: Neurosci.* 2003;4:231–234.
27. Dweck CS. Caution—praise can be dangerous. *Am Educator.* Spring 1999:1–5.
28. Gilbert D. *Stumbling on Happiness.* Toronto, Canada: Knopf Canada; 2006.
29. Caulford PG, Lamb SB, Kaigas TB, Hanna E, Norman GR, Davis DA. Physician incompetence: Specific problems and predictors. *Acad Med.* 1994;69(suppl 10):S16–S18.
30. Ward M, Gruppen L, Regehr G. Measuring self-assessment: Current state of the art. *Adv Health Sci Educ Theor Pract.* 2002;7:63–80.
31. Schön D. *The Reflective Practitioner. How Professionals Think in Action.* London, England: Temple Smith; 1983.
32. McDaniel MA, Roediger HL, McDermott KB. Generalizing test-enhanced learning from the laboratory to the classroom. *Psychon Bull Rev.* 2007;14:200–206.
33. Eva KW. Putting the cart before the horse: Testing to improve learning. *BMJ.* 2007;334:535.