evidence for the gender-based stereotype is that recall of biographical details regarding a male (stereotype-consistent) victim is superior to recall of such details regarding a female heart attack victim (R. Martin et al, 1998, Study 4).

An important question is whether gender stereotypes about cardiac-related symptoms influence whether women experiencing the symptoms of an evolving heart attack seek treatment. R. Martin et al. (2004) explored this issue in a sample of heart attack survivors. Although the men and women patients were of comparable age and risk status and had similar medical histories and symptoms, the women were less likely than the men to attribute their prehospitalization symptoms to cardiac causes. Also, among the patients who received opinions and advice about their symptoms from friends and relatives, women were less likely than men to be told that their symptoms might indicate a heart attack and were also less likely to be advised to seek medical attention. In an attempt to correct these tendencies, the American Heart Association currently sponsors the "Go Red for Women" campaign, designed to raise consciousness about the prevalence and early signs of heart disease in women.

EDUCATION

Similar issues in self-assessment arise in the classroom. Students, obviously, profit from self-assessment that is accurate. Whether it be a junior high school student wondering whether to crack open his social studies textbook one more time before the test or a medical student deciding whether to practice her intubating technique before the next shift, students make more effective decisions about where to apply their learning efforts when they can accurately discern their strengths and weaknesses (Thiede, 1999; Thiede, Anderson, & Therriault, 2003). Accurate self-assessments allow students to become more autonomous agents in their education, taking responsibility for gaining and improving on their knowledge and skill (Boud, 1995; Dochy, Segers, & Sluijsmans, 1999; Topping, 1998).

Accurate self-assessment is valuable all the way up the educational ladder. However, it is especially crucial in higher education and professional school settings, particularly as some schools move to a problem-based or case-based model of instruction (Barrows & Tamblyn, 1980; Boud & Feletti, 1991). This approach to education has gained numerous adherents in medical schools, and its prevalence has recently increased in many other settings in higher education (e.g., Duch, Gron, & Allen, 2001). In problem-based learning, students working in groups learn by confronting real-world cases and are responsible for figuring out how to approach each case to bring it to a reasonable resolution. This approach places a premium on problem-solving skills, professionalism, and learning in hurly-burly circumstances that mimic real life.

An essential component of problem-based learning is that students must identify what skills they need to acquire and what knowledge they must gain—in short, they must make correct self-assessments of strengths and deficits (Boud, 1995). Accurate self-assessment is also crucial for education to be a lifelong enterprise that continues far after the student has left the classroom (Guest, Regehr, & Tiberius, 2001; Sambell & McDowell, 1997). Company executives must continue to educate themselves in a changing business environment; doctors must refresh and enhance their treatment skills; and airline pilots must continue to monitor their flying proficiency.

Against this background, it is sobering to see that although self-assessment of skill and knowledge does tend to bear some relationship to objective performance, the relationship tends to be meager to modest. In a review on academic performance, Hansford and Hattie (1982) discovered that the average correlation between academic self-views and actual performance lay around .21. In a more extensive review, Falchikov and Boud (1989) found that, on average, the grades that college students would give to their work correlated .39 with the marks their teachers would give. Students also showed signs of overinflated self-views, in that 68% of the time they gave themselves higher marks than their teachers would.

To be sure, some specific circumstances did lead to improved, albeit still imperfect, self-assessments. Self-assigned grades, for example, were slightly more related to teachers' evaluations when the solutions to course assignments became more well-defined (e.g., in science classes vs. other areas of study). In addition, the correlations between grades that students gave themselves and teachers' grades were higher in advanced classes than in introductory courses. Studies containing more rigor in their design revealed more student-teacher agreement than those of lesser quality (Falchikov & Boud, 1989).[6]

Three observations should be made about the fact that advanced students provided more calibrated self-assessments than introductory students. First, this increase in calibration was not simply due to students' age. Older students were no more calibrated than younger ones. Second, this increase was observed for students who were assessing their classroom performance, and there is suggestive evidence that as students leave the classroom to face more unsettled and demanding training settings, their self-assessments do not necessarily become more accurate. For example, Arnold, Willoughby, and Caulkins (1985) tracked self-assessments of medical students from their first year, spent primarily in the classroom, to their final year, spent mainly in clinical settings. The relationship between students' self-ratings and supervisors' evaluations fell as students progressed through their studies. In addition, during the final year, students' self-ratings failed to correlate at all with their board scores.

[6]An anonymous reviewer of this manuscript made the insightful point that student-generated grades were occasionally used as inputs into formal grades in some of the studies examining student-teacher agreement. To date, the impact of this circumstance on student-teacher agreement is unknown, but it would be interesting to examine whether attaching consequences to student-generated grades prompts flaws or accuracy in the evaluations students produce.

Third, although experience prompts students to provide more accurate self-assessments, these assessments are still far from perfect—and not all students profit from experience, even when that experience is repeated and provides clear-cut feedback. Hacker, Bol, Horgan, and Rakow (2000) tracked students as they took repeated exams in a course, asking them before each test how well they thought they would perform. Better students provided more accurate predictions as the semester wore on, but poorly performing students did not, remaining dramatically overconfident despite the feedback they had received on previous exams.

The modest accuracy of self-assessments of academic performance is also striking given data on peer assessment, which indicate that assessments by peers are more accurate than self-assessments. In an extensive meta-analysis, Falchikov and Goldfinch (2000) found that an evaluation from a single fellow student correlated .72 with the teacher's evaluation, and that grades given by peers on average tended to be no higher than those given by teachers. To be sure, one should compare these results and those for self-assessment with caution, in that this meta-analysis on peer assessment involved different studies and circumstances than the one on self-assessment (Falchikov & Boud, 1989); students may have taken assessments of their peers more seriously than they did self-assessments.

However, in studies that specifically focus on self- and peer assessments in the same class, peer assessments tend to correlate more highly with instructors' evaluations and objective performance measures than do self-assessments (e.g., Lennon, 1995; Sullivan, Hitchcock, & Dunnington, 1999). For example, among surgical residents, Risucci et al. (1989) found that peer assessments correlated more highly with supervisors' evaluations and performance on an objective test of surgical skills than did self-assessments. Indeed, self-assessments did not correlate whatsoever with performance on the objective test.

These results suggest that when it comes to self-assessments among students, there is a good deal of room for improvement. The issue of self-assessment in education is complex because imperfections in self-assessments are produced by many different problems, depending on the circumstances. However, recent evidence from the literatures on cognitive and educational psychology suggests two general themes that explain erroneous self-judgments. One theme has to do with common educational practices that may confer the appearance but not the reality of skill. The other has to do with people's ability to judge whether they comprehend what they have just read. Both the cognitive and the educational psychology literatures also provide hints of procedures that students and instructors might follow to improve self-assessment.

Educational Methods That Undermine Accurate Self-Assessment

The goal of education is to impart learning—to give students knowledge and skill that they remember and can call upon at some later date in circumstances that may differ from those of the classroom. Learning, thus, possesses two components. One is *retention*, the ability to recall information or perform a skill over the long term. The second is *transfer*, the ability to apply the knowledge or perform the skill across a number of relevant situations. Whether the topic is American history, playing the cello, or interpreting the nuances of tax law, instructors train students with the goal that those students will be able to retain and transfer the lessons learned to some unknown but relevant circumstance at some date in the future.

With this goal in mind, one common instructional method provides a paradox. The method that most effectively promotes the rapid acquisition of knowledge and the highest levels of proficiency at the end of the lesson—and thus the appearance of learning—is the one that ensures that whatever is learned in the classroom will be forgotten rapidly. The method of instruction that produces these effects is *massed training*, in which instructors train students in one or a few intense sessions. Massed training has advantages. Students undergoing intense training quickly obtain the relevant skill and then display it at a high level. Indeed, much research shows that massed training is more efficient than any alternative for bringing students quickly to a high level of performance (Dempster, 1990; Glenberg, 1979, 1992). In addition, other practices can be combined with massed training to further promote rapid knowledge gain and high performance. For example, students learn more quickly if instructors keep the conditions of learning constant and provide continuous feedback (Bjork, 1999), and if instructors model a solution or performance for students before they begin to generate their own (such as when a tennis instructor models a correct backhand stroke; Bjork, 1988; Jacoby, 1978).

There is only one central problem with this recipe. Although massed training produces quick learning and high performance in the short term, the knowledge and skill imparted tend to be forgotten rapidly—a result that has been known since the time of Ebbinghaus (1885/1964). In essence, although massed training is efficient for acquiring skills, it is not optimal for retaining them. Nowhere might this problem be more evident than in training people how to drive cars. Although millions of dollars are spent on formal driver-education courses, research data suggest that such courses do not produce safer drivers (Mayhew & Simpson, 1996). Indeed, evidence tenuously suggests that young people who complete formal driver-education courses experience more frequent accidents and injury than those who learn more informally from their friends and family (Skelly, 1968; Stock, Weaver, Ray, Brink, & Sadoff, 1983). Similarly, in Norway, truck drivers who were formally trained in skid control on slippery surfaces experienced more crashes, not fewer, than those not trained (Christensen & Glad, 1996), presumably because formal training did more to raise drivers' confidence than their competence (Wilde, 1998).

Retaining knowledge and skill requires a different recipe. Students retain information and skill better when they acquire

information and practice their skill via *spaced*, or *distributed*, training, in sessions that are divided over several occasions, even though this means that students initially learn more slowly and with more difficulty (Dempster, 1990; Glenberg, 1992). Retention is enhanced even further if instructors reverse other practices that are associated with rapid acquisition. Student retain information more successfully if instructors "change up" the circumstances under which material is learned, adding some variability and unpredictability to the presentation of material (e.g., Reder, Charney, & Morgan, 1986). For example, baseball batters learn to hit curveballs, fastballs, and sliders better if those pitches are thrown to them in a random sequence rather than in blocks (e.g., first fastballs, then curveballs, then sliders) during practice (Hall, Domingues, & Cavazos, 1994). Students also retain information better if instructors at times withhold feedback and avoid providing a model that students can imitate (Bjork, 1999).

But if distributed training is superior to massed training when it comes to retention, why does massed training continue to be popular? Part of the answer lies in a common illusion shared by instructors and students. People confuse speed and ease of learning with competence—and this confusion leads to errors in self-assessment. Students and instructors both assume that if a skill has been learned quickly and the student finds it easy to perform, then the student will maintain the skill in the long term (Bjork, 1994, 1999). Short-term excellence is mistaken for long-term competence.

Two examples show how massed training misleads self-assessment. First, Baddeley and Longman (1978) taught postal workers in Britain how to type so that they could use a new mail-sorting system. Workers learned to use the system under one of four different training schedules that ranged in how massed versus distributed they were. In the most massed schedule, employees learned to type in two 2-hour sessions per day for 20 days. In the most distributed schedule, employees trained for one 1-hour session per day for 60 days. Trainees were more satisfied with massed training than they were with distributed training, even though the most distributed schedule produced the best long-term retention of typing skill.

Second, D.A. Simon and Bjork (2001) asked college students to learn keystroke patterns either in a blocked format or in a random one that interweaved the patterns. Participants learned the keystroke patterns more efficiently with blocked presentation than with random presentation—and as a consequence were more optimistic about what their performance would be later on. However, when participants were tested the next day, participants trained on the random schedule outperformed those who had learned under a massed format.

Problems in Reading Comprehension

Much of learning involves picking up a textbook and reading it over, taking care to commit to memory important facts and principles presented in the text. Obviously, good study skills require knowing when one has comprehended the material in the text and can remember it later when tested, but research presents a cautionary tale about the ability of students to know when they have understood text material adequately and committed it to memory.

People are not very good at assessing their comprehension of written materials. They think they have understood a piece of text when they have not. Glenberg, Wilkinson, and Epstein (1982; see also Epstein, Glenberg, & Bradley, 1984) asked college students to read several passages and to rate their understanding of each. Various sentences in the passages directly contradicted one another, yet students failed to recognize this fact—leading them to express high but unwarranted confidence in their comprehension of the materials.

Other studies affirm this pattern. In a common experimental paradigm, students are asked to read a number of short texts and to predict how well they could answer questions about each. These predictions are then compared with actual performance on quizzes probing comprehension of and memory for the material. Researchers then correlate each participant's predictions and performance. In 1998, Maki identified 25 studies that had used this paradigm; the average correlation between prediction and performance was .27—statistically significant, but hardly strong. Furthermore, poor readers and lackluster students have an especially difficult time knowing when they have understood a text (Bol & Hacker, 2001; Maki & Berry, 1984; Maki, Jonas, & Kallod, 1994; Moreland, Miller, & Laucka, 1981; Shaughnessy, 1979; Sinkavich, 1995; although see Glenberg & Epstein, 1987, for an exception).

Blind spots in assessing comprehension create problems because students regulate their study habits on the basis of what they think they have (or have not) understood and committed to memory (for a review, see Son & Metcalfe, 2000). If students think they have successfully absorbed what is in chapter 3 of the textbook, they will put it aside to stare down chapter 4 (the one that confused them last time they read it). Students who know what they understand study more effectively than those who misjudge their comprehension, and they perform better on exams (Dufresne & Kobasigawa, 1988, 1989; Thiede, 1999). For example, Dufresne and Kobasigawa (1989) asked children in first, third, fifth, and seventh grade to study booklets filled with paired-associate tasks (i.e., the students had to memorize a target word that was paired with each cue word), and they were supposed to study the booklets until they could remember all the word pairs perfectly. Some of the booklets were designed to be easy (i.e., the words in each pair were related to one another), and some were difficult (e.g., the words were not at all related). Fifth and seventh graders studied the difficult booklets more than the easy ones, but their younger counterparts studied the two kinds of booklets equally. When tested later, older children outperformed their younger peers—more often achieving perfect scores—exactly because they had spent their study time more prudently on the harder booklets.

Improving Self-Assessment—and Thus Enhancing Learning

In sum, extant research suggests that students face two general problems when it comes to assessing their own skills. One is that massed training may mislead them into thinking they have acquired a skill that they will retain. The other is that it is difficult for them to assess accurately whether they have understood and can remember what they have read. More recent research, however, has begun to suggest some practices that instructors or students can use to improve self-assessment, as well as actual academic achievement.

Introducing "Desirable Difficulties" to Instruction

Massed training, along with associated practices, produces rapid learning that is difficult to retain, leaving students potentially overconfident in their knowledge and ability. One simple way to alleviate this problem is to introduce "desirable difficulties" that harm the speed with which students learn but that leave them better able to retain what they have learned and to transfer it to different situations in the future. Bjork (1994, 1999) discussed what "obstacles" to place in front of students to enhance long-term proficiency. These obstacles include spreading training over several sessions, varying the circumstances of the training, reducing feedback, and providing "contextual interference" (i.e., practicing different subskills under a random schedule rather than under a blocked one).

Forgoing massed training and introducing such desirable difficulties, however, might be somewhat difficult for an educational institution to do (Bjork, 1994, 1999). The rapid improvement and high short-term performance that massed training provides look impressive to instructors and their administrators. When formally evaluated against alternatives, massed training wins out on many desirable criteria—performance is high, time is used effectively, and students express high levels of satisfaction. Instructors and administrators fail to recognize the long-term shortcomings of massed training because instructional programs based on this method tend not to be evaluated over the long term. Providing long-term evaluations of retention and transfer would go a long way toward identifying when massed training provides more the illusion of competence than the fact (Bjork, 1994, 1999).

Long-term evaluation of retention would also go a long way toward identifying potentially important instances when enduring learning is not possible and therefore skill training should not be attempted. For example, some scholars have proposed that some forms of driver's education might do more harm than good, making drivers more confident about their ability to handle difficult driving situations than is warranted. As a substitute, they have proposed that drivers forgo skills training and instead undergo "overconfidence training," in which students are shown just how difficult it is to handle adverse driving conditions so that they drive more carefully—or just stay home when it is wise to do so (Wilde, 1998).

Improving Reading Comprehension via Self-Testing

Students can also take several steps on their own to improve their self-assessment of their reading. One is to test their understanding as they read text, rating what they believe their understanding to be before they test that understanding. For example, Koch (2001) asked physics students to stop periodically what they were reading to rate how well they thought they had understood material they had just read. The students then answered a few questions about what they had read and noted any discrepancies between their ratings and actual performance. Students who went through this exercise posted higher scores on a subsequent physics test than students who did not engage in this form of self-questioning (see also Walczyk & Hall, 1989, for similar results). However, self-testing can sometimes be hazardous to one's academic health if the self-tests do not accurately represent the subsequent performance situations. Bol and Hacker (2001) had students take a practice test before taking an actual exam a few days later. Relative to a control group, students who had taken the practice test performed worse, presumably because the practice test was not representative of the material covered on the actual exam.

Self-testing is particularly valuable if students and instructors take care to place a delay between study and the self-testing exercise. Self-testing done immediately after study tends to lead to inaccurate self-assessments. Why might this be so? Students tend to make assumptions about their memory on the basis of the accessibility or fluency of material (Benjamin, Bjork, & Schwartz, 1998) even when that fluency is not diagnostic of long-term memory. Just after studying, material is quite accessible and fluent, and people are not in a good position to predict what material they might forget. However, after a delay, students need not lean on their intuitions about memory. Instead, they have actual data about the information they have retained and what they have forgotten.

Several studies show that self-assessments are more accurate if collected after a delay rather than right after study. For example, Thiede and Dunlosky (1994) asked college students to remember the translations of several Swahili words. The students were presented with the words again either immediately after studying them or after approximately a 30-second delay and asked whether they would recall the translations later in the experimental session. Self-assessments provided after the 30-second delay more accurately predicted subsequent recall than did self-assessments collected immediately (see Dunlosky & Nelson, 1992; Kelemen, 2000; Nelson & Dunlosky, 1991), a result that has been found to be just as true of kindergartners as it is of college students (Schneider, Vise, Lockl, & Nelson, 2000). Indeed, the more the delay the better (Kelemen & Weaver, 1997).

The fact that judgments of learning are more accurate after a delay should inform other interventions for improving the accuracy of self-assessment. Thiede and Anderson (2003) discovered that asking students to summarize a passage they had recently read led to more accurate self-assessments of learning—but only

if students delayed a short while before summarizing the text. Similarly, Thiede, Anderson, and Therriault (2003) found that simply asking students to write down five keywords from material they had studied improved self-assessments of learning, but again only if there was a delay between study and the keyword exercise. Moreover, among students allowed to study the material a second time before the self-assessment, those who had undergone the delayed keyword exercise outperformed their peers who did not do the exercise or had done it right after their initial study session.

Reviewing Past Performance

Reviewing one's performance has been shown to lead to better self-assessment of a variety of skills. For example, medical students reach a more appropriate appreciation of their interviewing skills if they watch videotapes of their performance (Ward et al., 2003). This videotape exercise can be enhanced when students review the videotapes with faculty (Lane & Gottlieb, 2004; Scherer, Chang, Meredith, & Battistella, 2003). The review of past work need not necessarily involve videotape. Cochrane and Spears (1980) had students taking a clinical dietitian course periodically rate their own performance and then meet with a faculty member who had also rated their performance. Over time, students' self-assessments came to correspond more closely with those of their supervisors.

Benchmarking

Self-assessments may also be improved when people take pains to compare their choices and performance against those of others, a practice known as benchmarking. For example, Farh and Dobbins (1989) asked college students to play the role of editor, correcting an error-strewn report that someone else had putatively written. After completing the task, students rated their performance. Some students, however, were shown other students' efforts before being asked to provide self-assessments. The self-ratings of these students correlated more strongly with objective markers of performance than did the self-ratings of the students who had no benchmarking opportunity.

Benchmarking has also been found to improve the self-ratings of medical students. In one study, family-practice residents completed a standard exercise in which they interviewed a mother who might have physically abused her child and subsequently rated how well they thought they had performed along several dimensions. They then watched their own videotaped interview, along with four benchmark interviews that displayed a wide range of competence in interviewing technique (D. Martin, Regehr, Hodges, & Mcnaughton, 1998). Correlations between self-ratings of performance and supervisors' ratings were higher after students had viewed these benchmark interviews than before.

An important caveat regarding benchmarking is that it may improve the self-ratings of high-performing students, permitting them to see just how special their skills are, but not the self-assessments of poor-performing students. Kruger and Dunning (1999, Study 3) suggested this when they asked students with excellent grammar skills and those with poor grammar skills to review grammar tests that had been filled out by other students. After this benchmarking exercise, top performers appropriately raised their estimates of how distinctive their grammar skills were relative to those of other students, but poor performers did not revise their self-estimates in any meaningful way. Echoing this result, Hodges et al. (2001) reanalyzed previous data on medical residents and discovered that benchmarking exercises helped top performers to recognize the excellence of their interviewing skills—in that their self-ratings after benchmarking more closely corresponded to their instructors' ratings—but did not prompt poor performers to recognize deficits in their interviewing skills. In short, benchmarking might help students in general gain insight into their skill, but it does not help the poor performers who most need to adjust their self-assessments.

Peer Assessment

Recent work also suggests that students benefit from peer assessment, presumably because such feedback informs students of shortcomings they do not necessarily know they have. Academic feedback from peers can take on many forms, from informal criticism to formal grades. Studies of peer assessments have examined a wide variety of academic tasks, although work in this area is still in its infancy (Topping, 1998). Researchers have looked at peer assessments for writing, oral presentations, group projects, and professional skills. The research results suggest that peer assessments are generally reliable and correlate highly with teachers' evaluations (Falchikov & Goldfinch, 2000; Topping 1998). Students whose work is evaluated by their peers tend to achieve higher academic marks than those who do not go through peer-assessment exercises. Students also tend to possess more positive attitudes toward the learning experience after peer assessments exercises (Topping, 1998).

The success of peer assessment in raising achievement should perhaps not come as a surprise. Peer assessment might improve students' accomplishment via many different mechanisms. Beyond improving students' imperfect evaluations of their own work, peer assessment creates more "time on task," allows students to practice their skills, engages them more fully in their assignments, prompts them to reflective more on what they have done, and assigns them more responsibility and accountability (Topping, 1998). Important side benefits may include enhanced social skills, in that students learn to give and receive feedback (Marcoulides & Simkin, 1991).

Peer assessment may prove especially valuable in cases in which structured and formal education is neither preferred nor even available. Such a case arises in schools that use a problem-based learning model, which leaves students more on their own during their studies. Perhaps a more important case arises when people leave formal schooling, yet still must engage in continuing or lifelong learning. In both these circumstances, students must decide for themselves whether they need further education

(Topping, 1998). In the medical realm, peer assessment, coupled with self-assessment exercises, may prove a significant help with health care training in developing countries where formal supervision is expensive to maintain. In the late 1990s, the Indonesian government, in concert with several international organizations, conducted a training program with local health-clinic counselors to improve their communication skills when discussing contraception with clients. Counselors attended a 5-day refresher seminar and then were assigned to one of three groups. For the next 16 weeks, one group completed self-assessment exercises on their communication skills. The second group completed self-assessment exercises, plus attended weekly peer-review sessions. The third group served as a control and did not receive any intervention.

Six months after training, researchers taped and coded two interviews involving each counselor, to assess the extent to which each counselor provided medical information and exhibited a facilitative and active communication style. Both interventions increased facilitative communication—with the intervention involving peer assessment producing more improvement in how much information (e.g., about lifestyle consequences) clients were given. As a consequence, clients talked more actively with counselors who had received the intervention that combined self- and peer assessment (Kim, Putjuk, Kols, & Basuki, 2000). The costs associated with these self- and peer-assessment interventions turned out to be low given the effects these interventions brought about.

Peer assessment, however, is not without difficulties. It can be demanding and time-intensive. It can provoke anxiety (Topping, 1998), as well as fears that grades will be more a product of illegitimate influences (such as friendship) than of objective attainment (Dancer & Dancer, 1992). Poor students do not provide as accurate assessments as their more skilled peers do (Saavedra & Kwun, 1993). Thus, using peer assessment is not guaranteed to improve student performance unless the assessments are conducted carefully. Peer assessments become more valid as they are based on a larger number of observations and a greater number of dimensions of skill. They are also most helpful when standards are clear and more than one peer provides an assessment. Peer-assessment exercises are also enhanced if instructors communicate the purpose of the exercises clearly, articulate the dimensions of judgment clearly, provide training when necessary, and monitor students' evaluations, intervening when they are too harsh or too lenient (Norcini, 2003). Peer assessments better match teachers' evaluations when students provide global evaluations based on a number of well-defined criteria rather than a number of separate evaluations along several individual dimensions (Falchikov & Goldfinch, 2000).

THE WORKPLACE

Given the research reviewed earlier in this monograph, it is not surprising that biased self-views also affect success and social relationships in the workplace. When one looks at the accuracy of self-assessment in the workplace, from the office cubicle to the executive boardroom, one sees that people tend to hold overly inflated self-views that are modestly related to actual performance. One also sees reverberations of other themes (overconfidence in judgment, egocentric neglect) that are apparent in the domains of health and education.

If people often have biased self-views, it is not surprising that these biases affect workplace outcomes ranging from the ability to anticipate performance evaluations to the ability to gauge promotion opportunities. What is perhaps less apparent is that organizations can adopt many routines and procedures that keep biased self-views from affecting performance in many situations. As a result, biased self-views may have their most systematic and damaging effects in the circumstances in which individuals are least likely to face routines and procedures that could repair their biased decisions, for example, when people confront new projects that have few precedents, or when CEOs and management teams at the very tops of organizations, where there are few routine correctives, make important, large-stakes decisions about which new markets to enter and which companies to acquire.

Accuracy of Self-Knowledge in Organizations

In organizational life, the largest surprises generated by lack of self-knowledge may be those that are produced when self-evaluations are not echoed by supervisors who set raises and hand out promotions. If employees overrate their own performance, it is difficult to imagine how people could not wind up disappointed at least some of the time. For example, Zenger (1992) studied several hundred engineers at two high-tech companies and found that 32% of the engineers in one company and 42% in the other rated their own performance in the top 5% of all engineers. Imagine the difficulty of conducting honest performance evaluations for these engineers.

Although workers may find it tempting to blame their supervisors when they receive a less-than-stellar evaluation, research suggests that they should trust their supervisors' views more than their own. Typically, the views of other people—subordinates, peers, and superiors—agree with each other more often than with self- views. In one typical study, Bass and Yammarino (1991) studied U.S. Navy officers who had graduated from the U.S. Naval Academy and were on active duty on ships ranging from tenders to aircraft carriers. The researchers collected leadership ratings from the officers' subordinates and the officers themselves. They also retrieved some measures of leadership from the officers' superiors—each officer's performance on regular "fitness reports" and promotions relative to peers. On all seven positive leadership dimensions studied (e.g., charisma, individual consideration, intellectual stimulation), the officers rated themselves more positively than did their subordinates. But although the officers' self-ratings displayed almost no correlation with