Performance Effects of Setting a High Reference Point for Peer-Performance Comparison

Henry Eyring and V.G. Narayanan

Harvard Business School

June, 2016

We thank Dennis Campbell, Ian Gow, Tatiana Sandino, Rajesh Vijayaraghavan, Bob Kaplan, Suraj Srinivasan, Bob Simons, and workshop participants at Harvard Business School for their helpful comments. We also thank HarvardX research directors and course managers, HarvardX-associated faculty, and Harvard Business School research assistants for their support of the proposed field experiment.
Performance Effects of Setting a High Reference Point for Peer-Performance Comparison

June, 2016

Abstract

The proposed study addresses performance effects of setting a reference point for peer comparison above that of peer-median performance. Evidence that relative performance information (RPI) affects performance even when performance is not linked to pay or made visible to others is well established, yet evidence of how the performance effects depend on the height of commonly included reference points for peer-performance comparison is lacking. We propose addressing this gap with a field experiment testing performance effects of providing peer-top-quartile as opposed to peer-median performance as an RPI reference point. We will also test how the performance effects depend on initial performance. We predict a concave relationship between an individual’s initial performance before RPI display and the performance effect of providing the higher as opposed to lower RPI reference point. We will address the nature of the concavity in explaining the resulting average performance effect.
1. INTRODUCTION

The proposed study addresses performance effects of setting a reference point for peer comparison above peer-median performance. Providing individuals with relative performance information (RPI), or information for comparing one’s own performance to that of peers, elicits performance improvement in a variety of compensation settings, including those in which performance is not linked to pay or made visible to others (Allcott [2011], Hannan, Krishnan, and Newman [2008], Tafkov [2013]). Theories of social comparison, reference points, expectancy, and goals could guide inquiry into the performance effects of the height of reference points for peer-performance comparison that are commonly displayed along with RPI.¹,² Such inquiry could inform the many government, non-profit, and corporate administrators who are using RPI reference points to influence constituents’ performance.³ However, evidence of performance effects of RPI reference point height is lacking.

We propose providing such evidence through a field experiment in online education. We will compare the performance effects of providing the peer top quartile as opposed to the peer median as a reference point for peer comparison within RPI. The data include measures of a range of actions taken in online courses, as well as a log of each instance of a student who receives RPI accessing it, over multiple months. The experimental setting


² We use the term “RPI reference point height” to refer to the magnitude of a percentile of peer performance that is included as a reference point in a report of one’s own performance relative to that reference point.

and intervention do not involve explicit rewards for or public visibility of one’s own performance, allowing for the identification of the distinct effects of RPI reference point height. Identifying these effects forwards accounting literature on how comparison of one’s performance to that of peers affects task performance without altering rational expectations of pay for marginal effort (Murthy [2010], Hannan et al. [2008, 2013], Tafkov [2013]). The proposed effect-identification would also show how comparison to high-performing peers influences performance through reporting, as opposed to through peer-group composition. The latter avenue of performance effects is addressed in a related body of economic research (Hanushek et al. [2003], Lavy, Silva, and Weinhardt [2012]).

Multiple disciplines offer insight into how RPI and reference points increase performance apart from the rate at which performance is compensated or the visibility of one’s performance to others. First, social comparison theory and related accounting research explain that RPI facilitates a reward for effort-related performance in terms of favorable comparison to peers that enhances one’s self-image (Brown et al. [2007], Garcia and Tor [2007], Tafkov [2013]). Second, economic research shows that providing higher reference points for total pay increases willingness to exert effort at a given piece rate, lending evidence to reference-dependent preferences for effort provision (Abeler et al. [2011]). Third, RPI reference points that are portrayed as standards for success, as in the proposed study, may exhibit characteristics of goals. Goal theory states that goals energize and focus effort so as to increase performance (Locke and Latham [2002]). Finally, expectancy theory notes that a moderating factor of motivation to work toward a level of performance is one’s belief in the attainability of that performance level (Atkinson [1957]). We draw from these lines of research on social comparison, reference points,
goals, and expectancy in predicting the effect of providing a higher RPI reference point than the median.⁴

Forming and testing those predictions extends recent economic and psychology field studies showing generally positive performance effects of providing RPI reference points (Azmat and Iriberri [2010], Allcott [2011], Harper et al. [2013], Schultz [2007]). These studies report heterogeneity in performance effects of RPI reference points: individuals underperforming the RPI reference point are positively affected by its provision, while those outperforming are less positively or even negatively affected. These studies do not, though, test the performance effects of varying the height of the RPI reference point provided. We draw from relevant theory to predict a concave relationship between an individual’s performance before RPI reference point display and the positive performance effect of providing them with a higher RPI reference point, whereby those initially performing between the alternative reference points are most positively affected. We will address the nature of the predicted concavity in explaining the average performance effect of providing the higher RPI reference point. As additional analyses, we propose assessing the moderating effects of gender, confidence in one’s ability to reach the reference point, and interest in the comparison facilitated by the reference point on the performance effects of RPI reference point height.⁵

---


⁵ Each of these has been shown to influence responses to peer comparison and/or performance feedback, and we would address how they moderate the effect of RPI-based feedback that includes a higher, as opposed to lower, reference point. See Lavy et al. [2012] regarding gender, Bandiera et al. [2013] regarding confidence, and Tafkov [2013] regarding interest in comparison.
The proposed study would make three main contributions. First, exploring the performance effect of RPI reference point height in a field experiment setting without pay for or external visibility of relative performance allows identifying the distinct effects of RPI reference points. This would speak to the growing body of economic, psychology, accounting, and management research on the application of RPI as a self-sufficient tool for influencing performance and behavior. Studies of such an effect include both the private and public sector, with outcomes including retail service, education, energy consumption, web-site content contribution, and taxpaying. Research on the importance of the reference point displayed along with RPI could accordingly have significant policy implications for a variety of corporate and other societal settings.

Second, understanding the effects of RPI reference points in isolation may also inform theory and empirical work in a variety of other accounting and economic streams of research. Research on RPI-related accounting and economic mechanisms might draw from the current paper’s results in a number of ways. For example, finding weaker performance returns to RPI reference point height than to goal difficulty could partially explain the prevalence of much easier targets than goal theory suggests are optimal given that targets sometimes derive from or communicate RPI (Butters [2000], Fisher, Peffer, and Sprinkle [2003], Merchant and Manzoni [1989]). Understanding returns to RPI reference point height would relatedly help in predicting the performance effects of

---


7 For examples of such RPI use, see Aranda, Arellano, and Davila [2014], Bol et al. [2010], and Murphy [2000] regarding target setting, and Securities and Exchange Commission [2015] regarding disclosure and monitoring, and Gibbons and Roberts [2013] regarding contracting.
supervisor discretion in target setting that involves relative performance (Bol, et al. [2010]). Finding a concave relationship between initial performance and performance returns to reference point height would suggest limited benefits of RPI reference point height in the problem areas of tournaments – those at either extreme relative to the rewarded cutoff (Casas-Arce and Martinez-Jerez, [2009], Hannan et al. [2008]). RPI reference points, especially the median and top-quartile, are also influential in measuring corporate performance, gauging whether executive compensation is appropriate based on corporate performance, and evaluating employees.\(^8\) Research in those settings could benefit from illustration of the behavioral responses to viewing RPI reference points that are differentially challenging to reach.

Third and finally, a substantial amount of accounting research shows that the format of information display influences decisions ranging from stock trading to assigning employee bonuses (Bloomfield, Nelson, and Smith [2006], Dilla and Steinbart [2005], Maines and McDaniel [2000]). The proposed study extends this research by showing how the height of the reference point included in RPI influences performance. Given the common use of RPI and associated percentiles in information displays, the height of reference points set for the purpose of peer-performance comparison is a salient feature to provide evidence on (Song et al. [2015], Bizjak et al. [2011], Gibbons and Roberts [2012] p. 67).

---

2. THEORY AND HYPOTHESIS DEVELOPMENT

Economics-based research addresses multiple functions of RPI that account for its widespread use (Gibbons and Roberts [2012] p. 67). Incorporating RPI in incentive contracts reduces the risk imposed on agents by filtering out common noise from a performance measure linked to incentives (Holmstrom [1982], Lazear and Rosen [1981]). The performance measure is then a more precise signal of effort, and attaching incentives to it imposes less risk on the agent in the form of uncontrollable events that also influence the measure (Banker and Datar [1989]). RPI is also of use to agents in forming expectations of pay for marginal effort in nonlinear incentive schemes. When pay is contingent upon reaching certain levels of relative performance, agents can use information on their proximity to those levels to form such expectations. Empirical studies show that agents subject to nonlinear-incentive schemes expend effort according to their updated rational expectations of pay for marginal effort (Asch [1990], Hannan et al. [2008], Casas-Arce and Martinez-Jerez [2009]).

An emerging stream of accounting and economic research explores an additional role of RPI. Agents incorporate the information into decisions about effort provision even when it does not provide information on the rate of pay for marginal effort (Azmat and Iriberri [2010], Hannan et al. [2008], Murthy [2010], Tafkov [2013]). These studies show generally positive performance effects of providing RPI in fixed-wage, individual-performance piece-rate pay, and in no-pay contexts. In many studied applications, RPI includes reference points for peer comparison, oftentimes peer-median performance (Blanes i Vidal and Nossol [2011], Harper et al. [2013]). Recent economic studies have incorporated the role of reference points in influencing effort provision, termed
“reference-dependent preferences,” into models of utility traditionally formulated as separable in and comprised only of monetary pay-off and cost of effort (Abeler et al. [2011], Farber [2008]). Those models and empirical results provide evidence that reference points for total pay and for peer-performance comparison influence effort provision above and beyond the rate of pay for marginal task performance. Although RPI and reference points both exhibit power in motivating effort, little is known regarding the effect of RPI combined with reference points of varying heights.

A number of field studies report generally positive effects of RPI performance display, but do not test the performance effects of varying the height of included reference points (Allcott [2011], Allcott and Rogers [2014], Azmat and Iriberri [2010], Harper et al. [2013], Schultz et al. [2007]). The results show heterogeneous performance effects depending on initial performance relative to the reference point; individuals initially underperforming a reference point exhibit a more positive effect than those initially outperforming it. Offered explanations include the difficulty of achieving beyond an already high level, as well as a downward psychological attractive power of reference points for those performing above them (Allcott [2011]), Schultz [2007]). A potential implication of the observed heterogeneity is that a higher reference point, relative to which individuals are situated differently, would yield different performance effects. We will test this by displaying and accurately labeling either the top quartile or median reference point within RPI. Our proposed analysis addresses the performance effects of these reference points of alternative height on average as well as within the groups initially below, between, and above each alternative.
Theory and evidence on social comparison, reference points, goals, and expectancy explain forces underlying responses to RPI and reference points. Identifying these forces facilitates predicting overall and initial-performance-dependent effects of reference point height.

Regarding RPI, social comparison theory explains utility resulting from providing effort necessary to improve one’s standing on desirable abilities or accomplishments in comparison to peers. Favorable comparison to peers validates a positive self-image (Smith [2000], Brown et al. [2007]). Interest in the performance of peers and performing well relative to them holds in settings in which peers are identifiable and when they are anonymous, as in the proposed study (Klar and Giladi [1998], Tafkov [2013], Xiao and Lucking [2008]). Providing RPI enables social comparison, thereby providing a reward for effort expended toward improving one’s performance relative to peers. Accounting research offers evidence that providing RPI works through that means to elicit performance improvement in settings without relative-performance-based pay and when one’s performance is not visible to others (Hannan et al. [2008], Murthy [2010], Tafkov [2013]).

Regarding reference points, subtly implied reference points for rational expectations of total pay influence effort without altering the rate of pay for performance (Abeler et al. [2011]). Individuals exhibit loss-aversion to pay below the reference point. Utility from eliminating negative deviation from the reference point explains the ability of reference points to induce higher effort without changing the rate of pay (Abeler et al. [2011], Farber [2008]). Goals are more explicitly set standards for performance or outcome-achievement, and may be suggested by others or set completely by personal volition. RPI
reference points may assume properties of goals to the extent that individuals accept them as standards for success. Goal theory literature has established the ability of assigned goals to improve performance, without rewards for their achievement, through forces including the following: 1) directing attention, 2) energizing activity, 3) affecting persistence, and 4) leading to the arousal, discovery, and/or use of task-relevant capabilities (Locke and Bryan [1969], Locke and Latham [2002]).

Applicable to both RPI and reference points is that motivation to achieve an outcome, such as favorable peer comparison or reaching a performance reference point, depends on perceived attainability of that outcome. Expectancy theory states that motivation is increasing in “expectancy,” or the belief that effort will lead to performance necessary to achieve a desired outcome (Atkinson [1957], Lawler and Suttle [1973], Vroom [1964]). Goal theory similarly states that belief that a goal is attainable is essential to its motivational effect (Erez and Zidon [1984], Locke et al. [1986], Locke and Latham [2002]). Perceived attainability is particularly relevant to the provision of the RPI reference points addressed in this study, which are inherently unattainable for either a large portion or the vast majority of the entire group.

An additional motivating force present in our study and in a variety of corporate and public-sector contexts is a visual indication of approval for performing well relative to a peer reference point (Allcott [2011], Campbell [2002], Vanek Smith [2015]). An example of a visual used in practice is color-coding, with green indicating high and red indicating low performance. Another example is a smiley face for outperforming a reference point. We adopt smiley faces, as used in field experiments from psychology and economics, in
order to test the performance effects of RPI reference points when outperforming them is visually congratulated (Allcott [2011], Schultz et al. [2007]).

The development of Hypotheses 2-4, regarding RPI reference point height, take into account the motivating forces of RPI and reference points described. Hypothesis 1 addresses the effect of providing RPI with a reference point in our study to test the intervention’s validity as a tool for eliciting performance improvement. All of our hypotheses are stated in the alternative form and will be assessed using a two-tailed test.

H1: Providing relative performance information with a congratulated descriptive norm reference point for peer comparison positively affects performance.

To arrive at a prediction of the overall effect of RPI that displays the peer top quartile as opposed to the peer median as a reference point, the development of Hypotheses 2a-c and Hypotheses 3a-c below address its effect in each cross-section of initial performance segmented by the alternative reference points. We predict a concave relationship between an individual’s performance before RPI reference point provision and the performance returns from providing a higher RPI reference point.

That hypothesized concave relationship is based in part on our expectation that the top-quartile reference point will not have an incrementally positive effect relative to the median reference point for initially below-average performers. In fact, expectancy theory and forces of reference points and social comparison even suggest the possibility of
negative returns to a higher reference point for this group. The higher reference point would impose a lower value on expectancy, a construct positively related to motivation, for individuals below both alternatives (Atkinson [1957]). Also, loss-aversion from negative deviation from reference points is greater the closer one is to the reference point (Kahneman and Tversky [1979]). Loss-aversion is an effort-motivating force that would be weaker with a more distant reference point as long as the individual is similarly interested in surpassing either (Abeler [2011], Heath et al. [1999]). In terms of utility from congratulations for exceeding an RPI reference point, the lower reference point would offer greater returns to marginal effort to individuals who would otherwise maintain their percentile rank or change it less than enough to reach the top-quartile. Individuals would be able to project this by viewing a legend provided with the RPI indicating that a smiley face is displayed upon exceeding the reference point.

From a social comparison theory standpoint, raising the reference point will increase the height of the upward social comparison for this group. Social comparison of performance involving great upward distance has been shown to cause discouragement and hurt performance (Rogers and Feller [2015]). Individuals who are below average because they struggle to interact effectively with the course might also resort to ineffective or unsustainable strategies for doing so (Hannan et al. [2008]). Research on tournaments and rank-based-pay also suggests that individuals are discouraged to the point of giving up when they feel that an explicitly rewarded reference point is too high (Bandiera et al. [2013]). A similar force may apply to peer comparison reference points. Further, the upward attractive force of the peer median reference point may be uniquely powerful due
to its relevance to comparison of oneself to what is average given the common desire to consider oneself above-average (Dolan et al. [2012], Larrick et al. [2007]).

On the other hand, goal and reference point research implies some advantage of the higher reference point among initially below-average performers. Goal theory shows that, subject to goal commitment, a higher goal elicits greater performance (Locke and Latham [2002]). Also, individuals who surpass the median might be drawn higher still by the higher reference point. This concept is in line with utility from eliminating feelings of loss-aversion for performance below a reference point, and with desires to reach a congratulated level of performance (Abeler [2011], Heath et al. [1999], Reno, Cialdini, and Kallgren [1993]). Although theoretical implications for the effect of providing the top-quartile as opposed to median point for individuals initially below both are mixed, we predict a slightly negative performance effect among these individuals.

In the partition between the two alternative reference points, forces from goal theory and expectancy theory are more aligned with a positive effect of a higher reference point than in the case of below-average performers. The in-between performers are more likely to view the higher reference point as attainable, and therefore to be committed to it as a goal (Locke and Latham [2002]). In terms of expectancy theory, the belief that effort will lead to performance necessary to reach the reference point would be higher for these individuals than for those who are below average (Atkinson [1957]). Expectancy theory also explains that one of the necessary components of motivation to exert effort is “valence,” or satisfaction from achieving a level of performance necessary to receive a reward (Atkinson [1957], Lawler [1968]). Valence, in the case of RPI reference points, would most closely map to utility from reaching reference points and goals (Abeler et al.
The prospect of receiving a congratulatory smiley face would plausibly enhance the motivational force of that valence. For in-between performers, valance from reaching a reference point would be available for the entire group when shown the higher of the two, but only for those who fall beneath the lower when shown the lower of the two. A possible countervailing force from social comparison theory that might work to the advantage of providing in-between performers with the median reference point is nearly universal interest in outperforming the median (Dolan et al. [2012], Sedikides, Gaertner, and Toguchi [2003], Svenson [1981], Windschitl, Kruger, and Simms [2003]). Such interest in the median reference point might result in more engagement in social comparison. Engagement in social comparison generally drives positive performance effects of RPI (Tafkov [2013]). We propose analyzing individuals’ interest in comparing themselves to the median and top-quartile reference points as part of the additional analysis that would draw from a survey and from data on RPI access. The balance of related theory and empirical work, though, suggests that the effect of providing the top-quartile reference point as opposed to the median will be positive for this group, and especially more positive than for initially below-average performers.

Individuals in the top quartile of performance view a reference point below their position in the distribution in both the case of median and top-quartile RPI reference point display. Neither the top-quartile nor the median reference point, then, offers valence in the form of changing one’s initial performance relative to the RPI reference point. A few qualified forces would work to the positive performance effect of providing the higher reference point to this group. Top-quartile performers are more likely to fall beneath the
higher than the lower reference point. The greater prospect and instance of that event might incrementally motivate this group. Also, to the extent that peer-comparison reference points carry downward attractive power, a higher reference point might act as a bulwark to mitigate a resulting performance decline (Schultz et al. [2007]). Evidence of a downward attractive force, though, is mixed (Allcott [2011], Harper et al. [2013]). As with the hypotheses regarding the bottom three quartiles, greater innate interest in the median may drive social comparison and related performance among top-quartile performers when shown the median reference point. We will test in the additional analyses whether such interest persists among top-quartile performers. We predict a positive effect of providing the top-quartile as opposed to the median reference point among initially top-quartile performers despite the noted limitations of forces working toward that effect. We predict that performers initially between the median and top-quartile, those theoretically more uniformly benefitted by viewing the top-quartile reference point, will exhibit greater positive effects from its display than will performers initially in the top quartile.

Hypotheses 2a-c predict the effect of the top-quartile as opposed to median reference point in each segment of initial performance addressed above. These predictions include effect directions despite the noted possibility of forces working in the opposing direction. Identifying the direction and significance of effects among segments helps in understanding the nature of the concavity that Hypotheses 3a-c outline. Identifying the nature of any observed concavity, in turn, deepens understanding of the overall effect of the higher reference point, which Hypothesis 4 addresses.
H2a: Presenting the peer top quartile, as opposed to the peer median, as a congratulated reference point for performance negatively affects the performance of individuals who are initially below both reference points.

H2b: Presenting the peer top quartile, as opposed to the peer median, as a congratulated reference point for performance positively affects the performance of individuals who are initially between both reference points.

H2c: Presenting the peer top quartile, as opposed to the peer median, as a congratulated reference point for performance positively affects the performance of individuals who are initially above both reference points.

H3a-c address a concave relationship between initial performance and returns to a higher reference point, which H2a-c collectively imply. H3a predicts a relatively more positive effect of the top-quartile as opposed to the median reference point in the segment of performers initially between these alternative reference points than in the outer two segments. H3b (H3c) predict a more positive effect in the in-between segment than in the lower (higher), helping to further define the shape of performance returns to a higher reference point along the scale of initial performance.
H3a: The performance effect of presenting the peer top-quartile, as opposed to the peer median, as a congratulated reference point for performance is more positive for those who are initially between median and top-quartile performance than for those who are not.

H3b: The performance effect of presenting the peer top-quartile, as opposed to the peer median, as a congratulated reference point for performance is more positive for those who are initially between median and top-quartile performance than for those who are initially below median performance.

H3c: The performance effect of presenting the peer top-quartile, as opposed to the peer median, as a congratulated reference point for performance is more positive for those who are initially between median and top-quartile performance than for those who are initially above top-quartile performance.

H4 addresses the average effect of providing the top-quartile as opposed to median reference point. H2a predicted a negative effect among the half of individuals initially below average. H2b predicted a positive effect among the quarter of individuals who are initially between alternative reference points. H2c and H3c collectively predicted a positive effect for the quarter of individuals above both alternative reference points that is less positive than it is for the quarter of individuals initially in-between the alternatives. H2a-c and H3a-c might involve effects in segments of initial performance that balance out, or alternatively that yield a directional net effect. We test for a net effect, but do not
predict its direction given the directional opposition of predicted effects along the scale of initial performance.

H4: Presenting the peer top quartile, as opposed to the peer median, as a congratulated reference point affects performance.

The proposed study would most directly contribute to a growing body of economic, psychology, accounting and management literature on the distinct effects of RPI, or those separate from pay for or visibility of relative performance (Blanes i Vidal and Nossol [2011], Hannan et al. [2008], Harper et al. [2013], Tafkov [2013]). While such applications of RPI often include a reference point for peer comparison, little, if any, empirical evidence has established the effects of RPI reference point height (Allcott [2011], Harper et al. [2013]). Testing the performance effects of alternatively high reference points, and how the effects depend on whether one’s initial performance falls beneath, between, or above both alternatives would help identify the importance of reference point choice. Identifying heterogeneity in the effect of alternatively high reference points depending on initial performance would also inform how to effectively assign reference points based on initial performance when varying reference points among individuals is possible.

The results would also isolate effects of comparison to peers through reporting from comparison through social interaction. A body of economic research addresses
comparison through social interaction (Hanushek et al. [2003], Lavy et al. [2012], Lin [2010], Lyle and Smith [2014]). Analyses of plausibly exogenous changes in peer group composition show that exposure to high performing peers leads to improvement in one’s own performance, with some evidence that very high or very low performing peers carry disproportional weight in influencing one’s own performance (Lavy et al. [2012], Lazear [2001], Hoxby and Weingarth [2006]). Fundamental to the performance effects in these studies are social interactions that involve assistance from high-performers, networking, and learning through observation (Lavy et al. [2012], Lyle and Smith [2014]). The proposed study focuses on displaying an anonymous standard of peer performance as opposed to altering one’s peer-group or mentors. We will compare the results of reporting anonymous reference points of high peer-performance tested in our study with results from this related line of research on the effects of interacting with or observing peers and mentors. Doing so will offer insight into whether as strong a relationship between high peer-performance and own performance holds through reporting alone.

Identifying the distinct performance effects of RPI reference point height might also inform theory and empirical work in economic and accounting research regarding RPI-related mechanisms. For example, target-setting literature notes the apparent disconnect between the prescriptions of goal theory that goals should be attainable infrequently with the prevalence of frequently achievable targets set in organizations (Ioannou, Serafeim, and Li [2014], Merchant and Manzoni [1989]). Targets often either explicitly contain relative performance information or at least allow inference of one’s relative performance (Aranda et al. [2014], Bol, et al. [2010], Merchant and Manzoni [1989], Murphy [2000]). A partial explanation for the prevalence of highly attainable targets, then, might be that
RPI reference points have optimal performance effects at more attainable levels than goals. The current proposal would assess performance effects of raising RPI reference point to levels of attainability beyond which goals exhibit positive returns to incremental difficulty (Erez et al. [1985]). Finding weaker returns would suggest less of a disagreement between behavioral responses to RPI reference points and the prevalence of attainable targets found in practice than has been identified as existing between such targets and the prescriptions of goal theory.

Also, illustrating the distinct performance effects of RPI reference point height could help in identifying when the use of supervisor discretion to impose easier targets that involve RPI would be either damaging or beneficial. For instance, supervisors set lower RPI-based targets to mitigate fairness concerns and to avoid confrontation costs with higher-level managers (Bol et al. [2010]). Information on the initial relative performance of these groups, paired with the results of the current study on RPI reference point height, would indicate whether a lower reference point is likely to improve or attenuate performance of individuals for whom supervisors use discretion to provide easier RPI-based targets.

The literature on tournaments and rank-based pay evidences that those who are performing very well or poorly compared to a rewarded relative performance mark do worse when they notice the distance (Asch [1990], Hannan et al. [2008], Casas-Arce and Martinez-Jerez [2009]). The current study’s predictions would imply that increasing the height of an RPI reference point included in interim performance information would have limited benefits and perhaps even some negative performance effects for individuals at extremes of the performance distribution.
Research on employee evaluations notes the prevalence of forced-curves, and associated RPI reference points. 29% of corporations in a Corporate Executive Board survey reported using forced-curve employee rankings for performance management (McGregor [2013]). Some systems include peer quartiles as RPI reference points, as are tested in the proposed study (Grote [2005]). Findings of how individuals respond to viewing alternatively high standards of performance could inform how changing peer-performance reference points for ordinal evaluation affects performance through means other than explicit incentives attached to the reference points.

At the corporate executive level, financial statements list peer-group composition along with executive pay relative to target percentiles of the peer group (Bebchuk and Fried, [2004], Bizjak et al. [2011]). The SEC has proposed requiring companies to disclose both the percentile of the CEO’s pay and a standardized measure of the company’s performance relative to the compensation peer group (Securities and Exchange Commission [2015]). While the disproportional prevalence of targets above the peer median has drawn widespread criticism, pay and compensation targets set above the peer top-quartile might elicit performance improvement worth compensating at that level at least for some members of the performance distribution (Bizjak et al. [2011]). Research on behavioral responses of comparison to the peer median and peer top-quartile for individuals in cross-sections of initial performance would help in identifying categories of initial performers that might most positively be affected by being held to the higher reference point, although those dynamics should be considered along with financial and career concerns.
Finally, an additional body of accounting literature to which the proposed study contributes is that regarding the format of performance information reports. Studies show that order, categorization, visibility, and other display characteristics of information included in performance reports influence decisions made in stock trading and in employee evaluation (Bloomfield et al. [2006], DeBusk, Brown, and Killough [2003], Dilla and Steinbart [2005], Maines and McDaniel [2000]). Although performance reports are also often provided to individuals to aid in improving their own performance, little evidence shows how formatting the information differently alters performance effects (Yigitbasioglu and Velcu, [2012]). Identifying the effect of providing RPI along with alternatively high reference points would provide initial evidence on affecting performance through performance report formatting.

3. DATA

HarvardX, an online course provider, serves as the setting for the proposed field experiment. In 2012, MIT and Harvard University jointly founded edX, an online course platform that offers free online courses, assessments, and certificates of completion for higher-education-level courses. HarvardX provides course offerings led by Harvard University faculty members on the edX platform. Enrollment is open globally and with no prerequisites or application process, and all instruction occurs online. Course topics range from literature to statistics, and are open for periods ranging between a few weeks to a full year.
Four courses approved for the experiment range in typical enrollment from 6,000 to 25,000 students. A large proportion of students do not participate after enrolling in these freely offered courses; roughly 40% in the last completed round of each course never accessed it. We exclude from our study students who enroll in but do not access the course to avoid including in our estimated effect those who were unlikely to receive the treatment. The approved courses run between the spring and fall of 2016. Three are available both in the spring and late summer. Running the experiment as proposed in this section and as necessary to test the main hypotheses, listed in section 2, on students in the spring provides several advantages: increased sample size, establishing the consistency of results, and identifying the potential advantage of restricting our sample to a smaller group of more-active students in some courses to look for potentially sharper results. We would postpone data analysis until after the in-principle acceptance determination. We project having a sufficient number of students, roughly 12,000, to provide 80% power for testing for testing our main hypotheses in the late summer and fall after receiving and seeking IRB approval for any suggested changes to the experiment design.

The proposed experiment includes a control group, which will receive no RPI display, and two treatment groups, one that will receive an RPI display with the peer median reference point, and one that will receive an RPI display with the peer top-quartile reference point. The reference point in a display will be labeled either “classmate median” or “classmate top quartile” to accurately reflect the point in the performance distribution that it represents. A student receiving a display will thus be aware which of those two standards he or she is being compared to. The experimental intervention is to provide students access for a period of at least two months to a personal RPI display with a
reference point. Waiting for data from at least one week of student activity in each course before the experimental intervention begins allows establishing top-quartile and median reference points. We will deliver the RPI to each of the two treatment groups using weekly emails with a link to the RPI data displays. The displays will be updated and available daily. We have included, as a separate document, sample RPI displays and associated emails and surveys. The document contains an example RPI display in a webpage. The background, text colors, and text position around the display may change slightly, but will be standard throughout the experiment and for both treatment groups.

To increase exposure to the intervention, we have permission in three courses to place links reading, “Check your progress” within the course platform. Control group students clicking the link will be directed to the default HarvardX progress chart for the course, showing completion of individual assignments. Treatment group students clicking the link will be directed to their personal RPI display, below which sits a link to the default HarvardX progress chart for the course.

Over the previous academic year, we have conducted five field experiments in HarvardX courses of similar size in which we tested other elements of performance feedback other than those addressed in this proposal. We are thereby familiar with the technology for gathering and displaying performance data and performing analyses of that intervention’s effects. We have received IRB approval of the experimental intervention of varying RPI reference point height to test H1-H4. We are currently seeking approval for one of the interventions described in the additional analysis section, as explained in detail there.
Available quantitative data include each student’s number of clicks on course content, number of days on which they were active, number of video views, number of discussion forum posts, and several other measures of activity in the course. Qualitative data include students’ responses to registration information as well as pre-course surveys, elements of which we incorporate in the additional analyses.

Definitions of dependent, independent, and moderator variables appear in Table 1. We will use “Activity Level” as opposed to grades, which are based on self-assessment in some of the experiment’s host courses, as the measure shown in RPI displays in order to keep the main dependent variable objective. We define “Activity Level” in Table 1. Our main dependent variable is “Δ Activity Level,” defined as Activity Level for the student at the experiment end date minus Activity Level for the same student at the experiment start date. Following is a restatement of each hypothesis with a list of the variables and samples used for its testing. In the case of the main explanatory variable being an interaction term, we will include its component variables in the regression. Unless otherwise indicated, the sample for each test contains all students who were active in the course in both the control group and the treatment groups. All tests include course fixed effects to filter out the effect of any imbalances in subsample size across treatment and control groups due to variance in the number of students assigned to an experimental group who access the course and thereby qualify for inclusion in the analysis.

H1: Providing relative performance information with a congratulated descriptive norm reference point for peer comparison positively affects performance.
Dependent Variable: $\Delta$ Activity Level

Explanatory Variable: Relative Performance Information = 1

Sample: All experiment groups

H2a: Presenting the peer top quartile, as opposed to the peer median, as a congratulated reference point for performance negatively affects the performance of individuals who are initially below both reference points.

Dependent Variable: $\Delta$ Activity Level

Explanatory Variable: Top-quartile Reference Point

Sample: Relative Performance Information = 1 & Initially Below Average = 1

H2b: Presenting the peer top quartile, as opposed to the peer median, as a congratulated reference point for performance positively affects the performance of individuals who are initially between both reference points.

Dependent Variable: $\Delta$ Activity Level

Explanatory Variable: Top-quartile Reference Point

Sample: Relative Performance Information = 1 & Initially Third Quartile = 1
H2c: Presenting the peer top quartile, as opposed to the peer median, as a congratulated reference point for performance positively affects the performance of individuals who are initially above both reference points.

Dependent Variable: $\Delta$ Activity Level
Explanatory Variable: Top-quartile Reference Point
Sample: Relative Performance Information = 1 & Initially Top Quartile = 1

H3a: The performance effect of presenting the peer top-quartile, as opposed to the peer median, as a congratulated reference point for performance is more positive for those who are initially between median and top-quartile performance than for those who are not.

Dependent Variable: $\Delta$ Activity Level
Explanatory Variable: Top-quartile Reference Point * Initially Third Quartile
Sample: Relative Performance Information = 1

H3b: The performance effect of presenting the peer top-quartile, as opposed to the peer median, as a congratulated reference point for performance is more positive for those who are initially between median and top-quartile performance than for those who are initially below median performance.
H3c: The performance effect of presenting the peer top-quartile, as opposed to the peer median, as a congratulated reference point for performance is more positive for those who are initially between median and top-quartile performance than for those who are initially above top-quartile performance.

H4: Presenting the peer top-quartile, as opposed to the peer median, as a congratulated reference point affects performance.
4. ANALYSIS

The proposed analytical strategy to make best use of the data, including null results, draws both from Null Hypothesis Significance Testing (NHST) and Bayesian analysis. We will conduct NHST using ordinary least squares regressions for each hypothesis. In the event of failure to reject a null hypothesis, or an alternative that predicts no relation, we will conduct Bayesian analysis indicating how much more likely a significant relation is than it was before the realization of the data.

Research on data analysis in online education addresses the challenge of zero-inflation from the large percentage, nearly half, of enrollees who do not access the course, as well as the influence of high outliers who utilize the course approximately ten times more than the 99th percentile (Lamb et al. [2010]). We exclude from our study students who enroll in, but do not access, the course. We will also winsorize values for the dependent variable, Activity Level, in its 99th percentile to ensure that a small number of extreme outliers do not drive or offset results. We will cluster standard errors at the student level to correct for auto-correlation from students who enroll in more than one course hosting the experiment. Any students present in more than one course hosting the experiment will be included in the same experimental group in all courses, and their experimental group membership will, as with all students, be set through random assignment.

We project sample sizes that will provide sufficient power (0.80) for finding significance at the 5% level. We base anticipated effect sizes on previous field experiments in which we sent similar RPI to students. We use the most recent instances of
the courses for which we have approval in estimating variance. We apply our sample
selection to those courses and use the resulting sample sizes as our per-course projections.

*Power Analysis*

H1:
Projected N >= 12,000
Projected N for Sufficient Power = 4,113 [std. dev. = 54, mean difference = 5, sampling
ratio = 2:1]

H2a:
Projected N >= 4,300
Projected N for Sufficient Power = 3,456 [std. dev. = 42, mean difference = 4, sampling
ratio = 1:1]

H2b:
Projected N >= 1,850
Projected N for Sufficient Power = 1,592 [std. dev. = 57, mean difference = 8, sampling
ratio = 1:1]

H2c:
Projected N >= 1,850
Projected N for Sufficient Power = 1,822 [std. dev. = 61, mean difference = 8, sampling ratio = 1:1]

H3a:
Projected N >= 8,000
Projected N for Sufficient Power = 6,800 [std. dev. = 51, mean difference = 7, sampling ratio = 3:1]

H3b:
Projected N >= 6,150
Projected N for Sufficient Power = 3,852 [std. dev. = 47, mean difference = 9, sampling ratio = 2:1]

H3c:
Projected N >= 3,700
Projected N for Sufficient Power = 3,608 [std. dev. = 59, mean difference = 11, sampling ratio = 1:1]

H4:
Projected N >= 8,000
Projected N for Sufficient Power = 5,094 [std. dev. = 51, mean difference = 4, sampling ratio = 1:1]
5. ADDITIONAL ANALYSIS

Proposed additional analyses address the dependence of the effect of reference point height on gender, interest in comparison to the reference point, and confidence in one’s ability to reach the reference point. Each of these factors has been shown to influence performance responses to performance feedback and/or peer comparison. The proposed experiment’s intervention allows tracing the moderating effect of each factor under conditions unique relative to extant empirical evidence. In particular, our setting allows testing how the theories underlying moderating effects apply without altering peer groups and in the absence of financial rewards tied to relative performance. We also propose testing provision of multiple reference points in the same display relative to providing only one.

The additional analyses regarding gender, confidence in ability to reach the reference point, and interest in comparison of one’s performance relative to the reference point draw from a subsample of experiment subjects who provide information necessary for each in the registration or survey process. The ability to test the provision of multiple reference points in the same display would be subject to sample size that might be gained through approval to run the experiment in additional courses. These topics are addressed in the additional analysis section because we are less certain of achieving the effect size and necessary statistical power to detect them.
Gender

Studies of social interaction point to differences in the cooperativeness and benefit from interactions dependent on gender (Eagly [1978], Cross and Madson [1997]). Recent evidence suggests females exhibit more positive performance effects from interaction with high performing peers than do males (Lavy et al. [2008]). We propose testing whether females similarly benefit more from comparison to a high reference point of peer-performance, or whether such gender-based heterogeneity is weaker when the element of social interaction in peer comparison is weaker. Three other background variables that the majority of students provide information on are age, country (which we code as developed or developing), and level of education. We would include these and their interactions with RPI reference point height as controls in testing the moderating effects of gender.

Visibility of Additional Reference Points for Peer Performance

We would create an additional experiment group to be shown performance relative to the bottom quartile, median, and top quartile of peer performance. This would identify how individuals respond to viewing their position in a performance distribution separate from having a certain reference point set and congratulated. In order to do so, we would not provide a congratulatory signal (smiley face) to this group for surpassing a percentile. We would compare the performance of this treatment group to that of the top-quartile and median reference point treatment groups. If we were able to obtain a sample size of a few thousand students in each group, we would also conduct the comparison for individuals initially below, between, and above the median and top-quartile reference points.
Confidence in Ability to Reach the Reference Point

Research on explicitly rewarded performance cutoffs shows that individuals who determine they are very far from reaching the cutoffs then reduce performance. One source of the effect is a sense of futility that leads to reduced effort (Bandiera et al. [2013]). Another is switching to riskier or unsustainable strategies (Hannan et al. [2008]). Marginal utility from pay itself would be zero for individuals who near a rewarded cutoff and fall short of it. Reducing distance from an RPI reference point that is not reached, though, may yield marginal utility in the form of a more positive self-image (Garcia and Tor [2007]). When reference points are RPI-based and not explicitly rewarded, then, individuals may not so readily give-up or resort to all-or-nothing strategies. In such contexts, confidence in one’s ability to surpass a reference point may not be so critical to improvement upon viewing it.

We propose distributing a survey regarding confidence in one’s ability to outperform the peer median and top-quartile reference points. We would test whether perceived attainability is a strong moderator of positive performance responses. We would also test whether the higher reference point is perceived as significantly less attainable. Doing so would evaluate a potential downside of providing a high reference point in the form of diminished confidence in its attainability, and would also indicate whether individuals’ level of confidence is important in assigning the most effective RPI reference point to them.
Interest in Comparison of One’s Own Performance to the Reference Point

A great deal of evidence shows that individuals are interested in comparing themselves to what is average and desire to consider themselves above average (Dolan et al. [2012], Sedikides, Gaertner, and Toguchi [2003], Svenson [1981], Windschitl, Kruger, and Simms [2003]). However, less has been shown regarding comparison to the top quartile. Gauging relative interest in comparison of oneself to the top quartile could help in explaining the performance effects of the alternative reference points in our study. We would administer a survey regarding interest in comparison to, and in surpassing, the median and top-quartile reference points. The data also include every time that a student accesses their RPI display. We would use the survey responses and the measure of RPI access in order to assess relative interest in the two reference points. We would also seek to identify variation in that interest by whether an individual’s performance was initially below, between, or above the top-quartile and median reference points.

6. PILOT STUDY

We offer field experiments that we ran during the 2014-2015 academic year at HarvardX as proof of concept for this experiment’s operationalization. We provided students with representations of their performance compared to their peers. We tested reference point comparisons other than those in this proposal, the number of measures provided, providing students with information on their relatively strong or weak points, and student choice of measures.
As we will for the proposed experiment, we created code for generating displays of RPI and emailed display links to students. We will enhance the code to operate more independently of manual data curating, but we will otherwise apply the same operations used in our prior experiments.

In addition to providing much of the technological infrastructure for the current proposed experiment, these experiments helped us in gauging likely sample sizes and effect sizes. While we will reserve announcing their results for working papers, the relative performance information displays in those studies showed statistically significant effects on performance by the analytical methods in the proposed study.
References


CASAS-ARCE, P. and F. A. MARTINEZ-JEREZ. 'Relative Performance Compensation,
GORMAN, A. 'How One Hospital Reduced Unnecessary C-Sections.' The Atlantic, 2015. Available at: http://www.theatlantic.com/health/archive/2015/05/how-one-hospital-reduced-unnecessary-c-sections/392924/
HANNAN, L.; G. MCPHEE; A. NEWMAN; and I. TAFKOV. 'The Effect of Relative


Table 1: Variable Definitions

<table>
<thead>
<tr>
<th>Dependent Variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity Level</td>
<td>The following weighted sum, that scales each type of action’s mean to approximately the mean of video views: video views + 1.5 x problem attempts + 20 x forum posts + 2.5 x other forum actions + 5 x number of days active in the course + 0.1 x total actions</td>
</tr>
<tr>
<td>Δ Activity Level</td>
<td>Activity Level at the experiment’s end minus Activity Level at the experiment’s beginning</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dependent Variable Components</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video Views</td>
<td>The number of times a student started watching a video</td>
</tr>
<tr>
<td>Problem Attempts</td>
<td>The number of times a student entered an answer to any problem</td>
</tr>
<tr>
<td>Forum Posts</td>
<td>The number of posts a student made in discussion forums</td>
</tr>
<tr>
<td>Other Forum Actions</td>
<td>The number of actions (e.g., voting for a post, responding with a comment to an original post) a student took in discussion forums</td>
</tr>
<tr>
<td>Number of Days Active in the Course</td>
<td>The number of calendar days on which a student accessed the course</td>
</tr>
<tr>
<td>Total Actions (component of Activity Level)</td>
<td>All actions in the course that are recorded electronically; these include video views, problem attempts, forum posts, and other forum actions, but are not limited to them</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>An indicator variable equal to one if the individual is a member of the control group, which does not receive RPI displays.</td>
</tr>
<tr>
<td>Median Reference Point (RPI_M)</td>
<td>An indicator variable equal to one if the individual is a member of the treatment group that receives an RPI display with the peer median reference point; the peer median is the median activity level of individuals who have accessed the course</td>
</tr>
<tr>
<td>Top-quartile Reference Point (RPI_T)</td>
<td>An indicator variable equal to one if the individual is a member of the treatment group that receives an RPI display with the peer top-quartile reference point; the peer top-quartile is the top-quartile activity level of individuals with who have accessed the course</td>
</tr>
<tr>
<td>Relative Performance Information (RPI)</td>
<td>An indicator variable equal to one if either Top-quartile Reference Point = 1 or Median Reference Point = 1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Moderator Variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category</td>
<td>Description</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Initially Below Average</td>
<td>An indicator variable equal to one if the individual’s activity level was less than or equal to the median of all individuals who had accessed the course at the experiment’s start</td>
</tr>
<tr>
<td>Initially Third Quartile</td>
<td>An indicator variable equal to one if the individual’s activity level was greater than the median and less than the top-quartile of all individuals who had accessed the course at the experiment’s start</td>
</tr>
<tr>
<td>Initially Top Quartile</td>
<td>An indicator variable equal to one if the individual’s activity level was greater than the top-quartile of all individuals who had accessed the course at the experiment’s start</td>
</tr>
<tr>
<td>Gender</td>
<td>An indicator variable equal to one if the individual indicated their gender in the course registration process and chose Female, and equal to zero if the individual indicated their gender in the course registration process and chose Male.</td>
</tr>
<tr>
<td>Developed Country</td>
<td>An indicator variable equal to one if the individual indicated their country of residence and the country is of UN Developed Nation status, and equal to zero if the individual indicated their country of residence and the country is of UN Developing Nation Status</td>
</tr>
<tr>
<td>Level of Education</td>
<td>An indicator variable equal to one if the individual indicated their level of education and has a bachelor’s degree or higher degree.</td>
</tr>
<tr>
<td>Age</td>
<td>The age, if any, that an individual indicated during registration, truncated at 5 and 100.</td>
</tr>
<tr>
<td>Attention to Relative Performance Information</td>
<td>The number of times an individual follows a link to view their RPI display (used in additional analysis regarding students’ interest in the RPI reference point displayed)</td>
</tr>
<tr>
<td>Survey Response Variables</td>
<td>Answers to standard HarvardX, and to experiment-specific, surveys are coded as shown in the self-contained Experiment Instrument Samples document.</td>
</tr>
</tbody>
</table>
The following are outlines of tables and figures for basic descriptive statistics and for the tests described in Section 2: Theory and Hypothesis Development. Additional tables and figures would be added for tests described in Section 5: Additional Analysis, and potentially beyond that to more fully illustrate the data in general.

<table>
<thead>
<tr>
<th>Panel A: Sample Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Enrollment</td>
</tr>
<tr>
<td>Exclude Students who did not Access the Course</td>
</tr>
<tr>
<td>Final Sample</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Descriptive Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>--------------------------------</td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Level of Education</td>
</tr>
<tr>
<td>Developed Country</td>
</tr>
<tr>
<td>Familiarity With Subject</td>
</tr>
<tr>
<td>Commitment to Complete Course</td>
</tr>
<tr>
<td>Number of Online Courses Previously Enrolled In</td>
</tr>
<tr>
<td>Number of Online Courses Previously Completed</td>
</tr>
<tr>
<td>Activity Level at Course End</td>
</tr>
<tr>
<td>Δ Activity Level</td>
</tr>
<tr>
<td>Table 3: Correlation of Descriptive Statistics</td>
</tr>
<tr>
<td>----------------------------------------------</td>
</tr>
<tr>
<td>I    II   III  IV   V   VI   VII  VIII IX  X</td>
</tr>
</tbody>
</table>
| I Gender  
II Age  
III Level of Education  
IV Developed Country  
V Familiarity With Subject  
VI Commitment to Complete Course  
VII Number of Online Courses Previously Enrolled In  
VIII Number of Online Courses Previously Completed  
IX Activity Level at Course End  
X Δ Activity Level |
<table>
<thead>
<tr>
<th>Country</th>
<th>Number of Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developed</td>
<td></td>
</tr>
<tr>
<td>U.S.A.</td>
<td></td>
</tr>
<tr>
<td>Developing</td>
<td></td>
</tr>
<tr>
<td>Ghana</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Tabulation by Country
<table>
<thead>
<tr>
<th>Init. Below Average</th>
<th>Init. Third Quartile</th>
<th>Init. Top Quartile</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPI_M</td>
<td>RPI_T</td>
<td>Control</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

**Fig. 1**

**Δ Activity Level**
Table 5: Distribution of Δ Activity Level

<table>
<thead>
<tr>
<th></th>
<th>Number of Students</th>
<th>Δ Activity Level</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RPI_M</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initially Below Average</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initially Third Quartile</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initially Top Quartile</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>RPI_T</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initially Below Average</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initially Third Quartile</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initially Top Quartile</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initially Below Average</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initially Third Quartile</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initially Top Quartile</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 6: Average Effect of RPI and Partitioned Effects of Reference Point on \( \Delta \) Activity Level

<table>
<thead>
<tr>
<th>RPI</th>
<th>RPI_T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course Fixed Effects</td>
<td>Yes</td>
</tr>
<tr>
<td>Bayes Factor</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Clustering</td>
<td></td>
</tr>
<tr>
<td>Sample</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

\( \Delta \) Activity Level

- All
- RPI & Init. Below Average
- RPI & Init. Third Quartile
- RPI & Init. Top Quartile
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>(\Delta) Activity Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPI_T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Init. Third Quartile</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPI_T \times Init. Third Quartile</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course Fixed Effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Bayes Factor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clustering</td>
<td>Student</td>
<td>Student</td>
<td>Student</td>
<td>Student</td>
<td></td>
</tr>
<tr>
<td>Sample</td>
<td>RPI</td>
<td>RPI</td>
<td>RPI &amp; (Init. Third Quartile or Init. Below Average)</td>
<td>RPI &amp; (Init. Third Quartile or Init. Top Quartile)</td>
<td></td>
</tr>
</tbody>
</table>

Table 7: Average Effect of Top-Quartile as Opposed to Median Reference Point, and Concavity of that Effect in Initial Performance