Discouraged by Peer Excellence:
Exposure to Exemplary Peer Performance Causes Quitting

Todd Rogers

Harvard Kennedy School

Avi Feller

University of California, Berkeley

To cite:

Corresponding Author:

Todd Rogers, Management, Leadership, Decision Sciences, Harvard Kennedy School, 79 JFK St., Cambridge, MA 02138

E-mail: todd_rogers@hks.harvard.edu

Abstract

People are exposed to exemplary peer performances often (and sometimes by design in interventions). In two studies, we showed that exposure to exemplary peer performances can undermine motivation and success by causing people to perceive that they cannot attain their peers' high levels of performance. It also causes de-identification with the relevant domain. We examined such *discouragement by peer excellence* by exploiting the incidental exposure to peers' abilities that occurs when students are asked to assess each other's work. Study 1 was a natural experiment in a massive open online course that employed peer assessment ($N = 5,740$). Exposure to exemplary peer performances caused a large proportion of students to quit the course. Study 2 explored underlying psychological mechanisms in an online replication ($N = 361$). Discouragement by peer excellence has theoretical implications for work on social judgment, social comparison, and reference bias and has practical implications for interventions that induce social comparisons.

Keywords

Leaders and organizations often expose people to selective information about their peers as a motivational tool (e.g., photos on lobby walls of exemplary employees' smiling faces , graphs showing customers that they are less energy efficient than their neighbors). In this article, we show that such practices can backfire when they lead people to perceive that the level of performance of their exemplary peers is out of reach. Such discouragement, which we refer to as *discouragement by peer excellence*, undermines motivation and success and causes de-identification with the relevant domain. In two studies, we examined this discouragement by exploiting the incidental exposure to information about peers' abilities that occurs when students assess each other's work (Topping, 1998). Study 1 was a natural experiment in a massive open online course (MOOC) that employs peer assessment ($N = 5,740$). Study 2 was an online replication exploring the underlying psychological mechanisms ($N = 361$).

People frequently conform to what they perceive to be the typical behaviors of other people (see Cialdini & Goldstein, 2004). For example, when people perceive that they use more energy than their neighbors, they tend to reduce their energy use (Allcott & Rogers, 2014; Schultz, Nolan, Cialdini, Goldstein, & Griskevicius, 2007); when people hear that most people vote, they become more motivated to vote (Gerber & Rogers, 2009); when people learn that people similar to themselves reuse resources, they tend to reuse resources (Goldstein, Cialdini, & Griskevicius, 2008); and when students see reductions in school harassment, they become less likely to engage in harassment (Paluck & Shepherd, 2012). Nearly all of the research on people's perceptions of these norms, though, has drawn people's attention to attainable social comparisons—that is, standards of performance that people believe they can achieve.

In the current research, we explored the impact of drawing attention to seemingly unattainable social comparisons. Some past research has shown that comparing oneself to "superstars" can enhance self-views when people perceive that they could attain comparable success. However, if such success seems unattainable, such social comparisons can undermine how positively people feel about themselves (Z. Lockwood & Kunda, 1997). The behavior outcomes we used in the current research differ from the subjective, self-focused measures used in most research of this kind (P. Lockwood, 2006; P. Lockwood, Jordan, & Kunda, 2002; Morse & Gergen, 1970).

One might imagine that people would recognize that excellent performances are outliers in a distribution of all performances. But research on reference bias suggests that people tend to adopt the most salient comparators as their reference groups (Duckworth & Yeager, 2015; Heine, Lehman, Peng, & Greenholtz, 2002). We argue that this shifting of reference group explains why, in our studies, students who assessed peers' excellent essays were more likely to quit than those who assessed peers' lower quality essays. The excellent essays, we suggest, shifted students' beliefs about their peers' abilities, and therefore how their own abilities compared with their peers'.

The *self-evaluation model*, which reflects many insights from social-comparison research (Festinger, Riecken, & Schachter, 1956; Garcia, Tor, & Schiff, 2013), predicts how people behave when they believe that they are not able to perform as well as their peers (Tesser, 1988). The model proposes that people aim to maintain positive self-evaluations when they compare themselves to other people. In the context of the current research, the model offers a key prediction: When people compare their own behavior with that of what they believe is a superior reference group, they will disengage from the behavior. This disengagement may be manifested as a reduction of effort (Karlsson, Loewenstein, & Seppi, 2009; Tesser & Cornell, 1991)[1] and as

a de-identification with the domain ([Tesser & Paulhus, 1983](#)). In the current research, we tested whether discouragement by peer excellence involves both forms of disengagement.[2]

Study 1: Does Peer Assessment in a MOOC Induce Discouragement by Peer Excellence?

Study 1 examined discouragement by peer excellence among students in a MOOC. Peer assessment is often used in MOOCs because the massive enrollment in such courses—often thousands—renders personalized instructor assessment impractical ([Piech et al., 2013](#)).

Method

Design. Study 1 used data from a large MOOC (the proprietors prefer to retain anonymity). Initial enrollment was more than 150,000 students; 3,894 students (2%) completed the course with a passing final grade—a persistence rate that is not uncommon in MOOCs ([DeBoer, Ho, Stump, & Breslow, 2014](#); [Ho et al., 2014](#)). To examine student motivation and success, we focused on the 5,740 students who completed the one evaluated writing assignment in the course and assessed the essays of their classmates. This assignment was due during the middle third of the course and was a requirement for obtaining a passing final grade. Each student was also required to read and assess at least three randomly assigned essays written by classmates. Students also evaluated their own essays. The final assessment scores that students earned for their essays were determined by calculating the median peer-assessment score. Students whose self-evaluation scores were within 5% of the median peer-assessment scores earned the higher of the two scores. Because the essays were randomly assigned, we were able to examine how the quality of those essays affected the students' course success. Put differently, we could study whether discouragement by peer excellence arose when students read and assessed particularly high-quality portfolios of essays.

Data. Overall, 6,039 students submitted the written assignment for the course. We restricted our analysis to the subset of 5,740 students who also read and assessed the required minimum of three of their peers' essays. Of these, 83% completed a course presurvey that asked about basic demographic information; 46% of these participants were female and their median age was 34.

Each essay was assessed on a scale of 0 to 9. For all assessed essays, the average score was 7.2 (median = 8); for the essays of students in the analysis sample, the average score was 7.5, with a median of 8. [Figure 1b](#) shows the global distribution of essay scores, which, unsurprisingly, has a long left tail.
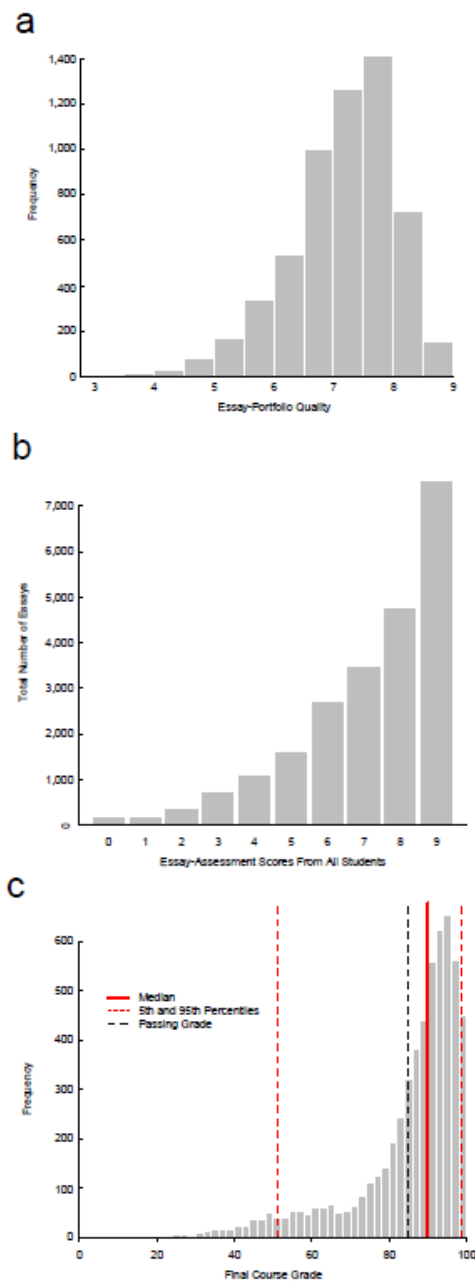
a

b

c

Fig. 1. Distributions of (a) essay-portfolio quality (i.e., the average score of the first three essays assessed), (b) essay-assessment scores from all students, and (c) students' final course grades.

The *essay-portfolio quality* for each student was defined as the average assessment score of the essays read and assessed by that student, excluding the assessment score given by that student to those essays. Roughly one quarter of the students assessed more than the minimum requirement of 3 essays; 0.2% of students assessed 50 or more essays (see Table S1 in the Supplemental Material available online). Including a greater number of essays in the average reduced the variance of essay-portfolio quality but did not change the mean (see Fig. S1 in the Supplemental Material). We therefore restricted our measure of essay-portfolio quality to the first 3 essays that each student assessed. The results were virtually unchanged when we relaxed this restriction. We used the times at which a student began assessing essays to determine which were assessed first. Because students could assess multiple essays at once, there was not a perfect one-to-one correspondence between the order in which students started assessing essays and when they finished assessing them. Figure 1a shows the distribution of essay-portfolio quality. For ease of interpretation, we used the *z*-score for essay-portfolio quality (as with the overall assessment score distribution, there is a long left tail). Finally, the random assignment of essay-portfolio quality across students appeared to be valid. The distribution of essay-portfolio quality by student's essay assessment score was almost perfectly flat, which is consistent with a valid randomization (see Table S2 in the Supplemental Material).

Results

There were two measures of course success. The first was the final course grade; the distribution of grades (Fig. 1c) is highly skewed, with a mean grade of 85 and a median grade of 90. Roughly

a quarter of students received a grade of 95 or higher; 3% received a grade of 99 or higher. The second outcome was whether a student received a certificate of course completion, which required a final course grade of 85 or higher. This threshold was explicitly stated at the beginning of the course and on the syllabus. Students were told in advance that there would not be a curve: Earning a course certificate was based on absolute performance. Of the 5,740 students in the study sample, roughly two thirds (3,857) completed the course.

We predicted that there would be a linear effect such that students with higher essay-portfolio-quality values would earn certificates at a lower rate and would earn lower final course grades. The results of the primary linear regression were consistent with these predictions (Table 1). In general, a 1-standard-deviation increase in essay-portfolio quality decreased by roughly 1 percentage point the probability that a student would earn a certificate in the course ($p = .03$ with no covariates; $p = .02$ when we controlled for a student's own essay score). The effect on final course grade was substantially smaller—a 1-standard-deviation increase in quality decreased the final grade by 1/3 point—and was marginally statistically significant ($p = .08$ with no covariates; $p = .09$ controlling for a student's own essay score).

Table 1. Linear Effects of Essay-Portfolio Quality on the Primary Outcomes

Predictor	Earning a certificate		Course grade	
	Model without writing score (R^2 = .001)	Model with writing score (R^2 = .45)	Model without writing score (R^2 = .001)	Model with writing score (R^2 = .44)
Essay portfolio quality	−0.013*	−0.011*	−0.35[†]	−0.25[†]
	(0.006)	(0.004)	(0.20)	(0.15)

Note: The table presents unstandardized regression coefficients with heteroscedastic robust standard errors in parentheses. [†]$p < .1$. *$p < .05$.

Further inspection of the data revealed a powerful, nonlinear relationship between essay-portfolio quality and student success. Figures 2a and 2b show the effects of essay-portfolio quality on earning a certificate and final course grade, estimated via a cubic regression spline with knots at the quintile boundaries (Wood, 2006). Both figures show a striking decrease at the top of the distribution, and this result is highly significant—for completion, approximate $F(3.8, 5736.2) = 3.6$, $p = .007$; for grade, approximate $F(4.2, 5735.8) = 3.9$, $p = .005$. In other words, there was strong evidence against the null hypothesis of no relationship between average essay score and the student success. This result was robust to different spline implementations and was essentially unchanged after we adjusted for students' own essay scores.
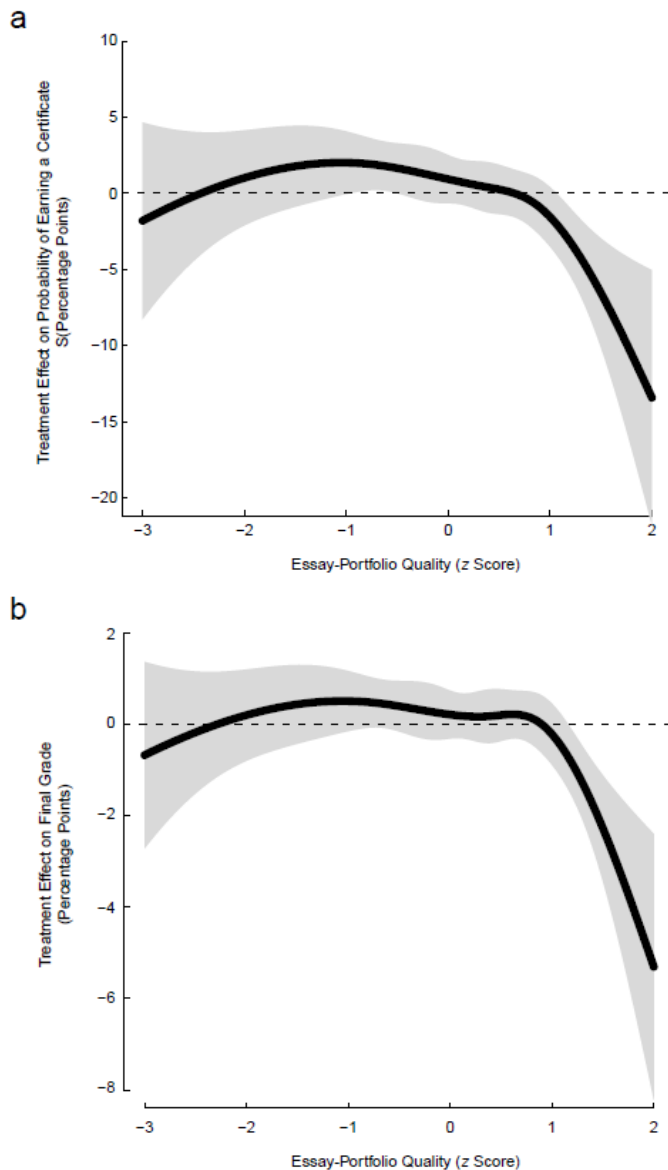
Fig. 2. Results from Study 1. Essay portfolio quality is graphed as a function of (a) probability of earning a certificate and (b) final grade. The dark lines are the cubic-regression-spline estimates; the gray areas around the lines are the 95% confidence intervals. Estimates are less certain at the extremes of essay-portfolio quality than at the center because fewer students assessed extremely high- or low-quality papers. These estimates were not adjusted for the students' own writing scores. Note that the underlying distribution was skewed.

Note that around 68% of students with an average essay-portfolio quality earned a certificate for passing the course. By contrast, among students with an essay-portfolio quality more than 1 standard deviation above the class mean, only 64% earned a certificate. For students with the top 100 essay-portfolio-quality scores—roughly, essay-portfolio quality greater than 1.6 standard deviations above the mean—the certificate-earning rate was a mere 45%. To put this effect size in context, consider that 93% of students who wrote a "perfect" essay (score of 9 of 9) earned a

certificate, whereas 75% of students who wrote an essay that was at the mean (score of 7.5 of 9) earned a certificate. Therefore, assessing the highest-quality essays (relative to assessing average essays) had as great an effect (or greater) on the rate of earning a certificate as did writing the highest-quality essays (relative to writing an average essay): –23 percentage points vs. 18 percentage points.

Figure S2 in the Supplemental Material shows the simple outcome means by quintile of essay-portfolio quality, a secondary check on the spline estimates. Although these unadjusted estimates are inherently noisier than the spline estimates in Figure 2, they show the same overall pattern and clearly demonstrate that the primary impact is in assessing the highest-quality essays. In particular, the course-completion rate was 64% in the highest quintile but 68% in the other four quintiles ($p = .02$); the average grade was 83.9 in the highest quintile but 85.2 in the other four quintiles ($p = .01$).

We also addressed our research question using a standard moderation analysis, testing whether the impact of assessing an excellent essay varied with the quality of the essay written by that student. Intuitively, we expected students who had written essays of low quality to be more affected by exposure to excellent essays, relative to students who had written essays of high quality, because they would be most likely to experience discouragement when contrasting their perceptions of the quality of their own essays with the quality of the most excellent essays. However, even with a linear specification, as used in Table 1, our analysis was severely underpowered: We were able to detect interaction effects only if they were at least of roughly the same magnitude as the main effect. Unsurprisingly, we found no meaningful moderation ($p > .8$ for either outcome).

Another analytic approach might have been to look at the relationship between course outcomes and the difference between a student's essay-portfolio quality and the quality of the essay written by that student. As noted, we would expect that discouragement by peer excellence would be greatest when this difference was greatest. However, the quality of the essay written by a student was strongly related to two other factors: First, it was an extremely strong predictor of final course grade ($r = .7$) because it was a component of the final grade and because it probably reflected students' abilities to succeed in the course. Second, because of the random assignment to essay-portfolio quality, there was likely to be a greater difference between essay-portfolio quality and the quality of students' own essay among students who had written essays of either low or high quality. Imagine students who had the worst possible self-essay-quality rating of 0 (on a scale ranging from 0 to 9). Because essay-portfolio quality was randomly assigned, these students evaluated portfolios ranging from 0 to 9 in quality, which meant that they would nearly always assess essays that were superior to their own. On the other hand, imagine students with the best possible quality rating of 9 for their own essays. Because essay-portfolio quality was randomly assigned, these students would evaluate portfolios ranging from 0 to 9 in quality, which means that they would nearly always assess essays that were inferior to their own. In other words, essay-portfolio quality was randomly assigned—but the difference between essay-portfolio quality and the quality of a student's essay was not. Putting these two factors together explains why this approach did not lead to a valid assessment of discouragement by peer excellence.

We also attempted to collect preassignment and postassignment survey questions about students' predicted performance relative to their peers. Because of issues with implementation and data collection, we were unable to compile or analyze these data.

Discussion

Student success declined when students were randomly assigned to read and assess truly excellent essays written by their peers. This provides strong field evidence that peer assessment can cause discouragement by peer excellence. Several findings are of particular note. First, the discouragement-by-peer-excellence effect was sizable, but only for people exposed to the most excellent essays. Second, reading poor-quality essays written by peers did not boost students' success. Because exposure to excellent essays by peers discourages students, one might expect that exposure to poor essays by peers could encourage students. This is not what we found, and we can only speculate as to why this asymmetry arose. Perhaps students entered the course confidently believing that they were capable of writing essays at least as well as their classmates were. This would have rendered exposure to poor peer essays inconsequential. We hope future research explores this.

Study 2: What Causes Discouragement by Peer Excellence?

In Study 2, we addressed why students' success declined when they read and assessed their peers' excellent essays.

Method

Participants. We recruited participants via Amazon's Mechanical Turk using an announcement that offered to pay participants U.S. $1.20 for a 10-min survey. We collected data from 361 participants (48% female; mean age = 34 years). We decided to collect data from at least 350 participants on the basis of the results of a pilot study.

Design. Study 2 used the Qualtrics survey platform (http://www.qualtrics.com/). Participants were presented with a quote and essay prompt previously used in the writing section of the SAT (McGinty, 2012). For the quote, writing prompt, and other study materials, see the Supplemental Material. Participants were asked to answer the prompt carefully and were told that the best responses would be entered into a lottery to receive a $1.00 bonus. Participants were kept on this writing page for a minimum of 3 min and were required to write a minimum of 500 characters (approximately 100 words).

Participants were then randomly assigned to one of two conditions. In the excellent-essay condition, participants were asked to assess two well-written essays from peers (averaging 195 words each) using a 6-point scale (1 = *poor*, 2 = *fair*, 3 = *good*, 4 = *very good*, 5 = *excellent*, 6 = *extraordinary*). In the poor-essay condition, participants were asked to assess two poorly written essays from peers (averaging 199 words each) using the same scale. All four essays were written by the researchers and were pilot-tested to confirm that the excellent essays were better than the poor essays. As a manipulation check, all participants were presented with their own essays and asked to rate how much better or worse their own essays were compared with the two essays they assessed (0 = "My written answer is WORSE than the two I read," 1 = "My answer is BETTER than the two I read"). Participants were then asked "How capable are you of writing a response on Mechanical Turk that is as good as the one you just read?" (1 = "I CANNOT write a response that is that good," 7 = "I DEFINITELY CAN write a response that is that good").

Participants then answered two questions adapted from the Centrality subscale of the Multidimensional Inventory of Black Identity ([Sellers et al., 1997](#)). Participants were asked how strongly they agreed with the following two statements about their ability to write an excellent short answer to a philosophical question like the one they just answered: (a) "It has very little to do with how I feel about myself" and (b) "It is an important reflection of who I am." Participants responded using a 7-point scale (1 = *strongly disagree*, 7 = *strongly agree*). The first question was reverse-scored. Responses to the two questions were highly correlated with each other (r = .481, $p < .001$) and were averaged to create a domain-identification index. Finally, participants were offered the opportunity to write a short essay response to another SAT question. Participants were told that if they chose to write a response, their answers would be assessed by other Mechanical-Turk users and that essays assessed in the top 25% would receive a $0.50 bonus.

To validate that the 2 essays in the poor-essay condition were in fact poor and the 2 essays in the excellent-essay condition were in fact excellent, we conducted a 4-min follow-up Mechanical Turk survey, offering to pay participants $0.25 for their time. The 361 people who participated in Study 2 were excluded from participating in the validation survey. After respondents passed an attention check (3% failed it), they were presented with Study 2's quote and essay prompt and were asked to use a 7-point scale (1 = *very poor*, 7 = *excellent*) to rate 4 essays randomly selected from among the 361 essays written by Study 2 participants and the 2 poor and 2 excellent essays used for the Study 2 manipulation. Of the 894 respondents (57% male; mean age = 33 years), 34 participated in the survey twice, and 2 participated three times. We included all 932 observations because the respondents who participated more than once graded different essays each time they participated.

Results

Manipulation checks. Participants in the excellent-essay condition assessed the two essays they were assigned as being of significantly higher quality (mean assessed score = 4.27, $SE = 0.075$) than did participants in the poor-essay condition (mean assessed score = 2.11, $SE = 0.064$), $t(359) = 21.88$, $p < .001$. In addition, only 30% of participants in the excellent-essay condition thought their own essay was better than the ones that they had assessed, whereas 97% of participants in the poor-essay condition thought that their own essay was better, $\chi^2(1, N = 361) = 169.26$, $p < .001$.

Perceived attainability and domain identification. Participants in the excellent-essay condition felt significantly less able to write an essay that would be as good as the ones that they read and assessed ($M = 4.66$, $SE = 0.128$) compared with those in the poor-essay condition ($M = 6.53$, $SE = 0.074$), $t(359) = -12.47$, $p < .001$. Further, participants in the excellent-essay condition had a significantly lower domain-identification index ($M = 3.97$, $SE = 0.103$) compared with those in the poor-essay condition ($M = 4.53$, $SE = 0.112$), $t(359) = -3.68$, $p < .001$. This effect suggests that exposure to excellent essays (relative to exposure to poor essays) caused a reduction in the extent to which participants thought that the ability to write such essays was self-relevant.

Motivation, success, and mediation. Reading and assessing excellent essays, relative to poor essays, caused discouragement by peer excellence. Of participants in the poor-essay condition, 43% ($SE = 3.7\%$) chose to write a second essay, whereas only 27% ($SE = 3.3\%$) of participants in the excellent-essay condition chose to write a second essay, $\chi^2(1, N = 361) = 9.87$, $p = .002$.

Figure 3 shows this pattern of results. Statistical mediation tests (Preacher & Hayes, 2004) showed that participants' perceived ability to write an excellent essay statistically mediated the effect of reading excellent or poor essays on motivation to write a second essay. The domain-identification index did not mediate the treatment effect. See the Supplemental Material for complete details.
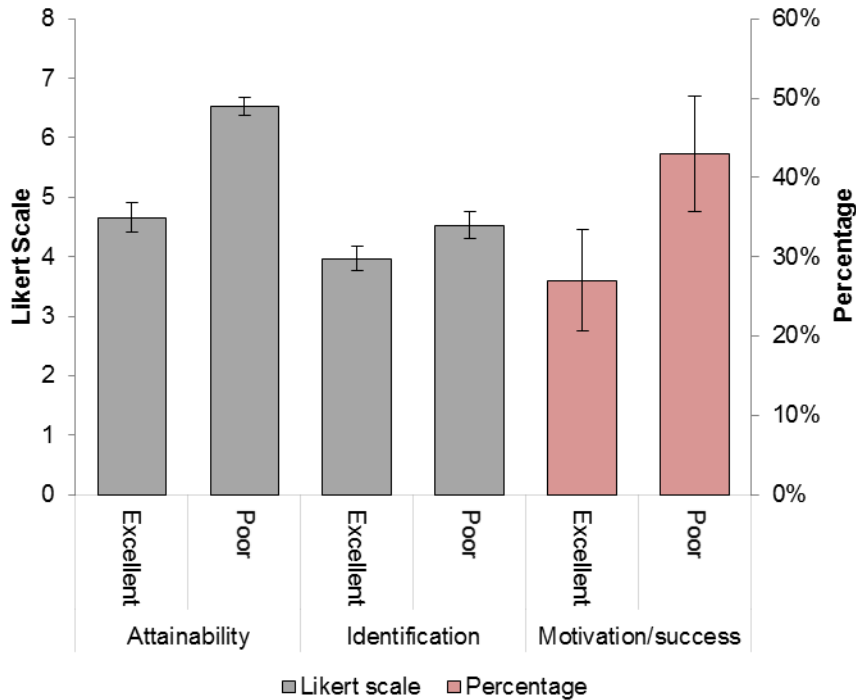

Fig. 3. Results for the excellent-essay and poor-essay conditions: Likert-scale scores for (a) perceived attainability and (b) domain identification and (c) the percentage of students who wrote a second essay (a measure of motivation and success. Error bars represent 95% confidence intervals.

Validation and moderation. The excellent essays were, in fact, excellent, and the poor essays were, in fact, poor. All 365 essays were assessed approximately 10 times. (We initially examined the validation-survey data after each response had been assessed an average of 4 times; because assessments of the same essays varied widely across participants, we decided to increase the number of assessments per essay to 10 before examining the data again.) The 2 essays in the excellent-essay condition were assessed to be in the top decile of all 365 essays, whereas the 2 essays in the poor-essay condition were assessed to be in the bottom decile of all 365 essays.

As with Study 1, we explored whether discouragement by peer excellence was moderated by a student's own essay quality—that is, whether the impact of assessing excellent essays was smaller for students who themselves wrote high-quality essays. Using the validation assessments to measure the quality of the 361 study participants' essay-writing abilities, we did not find that essay-writing ability moderated susceptibility to discouragement by peer excellence, log-likelihood = 0.001, $SE = .23$, $p = .995$. This null finding was consistent with the results of the underpowered moderation analysis reported in Study 1. We believe that the most sensible

interpretation is that the studies were poorly powered to detect anything but large interaction effects; if such moderation existed, it was not large.

Discussion

Study 2 replicated the discouragement-by-peer-excellence effect observed in Study 1 and provided evidence for how it happened. Exposure to other peoples' excellent essays undermined people's sense that they were capable of producing essays that were as good as those to which they had been exposed, and this changed belief statistically mediated the discouragement effect.

What does it mean that perceived attainability mediated discouragement by peer excellence? Participants responded to the question "How capable are you of writing a response on Mechanical Turk that is as good as the one you just read?" Answers to this question reflected participants' assessments of their own abilities as well as their assessments of their reference groups' abilities. One might be concerned that this would muddy any interpretation of mediational results. We argue that both of these dimensions—self-assessment and reference-group assessment—combine to shape people's subjective experience of evaluating peer essays. Study 1's students who read excellent essays, we argue, believed that they could not write high-quality essays like those of their peers—an argument supported by Study 2. This changed belief decreased their motivation and subsequent success in the MOOC—also supported by Study 2. Although we hope that future research teases apart these two cognitive processes underlying discouragement by peer excellence, they combine in the current studies to create a state that undermines student motivation and performance.

Because Study 2 involved only two conditions, we are not able to rule out one possible alternative interpretation: Our finding of lower motivation to write a second essay in the excellent-essay condition compared with the poor-essay condition could have been due to the inferiority of the poor essays (i.e., the low quality encouraged participants to write a second essay) rather than to the superiority of the excellent essays (i.e., the high quality discouraged participants from writing a second essay). Although plausible, such an explanation would be inconsistent with the strong field results in Study 1.

General Discussion

Being exposed to peers' excellent performance makes people feel less capable of performing at the level of those peers. This changed belief appears to decrease student performance. Exposure to peers' excellent performance also reduces the extent to which people self-identify with the relevant domain. The discouragement-by-peer-excellence effect is powerful: Real students who assessed peers' excellent essays were substantially less likely to earn course credit than those who assessed peers' less exceptional essays. The discouragement-by-peer-excellence effect is inconsistent with research on descriptive social norms. Our findings have theoretical implications for work on social judgment, social comparison, and reference bias and have practical implications for interventions that induce social comparisons.

To what extent are people exposed to other peoples' excellent performances in their daily lives? As discussed in the introduction, leaders and organizations often direct attention to peers' excellent performance as a motivational strategy. Sometimes this is transparent. Employee-of-the-year awards may create explicit incentives to perform well, but they are unlikely to shift

people's perceptions of their typical peers' abilities very much. After all, employees of the year are celebrated exactly because they are excellent. Other times, though, peers' excellent performances are highlighted in more subtle ways. Consider managers off-handedly mentioning specific outstanding behaviors by specific employees during team meetings. We predict that the more subtle these appeals, the more likely they will be to shift perceptions of peers' abilities and, consequently, the more likely they will be to induce discouragement by peer excellence. In addition, because people tend to self-promote in social contexts (Leary et al., 1994), they are more likely to talk about when they were excellent than when they were subpar. Of course, most praise from other people (and from one's self) is not for truly excellent performances, but rather for above-average performances. Such praise is unlikely to induce discouragement by peer excellence. However, praise for truly excellent performances may discourage. One might neutralize discouragement by peer excellence in these situations by noting how unusually excellent the praised performances are. This may preserve people's motivation for future public praise without inducing disengagement.

In addition to the contribution of our findings to research on motivation, social comparison, and social judgment, the discouragement-by-peer-excellence effect is of practical importance. For example, interventions that induce social comparisons should be careful to not use truly excellent performances for comparisons. Consider the home-energy reports sent by some public utility companies in which consumers' energy use is compared with that of their neighbors. It has been shown to enduringly and persistently reduce home energy use (Allcott & Rogers, 2014). Future research should explore whether such interventions can be made more effective by avoiding comparisons with peers' excellent performances.

Peer assessment is a popular practice in both online and offline educational settings (Piech et al., 2013; Sadler & Good, 2006; Topping, 1998). It is part of the movement to increase active learning in the classroom, which has been shown to increase student success (Freeman et al., 2014; Haak et al., 2011). It is used in MOOCs as well as in brick-and-mortar classrooms. The current research suggests that educators using peer assessment should ensure that peers' excellent performances are not interpreted as typical. Although there are many ways that one might prevent such interpretations, we speculate on two. First, discouragement by peer excellence during peer assessment may be neutralized by balancing exposure to peers' excellent performances with exposure to less excellent performances. Second, discouragement by peer excellence may be neutralized by swiftly acknowledging the distinction of truly excellent performances so as to minimize the risk that the excellent performances are interpreted as typical. The discovery of the discouragement-by-peer-excellence effect is one of the first examples in which research conducted with a MOOC has implications for both online and brick-and-mortar learning (Ho et al., 2014; Martin, 2012; Pope, 2014).

Author Contributions

Both authors contributed to the design of Study 1. A. Feller conducted and oversaw data analysis for Study 1. T. Rogers conducted and oversaw design and analysis of Study 2. Both authors contributed to writing and revising the final manuscript.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Supplemental Material

Additional supporting information can be found at http://pss.sagepub.com/content/by/supplemental-data

Open Practices

All data and materials have been made publicly available via Open Science Framework and can be accessed at https://osf.io/34cd8/. The complete Open Practices Disclosure for this article can be found at http://pss.sagepub.com/content/by/supplemental-data. This article has received the badges for Open Data and Open Materials. More information about the Open Practices badges can be found at https://osf.io/tvyxz/wiki/1.%20View%20the%20Badges/ and http://pss.sagepub.com/content/25/1/3.full.

Notes

1. This is also consistent with Moore and Kim's (2003) finding that people are less likely to gamble on objectively difficult tasks relative to easy tasks.

2. The self-evaluation model is one of many lines of social-comparison research. This literature tends to focus on social comparison's implications for self and attitudes rather than performance (Garcia et al., 2013). One relevant finding, which may be seen as being inconsistent with discouragement by peer excellence, is that people who are prone to making upward social comparisons tend to have better educational outcomes (Blanton, Buunk, Gibbons, & Kuyper, 1999). The current research differs from that work in that it focuses on the consequences of being randomly (situationally) exposed to excellent peer work. People who are dispositionally prone to seeking out and making upward comparisons with excellent peer work are likely to be aware that the work is excellent and so may not interpret it as being representative of their peers work—which is, as we suggest, a key condition underlying discouragement by peer excellence. In addition, rather than causing the educational success, this dispositional tendency may covary with other unmeasured attributes that correlate with educational success.

Psychological Science, 27(3), 365-374.

References

Allcott, H., & Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review*, *104*, 3003–3037. doi:10.1257/aer.104.10.3003

Blanton, H., Buunk, B. P., Gibbons, F. X., & Kuyper, H. (1999). When better-than-others compare upward: Choice of comparison and comparative evaluation as independent predictors of academic performance. *Journal of Personality and Social Psychology*, *76*, 420–430.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, *55*, 591–621. doi:10.1146/annurev.psych.55.090902.142015

DeBoer, J., Ho, A., Stump, G., & Breslow, L. (2014). Changing "course": Reconceptualizing educational variable for massive open online courses. *Educational Researcher*, *43*, 74–84. doi:10.3102/0013189X14523038

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, *44*, 237–251. doi:10.3102/0013189X15584327

Festinger, L., Riecken, H. W., & Schachter, S. (1956). *When prophecy fails*. Minneapolis: University of Minnesota Press.

Freeman, S., Eddy, S., McDonough, M., Smith, M., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *USA*, *111*, 8410–8415. doi:10.1073/pnas.1319030111

Garcia, S. M., Tor, A., & Schiff, T. M. (2013). The psychology of competition: A social comparison perspective. *Perspectives on Psychological Science, 8*, 634–650.

Gerber, A. S., & Rogers, T. (2009). Descriptive social norms and motivation to vote: Everybody's voting and so should you. *The Journal of Politics*, *71*, 178–191. doi:10.1017/S0022381608090117

Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*, *35*, 472–482.

Haak, D., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science*, *332*, 1213–1216. doi:10.1126/science.1204820

Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales? The reference-group effect. *Journal of Personality and Social Psychology*, *82*, 903–918. doi:10.1037/0022-3514.82.6.903

Ho, A. D., Reich, J., Nesterko, S. O., Seaton, D. T., Mullaney, T., Waldo, J., & Chuang, I. (2014). *HarvardX and MITx*: *The first year of open online courses, fall 2012-summer 2013* (HarvardX and MITx Working Paper No. 1). Retrieved from Social Science Research Network Web site: http://ssrn.com/abstract=2381263

Psychological Science, 27(3), 365-374.

Karlsson, N., Loewenstein, G., & Seppi, D. (2009). The ostrich effect: Selective attention to information. *Journal of Risk and Uncertainty*, *38*, 95–115. doi:10.1007/s11166-009-9060-6

Leary, M. R., Nezlek, J. B., Downs, D., Radford-Davenport, J., Martin, J., & McMullen, A. (1994). Self-presentation in everyday interactions: Effects of target familiarity and gender composition. *Journal of Personality and Social Psychology*, *67*, 664–673.

Lockwood, P. (2006). Someone like me is successful: Do college students need same-gender role models? *Psychology of Women Quarterly*, *30*, 36–46. doi:10.1111/j.1471-6402.2006.00260.x

Lockwood, P., Jordan, C. H., & Kunda, Z. (2002). Motivation by positive or negative role models: Regulatory focus determines who will best inspire us. *Journal of Personality and Social Psychology*, *83*, 854–864. doi:10.1037/0022-3514.83.4.854

Lockwood, Z., & Kunda, P. (1997). Superstars and me: Predicting the impact of role models on the self. *Journal of Personality and Social Psychology*, *73*, 91–103. doi:10.1037/0022-3514.73.1.91

Martin, F. G. (2012). Will massive open online courses change how we teach? *Communications of the ACM*, *55*(8), 26–28. doi:10.1145/2240236.2240246

McGinty, S. M. (2012). *The college application essay* (5th ed.). New York, NY: The College Board.

Moore, D. A., & Kim, T. G. (2003). Myopic social prediction and the solo comparison effect. *Journal of Personality and Social Psychology*, *85*, 1121–1135.

Morse, S., & Gergen, K. (1970). Social comparison, self-consistency, and the concept of self. *Journal of Personality and Social Psychology*, *16*, 148–156. doi:10.1037/h0029862

Paluck, E. L., & Shepherd, H. (2012). The salience of social referents: A field experiment on collective norms and harassment behavior in a school social network. *Journal of Personality and Social Psychology*, *103*, 899–915.

Piech, C., Huang, J., Chen, Z., Do, C., Ng, A., & Koller, D. (2013, July). *Tuned models of peer assessment in MOOCs*. Paper presented at the 6th International Conference on Educational Data Mining, Memphis, TN.

Pope, J. (2014). *What are MOOCs good for*? Retrieved from the MIT Technology Review Web site: http://www.technologyreview.com/review/533406/what-are-moocs-good-for/

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods*, *Instruments*, *& Computers*, *36*, 717–731. doi:10.3758/BF03206553

Sadler, P., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment*, *11*, 1–31.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, *18*, 429–434.

Psychological Science, 27(3), 365-374.

Sellers, R. M., Rowley, S. A., Chavous, T. M., Shelton, J. N., & Smith, M. A. (1997). Multidimensional Inventory of Black Identity: A preliminary investigation of reliability and constuct [sic] validity. *Journal of Personality and Social Psychology, 73*, 805–815.

Tesser, A. (1988). Toward a self-evaluation maintenance model of social behavior. In L. Berkowitz (Ed.). *Advances in experimental social psychology* (Vol. 21, pp. 181–227). San Diego, CA: Academic Press. doi:10.1016/S0065-2601(08)60227-0

Tesser, A., & Cornell, D. P. (1991). On the confluence of self processes. *Journal of Experimental Social Psychology*, *27*, 501–526. doi:10.1016/0022-1031(91)90023-Y

Tesser, A., & Paulhus, D. (1983). The definition of self: Private and public self-evaluation management strategies. *Journal of Personality and Social Psychology*, *44*, 672–682.

Topping, K. (1998). Peer assessment between students in college and universities. *Review of Educational Research*, *68*, 249–276. doi:10.3102/00346543068003249

Wood, S. N. (2006) Generalized additive models. Boca Raton, FL: CRC Press.