

1 MYTH: STUDENTS ARE ACCURATE JUDGES OF HOW MUCH THEY KNOW

Most teachers have probably had the experience of asking students whether they have any questions on a particular topic and receiving confirmation from the students that they understand the material, only to learn from later exam results that this was not the case. Sometimes students may be too shy or anxious to speak up, but often they genuinely believe that they know more than they do. Students often express a great deal of confidence in the degree to which they have learned something (e.g., Shaughnessy, 1979; Sinkavich, 1995). However, students' evaluations of their own learning can be extraordinarily inaccurate. Bjork, Dunlosky, and Kornell (2013) assert that students' overconfidence arises because they misinterpret information about their learning and have inaccurate views about what learning strategies are most effective. It is therefore possible for students to be confident that they know something without actually knowing it. One team of researchers even found that students' predictions regarding how well they would remember information they had studied were negatively correlated with their actual memory (Benjamin, Bjork, & Schwartz, 1998). That is, students had poorer memory for information they were more confident they would remember than for information about which they were less confident. Students' ability to accurately assess their own knowledge has enormous implications for their capacity to select appropriate study strategies, effectively allocate their study time, and know when they have reached an appropriate level of mastery (Nelson & Dunlosky, 1991; Bjork et al., 2013).

Researchers have used two types of studies to test the accuracy of students' estimates of their own knowledge pertaining to academic

information. In some studies, students judge their own performance relative to a given standard by estimating how well they did on an exam or how many items they answered correctly. In other studies, students judge their knowledge or performance relative to other students. As demonstrated by the research results reported below based on both types of studies, students' judgments of their own learning are often quite inconsistent with objective measures of that learning. However, the accuracy of self-judgments of learning is not consistent across students. Specifically, high-performing students are much more accurate than low-performing students in judging their own knowledge. Moreover, high-performers tend to underestimate their own performance, whereas low-performing students tend to exhibit overconfidence in their performance.

In one illustrative study (Langendyk, 2006), advanced medical students in Australia completed an assignment requiring them to make a complex diagnostic assessment. The assignments were then evaluated according to specific criteria by the students themselves, by student peers, and by faculty. Low-achieving students tended to give themselves and their peers higher ratings than those provided by faculty, but high-achieving students gave themselves lower ratings than those provided by faculty. According to Langendyk, students who were low achievers with respect to the assignment were simply "unable to assess accurately the quality of their own work" (p. 173). Because the students in this study were advanced medical students, most of them performed adequately in an absolute sense; however, the study shows that even academically advanced graduate students do not always have insight into their own performance and are sometimes unable to distinguish high-quality from low-quality work. The low-achieving students were unable to accurately judge the quality of their own performance or the performance of higher-achieving peers.

The tendency for lower academic performers to have difficulty judging the quality of their own performance has more frequently been the subject of research involving undergraduate students. Shaughnessy (1979) studied introductory psychology students as they completed four multiple-choice exams over the course of a semester. As students responded to each exam item, they also rated their degree of confidence that their answer was correct. For the first three exams, students later studied their answers and their confidence judgments; therefore, they received feedback both on their test performance and the accuracy of their judgments. Shaughnessy reported that students' self-judgment accuracy was positively correlated with test performance. That is, students who knew more information were much more capable of evaluating how much they knew.

Similarly, Sinkavich (1995) assessed students' confidence in their responses on multiple choice exams. Students rated their confidence in their responses and later received individualized feedback, compared their feedback with that of other students, and received encouragement to try to improve their ability to identify what they did and did not know. Consistent with earlier findings, and despite repeated individualized feedback, students who did well on the exams (those in the top third of the class in terms of exam score) judged their level of performance much more accurately than did poor performers (those in the bottom third of the class). In a more recent study (Ehrlinger, Johnson, Banner, Dunning, & Kruger, 2008), college students completed a difficult exam in class and then rated their performance immediately afterward. Students in the bottom quartile in terms of exam performance rated their performance at the 61st percentile, and their estimates of their own raw scores were inflated by an average of 20%. In contrast, those in the top quartile were more accurate, but tended to underestimate their performance both in terms of test score and standing relative to other students.

In a more complex classroom study (Hacker, Bol, Horgan, & Rakow, 2000), researchers again had undergraduates estimate their exam performance – this time both before and after taking exams. Immediately prior to taking an exam, students estimated the proportion of items they expected to get correct. Immediately following the exam, students reported the proportion of items they believed they had answered correctly. This procedure was repeated twice as the semester progressed. Throughout the course, the instructor emphasized the importance of accurate self-assessment and provided instruction on how to accomplish it. The week before each exam, students also completed practice tests on which they received feedback. The researchers replicated the results of other studies and provided even greater detail: students earning As and Bs were most accurate in their judgments; students earning Cs and Ds were highly overconfident in their predictions before the exam, but were much more accurate in their self-judgments after they had completed exams; and students whose exam scores were below 50% were grossly overconfident in their self-judgments both before and after taking the exams. Students in this lowest-performing category overestimated their actual exam performance by as much as 31 percentage points, and the lower their exam scores, the greater their overconfidence.

Laboratory studies of student self-knowledge provide additional insight into the findings from classroom research cited above. Kruger and Dunning's (1999) research allowed them to evaluate student self-knowledge in a more controlled environment than that of a conventional

classroom. In one of their studies, college students completed a logical reasoning test. The students then estimated the number of items they had answered correctly and reported how they believed they had performed relative to other students. Similar to classroom studies, students in the bottom quartile of test performance greatly overestimated their performance on the test itself as well as their performance relative to others. Not only did these low-performing students overestimate their performance, they also estimated their performance as above average: on average rating their performance at the 62nd percentile when it was actually at the 11th. Again mirroring classroom studies, students in the top quartile were more accurate and tended to underestimate their performance. Kruger and Dunning reported similar findings with respect to grammatical skills. Students in the bottom quartile of performance on a grammar test grossly overestimated their performance – rating themselves at the 61st percentile when their performance fell at the 10th percentile. Students in the second and third quartiles also overestimated their performance, but were more accurate than the lowest-performing students. Only students in the top quartile were accurate in their estimates of their absolute test performance, but, again, they tended to underestimate their performance relative to other students.

It is interesting to note that judgments of students' own knowledge and performance – particularly among the majority of students whose performance is at or below the level that would earn them a B according to conventional grading standards – tend to be quite inaccurate whether the students predict their performance before or after taking an exam. Kruger and Dunning (1999) explained the inaccuracy of self-judgments, in particular those made by low performers, by asserting that “incompetence ... not only causes poor performance but also the inability to recognize that one's performance is poor” (p. 1130). To illustrate, they cited the ability to write grammatically correct sentences which, they observed, requires the same skills necessary to recognize grammatical errors. In other words, someone who is incapable of good writing will be unable to recognize and correct bad writing. Dunning and his colleagues referred to this as a “double curse” because “in many intellectual and social domains, the skills needed to produce correct responses are virtually identical to those needed to evaluate the accuracy of one's responses” (Dunning, Johnson, Ehrlinger, & Kruger, 2003: 84–85). Skill or knowledge deficits prevent students from knowing whether their answers are correct, and also from recognizing that other students' performance is superior.

High-performing students sometimes misjudge their own performance, but to a lesser degree. Moreover, high performers tend to underestimate

their performance – at least relative to that of other students. Dunning (2005) explained that strong students underestimate the uniqueness of their performance. Because they are more knowledgeable, they are better able to accurately evaluate the quality of their work. Therefore their self-evaluations tend to be more accurate than those of low-performing students with respect to the proportion of test items answered correctly. Because they are more knowledgeable, they are likewise better at recognizing when they do not know something. However, strong students often make the false assumption that because they know something, most other students must know it as well. This leads them to overestimate the performance of other students (Ehrlinger et al., 2008).

Yet another factor contributing to students' difficulty in making accurate judgments of their own knowledge is hindsight bias: the tendency to assume once something happens that one knew all along that it was going to happen (Fischhoff, 1975; see also Hawkins & Hastie, 1990, for a review). When students receive feedback suggesting that their knowledge is incomplete, such as getting an exam item incorrect, they may respond by telling themselves that they actually did know the information. Although they do not have a strong grasp of the material, they feel as if they do because they recognize something about the item content. Looking back, once they know the answer, the solution seems obvious. This feeling of familiarity can lead students to have an exaggerated sense of what they know. Hindsight bias therefore reinforces the feeling that their failure was due to the nature of the assessment rather than the nature of their knowledge – which makes it more difficult for them to learn from feedback.

Koriat and Bjork (2005) postulated a contrasting phenomenon that they termed foresight bias, which leads people to overestimate how well they will recall information when they predict their future performance at a time when the information to be learned is available to them. That is, people fail to account for the fact that the memory cues available to them while studying will not be available when they are asked to recall the information. The relevance to academic performance is clear, in that students often judge their own learning and make decisions about additional studying at times when they have the relevant academic material available to them. Bjork and colleagues (2013) similarly explained that learners often mistake their sense of fluency regarding information to be learned as evidence of actual learning. When information seems easy to learn, or seems to come to mind easily in the presence of specific memory cues, students believe that they genuinely understand the information even when they do not.

Ehrlinger (2008) pointed out that one's motivation also plays a pivotal role in the accuracy of one's self-judgments. She noted that people will be motivated to recognize the limits of their knowledge only if their primary objective is to increase that knowledge. If, instead, one's primary goal is to see oneself in a positive light, the person will tend to avoid or distort feedback that suggests a lack of knowledge. Ehrlinger suggests that people motivated primarily by a desire to maintain a positive self-image will have difficulty acknowledging and learning from feedback indicating that they are not doing well. This observation is consistent with the finding that despite repeated testing and ongoing feedback and reflection on their performance, students tend to base their self-assessments on their beliefs and expectations about themselves, rather than on their past performance (Hacker et al., 2000).

There is mixed evidence concerning the extent to which students can improve the accuracy of their self-evaluations. As cited earlier in this chapter, Kruger and Dunning (1999) gave students a test of grammar and had the students rate their own performance. Several weeks later, the researchers invited participants who had scored in the top and bottom quartiles on the grammar test to return to the lab to grade tests completed by five other participants, and then to rate their own performance once again. Students in the top quartile became more accurate in their self-judgments after seeing the work of other students. Those in the bottom quartile failed to gain insight into their poor performance even after seeing the work of stronger students. Hacker and colleagues (2000) likewise found that although high- and low-performing students were inaccurate in their self-judgments at the start of a course, the high-performing students became much more accurate over time while the low performers showed no improvement in accuracy. Kruger and Dunning found that it might be possible to train students to judge their work more accurately. The catch is that the way to do this is simply to help them improve their skills on the relevant task. That is, students rated their skills more accurately as their skills increased. Nonetheless, they still overestimated their performance relative to other students.

A slightly different pattern of results emerged in another classroom study. Miller and Geraci (2011) noted that improving student metacognition (i.e., knowing what they know) is more challenging in the classroom than in the laboratory. These researchers had students predict their own exam performance immediately prior to completing each of four exams. High scorers were again more accurate than low scorers, and accuracy did not improve over time despite the incentive of extra credit for making accurate predictions. In a second study, Miller and Geraci

provided students with more explicit feedback on the accuracy of their self-judgments. This time, low-performing students demonstrated some increase in accuracy over time, but appeared to reach an accuracy ceiling. The researchers speculated that there may be a limit to how much low-performing students can improve their self-evaluations. More importantly, however, the increase in accuracy did not lead to an improvement in exam performance. Low-scoring students improved their accuracy by lowering their predicted scores, rather than by improving their test scores.

Other researchers have similarly investigated whether students can improve the accuracy of their self-judgments if provided with adequate incentives. As noted above, Miller and Geraci (2011) found that offering extra credit for accurate predictions did not lead to increased accuracy. In a more complex test of the effects of incentives (Hacker, Bol, & Bahbahani, 2008), researchers again found that offering points for accuracy had no overall effect on judgment accuracy. However, the researchers qualified this conclusion because high-performing students were accurate throughout the course so a ceiling effect would have prevented significant improvement. In contrast, low performers were less accurate, but improved slightly in their ability to judge their performance after taking an exam. Unfortunately, there was no such improvement in their ability to predict their performance before the exam, which is arguably more important because it is this factor that would help them to determine whether they were sufficiently prepared.

The findings from laboratory studies on the effectiveness of incentives for increasing the accuracy of student self-judgments parallel the findings from classroom research. Ehrlinger and colleagues (2008) tested the impact of a particularly strong incentive. The researchers had students complete a 20-item multiple-choice test of logical reasoning ability and then predict the number of items they answered correctly. Students were offered \$100 if their predicted score exactly matched their actual score, and \$25 if their predicted scores were within 5% of their actual scores. Consistent with other studies, low scorers overestimated and high scorers underestimated their own performance. The large monetary incentive had no effect on the accuracy of self-judgments. The researchers reported similar results with respect to social incentives. Students taking a test of logical reasoning ability who were told that they would be interviewed by a professor regarding their rationale for their responses on the test were no more accurate in judging their performance than students with no such incentive.

Aside from finding ways to help students learn more, which is always a priority in education, the best hope for helping students to become better

judges of their own knowledge and performance is to have them engage in frequent and repeated self-assessment. Lopez and Kossack (2007) conducted a study in which students in one class did not engage in self-assessment of their knowledge, students in a second class self-assessed on the first and last days of class, and students in a third class self-assessed on the first day of class and also following each of four exams. Only students who self-assessed after each exam became more accurate in their judgments by the end of the course. The researchers concluded that students can improve their ability to gauge their own knowledge if they do so repeatedly and systematically. Unfortunately, the findings do not permit a comparison of students at various performance levels, which is an important consideration given research (e.g., Kruger & Dunning, 1999; Miller & Geraci, 2011) suggesting that interventions to improve students' self-assessment accuracy have only modest effects on low-performing students whose self-ratings tend to be the least accurate.

Accurate self-evaluation can play an important role in student learning. As Shaughnessy (1979) points out, students who cannot judge their knowledge accurately are likely to study less efficiently and effectively: spending too much time reviewing familiar content and failing to recognize and review content they do not know well. Existing research suggests that students often misjudge their level of understanding and performance with respect to academic material, and that low-performing students are at particular risk because they tend to grossly overestimate their performance. This pattern creates a paradox, in that low-performing students might improve their self-evaluation skills if they more effectively mastered course content, but their deficient self-evaluation skills make learning more difficult. Current research suggests that, although the overall effects may be small, the most promising strategy for improving the accuracy of students' self-judgments is to have them engage in ongoing, systematic self-evaluation and to provide them with feedback on the accuracy of their evaluations. Moreover, students should not assess their knowledge immediately after studying because self-evaluations become more accurate after a short delay (Dunlosky & Nelson, 1992). With continued practice and feedback, students may learn to be better judges of what they know.

References

- Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, 127, 55–68.

- Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual review of Psychology*, 64, 417–444.
- Dunlosky, J. & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition*, 20, 374–380.
- Dunning, D. (2005). *Self-insight: Roadblocks and detours on the path to knowing thyself*. New York: Psychology Press. doi: 10.4324/9780203337998.
- Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, 12(3), 83–87. doi: 10.1111/1467-8721.01235.
- Ehrlinger, J. (2008). Skill level, self-views, and self-theories as sources of error in self-assessment. *Social And Personality Psychology Compass*, 2(1), 382–398. doi: 10.1111/j.1751-9004.2007.00047.x.
- Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes*, 105(1), 98–121. doi:10.1016/j.obhdp.2007.05.002.
- Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 288–299.
- Hacker, D. J., Bol, L., & Bahbahani, K. (2008). Explaining calibration accuracy in classroom contexts: The effects of incentives, reflection, and explanatory style. *Metacognition and Learning*, 3(2), 101–121. doi: 10.1007/s11409-008-9021-5.
- Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology*, 92(1), 160–170. doi:10.1037/0022-0663.92.1.160.
- Hawkins, S. A. & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin*, 107(3), 311–327. doi: 10.1037/0033-2909.107.3.311.
- Koriat, A. & Bjork, R. A. (2005). Illusions of competence in monitoring one's own knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 187–194.
- Kruger, J. & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134. doi: 10.1037/0022-3514.77.6.1121.
- Langendyk, V. (2006). Not knowing that they do not know: Self-assessment accuracy of third-year medical students. *Medical Education*, 40, 173–179.
- Lopez, R. & Kossack, S. (2007). Effects of recurring use of self-assessment in university courses. *International Journal of Learning*, 14, 203–214.
- Miller, T. M. & Geraci, L. (2011). Training metacognition in the classroom: The influence of incentives and feedback on exam predictions. *Metacognition and Learning*, 6(3), 303–314. doi: 10.1007/s11409-011-9083-7.

- Nelson, T. O. & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "Delayed-JOL Effect." *Psychological Science*, 2, 267–270.
- Shaughnessy, J. J. (1979). Confidence–judgment accuracy as a predictor of test performance. *Journal of Research in Personality*, 13(4), 505–514. doi: 10.1016/0092-6566(79)90012-6.
- Sinkavich, F. J. (1995). Performance and metamemory: Do students know what they don't know? *Journal Of Instructional Psychology*, 22(1), 77–87.