Difficulties with Self-Assessment in Computer Programming

Rowland Pitts
George Mason University, Information Technology
rpitts@gmu.edu

Abstract

Research indicates that people tend to incorrectly assess their own abilities in a variety of social and intellectual domains, and that the lower one's level of competence is, the larger the error in self-assessment tends to be. This paper presents the results of a preliminary study that indicates the same phenomena hold true in the domain of computer programming.

Introduction

People tend to incorrectly assess their own abilities in a variety of social and intellectual domains. Kruger and Dunning predicted that less competent individuals "will dramatically overestimate their ability and performance relative to their competent peers," and conducted a series of tests to assess the "metacognitive skills of the incompetent to explain, in part, the fact that people seem to be so imperfect in appraising themselves and their abilities" (Kruger and Dunning 1999).

Kruger and Dunning conducted tests in multiple social and intellectual domains. Figure 1 depicts the results of a logic test. Participants were divided into quartiles based on their actual performance, and compared to their perceived ability and perceived performance levels. Although the test domain was logic, it is worth noting that the resulting observations were consistent across all test domains.

Their findings, in part, included that less competent performers tend to overestimate their perceived ability as well as their actual level of performance, and that the lower the level of competence, the greater the errors tend to be. The authors point out that "perhaps the best illustration of this tendency is the 'above-average effect,' … the tendency of the average person to believe he or she is above average, a result the defies the logic of descriptive statistics." Additionally, the most competent performers tended to underestimate their ability, a phenomenon that Kruger and Dunning attribute to a *false consensus effect* (Ross, Green and House 1977).

This paper presents the results of conducting an analogous study in the domain of computer programming.

Method

Participants

The experiment was conducted in a second-semester, undergraduate computer programming class. The class started out with thirty students, twenty-four of whom remained at the end of the semester, and twenty of whom volunteered to participate in the study. Participation was anonymous, but students were able to self-identify, and seven participants chose to do so.

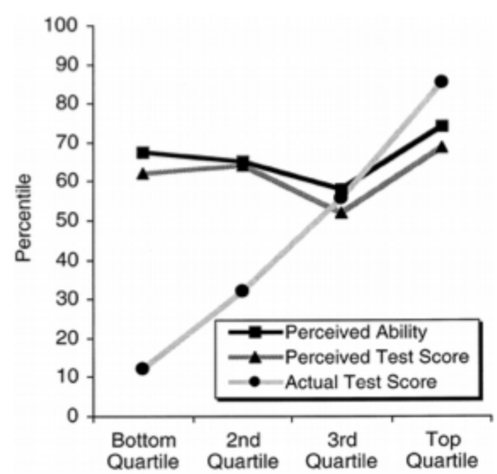

Figure 1

Copyright © 2015, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Procedure

In order to parallel the Kruger Dunning study, the course final exam served as the test instrument.

Volunteers completed surveys to assess their scoring expectations prior to taking the exam, as well as after having completed it but prior to receiving their grades. These questions served to measure their perceived ability, as well as their perceived test score after having taken the exam. As in the original study, the participants were grouped into quartiles based on their actual performance, and the results graphed with the responses for perceived and actual ability.

Additionally, to test specifically for the above-average effect, participants completed an extra question assessing their ability, with possible answers ranging from well-below average, to well-above average on a Likert scale.

Preliminary Data Analysis

All the observations noted in the logic test results above were present in the final exam test performance and survey data. As depicted in Figure 2, those in the lowest two quartiles grossly overestimated both their perceived ability and their actual performance (expected score), while the top performers underestimated both of the same. Regarding the above-average effect, ten respondents indicated that they were average, eight indicated that they were above average, and two indicated that they were well above average. Not a single respondent indicated that they were below average, let alone well below average.

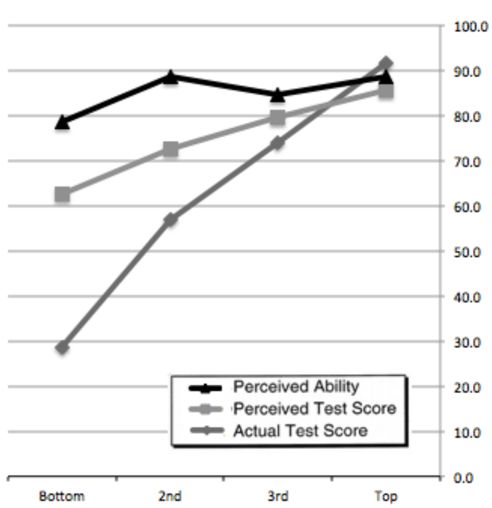

Figure 2

Conclusion and Future Work

The results from this experiment concur with the results of Kruger and Dunning's original experiment; however, this research was limited in multiple ways.

The original study included two interesting components that were not fully considered herein. In particular, the participants were 1) provided training after the initial test and prior to retesting, and 2) they were exposed to each other's work in order to familiarize each with the performance of others, so that they might ideally adjust their self-assessment accordingly.

Exposing students to each other's final exams is not practical, and neither is retraining, which effectively means repeating a class. Of the three who both self-identified and indicated that they had repeated the class, two remained in the bottom quartile with final exam scores of 23% and 24%. Both had expected to earn a B+ on the exam, and after taking it had adjusted their expectations only marginally downward. In spite of "retraining," these two participants were still grossly overestimating both their ability and performance. The third overestimated in a similar fashion, but had actual performance one quartile up from the bottom.

Unlike the relative scores in the original experiments, these scores are true grades, which highlights some questions germane to the academic arena. 1) Can those in the 2nd quartile effectively improve enough to move up to a passing grade (if so, how so), and 2) is it possible to identify earlier those who will always remain in the bottom quartile? These could have a significant impact on course scheduling and staffing.

Beyond academia, it would be interesting to determine if these issues exist in industry as well. Accurately identifying competency could improve staffing and promotional decisions, making organizations more effective. Kruger and Dunning make the point that competence is necessary to assess the performance of others. Further, "if incompetents have people reporting to them, their poor judgment may damage careers besides their own" (Abrahams 2005).

References

Kruger, J., and 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.

Ross, L., Greene, D. and House, P. 1977. The False Consensus Effect: An Egocentric Bias in Social Perception and Attributional Processes. *Journal of Experimental Social Psychology* 13:279-301

Abrahams, M. 2005. Those Who Can't, Don't Know It. *Harvard Business Review* December 2005:19