DUNNING-KRUGER EFFECT FOR STUDENTS TAKING STATISTICS COURSES

Carl Lee and Aklilu Zeleke
Central Michigan University
Michigan State University,
Carl.lee@cmich.edu

Dunning-Kruger (D-K) effect refers to "low-performing people tend to overestimate their performance to a much greater extent than high-performing people underestimate their performance". Many studies have revealed similar as well as different, and even reverse results depending on culture and task. In this paper, we examine the D-K effect for students taking statistics courses at Michigan State University. An attitude survey, an assessment instrument and questions related to the D-K effect were administered in statistics classes. The D-K effect are investigated and compared between different genders and different learning styles.

BACKGROUND

Dunning-Kruger effect (D-K effect) was named after the paper by Kruger and Dunning (1999). The Dunning-Kruger effect refers to "*low-performing people tend to overestimate their performance to a much greater extent than high-performing people underestimate their performance"*. The reason of such an effect, according to Kruger and Dunning (1999), was attributed to a deficit in metacognitive skill; that is, the low-performance individuals tend to be lacking of the awareness of their incompetence, thus, overestimating their performance. Since their 1999 paper, many studies have been conducted that investigated the D-K effect. Some literature focused on the possible explanations of the D-K effect while others focused on investigating if similar D-K effect exists when people were given tasks of different levels of difficulty. Yet, some others focused on the D-K effect for people from different cultures.

Various explanations about the DK-effects have been proposed. Dunning and Kruger in their 1999 paper argued that this is the lack of metacognitive ability to provide proper estimate of their ability for solving the task, and argued that there is a relationship between task performance, metacognition, and judgmental accuracy. Krueger and Mueller (2002) proposed the "regression to the mean" Hypothesis : " *people at all skill levels are prone to similar difficulties in estimating their relative performance. Their subjective estimates of performance are imperfectly correlated with objective performance measures, so their estimates of relative performance regress toward the mean.*" Burson et al. (2006) studied how the level of difficulty of the task affect people's judgement of their performance, and proposed the "noise-plus-bias: hypothesis: "*the primary drivers of miscalibration in judging percentile are general inaccuracy due to noise and overall biases that arise from the task difficulty level"*. They argued that for the easy task, there is a positive bias, the worst performers do poorly but do not know it and overestimate their standing. There is higher chance that people make their "guesses of answers" regardless of their actual ability, since the difficult tasks may be beyond their knowledge or skill levels for most people. Thus, there is a higher possibility that best performers have succeeded purely by good luck on hidden variables (such as good luck with guesses). Therefore, the individual's estimates of their performances tend to subject to higher noise.

Similar to the D-K effect, there have been evidences presented in the literature focusing on the "unskilled-unaware-of-it" effect among the group of lower performers. Dunning et al. (2004) reported that *"people are unrealistically optimistic about their own health risks compared with those of other people; students seem largely unable to assess how well or poorly they have comprehended material they have just read, and tend to be overconfident in newly learned skills; in the workplace, employees tend to overestimate their skill, making it difficult from giving meaningful feedback."*

Literature on the comparison of 'unskilled-unaware' phenomenon between male and female has been pretty consistent. Kim et al. (2015) reported no difference between male and female on the inabilities of self-assessment for low performer group. Regarding the effect of the culture difference on the DK-effect, it is interesting to see the "unskilled-unaware phenomenon" is mirror opposite in another culture as pointed out by DeAngelis (2003) and in research comparing North American and East Asian self-assessments by Heine (2001) (see also Falks et al.,2009). They concluded that "*East*

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

Asians tend to underestimate their abilities, with an aim toward improving the self and getting along with others."

In this paper, we examine the D-K effect for students taking statistics courses at Michigan State University. An attitude survey, an assessment instrument and questions related to the D-K effect were administered in statistics classes. The D-K effect are investigated and compared between students in different genders. The relationship of the D-K effect and students' attitudes are investigated.

THE METHOD

Participants for this study were those who took a calculus based introductory statistics course for science majors at Michigan State University (MSU). All students were enrolled at the residential college within MSU. Students are pursuing a career in the sciences or the study of sciences. Two sections, each with 48 students, participated in this study. A total of 87 students completed the course. Among them, 75 were females (74.7%) and 22 males (25.3%).

The class is a flip class, where lecture notes were posted on a course website and students were expected to read them before coming to class. During the first twenty minutes of class, students worked on class activity worksheets in groups of four. The worksheet problems were designed to test students' understanding of the material they read before class.

The assessment performance collected included a pre and post assessment instrument. Three questions are added to the post-assessment instrument to collect data related to DK-effect. These three questions are

- When comparing with other students in this class, your performance in this assessment exam, based on your estimate, is at approximately _____ percentile.
- When comparing with other students in this class, your overall quantitative ability, based on your estimate, is at approximately _____ percentile.
- What is your estimated score of the assessment you just did (out of 20 points): _____

An attitude survey similar to the *Survey of Attitudes Toward Statistics (SATS) (*SATS-28; Schau, 1992) was administered to collect students' responses on attitudes towards statistics.

DOES DK-EFFECT EXISTS AMONG STUDENTS TAKING INTRODUCTORY STATISTICS?

We first summarize the assessment results and the D-K effect items by four variables: (1) the actual score (Actual Score), the post-assessment scores percentile scale , (2) self-estimated percentile of assessment test relative to the entire class (Self_Est_Exam_Percentile), (3) self-estimated percentile of overall quantitative ability relative to the entire class (Self_Est_Ability_Perventile) and (4) self-estimated exam score (Self_Est_Score) in percentile scale. The relationship among these variables are displayed in Figure 1 for female and male students, respectively. It shows that there is very little relation between Actual Score and the other three percentiles, in particular in students' self-estimated percentile of their exam score, regardless of gender. The positive relation between Est_Exam_percentile and Est_Ability_Percentile ($r = .749$ for female, .966 for male) suggests that students' self-assessment of performance for this course is highly consistent with self-assessment of their overall quantitative ability, regardless of gender. However, the correlations between Est_Exam_Score and Est_Ability_Percentile are different for female ($r=.320$) and for male ($r=.593$) students. A possible explanation may be that the levels of confidence of quantitative ability are different. Literature has suggested that gender is a significant factor associated with mathematics confidence; females tend to show lower confidence than males (e.g., Flanagan and Einarson, 2017, and references herein). Thus, female students' self-estimate of performance for quantitative ability is lower than male students.

Similar to Kruger and Dunning (1999), we construct a figure to illustrate the average percentiles within each quartile of the actual score. Unlike the finding from Kruger and Dunning (1999), Figures 1, 2 and Table 1 suggest that students' self-estimates of their performance are equally poor regardless of low or high actual performance. That is, the low performers highly overestimate their actual performance; while the top performers drastically underestimate their performance. In fact, the average percentiles of self-estimated results are almost the same, except the 2nd quartile. Statistics is considered as a difficult subject for most college students. Our finding agrees with the finding presented in Burson et al. (2006), who suggested that the level of difficulty

of the task affect people's judgement of their performance. According to Burson et al. (2006), when performing a very difficult task, the judgement of their performance are equally poor regardless their actual performance.

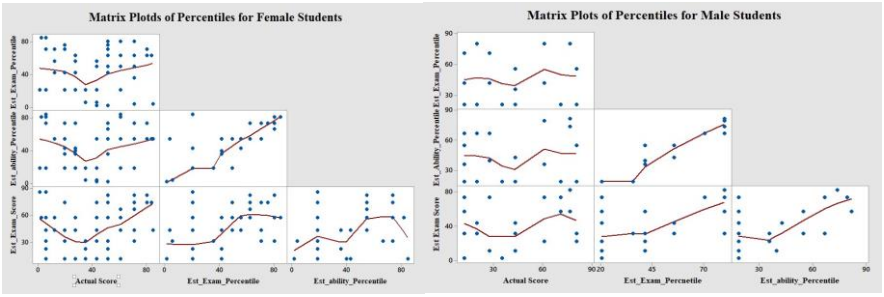

Figure 1: Matrix Plots of Percentiles for female and male students

Table 1: Comparison of Average percentiles between Actual Score and Self-judged percentile for each quarter of Actual Score

Based on Actual Score	Actual Score Percentile	Self-Judged Percentile	p-value of Paired-t
1st quarter (n=23)	12.00 (6.48)	47.09 (25.22)	.000**
2nd quarter (n=24)	35.50 (6.66)	32.00 (19.12)	.447
3rd quarter (n=19)	57.00 (4.82)	47.84 (24.72)	.148
4th quarter (n=19)	76.00 (5.23)	47.11 (25.45)	.000**
One-Way ANOVA, p-value of F-test		0.042*	

One-way ANOVA comparing self-judged percentiles among four quarters of Actual Score indicate there is a statistical significant difference at 5% level (p-value = 0.042). This is mainly due to the low average percentile self-judged performance of the 2nd quarter of Actual Score. The other three quarters are almost the same. When comparing the actual scores and self-judged percentile for each quarter of the Actual Score, we see the average percentiles of 1st quarter and the 4th quarter are significantly different, as shown in Table 1 and Figure 2.

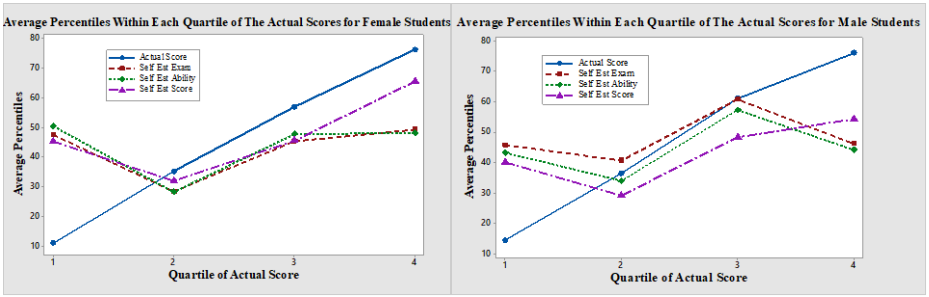

Figure 2: Average Percentiles within each quarter of the actual score for male and female students

Table 3: Comparison of self-judged percentiles between female and male students for each quartile of Actual Score

Based on Actual Score	Female: Mean (n,sd)	Male : Mean (n,sd)	p-value
1st quarter	47.6 (17, 26.1)	45.8 (6, 24.9,)	.883
2nd quarter	28.4 (17, 18.9)	40.7 (6, 19.8)	.222
3rd quarter	45.3 (16, 25.3)	60.8 (2, 27.2)	.586
4th quarter	49.3 (11, 25.4)	46.2 (6, 29.5)	.833
One-Way ANOVA, p-value of F-test	.039*	.860	

When comparing gender, Table 3 indicates that there is no significant difference on the self-judged percentile. This result is similar to the findings in the literature that there is no difference between male and female. However, when we compare the self-judged percentiles among four quarters of Actual Score for female students, statistical significant differences are noticed (p-value = 0.039). The difference is mainly due to very low average percentile for the 2nd quarter of the Actual Score.

IS DK-EFFECT DIFFERENT FOR DIFFERENT LEARNING STYLES?

We take further steps to investigate if students' learning styles have different effect on DK-effect. Six different learning styles were surveyed. Students were asked to choose 'Strongly Agree, Agree, Neutral, Disagree and Strongly Disagree. These questions are: I earn better by (1) having a lot of class notes, (2) reading the textbook, (3) cooperative learning, (4) taking a lot of notes, (5) working on real world projects, and (6) doing a lot of home work. Among these six learning styles, only the style of "Doing a lot of home work" shows significant difference. The average percentiles of self-judged performance are 38.29 with s.d. 23.03 for the Agree/Strongly Agree group, 46.43 with s.d. 26.35 for the neutral group and 59.98 with s.d. 21.12 for the Disagree/Strongly Disagree group. However, their actual scores (in percentiles) are 42.67 for the Agree/Strongly Agree group, 39.05 for the Neutral group and 47.17 for the Disagree/Strongly Disagree group, which are not statistically significant. This finding seems to indicate that students who do not like to do homework tend to over judge their performances, even though their actual performance is not significantly higher than those who want to do a lot of homework.

CONCLUSION

The DK-effect refers to low performers tend to over judge their performance, while high performers tend to underestimate their performance. This paper investigates the DK-effects for students taking an introductory statistics course. Our finding indicates that DK effect exists and that their self-judgement for low performance group are very similar to high performance group. In other words, low performance group appears to dramatically over estimate their performance; while the high performance group, as opposite, dramatically underestimates their performance. There is no statistical significant evidence that gender shows different Dk-effects. An investigation about the effect of learning styles suggests students who do not like to do a lot of homework tend to highly overestimate their actual performance when comparing with the students who choose to do a lot of homework.

REFERENCES

Burson KA, Larrick RP, & Klayman J (2006). Skilled or unskilled, but still unaware of it: how perceptions of difficulty drive miscalibration in relative comparisons. *Journal of personality and social psychology*, *90*(1), 60-77.

DeAngelis, T. (2003). Why we overestimate our competence. Social psychologists are examining people's pattern of overlooking their own weaknesses. *Monitor on Psychology*, *34*(2), 60-63.

Dunning D, Heath D & Suls J.M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, *5*,69–106.

Falk, C. F., Heine, S. J., Yuki, M., & Takemura, K. (2009). Why do Westerners self-enhance more than East Asians? *European Journal of Personality*, *23*, 183-209.

Heine, S. J. (2001). Self as a product of culture: An examination of East Asian and North American selves. *Journal of Personality*, *69*, 881-906.

Kim Y-H, Chiu C-Y & Bregant J (2015). Unskilled and Don't Want to Be Aware of It: The Effect of Self-Relevance on the Unskilled and Unaware Phenomenon. *PLoS ONE* 10(6): e0130309. https://doi.org/10.1371/journal.pone.0130309.

Krueger, J., & Mueller, R. A. (2002). Unskilled, unaware, or both? The contribution of socialperceptual skills and statistical regression to self-enhancement biases. *Journal of Personality and Social Psychology*, *82*, 180–188.

Kruger, J. M. & 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*, 1121–1134.

Flanagan, K. M. and Einarson, J. (2017). Gender, Math Confidence, and Grit: Relationships with Quantitative Skills and Performance in an Undergraduate Biology Course. *CBE—Life Sciences Education*, *16*(3), ar47.

Schau, C. (1992). Survey of Attitudes Toward Statistics (SATS-28). Online at *http://evaluationandstatistics.com*.