

Moderator/Discussant: Karen Mann, PhD

Self-Assessment in the Health Professions: A Reformulation and Research Agenda

Kevin W. Eva and Glenn Regehr

Many researchers and educators have identified self-assessment as a vital aspect of professional self-regulation.^{1,2,3} This rationale has been the expressed motivation for a large number of studies of self-assessment ability in medical education, health professional education, and professions education generally. Unfortunately, the outcome of most studies would seem to cast doubt on the capacity for self-assessment, with the majority of authors concluding that self-assessment is, in fact, quite poor.⁴ In a recent article, Ward and colleagues suggested that this conclusion must be questioned because the methodologies used to evaluate self-assessment are fraught with methodological weaknesses.⁴ However, even studies that have attempted to address the weaknesses within the methodological paradigm have produced little evidence for effective self-assessment.⁵ Thus, the health professional education community is left with a conundrum that can only be resolved by deciding either that the conclusions of the studies are wrong, or that a critical premise underlying the concept of “self-regulation” in the professions is unassailable.

The current paper addresses this conundrum by arguing that there is a problem with the literature on self-assessment, and that this problem is more fundamental than a list of easily correctable methodological flaws. Rather, the roots of the problem in the self-assessment literature involve a failure to effectively conceptualize the nature of self-assessment in the daily practice of health care professionals, and a failure to properly explicate the role of self-assessment in a self-regulating profession. Until such an articulation of self-assessment is elaborated, it is difficult to know even which literatures might be informative in addressing this issue, and impossible to develop programs of

research that operationalize the concept of self-assessment ability in a form that can be effectively studied. Thus, we will begin with a brief reflection on the various functions of self-assessment for a practicing health care professional and the manner in which these functions operate.

The Purposes of Self-Assessment in Practice

Self-assessment has been defined broadly as the involvement of learners in judging whether or not learner-identified standards have been met.⁶ While attractive due to their concise and encompassing nature, we fear that such simple definitions risk being misleading as they can cause underappreciation of the complexities of the construct. Self-assessment functions both as a mechanism for identifying one’s weaknesses and as a mechanism for identifying one’s strengths. Each of these mechanisms can be considered to have distinct, albeit complementary, functions. As a mechanism for identifying weaknesses or gaps in one’s skills and abilities, self-assessment serves several potential functions. First, in daily practice, the identification of one’s weaknesses allows the professional to self-limit in areas of limited competence. For example, in many circumstances the professional can quickly reject certain plans of action because she recognizes that she is unlikely to be able to complete the component tasks necessary to enact the plan. In other circumstances, a professional might recognize that he is “over his head” in a particular case and decide that it is time to recruit additional resources: to “look this up,” to obtain a consultation, to recruit additional support, or to refer the problem to another individual who is more competent in this domain. Second, in reflecting on one’s practice in general, the ability to identify weaknesses can serve the function of helping the professional set appropriate learning goals. That is, the traditional model of self-regulated continuing professional development presumes that an individual will select ongoing learning activities that fill

professional gaps, but this presumes that the professional can effectively self-assess. Thus, in this role, the identification of weakness can help a professional to decide what must be learned. As a corollary to this, effective self-assessment is vital for setting realistic expectations of oneself, to avoid setting oneself up for failure. Thus, the identification of weakness also helps the self-regulating professional to decide what *not* to try learning, what should be accepted as forever outside one’s scope of competent practice.

There is a complementary set of functions served by the ability to accurately self-assess one’s strengths. First, in daily practice, having a clear and accurate sense of one’s strengths allows the professional to act with appropriate confidence. For example, knowing one’s strengths provides the professional with the confidence to move forward on a fitting plan of action without inappropriate hesitation or trepidation. Similarly, it ensures that the individual will choose to persist on an appropriate plan of action in the face of initially negative feedback. The right path is not always smooth even if it is right, and early abandonment of an appropriate plan of action is as costly as selecting an inappropriate plan in the first place. Second, when reflecting on one’s practice in general, an appropriate assessment of one’s strengths ensures that one can set appropriately challenging learning goals, pushing the edges of one’s knowledge rather than choosing professional development courses that merely reiterate what one already knows. At the same time, by knowing one’s strengths, a professional can select learning objectives that are within her grasp, and therefore will be able to enjoy the motivational influence of attaining her goals and experience the satisfaction of a job well done.

Together, then, the ability to accurately assess one’s weaknesses and one’s strengths generates a capacity for finding an effective balance both in daily practice and in setting personal learning goals. In daily practice, it generates a balance of

Correspondence: Kevin W. Eva, PhD, Department of Clinical Epidemiology and Biostatistics, Program for Educational Research and Development, MDCL 3522, McMaster University, Hamilton, ON, L8N 3Z5, Canada; e-mail: (evakw@mcmaster.ca).

confidence and caution, of persistence and flexibility, of experimentation and safety, and of independence and collaboration. In establishing learning goals, it generates a balance of learning enough but not too much, of starting neither too high nor too low, of knowing what to tackle and what to abandon. And in reflecting on accomplishments, it generates a balance of satisfaction and incentive, of self reward without self delusion.

In order to fulfill these various functions, it seems that self-assessment must be effectively enacted in three forms: summatively, predictively, and concurrently. Enacting self-assessment summatively, a professional must reflect on completed performances both for the purposes of assessing the specific performance and for the purposes of assessing his abilities generally. When evaluating performance on a particular task, the professional can often assess the overall quality of the completed job as a question that may come in various forms. That is, the individual might ask how good this performance was relative to what she could have done; relative to what her peers might typically do; relative to the best that could have been done (a gold standard); or relative to some minimally acceptable standard. Alternatively, there are some situations where the mechanisms for objectively assessing the outcome are not immediately available, in which case the professional might ask herself how confident she is in the conclusion or outcome generated (is it right? will it stand up? could there have been a better solution given the situation?).

The professional might then use her assessment of the specific task to draw summative conclusions about herself or her abilities in this domain generally. Again, such conclusions may be in absolute terms (am I good enough in this domain? am I minimally competent?) or in relative terms (am I average, above average, or below average, and against whom should I be comparing myself?). In drawing general conclusions about her abilities from a particular performance, the professional must also make determinations about whether this particular episode should be taken as an appropriate reflection of her general skills: were there extenuating circumstances that led to a particularly

poor (or good) performance that might lead one to discount this outcome as reflective of overall ability?

In addition to these summative functions, self-assessment must be used predictively. Professionals are constantly required to assess their likely ability to manage newly arising situations and challenges. In this predictive role, self-assessment leads to questions such as: Am I up to this challenge? Should I be starting this task (now, alone, in this way)? What are realistic goals for accomplishment in this context (what would I consider to be a good or acceptable outcome for me)? How much better might I imagine performing with some additional preparation and is the increased preparation worth the anticipated increase in performance? What additional resources should I recruit (either internally or from the outside) to complement my strengths and shore up my weaknesses?

Finally, self-assessment plays a vital role in its concurrent mode of functioning. In this concurrent mode, self-assessment acts as an ongoing monitoring process during the performance of a task. It is self-assessment in its concurrent mode that leads to questions such as: Is this coming out the way I expected? Am I still on the right track? Am I in trouble? Should I be doing anything differently? Should I persist in the face of negative feedback from the situation (that things are not going the way I thought they would or as easily as I thought they would)? Do I need to recruit additional resources (internal resources such as attention or external resources such as advice/assistance)? Do I need to reassess my original goal or my original plan?

Thus, self-assessment is a complicated, multifaceted, multipurpose phenomenon that involves a number of interacting cognitive processes. It functions as a monitor, a mentor, and a motivator through processes such as evaluation, inference, and prediction. Given this elaborated description of self-assessment, it is unlikely that simplistic questions such as "are health professional trainees effective self-assessors?" will lead to insightful discoveries about the nature and value of self-assessment. Rather, researchers must ask questions such as: On what basis do individuals make these decisions? What factors affect their

reasoning? How fine tuned does the assessment need to be in order to be useful?

A first step toward addressing these questions must be to determine who is already asking them and what insights we may borrow from their discoveries and reflections. Our search has led us to several literatures that seem particularly relevant: self-efficacy and self-concept; cognitive and metacognitive theory; social cognition; models of expert performance and the development of expertise; and the concept of reflective practice. In the following sections we will briefly touch on each of these literatures and suggest how they might inform our understanding of self-assessment. Our intent here is not to provide a systematic review of each literature, but to provide an overview of questions being addressed by researchers outside medical education that should inform our conception of self-assessment as a regulatory strategy. For each new literature we will define the area, provide examples of the issues under consideration, and then summarize the implications for self-assessment in the professions. We will end with a proposal for a program of research that has the potential to move the field beyond our current paradigm of repeatedly concluding that self-assessment is generically poor.

Self-Efficacy and Self-Concept

In studying the accuracy of self-assessments, education researchers in the health professions have tended to focus conceptually on what we have labeled the summative function – the ability to draw general conclusions about one's skills or knowledge in specific domains: How well do I understand endometriosis? Am I able to communicate effectively with other members of the health care team? Practically, this has usually been operationalized in research studies as a request that students try to estimate how well they will/did perform on an immediately following/preceding task. Yet, there is an important distinction between general assessments of one's ability in an area and the more specific question of how one did on a particular task. Researchers in the field of personality theory, for example, usefully distinguish between judgments of self-efficacy and the development of self-concept. Self-efficacy is the belief in one's

capabilities to recruit the resources and execute the actions required to manage prospective situations. Self-concept is the relatively sweeping cognitive appraisal of oneself that is integrated across various dimensions.⁷ Thus, self-concept beliefs are context free, generalized judgments of self-worth that involve cognitive self-appraisals independent of a specific task or goal (but not necessarily independent of domain). By contrast, self-efficacy is a context specific assessment of competence to perform a specific task or range of tasks in a given domain (i.e., an individual's judgment of her capabilities to complete a given goal). Self-efficacy is, by its very nature, driven by an interaction between self-concept beliefs about one's skills or abilities and the specific context in which those skills or abilities will be applied for the attainment of the particular goal. It is concerned with the contextually embedded orchestration of skills that lead to performance.

Self-efficacy differs importantly from the concept of self-assessment as currently envisioned in the health professions education literature in that self-efficacy is not only influenced by direct and indirect feedback, but also influences the future performance of tasks (the choices we make, the effort we put forth, how long we persist when confronted with obstacles or in the face of failure). Thus, there is an important reciprocity between self-efficacy and success. Not only will success lead to a strong sense of self-efficacy, but self-efficacy will also lead to an increased likelihood of success. Self-efficacy beliefs are not merely passive reflections of performance, but part of a self-fulfilling prophecy that affects performance. As a result, there is an advantage to high self-efficacy beliefs even in circumstances where such beliefs may not be warranted by past performance. Clearly there is a logical disadvantage to continually overestimating one's abilities, but this obvious disadvantage must be balanced with the value of believing that one can achieve more than one has in the past and that one can manage the challenges that one will face.⁸

As a result, researchers in the field of self-efficacy appear to be less worried about the "accuracy of self-assessment" and more worried about its impact on impending problem solving situations. They unconcernedly alter the situational

self-efficacy of study participants through manipulations such as: varying the order in which people consider hypothetical levels of future performance,⁹ having subjects contemplate various positive or negative performance-related factors,¹⁰ altering the "anchor" values representing high or low levels of performance,¹¹ or providing false performance feedback.¹² Such manipulations regularly alter subjects' expectations of success on future events within the context of the study, suggesting that subjects will take contextual information into account when judging (either explicitly or implicitly) the likelihood of future success on tasks within that context. Again, for researchers engaged in the study of self-efficacy, the important point to be taken from these studies is that "trivial" factors alter self-efficacy and can affect future performance.¹³ For them, the fact that one can radically alter an individual's self-assessment of future performance appears to be simply taken for granted, rendering the question of "accuracy" somewhat nonsensical.

Early on, Bandura provided a taxonomy of origins from whence information that would influence self-efficacy could be received.¹⁴ It included personal experience, vicarious experience, verbal persuasion, and physiological state. In addition, Cervone has argued that fundamental cognitive mechanisms (including common heuristics, as will be discussed in the next section) will influence the extent to which information from any given source will be weighed.¹⁵ In general, Cervone argues that self-efficacy judgments are not simply driven by an active, motivated distortion of facts in the service of ego protection ("hot cognition"), but rather that fundamental cognitive processes (i.e., those regularly used for a wide variety of judgment tasks – "cold cognition") influence self-efficacy beliefs quite independently. Overall then, it appears that researchers in the self-efficacy literature offer several theoretical and methodological approaches that can inform research in self-assessment. They acknowledge, in fact presume, the instability and situational specificity of self-reflective judgments, they examine and explicitly manipulate the factors that affect these judgments, and they concern themselves with the consequences of these judgments for future behavior.

Cognitive and Metacognitive Theory

In contrast to the focus on "accuracy" in the self-assessment literature and the focus on "consequences" in the self-efficacy literature, cognitive psychologists interested in metacognition (knowledge of one's own knowledge) tend to focus on delineating the mechanisms that allow us to mentally supervise and control the way in which we process information. Of particular interest for our purposes are questions of how people form metacognitive judgments, and what cues influence people's judgments of how well they have learned something. It is a fundamental assumption of this work that we do not have direct introspective access to our own memories or knowledge base. Rather, just as we must infer others' level of knowledge and motivations from their behaviors and other cues, so too we must use peripheral cues to make inferences about our own level of knowledge and learning. In fact, it is argued that our judgments of our own abilities are often based on the same inferential cognitive strategies, or heuristics, that we use to judge others. For example, the easier it is to process a piece of information, the more likely we are to judge that we will remember that information later (a fluency heuristic).¹⁵ Such heuristics are cognitive short-cuts that make us extremely effective and efficient at operating within a complex world despite our limited mental resources. However, they can also bias us in a way that leaves us susceptible to errors in decision making and, when applied to ourselves, errors in trying to identify our own strengths and weaknesses.

Studies from this field suggest that, when trying to judge one's ability in a domain or when trying to judge the likelihood of success on a task, the accuracy of these metacognitive judgments is dependent on the extent to which the apparent difficulty of learning mimics the actual difficulty of eventually retrieving the learned material from memory. For example, research demonstrates that, when people are trying to learn a piece of information (such as a list of words) for later recall, several factors affect their judgments of having succeeded in their learning efforts. Metacognitive judgments are more accurate if the repetitions of each word are spaced apart and interspersed with other words than if

repetitions of each word are blocked together.¹⁶ People appear to use the cue of fluency (i.e., ease of understanding) in judging the extent to which they have learned material and, as such, overestimate the amount they have learned when fluency is increased by blocking repetitions together. Similarly, metacognitive judgments are more accurate when there is a delay between study of the words and efforts to recall during practice. In general, people overestimate their learning if the words are blocked or if recall follows too closely on study of the words, because these forms of the task are easier than the actual task they will eventually be expected to perform (recall after a long delay). The harder the retrieval task during the learning period, the better the predictions of the amount of learning that took place.¹⁷

Importantly, however, merely mixing the list and delaying recall during practice are often insufficient to improve metacognition if people are left to their own devices during learning. That is, in order to recognize one's inability to recall the words it is necessary to actually try to recall them and make explicit mistakes in retrieval. Without these explicit errors as feedback, people continue to overestimate their ability to recall the words. Interestingly, participants are unlikely to spontaneously induce in themselves the failures that enable better judgments of learning. For example, judgments of learning tend to be more accurate after participants are forced to provide a response and produce the wrong word than if they are allowed to say, "I don't remember."¹⁸ This suggests that, without external pressure to do so, the participants did not try and fail, but rather simply did not try, and in doing so, missed an important cue that they might have used to improve their self-assessments. This finding is consistent with the higher correlations between performance and self-assessment seen in the health professions literature when judgments are elicited postperformance relative to preperformance.⁵

Taken as a whole, the findings from this literature emphasize the importance of moving beyond questions of "can people self-assess accurately?" to ones that explicitly focus on the various factors that affect judgments of learning or knowledge or ability. In the absence of

direct access to our mental states, we are forced to make metacognitive judgments based on a variety of internal and external cues. Metacognitive judgments tend to be more accurate when these cues accurately reflect the factors that affect subsequent performance,¹⁹ but there are many instances in which the cues used for judgments of learning lack predictive validity or worse yet, induce systematic discrepancies between predicted performance and actual performance.²⁰ A better understanding of which cues are used and which ones should be used in health professional education contexts (as well as the impact these cues have on study habits) might better guide training strategies and improve our understanding of the concept of self-regulation. Some insight into the types of cues that are often misleading in real world situations (and reasons for the lack of insight into the inappropriate use of specific cues) has been gained from researchers working in the field of social cognition, the focus of the next section.

Social Cognition

Research in social psychology has led many to conclude that much of what we want to know about ourselves resides outside of conscious awareness.²¹ Each of us possesses an adaptive unconscious that guides much of our behavior, motivations, and feelings. This part of the mind is labeled unconscious because although we have privileged access to the contents (current thoughts, memories, and objects of attention), we do not enjoy such access to the mental processes that are engaged. We have a tendency to confabulate explanations for our behaviors,²² but these explanations are often inference-based and no more trustworthy than are introspections about the inner workings of our kidneys.²³ This unconscious is adaptive because there are benefits to naivety. Most people believe themselves to be more popular, a better driver etc., than the average person.²⁴ While it is logically impossible that we are all above average, at the individual level such positive self-deceptions can be beneficial in practice; individuals who maintain such illusions are less likely to be depressed and more likely to persist at (and succeed on) difficult tasks.²⁵ Gilbert and Wilson talk of the psychological immune system, highlighting the great lengths we will travel to maintain a sense of well being, rationalizing and justifying

threatening information.²⁶ How we rationalize is somewhat idiosyncratic, but Gilovich, as one example, offers a number of mechanisms (some motivated, some inherent in fundamental cognitive processes) by which intelligent, thoughtful people can develop and maintain erroneous beliefs, many of which are relevant to an appreciation of what is required to accurately self-assess one's own strengths and weaknesses.²⁷

As one example, Gilovich describes gambling tendencies and presents evidence that counters the common belief that gamblers think they can beat the odds because they ignore or forget their losses. On the contrary, gamblers focus more attention on their losses and remember them better than wins. They maintain the belief that they are successful, however, by discounting the losses, focusing on the reasons why they should have won if not for some fluke event (e.g., the quarterback being injured). As a result, gamblers come to think of losses as "near wins" and thus maintain the belief that they can beat the odds. Learners likely find themselves in a similar situation. It is very easy to maintain an inaccurate perception of one's own ability by making claims like "I knew the answer, but read the question wrong" or "Wow, I made a lucky guess in response to that question." This tendency to discount conflicting information, combined with the rarity of corrective feedback increases the likelihood that flaws in reasoning will be reinforced.

Given that the ultimate goal of self-assessment is actually to avoid such biased images of oneself, social psychologists suggest that it is necessary to look outward at one's own behavior and how others react to it rather than simply reflecting inward.²⁸ When reflecting on our knowledge and abilities we have a great deal of information available to us that is not available to anyone else (e.g., private knowledge/idiosyncratic theories), but our phenomenal capacity for discounting distortions that do not fit with our perception of reality can render illusory the feeling of triangulation that additional information provides, thereby resulting in a misleading feeling of confidence in the accuracy of our judgments.²⁸ Sources of such illusions include a tendency we have to find more exceptions than truly exist, placing undue

weight on apparently unusual factors,²⁹ and being more likely than external observers to overlook situational influences on our actions, the tendency to do so being broadly recognized as the fundamental attribution error.³⁰

This creates a paradox for self-regulating professionals in that it suggests one must systematically and intentionally elicit the views of others (both explicit opinion and implicit reaction) in order to fully develop an accurate impression of oneself. Without question the perceptions of others are also prone to distortions, but the more heterogeneous the sources of information, the less susceptible our self-concept might be to biased search for confirmation. This notion that self-assessment is insufficient for the evolution of accurate self-concept is consistent with the finding that peers tend to be better predictors of performance than do individuals rating themselves, both in health sciences,³¹ and social psychology.³² This view again raises questions about self-assessment quite distinct from the simple question of accuracy that has preoccupied self-assessment researchers in professional education. To what extent do health care practitioners seek out assessments from others? What prompts them to do so? How can we optimally supplement self-assessments with the views of others to create a coherent and appropriate sense of self? To what extent is coaching/mentoring/peer evaluation necessary/beneficial for such achievements to be reached? This latter question has been a major focus in determining the characteristics of expert performance.

Models of Expert Performance and the Development of Expertise

Some social psychologists have argued that the adaptive unconscious is largely a pattern detector whereas the conscious serves more as a fact checker.³³ Taken broadly, this is also consistent with current models of proficient clinical reasoning, a construct being characterized as the flexible adaptation of multiple approaches to reasoning including both a nonanalytic, Gestalt-like, consideration of new cases, and a more carefully controlled (i.e., analytic) consideration of specific features.^{34,35} In the current context, the question of

interest is what role, if any, does this conscious or analytic process of self-regulation play in the development and maintenance of expertise.

In narrowly focusing the study of expertise on replicable elite performance, Ericsson and colleagues have been able to demonstrate on repeated occasions (and in diverse domains) the importance of deliberate practice – effortful, individualized training on specific tasks *selected by qualified teachers*.³⁶ Deliberate practice is distinct from the enjoyable state of play (characterized as flow—giving up reflective control)³⁷ and work (leading to immediate monetary and/or social rewards). Central to its definition is the presence of an instructor who can push students beyond their current ability by pointing to problems or novel approaches that are likely to go undetected if one relies solely on self-direction. In fact, notable by its *absence* in all domains of expertise except the health professions is an emphasis on self-directed learning. Contrary to popular belief, the role of early instruction and maximal parental support appears to be much more important to the development of child prodigies in many areas than innate talent.³⁸ Of course, all learning need not be a social event. People must practice and that practice often takes place in isolation. And people can read and learn things on their own. However, as a general rule, the notion that one could advance far beyond current level of ability without feedback from others who themselves maintain expertise is somewhat foreign in other domains.

Within health professional education “self-directed learning” has become a common value despite debate over whether or not adult learners are in fact able to self-direct their own learning.^{39,40} It now seems clear that while content expertise does not guarantee teaching effectiveness, students are better off with tutors who maintain enough expertise in the content to be learned that assistance and direction can be provided.⁴¹ Even practicing clinicians appear to have difficulty guiding their own continuing education efforts,⁴² a phenomenon described by Metcalfe as being the result of a region of proximal learning – people will labour in vain because they distribute their learning time towards material that

seems within reach rather than material that is maximally beneficial.⁴³

Again, one of the challenges, we think, is that “self-directed learning” has been ill defined. There is good evidence that active manipulation of a problem before receipt of the solution yields better learning (i.e., higher rates of analogical transfer) than simple provision of the problem and solution.⁴⁴ However, it should be noted that such benefits are always drawn from active manipulation of cases that an experimenter (i.e., someone with knowledge of the solution) has deemed relevant. When the reward structure of learning is made clearer by providing a broader overview of the material to be learned, students appear capable of distributing their study time in a more rational manner than Metcalfe’s region of proximal learning would imply.⁴⁵ As indicated in the section on metacognition, the accuracy of self-assessments increase in many tasks upon provision of what Bjork has called “desirable difficulties” (i.e., strategic provision of challenges that need to be overcome).⁴⁶

So again, the question of importance becomes not one of “are students able to direct their own learning in a way that enables expert performance?” but rather, becomes “can students distribute their learning efforts within boundaries provided by expert tutors?” There is evidence from the health professional literature consistent with Metcalfe’s region of proximal learning,⁴² but whether desirable difficulties can be provided in a way that guides continuing education efforts more strategically remains to be tested. Doing so is likely to require a particular form of reflection that may or may not be teachable.

The Concept of Reflective Practice

In trying to determine the purpose of self-assessment, we have been drawn to a reformulation that emphasizes reflection-in-practice rather than reflection-on-practice (using the terminology of Donald Schön).⁴⁷ Indeed, others have suggested that professional education is only one tradition in which the basic concept of reflection has been used to describe the need to assess one’s own abilities.³ Schön speaks of indeterminate zones of practice that have uncertainty, uniqueness and values conflicts. These

indeterminate zones of practice elude the canons of technical rationality. In such cases, competent practitioners must not only solve technical problems by selecting the means appropriate to clear and self-consistent ends, they must also reconcile, integrate or choose among conflicting appreciations of a situation so as to construct a coherent problem worth solving. When a problematic situation is uncertain, the use of technical problem solving requires, first, the effective construction of a well-formed problem. This formation of a problem is, itself not a technical task. Unique situations cannot be handled by applying standard theories or techniques derived from the store of professional knowledge. Schön claims that these indeterminate zones of practice are central to professional practice and are the reason that reflection is important. To elaborate this, Schön describes three states.

Knowing-in-action is an unreflective capacity to “know” what is happening in a situation and to enact the right actions to respond to it without having a capacity to consciously describe or understand exactly what we are doing or how (procedural knowledge, tacit knowledge, intuition etc.). It may be described in terms of strategies, understanding of phenomena, and ways of framing a task or problem appropriate to the situation. It is dynamic and intelligent. This gets us through some important portion of our professional day, but not all of it. It yields intended outcomes as long as the situation falls within the boundaries of what we have learned to treat as normal. However, sometimes things start going wrong and this requires reflection. Two kinds of reflection are described by Schön.

Reflection-on-action is a postexperience reflection on what we did. Its purpose is to discover how our particular enactment of knowing-in-action might have contributed to the situation as it arose and how we might have dealt with the situation differently. It may also happen by stopping the event in the middle for a “stop and think” but it is not tied to action in the sense of knowing-in-action. Again, much of the literature on self-assessment appears to be centred conceptually around the concept of reflection-on-action (with the implicit expectation that corrective action is taken when such reflection finds oneself

wanting in some area), and is operationalized as a simple version of this reflection-on-action (a summative, often numeric assessment of one’s performance on a particular task).

Reflection-in-action, by contrast, is a task-bound reflective process in which we continue to act but reshape our action online through explicit cognition. It is a form of more effortful, guided problem solving in which more active cognition is required in the form of creativity and/or experimentation. As we understand it, the state of reflection-in-action is a border state. It is embedded in action and it is possible for it to be invoked without explicit recognition or metacognitive awareness, so there may be blurring in the lines between reflection-in-action and knowing-in-action. For example, when driving, one might without much conscious awareness, slow or stop one’s conversation as one begins to increase attention to another driver who seems to be inattentive or dangerously unpredictable. Nonetheless, reflection-in-action appears to involve a “stepping up” of cognitive resources when a situation evolves beyond the routine and, therefore, beyond the scope of knowing-in-action processes.

The ability to shift from knowing-in-action activity to reflection-in-action activity must require some form of monitoring of the unfolding situation. When it is going well, this monitoring itself may be relatively unreflective and in some cases may even be insufficient for minor issues (it is sometimes possible to drive past our exit if we are thinking about other things). Again, however, as a situation evolves into conditions that render knowing-in-action processes insufficient, whether through slowly evolving variation from the norm (as in the case of the dangerously inattentive driver above) or through surprise (the unfolding of an accident in front of you), some monitoring activity, even if unconscious, must be sufficiently active to alert one to the need for a shift to reflection-in-action. A similar form of “unconscious monitoring” can be seen in the work of Csikszentmihalyi, who discusses the phenomenon of “flow,” which involves a sustained pleasure arising from a sense of absorption, a sense of control, and a loss of self-consciousness due to heavy investment of mental resources.³⁷ Again, it seems that

for a sense of flow to occur, there must be some monitoring (even if somewhat unreflective monitoring) of the experience for it to provide pleasure.

With this description as background we would suggest that the emphasis in the self-assessment literature on reflection-on-practice has come at the cost of attention to self-assessment as reflection-in-practice. Unquestionably it is important that health care professionals engage in some form of reflection-on-practice in order to engage in systematic continuing professional development. There are times when it is important that a person consider their practice at a macro level and be able to make self-reflective statements such as, “I am no longer comfortable handling these types of cases” or “I should talk to Susan because she knows more than I do in this domain.” However, we would like to argue that, on a day-to-day basis, reflection-in-practice is a substantially more important mechanism for ensuring safe and effective performance. Rather than being confident that our students can rate their overall ability (or their ability relative to their colleagues) we would prefer to know that our students know when to stop and “look it up” because they recognize that they do not know something important about the specific case they are faced with at that moment. Largely ignored in the current self-assessment literature are questions of whether or not individuals accurately reflect-in-practice, and make appropriately self-reflective statements such as, “I have to get more information,” “I can’t proceed without confirmation,” or “I need to supplement my skills with resources from elsewhere.” What cues prompt such questions and what contextual variables inhibit such reflection-in-practice from being engaged (or acted upon when engaged) are important research questions that have yet to be addressed. Further, there is little or no research examining the extent to which the processes involved in reflection-in-practice are similar to or different from the processes involved in reflection-on-practice. However, there may be some data already available that would provide a sense of the conditions under which reflection-in-practice could be studied systematically. Unpublished data we have collected suggest that students are able to indicate which test questions they are most likely to get

wrong during the course of sitting an exam.⁴⁸ Allen et al. have provided evidence that people take more time to provide diagnoses when the eventual diagnosis they provide is wrong relative to when it is correct.⁴⁹ Many literatures are rife with examples like these that suggest people are capable of self-assessing in practice despite the poor correlations often observed when they are asked to provide overall judgments of their ability. These findings, however, are typically published with a focus distinct from the question of self-assessment accuracy and as such, have yet to be considered systematically in this context.

Discussion

For at least 40 years professional programs in the Health Sciences have emphasized the importance of being able to self-assess one's ability as the critical foundation on which to build self-directed/life-long learning skills and preserve the self-regulating nature of the professions. In recent years data have continued to accumulate that call into question the strength of the footings on which this tower of rhetoric has been built. Several authors have tried to address these findings by criticizing the literature on self-assessment from methodological,⁴ and theoretical^{50,51} perspectives. Despite these efforts to salvage self-assessment as a meaningful and useful construct, the subsequent findings are still equivocal at best.^{5,52-54} This review was completed in an attempt to step back from the literature on self-assessment within the health professions and reconsider the sufficiency of the operational and conceptual definitions of self-assessment that the health professional education community (us included) have utilized.

If we are to take the lessons from these various literatures seriously, we must conclude from the metacognitive and self-efficacy literatures that self-assessment is not a stable skill, but one that that will vary by content, context and perspective. Thus, to ask whether students are "good" self-assessors, or even to ask which individuals are "good" self-assessors in a particular content domain radically underestimates the situational influences on the answers to these questions. Further we must conclude from the social cognition and expertise literatures that we have no

special insight into our own abilities relative to outside observers, suggesting that it may be inappropriate to try to separate concepts of "accurate self-assessment" from concepts of "accurate assessment" or "conceptual understanding". Thus, the route to self-improvement is not through becoming a more accurate self-assessor, but through seeking out feedback from reliable and valid external sources (experts, self-administered tests etc.), and then, according to the self-reflection literature, making a special effort to take the resulting feedback seriously rather than discounting it: to reflect rather than ruminate. Indeed Boud has suggested that the phrase self-assessment should not imply an isolated or individualistic activity; it should commonly involve peers, teachers, and other sources of information.³

Finally, we must be aware that the purpose of self-assessment is more complicated than simply "finding gaps and learning more." As the self-efficacy literature suggests, at the very least there are times when accurate self-assessment is not always consonant with improved performance. There are moments when confidence and persistence in the face of negative feedback may in fact be functional, and the literature would tell us that such persistence is more likely when feelings of self-efficacy are high regardless of past performance.

But, perhaps most importantly for daily practice, the literature on reflective practice suggests that the education and research community must move beyond the conceptualization and operationalization of self-assessment as a *conscious* metacognitive, and usually post hoc, summative process. That is, it seems likely that most of the value that self-assessment provides for the practitioner does not happen at the level where the individual is consciously reflecting on her performance or ability at a time that is remote from the performance or use of that ability. While this "reflection-on-practice"⁴⁷ may be important to self-directed learning and continuous professional development, there is an important way in which it may not be vital to self-regulation and safe practice.

Safe practice in a health professional's day-to-day performance requires an awareness of when one lacks the specific

knowledge or skill to make a good decision regarding a *particular* patient (i.e., when more information and/or a consultation is required). This decision making in context is importantly different from being able to accurately rate one's own strengths and weaknesses in an acontextual manner. From this perspective, to ensure effective self-regulation, self-assessment should not be conceived of as a general, personal "reflection-on-practice" leading to self-concepts of ability that in turn lead to a leisurely decision to learn more about a particular domain. Rather, safe practice requires that self-assessment be conceptualized as repeatedly enacted, situationally relevant assessments of self-efficacy and ongoing "reflection-in-practice," addressing emergent problems and continuously monitoring one's ability to effectively solve the current problem. As an analogy, we assume that most individuals do not read the dictionary out of the recognition that they need to improve their vocabulary. Rather, they look up specific words that they encounter when they are uncertain of the definition.

Conclusions

Our goals in undertaking this theory-oriented literature review were (a) to refine understanding of the key competency of self-assessment (and more broadly, self-directed learning) and (b) to begin an exploration of potential educational/evaluation methods to develop and assess the skills necessary for lifelong learning. If issues raised by the literatures we have explored are to be taken seriously, they cast doubt on the extent to which data published in the health professions literature on self-assessment provide evidence relevant to the ability of health professionals to function as self-regulating professionals. The flaws in the way that self-assessment has been conceptualized and operationalized in the current literature are sufficiently fundamental that scale tweaking and refinement of the criterion variables will not correct them. At the core of these flaws has been our communal presumption of the importance of personally generated summary judgments of overall performance: a concern about whether individuals are able to rate themselves relative to their peers, or to rate their own strengths and weaknesses relative to one another, or to accurately estimate the

percent of items correctly answered on a test. We now believe that placing the burden of personal self-regulation on this “personally generated summary judgment” form of self-assessment is inappropriate for two reasons. First, the literature from a variety of fields would suggest that our literature’s findings are, in fact, correct: people *cannot* effectively engage in these actions in any regular and stable way. Thus, it is time to recognize that, when trying to identify and redress gaps in learning, seeking and incorporating external evaluations will be a better model for effecting self-awareness than any form of personally generated summative assessment. Second, and to us more important, the focus on self-assessment as “summary judgments” fails to capture the context to which self-assessment is, in fact, critical to self-regulation: the context of reflection in practice. Self-assessment as a mechanism of ongoing monitoring *must* take precedence over self-assessment as a mechanism for identifying and redressing gaps.

Thus, the community’s research agenda pertaining to self-assessment needs to be reformulated, with the various literatures outlined here serving as a signpost for questions of more relevance. We must move beyond what has been described as, “guess your grade” studies of self-assessment.⁵⁵ From a “self-improvement” perspective, we should begin to focus upon questions such as what is the pedagogical/practice-based impact of engaging in formal self-assessment activities; to what extent do practitioners seek feedback from external sources; from which sources is feedback sought; to what extent does such feedback impact upon practice. And from a “situational monitoring” perspective we should begin to focus on what situational cues are most likely to prompt spontaneous reflection-in-practice; what teaching strategies/cultural changes are required to enable such situational factors to be influential; what impact do formal assessment requirements have on self-efficacy? Of course, this list is by no means comprehensive, and accurately addressing even the questions provided here will require methodological creativity. Still, such effort will be necessary if the thinking about self-assessment is to catch up with literatures in the domains of personality theory, cognitive psychology, social psychology, and expertise, that take for granted that

self-assessment is not a stable skill, but rather treat it as a situationally bounded cognitive process that is context specific and dependent upon expertise.⁵⁶ We think these literatures may provide insights into the types of data and/or methodologies that can be fruitful in further specifying the concept of self-regulation, and we would strongly encourage the field to take advantage of the insights they might provide rather than pursuing the methodologically weak and conceptually flawed approaches that now dominate self-assessment research in the health professions.

The Society of Directors of Research in Medical Education provided financial support for the preparation of this paper. Dr. Regehr is supported as the Richard and Elizabeth Currie Chair in Health Professions Education Research. Both authors are grateful to Karen Mann, Brian Hodges, David Stern, David Rogers, and anonymous reviewers for critical commentary that led to the improvement of this paper.

References

- 1 Arnold L, Willoughby TL. EV. Self-evaluation in undergraduate medical education: A longitudinal perspective. *J Med Educ.* 1985; 60:21–28.
- 2 Gordon MJ. A review of the validity and accuracy of self-assessments in health professions training. *Acad Med.* 1991;66:762–69.
- 3 Boud D. Avoiding the traps: Seeking good practice in the use of self assessment and reflection in professional courses. *Soc Work Educ.* 1999;18:121–32.
- 4 Ward M, Gruppen L, Regehr G. Research in self-assessment: Current state of the art. *Adv Health Sci Educ.* 2002;7:63–80.
- 5 Eva KW, Cunningham JPW, Reiter HI, Keane DR, Norman GR. How Can I Know What I Don’t Know? Poor Self-Assessment in a Well Defined Domain. *Adv Health Sci Educ.* 2004; 9:211–24.
- 6 Boud D. *Enhancing learning through self assessment.* London: Kogan Page, 1995.
- 7 Bandura A. *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall, 1986.
- 8 Shapiro DH, Jr., Schwartz CE, Astin JA. Controlling ourselves, controlling our world: Psychology’s role in understanding positive and negative consequences of seeking and gaining control. *Am Psychol.* 1996;51:1213–30.
- 9 Berry JM, West RL, Dennehey DM. Reliability and validity of the memory self-efficacy questionnaire. *Devel Psychol.* 1989; 25:701–13.
- 10 Cervone D. Randomization tests to determine significance levels for microanalytic congruences between self-efficacy and behavior. *Cognit Ther Res.* 1989;9:357–65.
- 11 Cervone D, Peake PK. Anchoring, efficacy, and action: The influence of judgmental heuristics on self-efficacy judgments and behavior. *J Person Soc Psych.* 1986;50:492–501.
- 12 Weinberg RS, Gould D, Jackson A. Expectations and performance: An empirical test of Bandura’s self-efficacy theory. *J Sport Psych.* 1979;1:320–31.
- 13 Cervone D. Thinking about self-efficacy. *Behav Modif.* 2000;24:30–56.
- 14 Bandura A. *Self-efficacy: The exercise of control.* San Francisco: Freeman, 1997.
- 15 Whittlesea BWA. Illusions of familiarity. *J Exper Psych: Learn, Mem Cognit.* 1993;19: 1235–53.
- 16 Bahrick HP. The importance of retrieval failures to long term retention: A metacognitive interpretation of the spacing effect. Presented at Metacognition: Theory and Application meeting, Barnard College, New York, 2004.
- 17 Kimball DR, Metcalfe J. Delaying judgments of learning affects memory, not metamemory. *Mem Cognit.* 2003;31:918–29.
- 18 Dunlosky J, Thiede KW. Causes and constraints of the shift-to-easier-materials effect in the control of study. *Mem Cognit:* In press.
- 19 Koriat A. Monitoring one’s own knowledge during study: A cue-utilization approach to judgments of learning. *J Exper Psych: Gen.* 1997;126:349–70.
- 20 Heath L, Tindale SR, Edwards J, Posavac EJ, Bryant FB, Henderson-King E, Suarez-Balcazar Y, Myers J Applications of heuristics and biases to social issues. New York: Plenum Press, 1994.
- 21 Bargh JA, Chartrand TL. The unbearable automaticity of being. *Am Psych.* 1999;54: 462–79.
- 22 Nisbett RE, Wilson TD. Telling more than we can know: Verbal reports on mental processes. *Psych Rev.* 1997;84:231–59.
- 23 Bargh JA, Gollwitzer PM, Lee-Chai AY, Barndollar K, Trötschel R. The automated will: Nonconscious activation and pursuit of behavioral goals. *J Person Soc Psych.* 2001;81: 1014–27.
- 24 Alicke MD, Klotz ML, Breitenbecher DL, Yurak PJ, Vredenburg DS. Personal contact, individuation, and the better-than-average effect. *J Person Soc Psych.* 1995;68:804–25.
- 25 Armor DA, Taylor SE. Situated optimism: Specific outcome expectancies and self-regulation. In MP Zanna, Ed. *Advances in experimental social psychology.* Vol. 30, San Diego: Academic, 1998: 309–379.
- 26 Gilbert DT, Wilson TD. Miswanting. In J Forgas, ed. *Thinking and feeling: The role of affect in social cognition.* Cambridge: Cambridge University Press, 2000: 178–197.
- 27 Gilovich T. *How we know what isn’t so: The fallibility of human reason in everyday life.* New York: The Free Press, 1991.
- 28 Wilson TD. *Strangers to ourselves: Discovering the adaptive unconscious.* Cambridge, MA: The Belknap Press of Harvard University Press, 2002.
- 29 Dawes RM, Faust D, Meehl PE. Clinical versus actuarial judgment. *Science.* 1989;243: 1668–74.

- 30 Ross L, Nisbett RE. *The person and the situation: Perspectives of social psychology*. New York: McGraw-Hill, Inc., 1991.
- 31 Eva KW. Assessing Tutorial-Based Assessment. *Adv Health Sci Educ*. 2001;6: 243–57.
- 32 Kolar DW, Funder DC, Colvin CR. Comparing the accuracy of personality judgments by the self and knowledgeable others. *J Person*. 1996;64:311–37.
- 33 LeDoux J. *The emotional brain: The mysterious underpinnings of emotional life*. New York: Simon and Shuster, 1996.
- 34 Custers EJFM, Regehr G, Norman GR. Mental representations of medical diagnostic knowledge: A review. *Acad Med*. 1996;71: S24–S26.
- 35 Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ*. 2005;39: 98–106.
- 36 Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med*. 2004;79(10 suppl): S70–S81.
- 37 Csikszentmihalyi M. If we are so rich, why aren't we happy? *Am Psych*. 1999;54:821–27.
- 38 Ericsson KA, Krampe RT, Tesch-Romer C. The role of deliberate practice in the acquisition of expert performance. *Psychological Review*. 1993;100:363–406.
- 39 Knowles M. *The adult learner: A neglected species*. Houston, TX: Gulf Publishing, 1973.
- 40 Norman GR. The adult learner: A mythical species. *Acad Med*. 1999;74:886–89.
- 41 Dolmans DH, Gijsselaers WH, Moust JH, de Grave WS, Wolphagen IH, van der Vleuten CP. Trends in research on the tutor in problem-based learning: Conclusions and implications for educational practice and research. *Med Teach*. 2002;24:173–80.
- 42 Sibley JC, Sackett DL, Neufeld V, Gerrard B, Rudnick KV, Fraser W. A randomized trial of continuing medical education. *N Engl J Med*. 2002;306:511–15.
- 43 Metcalfe J. Is study time allocated selectively to a region of proximal learning? *J Exp Psych*. 2002;131:349–63.
- 44 Eva KW, Neville AJ, Norman GR. Exploring the Etiology of Content Specificity: Factors Influencing Analogical Transfer in Problem Solving. *Acad Med*. 1998;73(10 suppl):S1–S5.
- 45 Nelson TO, Dunlosky J, Graf A, Narens L. Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psych Sci*. 1994;5:207–13.
- 46 Bjork RA. Memory and metamemory considerations in the training of human beings. In AP Ahimamura, J Metcalfe (eds). *Metacognition: Knowing about knowing*. Cambridge, MA: The MIT Press, 1994.
- 47 Schön D. *The Reflective Practitioner. How professionals think in action*. London: Temple Smith, 1983.
- 48 Eva KW, Dore K. Accurate self-assessment among test-takers when reflecting-in-practice. In preparation
- 49 Norman GR, Rosenthal D, Brooks LR, Allen SW, Muzzin LJ. The development of expertise in dermatology. *Arch Derm*. 1989;125:1063–8.
- 50 Regehr G, Hodges B, Tiberius R, Lofchy J. Measuring self-assessment skills: An innovative relative ranking model. *Acad Med*. 1996;71(10 suppl):S52–S54.
- 51 Gruppen LD, Garcia J, Grum CM, Fitzgerald JT, White CA, Dicken L, Sisson JC, Zweifler A. Medical students' self-assessment accuracy in communication skills. *Acad Med*. 1997;72: S57–S59.
- 52 Harrington J, Murnaghan J, Regehr G. Applying a relative ranking model to the self-assessment of extended performances. *Adv Health Sci Educ*. 1997;2:17–25.
- 53 Fitzgerald JT, Gruppen LD, White CB. The influence of task formats on the accuracy of medical students' self-assessments. *Acad Med*. 2000;75:737–41.
- 54 Reiter HI, Eva KW, Hatala R, Norman GR. Self and peer assessment in a problem based learning (PBL) curriculum: Application of a relative ranking model. *Acad Med*. 2002;77: 1134–9.
- 55 Colliver JA, Verhulst SJ, Barrows HS. Self-assessment in medical practice: A further concern about the conventional research paradigm. *Teach Learn Med*. 2005;17:200–1.
- 56 Kruger J, Dunning D. Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J Person Soc Psychol*. 1999;77:1121–34.