

# Relative Performance Feedback with Bounded Peer Comparison: Evidence from a Field Experiment

Henry Eyring  
Duke University, Fuqua School of Business

August 2025

## *Abstract*

This study uses a field experiment set in online professional education to examine whether relative performance feedback (RPF) has a more positive effect when the scope of ranks provided is restricted. Providing RPF on a randomly selected subset of measures yields higher performance than providing no RPF or RPF on the full set of measures. Thus, there are negative returns to displaying additional measures for peer comparison. The benefit of constraining RPF is consistent with theory that information overload impedes learning; when RPF shows only a subset of ranks, this yields roughly twice the increase in accuracy. The results also document cross-sectional variation. When the subset of ranks shown are more predominantly the recipient's own lower ranks, this boosts (damps) grades for higher (lower) initial performers, and lower performers reduce their actions by roughly 18%. These results provide evidence on whether and how to restrict access to measures for peer comparison.

**Keywords:** Relative performance feedback; information overload; information processing; social comparison; motivation; performance

**JEL Classifications:** C93, D91, I21, M41

---

I thank V.G. Narayanan, Xu Jiang, Patrick Ferguson, Wim Van der Stede, Shelley Li, and participants in workshops at Harvard Business School and Duke University and at the Columbia Management Accounting Conference for their helpful comments. I thank Charles Zhang, Douglas Muhlestein, and Sean Bennett for their excellent research assistance. I am grateful to HarvardX for the permission to conduct this study's field experiment.

## I. INTRODUCTION

Many studies in accounting and economics have shown that peer comparison can promote motivation and learning (Jackson and Bruegmann [2009], Chen et al. [2010], Tafkov [2013], Song et al. [2017]). Although relative performance feedback (RPF) enables peer comparison and is therefore a valuable tool for organizational control (Hannan, Krishnan, and Newman [2008], Tafkov [2013]), economic models explain that high cognitive load from added information can impede decision-making (DellaVigna [2009], Bordalo, Gennaioli, and Shleifer [2022], [2015]). This study addresses the implications of cognitive bounds for the scope of ranks included in RPF. Using a field experiment set in online professional education, the main tests explore whether it is advantageous to restrict RPF to include only a subset of measures for peer comparison. Further tests address whether there are heterogeneous performance effects of this constraint; social comparison theory suggests that the effects will differ for higher and lower performers in a manner that depends on whether the narrowed subset of ranks is predominantly their own higher or lower ranks (Bandiera, Barankay, and Rasul [2013], Festinger [1954]). Collectively, this study's tests offer initial evidence on whether and how it is useful to place bounds on the measures provided for peer comparison.

This paper extends research on the economic effects of cognitive bounds (Calzolari and Nardotto [2017], Bordalo, Gennaioli, and Shleifer [2022]). The hypotheses draw on theory regarding the usefulness of peer comparison both for learning and for motivation (Chen et al. [2010], Tafkov [2013], Song et al. [2017]). The tests herein provide the first evidence regarding how learning and motivation depend on whether RPF design reduces the number of measures available for peer comparison by displaying only a randomly selected subset. Specifically, analyses regarding cognitive bounds that plausibly limit one's ability to learn from peers through

peer comparison focus on effects of narrowing the scope of displayed ranks per se. These tests document effects on improved accuracy in answering problems that allow multiple attempts. In addition, analyses of theory regarding motivation focus on how responses to peer comparison will differ based on whether the specific subset shown are the recipient's own higher or lower ranks. These analyses document effects related to both the accuracy of problem attempts and to the total actions taken in the courses.

The empirical design uses a field experiment on participants in online professional education to randomly vary which of their ranks they can view. Each participant is randomly assigned to receive one of the following: information on their ranks across a full set of measures; information on their ranks across a randomly selected subset of those measures; or no information on their ranks. The measures in the experimental RPF are expressed as percentile ranks, which is a common type of feedback used in practice and in research that this study builds on (Allcott and Rogers [2014], Grote [2005], Pogones [2017]). Research has documented positive performance effects when peers can observe their ranks, whether the ranks are on input provision, output attainment, or—as with the measures included in this study's intervention—both types of performance (Chen et al. [2010], Kolstad [2013], Tafkov [2013], Song et al. [2017], Guenther, Lehnen, and Rilke [2024]). For example, these positive effects appear when RPF shows ranks by input measures, such as the number of calls made by salespeople or posts made in an online community (Chen et al. [2010], Guenther, Lehnen, and Rilke [2024]), or output measures, such as grades for students, or patient mortality in health care (Azmat and Iriberry [2010], Blanes i Vidal and Nossol [2011], Kolstad [2013]). Moreover, these effects are rooted in social comparison concerns and arise both in the presence and absence of financial incentives (Chen et al. [2010], Blanes i Vidal and Nossol [2011], Tafkov [2013]).

There is substantial evidence that peer comparison facilitates learning and lifts performance (Jackson and Bruegmann [2009], Chan, Li, and Pierce [2014], Song et al. [2017]). Foundational models of learning hold that one of the principal formats of learning and developing aptitude at a task is the observation of information about one's peers (Arrow [1994]). Research has documented that learning occurs when peers observe information about each other, and that such learning can be more important for performance than "learning by doing" (Chan, Li, and Pierce [2014]). When there are observable actions and/or outcomes among peers, this allows an information user to more readily identify alternative strategies and helps best practices to spread (Jackson and Bruegmann [2009], Song et al. [2017]). Evidence on the effects of restricting the information on ranks to a subset, though, is sparse.

On one hand, such a restriction may keep the scope of information within cognitive limits and thereby benefit performance (Bawden [2001], DellaVigna [2009]). Economic models of decision-making predict that incremental information may hamper productivity if the information exceeds cognitive bounds (DellaVigna [2009], Matějka and McKay [2015]). These models posit that a large amount of information can cause high levels of mental stress and, relatedly, impede rather than assist learning (Kruschke and Johansen [1999], Bawden [2001], DellaVigna [2009]).

On the other hand, decision-makers may be rationally inattentive to incremental information, such that incremental information is at least not harmful (Sims [2006]). Furthermore, restricting RPF to a subset may curtail the opportunity to learn from a wider range of information on one's peers, thereby harming performance (Song et al. [2017]). Given the presence of cognitive bounds, it is an empirical question whether it will be beneficial to provide the broadest set of information, or to constrain that set, to best promote learning.

To examine this tension empirically, this experiment randomly assigns RPF to include either a full set of nine measures or a randomly selected subset of four measures. The number nine (four) falls at the upper reach of (within) the “7 +/- 2” limit that Miller [1956] suggests is a cap on the number of cues that fit in short term memory. Research in disciplines spanning economics, information systems, and psychology has drawn on theory regarding these limits on cognitive capacity (Baddeley [1994], Samuels [1994], Saaty and Ozdemir [2003]).

The analyses explore the implications of bounded cognition for peer comparison. Information overload would plausibly impede the learning that occurs from peers comparing themselves on observables. This, in turn, would work against the formation of aptitude at a task (Jackson and Bruegmann [2009], Song et al. [2017]). Theory regarding cognitive bounds raises the possibility that stress and disengagement upon receipt of a larger set of information in RPF could lead to less information processing and related learning than would occur in response to a narrower set (Eppler and Mengis [2004], Bordalo, Gennaioli, and Shleifer [2022]).

Granular data from this study’s field site that tracks every mouse click on course content sheds light on multiple dimensions of performance—both the total number of actions a student takes in the course, and also problem-attempt accuracy. Specifically with regard to accuracy, the data document the accuracy of their individual problem attempts on problems that allow multiple attempts. As students learn from peers and adopt more effective strategies, this would plausibly enable them to gain aptitude and be more accurate in responding to questions. The setting is therefore particularly suited for examining performance effects of constraining the information for peer comparison and whether this helps to enhance learning and related accuracy in attempts at a task.

Drawing on research regarding social comparison and motivation, it is plausible that effects of constraining RPF will vary depending on whether the randomly selected subset of ranks in RPF is more predominantly the recipient's own higher or lower ranks. Social comparison theory suggests that the direction of that focus, toward one's own higher or lower ranks, will matter and that the effects may differ for higher and lower baseline performers (Chen et al. [2010], Song et al. [2017]). Although prior work has examined differences in the response to peer comparison among lower and higher performers (Chen et al. [2010], Bandiera, Barankay, and Rasul [2013]), research has yet to explore how, for lower and higher performers, an increased focus of feedback toward one's own lower or higher ranks affects performance.

There is evidence that higher initial performers can become uninterested in learning or identifying new strategies if they see information on their high ranks relative to peers (Chen et al. [2010]). The analyses in the current study test whether it will be beneficial to provide higher performers with information on their relatively lower ranks, since this information could spur learning for higher performers (Campbell [2006]).

For lower performers, if RPF focuses on their own lower ranks, this could reveal areas with large room for improvement and significant opportunity to adjust strategy. Yet, low performers may become discouraged by this information and relatedly reduce effort provision (Bandiera, Barankay and Rasul [2013]), which would plausibly reduce opportunities for "learning by doing" and the related improvement in strategies that result from trial-and-error (Campbell, Epstein, and Martínez-Jerez [2011]). Thus, there is considerable tension in forming predictions for lower performers. Collectively, theory suggests that there are counterbalancing forces with regard to effects on accuracy for this group, potentially yielding a small or null effect on accuracy. Moreover, theory suggests a negative effect on effort due to discouragement (Bandiera, Barankay,

and Rasul [2013]). Through the latter avenue of reduced effort provision, there is thus some reason to expect a negative net performance effect of constraining feedback if the subset of ranks in RPF are more predominantly a lower performers' own lower ranks.

The main tests document the following results. First, RPF has a more positive performance effect when it consists of a randomly selected subset of four ranks rather than a full set of nine ranks. Second, and consistent with theory regarding information overload and its relationship to learning, individuals provided with the full set of ranks are less accurate in their problem attempts than those who receive a narrower set. Individuals who receive the narrower set of ranks significantly outperform and are more accurate than those who receive no ranks or those who receive the full set of ranks.

Further analyses offer evidence in line with theory regarding social comparison. In particular, higher initial performers achieve higher grades and greater accuracy on problem attempts when the restricted subset of ranks shown is more predominantly their own lower ranks. This is consistent with the idea that higher performers can identify areas of remaining, high returns to effort when they see their own lower ranks (Campbell [2006]). For lower initial performers, on the other hand, there is a negative effect on grade and on the total number of actions taken in the course when the subset shown is more predominantly the individual's own lower ranks. This latter result is consistent with theory that lower performers are more prone to be discouraged by social comparison (Bandiera, Barankay, and Rasul [2013]).

This paper makes three contributions. First, the results extend the accounting literature on the role of cognitive bounds in decision-making (Elliott, Hobson, and White [2015], Campbell et al. [2017], Casas-Arce, Lourenço, and Martínez-Jerez [2017]). Studies in accounting have explored how features of information display and formatting can aid or hamper information

processing (Blankespoor [2019], Blankespoor et al. [2019]). Additionally, accounting research on management control has documented that cognitive bounds can significantly influence the processing of information. For example, these studies have found benefits of providing information less frequently (Casas-Arce, Lourenço, and Martínez-Jerez [2017]), providing information in a single rather than in multiple formats (Eyring, Ferguson, and Koppers [2021]), and providing AI generated suggestions to help interpret information (Costello, Down, and Mehta [2020]). However, it is an open question whether it would be beneficial on net to provide performance RPF on only a subset of ranks from a larger set. The results of this field experiment on peer comparison add to the growing accounting literature on information processing by documenting a positive average effect of restricting the scope of RPF in this manner.

Second, this paper extends the accounting and management research on heterogeneous responses to peer comparison in performance feedback. Prior work has demonstrated that effects of peer comparison vary depending on where one sits in the distribution (Hannan, Krishnan, and Newman [2008], Chen et al. [2010]). This study addresses the unanswered empirical question of how presenting baseline higher (lower) performers with their own higher or lower ranks affects performance. The results demonstrate heterogeneity in responses for baseline higher and lower performers, suggesting that whether RPF should focus on one's own higher or lower ranks depends on the feedback recipient's baseline performance.

Finally, this paper adds to the literature on the difficulty of lifting performance for those with high baseline performance. Research in accounting, management, and economics has shown that this group has difficulty identifying paths to improve and can become complacent with their high rank relative to peers (Chen et al. [2010]). High performers tend to have fewer opportunities to learn from peers (Song et al. [2017]). Attempts to motivate high performers and help them to

learn often face significant challenges and can even backfire, as in the case of target ratcheting (Leone and Rock [2002], Bouwens and Kroos [2011]). Yet, research in accounting suggests that there are still likely opportunities for high performers to improve, especially if they invest in areas where they have invested less and where returns to effort have not diminished as significantly (Campbell [2006]). The analyses herein examine that idea and document a positive effect, both on performance and accuracy, when the subset of ranks included in a higher performer's feedback is more predominantly their relatively lower ranks. This result provides a tool to lift baseline high performers to even higher levels, whereas other studies have shown that numerous other control mechanisms struggle to do so (Leone and Rock [2002], Chen et al. [2010]).

## **2. THEORY AND HYPOTHESES**

There is a large and growing body of research, in disciplines including accounting, management, and economics, on the use of peer comparison in performance feedback to lift performance (Blanes i Vidal and Nossol [2011], Bandiera, Barankay, and Rasul [2013], Tafkov [2013]). This stream of literature has documented positive effects of peer comparison, on learning, motivation, and performance, in settings that include retail, health care, sports, web development, and education (Azmat and Iriberry [2010], Chen et al. [2010], Blanes i Vidal and Nossol [2011], Song et al. [2017]). These effects appear whether RPF shows ranks by input measures, such as the number of calls made by salespeople or posts made in an online community (Chen et al. [2010], Guenther, Lehnen, and Rilke [2024]), or output measures, such as grades for students, or patient mortality in health care (Azmat and Iriberry [2010], Blanes i Vidal and Nossol [2011], Kolstad [2013]). Moreover, there is evidence that peer comparison generates positive net effects on performance

both in the presence and absence of financial incentives (Chen et al. [2010], Blanes i Vidal and Nossol [2011], Tafkov [2013]).

Advances in information technology have led to a surge in data gathering and dissemination within organizations. Relatedly, administrators face more choices in terms of the measures to display for peer comparison (Mercer [2019], Casas-Arce, Lourenço, and Martínez-Jerez [2017]). Simon [1971] framed the paradoxical problem of readily available information this way:

*In an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it. – Simon (1971), pp. 40–41*

In line with this prediction, evidence from many industries suggests that administrators struggle to identify whether and how to narrow the set of information in feedback (Ittner, Larcker, and Meyer [2003], Choi, Hecht, and Tayler [2012], Casas-Arce, Lourenço, and Martínez-Jerez [2017]). If the information user could place zero weight on information that is not helpful for goal achievement, then more information should weakly dominate less information. This idea is consistent with traditional economic models of a Bayesian decision-maker, who has limitless cognitive capacity and can optimally weight the information they receive (Savage [1954], Viscusi [1985]). More recent theory based in behavioral economics, though, notes that cognitive limits influence responses to information (Bawden [2001]). These models hold that decision-makers make trade-offs, and divide their finite attention among information cues. In these models, the

overprovision of information can harm the quality of decisions, impeding the process of learning (Eppler and Mengis [2004]).

Recent research in accounting has drawn on theory regarding bounded cognition when exploring effects of feedback design. Studies in this stream have documented positive effects of providing information less frequently (Casas-Arce, Lourenço, and Martínez-Jerez [2017]), and also of presenting only one type of measure to make that measure more salient (Eyring, Ferguson, and Koppers [2021]). Empirical research has yet to test whether it is beneficial to provide only a subset of ranks for peer comparison in RPF, rather than that same information as part of a larger set of ranks. This is relevant as the range of measures in organizations continues to expand, and as administrators struggle to decide whether to distribute performance feedback on a full set of measures or on a subset (Glasgow, Finlayson, and Ray [2018], Mercer [2019], Zenger and Folkman [2014]). The field experiment in this study provides evidence to examine these issues.

The hypothesis development draws on theory regarding bounded cognition to motivate predictions of how restricting RPF to a subset of ranks will influence performance. These theories explain that, when forced to process quantities of information that exceed their cognitive capacity, people become stressed, lose track of information, and disengage from the source of the information (Bawden [2001]). Research in psychology demonstrates that, due to information processing constraints, large amounts of information impair pattern recognition (Bawden [2001], Eppler and Mengis [2004]), depth perception (Braunstein [1976]), and risky decision-making (Payne [1976]), and induce harmful physiological responses associated with stress and information avoidance (Ettema and Zielhuis [1971], Boyce [1974]).

Following the provision of information on peers, information overload could hamper performance by impeding learning via peer comparison. A significant body of research has

demonstrated positive effects of peer comparison on learning and performance (Blanes i Vidal and Nossol [2011], Tafkov [2013], Song et al. [2017]). Such information allows strategy comparison among peers, which aids strategy evaluation and improvement. Specifically, RPF that shows ranks on output measures can indicate whether peers have a more effective strategy, and this type of benchmarking can, in turn, generate information search and experimentation (Blanes i Vidal and Nossol [2011], Song et al. [2017]). Also, RPF that shows ranks on input measures can disseminate best practices and allow direct comparisons of underlying strategies, including investment in various inputs (Chen et al. [2010], Song et al. [2017]). Research in economics has described this as a type of learning from harnessing the “wisdom of the crowds” (Chen et al. [2010], Muchnik, Aral, and Taylor [2013]).

Arrow [1994] frames learning from observing information about peers as one of the main methods of gaining aptitude, and Chan et al. [2014] find evidence that this method can be more important than trial-and-error or “learning by doing”. There is evidence that peer comparison is an influential tool to aid learning in a wide range of economic and management settings (Bandiera and Rasul [2006], Bayer, Hjalmarsson, and Pozen [2009], Chan, Li, and Pierce [2014], Song et al. [2017]). To the extent that RPF with a full set of ranks overwhelms cognitive limits, the resulting information overload would plausibly appear empirically in the form of reduced learning (Gigerenzer and Goldstein [1996], Barber and Odean [2008]).

Hindered learning about a task would, by extension, reduce skill development (Eppler and Mengis [2004]). In line with prior research, this study measures skill development in terms of an increase in accuracy (Gigerenzer and Goldstein [1996], Campbell, Epstein, and Martínez-Jerez [2011]). In this study’s setting, data on the accuracy of individual problem attempts, where

problems allow more than one attempt, provide the measure of accuracy to proxy for changes in the construct of skill or aptitude at the task.

Yet, there is tension in predicting whether constraining RPF will be advantageous. The literature on rational inattention suggests that decision-makers choose to allocate their limited attention among various cues (Sims [2003], [2006], Matějka and McKay [2015]). If this process involves allocating attention to the most motivating or informative ranks, and ignoring the others, it may be optimal to provide the full set so that individuals can direct their attention to the most useful cues without constraints imposed on the set they can select from.

To examine this tension empirically, the first pair of hypotheses relate to the restriction of RPF to consist of a scope of ranks that fits within cognitive bounds. These hypotheses are motivated by the significant body of theory above to suggest that reduced information overload will enable learning and thereby increase skill and aptitude, as measured using problem-attempt accuracy. Hypotheses 1-2 address these ideas.

H1: There will be a positive effect on grades of restricting the ranks in RPF to a random subset.

H2: There will be a positive effect on problem-attempt accuracy of restricting the ranks in RPF to a random subset.

The choice to restrict ranks in feedback provides an opportunity to keep performance feedback within cognitive bounds. The theory and hypotheses above relate to the benefit of restricting the ranks available for peer comparison per se. To examine cross-sectional variation,

the analyses then take into account the fact that, if only a subset of ranks appear in feedback, this also affords the opportunity to direct attention toward an individual's own higher or lower ranks by selectively including either. Theory regarding social comparison and the weighting of information helps to predict whether a subset of information that more predominantly focuses on one's higher or lower ranks will yield a more positive performance effect.

Economic theory related to salience bias and limited attention suggests that decision-makers may not optimally weight information from a large set (DellaVigna [2009], Bordalo, Gennaioli, and Shleifer [2022]). If there is a more useful subset of cues, the addition of competing cues can reduce the cognitive weight on and associated learning from those more useful cues. This raises the possibility that there will be an incremental benefit from restricting the set of ranks in a report that occurs when the ranks are restricted to the most useful set. Theory regarding peer comparison helps to predict whether the best information to direct attention to will be a person's own lower or higher ranks and how this will differ for higher and lower initial performers.

Among higher initial performers, prior literature suggests two reasons why information for peer comparison can hamper performance improvement. First, high performers exhibit reduced interest in refining and improving on their strategy when they see that they sit well above their peers whether on input or output measures (Chen et al. [2010]; Song et al. [2017]). Second, high performers have fewer opportunities to identify areas where peers are investing more than them if a report does not reveal these. Research in management control has drawn on the theory regarding diminishing returns to effort, a common feature of economic models of effort provision (Demski and Feltham [1976], Holmstrom and Milgrom [1991]), to posit but not test that revealing areas of lower investment compared to peers should help a high performer to invest where returns have not diminished as significantly (Campbell [2006]). To the extent that a high performer's RPF shows

their own higher ranks, this would thereby exacerbate two impediments to performance that plausibly attend the delivery of feedback. Accordingly, H3 and H4 predict a benefit to focusing the subset of ranks more predominantly on higher performers' own lower ranks rather than their own higher ranks.

H3: For initially higher performers, there will be a positive effect on grades of focusing the subset of ranks displayed on their own relatively lower ranks.

H4: For initially higher performers, there will be a positive effect on problem-attempt accuracy of focusing the subset of ranks displayed on their own relatively lower ranks.

Among lower initial performers, there is significant theoretical tension in whether there will be a positive effect when the subset of ranks in a report are their higher or lower ranks. From the standpoint of learning, it may be important for lower performers to identify problem areas where they are significantly below their peers so that they can adapt their strategy (Song et al. [2017]). Yet, if this information is discouraging and relatedly reduces effort provision, this could hamper willingness to engage in trial-and-error, a process that aids learning (Campbell, Epstein, and Martínez-Jerez [2011]). Indeed, there is theory to suggest that reports on one's lower ranks with regard to inputs or outputs may be more likely to diminish beliefs about one's own resolve or ability with respect to the task (Kaur, Kremer, and Mullainathan [2010]). Expectancy theory of motivation suggests that these beliefs are a fundamental component of motivation and that a reduction in such beliefs discourages effort (Lawler and Suttle [1973]). If there is a null effect on learning and a negative effect on effort, these would plausibly combine to yield a negative effect

on performance that results from focusing the subset of ranks on lower performers' own lower ranks. These implications lead to H5 and H6 below.

H5: For initially lower performers, there will be a negative effect on grades of focusing the subset of ranks displayed on their own relatively lower ranks.

H6: For initially lower performers, there will be a negative effect on course activity of focusing the subset of ranks displayed on their own relatively lower ranks.

### **3. FIELD SETTING AND EXPERIMENT DESIGN**

This study's field experiment is set in professional education courses offered by EdX, one of the largest online education providers. Harvard University operated this experiment's host courses.<sup>1</sup> Faculty from Harvard University teach these courses to provide training for professions including data science, biology, law, and medicine.

The field site gathers log files that record each click on course content. This yields data on each student's activity by measures of progress in the course, such as accessing the course on a given day, watching videos, completing graded content, and posting comments in the discussion forum. The experimental RPF shows percentile ranks, an approach to feedback that is common in practice and prior field research (e.g., Allcott and Rogers, [2014], Grote, [2005], Pogones, [2017]).

The sample selection incorporates guidance from research on performance in large-scale online field experiments. Specifically, the sample excludes those who are auditing the course, or

---

<sup>1</sup> The field experiment was approved by an IRB.

those who do not submit graded content, in line with prior research regarding grades in online courses (Emanuel and Lamb [2015]). This restriction of the sample to the active portion of an online community is consistent with methodology in prior studies of feedback in similarly large online settings (Chen et al. [2010]).

The experiment consisted of a treatment group that received experimental RPF and a control group that did not. Roughly one-fourth of the sample was randomly assigned to the control group, which received no RPF. Students in this group received an email with a link to view their grade in the course. The remaining roughly three-fourths of the sample, or those assigned to the treatment group, received an email with that same link provided to the control group. They also received an additional link to the RPF specific to this experiment (i.e., the treatment condition). By preserving three-fourths of the sample to be treated, the study increases power to test variation in the types of treatment provided (e.g., provision of a constrained or unconstrained sets of ranks in RPF, and how the student is ranked on the specific subset provided). It is still valuable to have the one-fourth in the control group that received no ranks, so as to establish how these various RPF conditions compare not only to each other, but to a counterfactual of no RPF provision.

The three-fourths of the sample which received RPF treatment was then split randomly so that roughly one-half of them received an unconstrained set of nine measures included in RPF, and the other half received only a subset of four measures. The set of four was randomly assigned at the individual-level, such that the individuals assigned to receive RPF with only four measures did not all receive the same four measures in their RPF. As a result, the effects of constraining ranks are not due to selection of a particular subset but to the general paring back of a full set of ranks to a subset randomly assigned to that individual. The set a given student received remained constant during the experiment and was displayed in terms of percentile rank for the prior week.

The full set of nine were: the number of days active in the course, the number of clicks on course content, the number of videos viewed, the number of problems attempted, the number of course chapters accessed, the number of discussion forum posts, the number of comments made on others' discussion forum posts, grade in the course at the end of the prior week, and the change in grade in the course during the prior week.<sup>2</sup> Appendix B shows example feedback displays in the two feedback groups—the unconstrained group that received nine measures, and the constrained group with a randomly selected subset of four.

Research on human cognition suggests that, when the number of pieces of information presented is any more than four, this begins to induce negative effects (Miller [1956], Chewing and Harrell [1990]). Surveys of practitioners suggest that some leaders limit feedback to a small number of measures in order to avoid information overload and focus attention on key measures (McKinsey [2016]). To test responses to the number of ranks displayed in reports and the role of cognitive limits in determining these responses, the reports show either four or nine ranks, numbers that would plausibly differ in the degree to which they lead to any negative effects associated with information overload (Miller [1956]).

The experimental feedback displayed the recipient's prior week's performance. Each subsequent week, students received updated feedback showing their most-recent week's ranks. The experiment ran for four weeks. The tests measure effects on a given dependent variable, such as *Grade*, as the difference between that dependent variable as it stood for the student at the end of the experiment—the course's end—and as it stood at the beginning of the experiment. This

---

<sup>2</sup> The experiment design includes both grade and grade-change as measures. The latter provides a more direct means for students to compare inputs, measured over the prior week, and output, given that grade-change is reported as change over the prior week. Tests of effect moderation consider whether this would aid learning and thereby be important to include in the randomly selected subset of four. At the same time, the inclusion of this measure would displace a fourth input measure or information on one's overall grade, possibly inducing a negative effect, and so prediction is not straightforward. An untabulated analysis finds no difference in the effects of constraining the set of ranks in feedback when the set of four includes grade-change versus when it does not.

approach to effect estimation, of measuring the change in the dependent variable after an intervention, is a common method of testing for effects in similar field experiments from economic and accounting research (Ashraf, Karlan, and Yin [2006], Chen et al. [2010], Casas-Arce, Lourenço, and Martínez-Jerez [2017]). As a result, the estimated effects are reflective of ultimate outcomes and not transient effects.

#### **4. DATA**

Appendix A contains a list of this study's variables and their definitions. The data document demographics and progress in the course by a range of measures. Table 1 provides descriptive statistics regarding demographics and the full range of course-related measures observed. Students are, on average, roughly 45 years old and hold a college degree.

The descriptive statistics in table 1 show the distribution of the course-related measures as they stood at the end of course. At the course end, the average student had accessed the course on eleven days, taken 1,886 actions in the course, viewed videos 126 times, made 116 problem attempts, made two discussion forum posts, made 52 comments on others' discussion forum posts, accessed nine chapters, and earned a grade of 73 percent. The experiment ran for four weeks and concluded at the courses' end. *Grade* is shown in table 1 as it stood at the end of the courses. *Grade* is equal to the percentage of problems in the course that the student has answered correctly and generally increases as the course progresses. *Accuracy* reflects the student's accuracy in attempting problems, captured as the ratio of points earned to problem attempts. A student who on average earned a point every other time that they tried a problem would, for example, have an *Accuracy* of 0.5. Table 1 shows that *Accuracy* at the end of the courses stood at 0.57.

Table 2 provides tests of a pre-treatment relationship between explanatory variables and either demographics or pre-experiment course metrics. Column 1 shows P-values associated with T-statistics from OLS regressions using *Ranks* and *Ranks x Constrained* as the explanatory variables, with the given demographic or course metric variable of interest as the dependent variable. Columns 2-4 show P-values associated with the T-statistics from OLS regressions using *Lower Rank Focus* as the explanatory variable, with the given demographic or course metric variable of interest as the dependent variable. For column 2, the sample is the Ranks Constrained group, or students randomly assigned to receive a randomly selected four of their nine ranks. For columns 3 and 4, the sample is the Ranks Constrained group then further split into at-or-above median ("Higher Init. Performer") and below median ("Lower Init. Performer") categories, based on their *Grade Before*, which is their *Grade* at the start of the experiment prior to delivery of RPF. These tests demonstrate no statistically significant relationship between either student demographics or pre-experiment dependent variables and the explanatory variables at a baseline.

Collectively, the results displayed in table 2 help to mitigate concerns that measured effects on performance are due to a failure in randomization. Such a confound would appear as imbalance in covariates, or relationships between the explanatory variables and either the demographic or dependent variables before the start of the experiment. The analyses in table 2 fail to reject the null hypothesis of no relationship for all tests.

## **5. ANALYSIS**

The empirical tests use Ordinary Least Squares (OLS) regression analysis. Each of the models incorporates course fixed effects to control for static differences across host courses. Similar economics-based research that employs a field-experimental research design uses OLS regression

with control variables (Ashraf, Bandiera, and Jack [2014], Casas-Arce, Lourenço, and Martínez-Jerez [2017]). Moreover, and following precedent from prior field experimental research, the tests measure effects on the change in a variable of interest during the experiment (e.g., Ashraf et al., [2006], Casas-Arce et al., [2017], Duflo et al., [2008]). All of the tables and figures use significance levels based on two-sided tests calculated with heteroscedastic-robust standard errors. T-statistics from two-sided tests are in brackets below each coefficient in the tables.

The tests of the main effect of restricting the ranks included in performance feedback use the general specification of model 1, which is displayed below.

$$(1) \quad Y = \alpha + \beta Ranks + \varphi Ranks \times Constrained + \rho Controls + \lambda Course + \varepsilon$$

Y represents the dependent variable of interest. *Ranks* is an indicator equal to one if the student was in the treatment condition that received RPF and zero if not. *Constrained* is an indicator equal to one if the student was in the treatment condition that received RPF and if the feedback was constrained to a random subset of four measures. The coefficient on *Ranks x Constrained* represents the effect of constraining the ranks shown in feedback, from nine down to four. The specification does not include *Constrained* separately from its interaction with *Ranks* because *Constrained* can only take a non-zero value when *Ranks* is equal to one. *Constrained* is relatedly subsumed by the inclusion of *Ranks* and *Ranks x Constrained*.

Figures 1-3 use models 2a and 2b to visually depict estimates of the effect of: 1) *Ranks*, or providing any RPF versus providing no RPF, 2) *Ranks Unconstrained*, or providing RPF on the full set of nine measures versus providing no RPF, and 3) *Ranks Constrained*, or providing RPF on a randomly selected subset of four measures versus providing no RPF. Note that these particular

specifications do not include a term to test whether constraining the set of ranks is more effective than not doing so—i.e., testing for a significant difference in the coefficients on *Ranks Constrained* and *Ranks Unconstrained*. Model 1 as used in tables 3-6 accomplishes that comparison, which is necessary to test H1 and H2. Specifically, that difference is captured by the coefficient  $\phi$  on *Ranks x Constrained* in model 1. Figures 1-3 provide visual depictions to offer context for that comparison.

$$(2a) \quad Y = \alpha + \beta Ranks + \rho Controls + \lambda Course + \varepsilon$$

$$(2b) \quad Y = \alpha + \beta Ranks\ Constrained + \phi Ranks\ Unconstrained + \rho Controls + \lambda Course + \varepsilon$$

Finally, tests for the effect of a *Lower Rank Focus* in feedback use model 3 below. Within the *Ranks Constrained* treatment arm, *Lower Rank Focus* is set equal to one if the subset of ranks included in an individual’s RPF are, on average, initially lower than his or her initial average rank across the full set of nine listed in the “Rank-Based Feedback Variables” section of Appendix A. These tests are conducted within the Ranks Constrained treatment condition.

$$(3) \quad Y = \alpha + \beta Lower\ Rank\ Focus + \rho Controls + \lambda Course + \varepsilon$$

Tables 3 through 8 document effect estimates both in the presence and absence of demographic controls. Table 2 documented that there was not a statistically significant relationship between the main explanatory variables and any of the demographics or pre-treatment course metrics. For thoroughness, the regressions are nonetheless shown both with and without controls.

This follows the guidance in (Angrist and Pischke [2009]) that it is customary to include controls even when analyzing effects of randomly assigned treatment conditions to increase the precision of effect estimates and account for any chance associates between such a variable and the explanatory variable.

### *5.1 Estimated Effects of Constraining the Set of Ranks in Feedback to a Random Subset*

Column 1 of table 3 shows a positive and statistically significant effect of *Ranks*, which is equal to one if the student was assigned to receive RPF and zero if not, on  $\Delta$  *Grade*. As described in the *Field Setting and Experiment Design* section, the analyses estimate effects on a given variable of interest, such as *Grade*, in terms of the change in that variable during the experiment. This is in line with methodology from prior feedback-related field experiments in economic and accounting research (Ashraf, Karlan, and Yin [2006], Chen et al. [2010], Casas-Arce, Lourenço, and Martínez-Jerez [2017]). The results in columns 1 and 2 demonstrate the validity of the experimental RPF as a means of influencing performance (Blanes i Vidal and Nossol [2011], Tafkov [2013]). Column 2 includes the following controls—*Gender*, *Age*, and *College Educated*.

The tests then proceed to the topic of restricting the ranks shown in feedback. Columns 3 and 4 of table 3 show estimated effects on  $\Delta$  *Grade* of *Ranks x Constrained*. The coefficient on *Ranks x Constrained* is the estimated effect of restricting the set of ranks provided in feedback to a random subset. That coefficient documents that constraining RPF in this way leads to an estimated 4.04 percentage point increase in *Grade*, or roughly 0.16 of a standard deviation in *Grade*. As discussed in more detail in the section on economic magnitudes below, effects of this magnitude in economic research have been deemed effective and recommendable strategies by regulators who deploy the research insights in policy making (Cheung and Slavin [2016], Fryer

and Howard-Noveck [2020]). Notably, policies that yield effects of this magnitude often involve much higher cost than the minor and nearly costless intervention in this study—one that consists of making slight computer code changes to adjust the display of feedback (Cheung and Slavin [2016]). This favorable comparison of efficacy to cost highlights the attractiveness of this feedback design intervention as a tool to boost human capital development. The results in columns 3 and 4 of table 3 are consistent with H1, that there will be a positive effect on grades of reducing the ranks in feedback to a random subset.

Figure 1 illustrates the benefit of RPF on  $\Delta$  Grade and its dependence on whether the ranks included in feedback are constrained. The omitted control condition for comparison is the group that did not receive RPF. The figure depicts the estimated effect of *Ranks* from model 2a on the left-hand side of the figure and the estimated effects of *Ranks Constrained* and *Ranks Unconstrained* from model 2b on the right-hand side of the figure. For figures 1-4, when the lines around the points in the figure lie fully above (below) the horizontal line at zero, this represents that the given treatment condition in the referenced sample had a positive (negative) effect that is statistically significant at at least the 0.1 level when compared to the control condition that received no RPF. Note that overlap in confidence intervals of two estimates—herein, the estimated effects of *Ranks Constrained* and *Ranks Unconstrained* in comparison to the condition that received no ranks—can occur even when those two estimates are statistically significantly different from each other (Greenland et al. [2016]). The coefficient on *Ranks x Constrained* from model 1 captures this difference and its statistical significance, for each of  $\Delta$  Grade,  $\Delta$  Accuracy, and  $\Delta$  Actions, in tables 3, 4, and 5, respectively. Recall that, for  $\Delta$  Grade, the coefficients on *Ranks x Constrained* in table 3 columns 3 and 4 documented a statistically significant benefit of constraining ranks in feedback, consistent with H1.

Table 4 uses the same specification of model 1 as table 3 but with  $\Delta Accuracy$  as the dependent variable. Columns 1 and 2 demonstrate a positive effect of *Ranks* on  $\Delta Accuracy$ . Specifically, there is a 0.07 increase on a base of 0.57 points per question attempted, or about a 0.30 standard deviation increase in accuracy. The coefficient on *Ranks x Constrained* in columns 3 and 4 shows that there is a positive effect of constraining the set of ranks provided for comparison on  $\Delta Accuracy$ . The increase in accuracy is statistically significant and rises by an incremental 0.06 when the ranks provided are constrained, relative to a base of 0.04 when they are not, representing a more than two times increase in accuracy when ranks are constrained relative to when they are not. This is consistent with H2, that there will be a positive effect on problem-attempt accuracy of reducing ranks in feedback to a random subset.

For  $\Delta Accuracy$ , figure 2 serves the same purpose as figure 1 did for  $\Delta Grade$ . Specifically, the figure depicts the estimated effect of *Ranks* from model 2a on the left-hand side of the figure. The estimated effects of *Ranks Constrained* and *Ranks Unconstrained* from model 2b are on the right-hand side of the figure. For both  $\Delta Grade$  and  $\Delta Accuracy$  there is a positive effect of *Ranks*, as shown on the left-hand side of both figures. This derives predominantly from the portion in the *Ranks Constrained* condition and not those in the *Ranks Unconstrained* condition.

Table 5 tests for effects on  $\Delta Actions$  from constraining the ranks in RPF. Columns 1 and 2 show a positive effect of RPF on  $\Delta Actions$ , consistent with prior studies that show that social comparison leads to greater effort provision (Chen et al. [2010], Blanes i Vidal and Nossol [2011]). Yet, there is not an incremental increase in  $\Delta Actions$  from constraining the number of ranks in RPF, as evidenced by the lack of a statistically significant coefficient on *Ranks x Constrained* in columns 3 and 4. This suggests that the positive effect of constraining the number of ranks in RPF is driven by increased accuracy rather than increased effort alone. This is consistent with theory

that information overload inhibits learning and the related development of effective strategies that could improve the efficacy of effort (Bawden [2001], Barber and Odean [2008]), herein captured by accuracy in problem attempts, such that it is optimal to constrain the information on ranks to improve learning.

Figure 3 shows a positive and statistically significant the effect of *Ranks* on  $\Delta$  *Actions*. When broken down into the *Ranks Constrained* and *Ranks Unconstrained* treatment conditions, the effect estimates are of similar magnitude. The results from table 5, columns 3 and 4 document that there is not a statistically significant difference in the effect of the experimental feedback on  $\Delta$  *Actions* when ranks are constrained versus when they are not.

### 5.2 Estimated Effects of Providing a Subset of Ranks with a Lower Rank Focus

Table 6 shows the effect of *Lower Rank Focus* on  $\Delta$  *Grade*, as estimated using model 3, within the *Ranks Constrained* treatment arm. There is a positive and statistically significant effect for higher initial performers, or those with an at-or-above median *Grade* at the start of the experiment. This provides evidence in support of H3. There is a negative effect for lower initial performers, or those whose *Grade* was below the median *Grade* at the start of the experiment. This provides some support for H5, albeit significant at only the 0.1 level. Figure 4 provides a graphical depiction of the effects estimated in table 6 using model 3. These opposing effects in either partition yield a null effect of *Lower Rank Focus* when the sample is not partitioned by initial performance.

The results in table 7 are consistent with the idea that part of the effect of *Lower Rank Focus* on  $\Delta$  *Grade* is due to increased accuracy of individual problem attempts. The estimated effect of *Lower Rank Focus* on  $\Delta$  *Accuracy* is 0.10 (-0.09) for initially higher (initially lower) performers. The effect among initially higher performers is statistically significant at the 0.05 level

and the effect among initially lower performers is statistically significant at the 0.1 level. The former result provides support for H4, that, for initially higher performers, there will be a positive effect on problem-attempt accuracy of focusing the subset of ranks in their report on their own relatively lower ranks.

Table 8 presents estimates of the effect of *Lower Rank Focus* on  $\Delta$  *Actions*. Column 1 shows a negative and non-statistically significant estimated effect before partitioning the sample by initial performance. The effect becomes significant and negative at the 0.1 level after including controls. The results in partitions demonstrate a positive but non-statistically significant effect among higher initial-performers and a negative and statistically significant effect among lower initial-performers that represents a roughly 18% reduction in total actions relative to the full sample. The latter result is consistent with H6, that, for initially lower performers, there will be a negative effect on course activity of focusing the subset of ranks in their report on their own relatively lower ranks.

While the hypotheses did not specify a prediction for the effect of focusing RPF on one's own lower ranks for accuracy, there is some evidence from table 7 at the 0.1 level that this has a negative effect on accuracy among initially lower performers. This is represented by the coefficients on *Lower Rank Focus* in columns 5 and 6. Paired with the result that such a focus reduces the number of actions taken in the course for lower initial performers, this result is consistent with the idea that aptitude at a task rises in part from learning through trial-and-error (Campbell, Epstein, and Martínez-Jerez [2011]). Specifically, if feedback with a lower rank focus causes discouragement and reduced engagement with the course, this could inhibit trial-and-error and related gains in aptitude.

### 5.3 Effect Magnitudes

The estimated effect of constraining the ranks provided for feedback in this study was .16 of a standard deviation in *Grade*. This is equal to the average economic effect magnitude of a wide range of treatments in economic settings from a review of 197 randomized-controlled trials focused on human capital and performance (Cheung and Slavin [2016]). Many of these studies involved high-cost interventions. For example, a study in the *Journal of Labor Economics* wherein students were offered “high-dosage tutoring” led to a .13 SD improvement in grades (Fryer and Howard-Noveck [2020]). Government policymakers referred to this as an “effective” but “high-cost” intervention that they recommended for implementation (Office of the State Superintendent of Education [2021]). The effect estimates in the current study are of similar magnitude but have negligible cost. The cost to reduce the number of ranks included in feedback is a matter of making minor adjustments to computer code, and this can be scaled over thousands of feedback recipients.

The estimated effect of providing a *Lower Rank Focus* within the constrained set for higher initial performers is roughly .15 of a standard deviation, similarly in excess of the .13 level that policy makers have noted as effective. Again, this is a low-cost intervention compared to many other interventions that achieve a similar effect size. For lower initial performers, the negative effect of *Lower Rank Focus* is negative and of a larger magnitude, at roughly .25 of a standard deviation of *Grade*, albeit not quite significant at the .05 level. The effect in terms of *Actions* is also negative and statistically significant, with a magnitude of roughly .20 of a standard deviation. This is an indicator of lower engagement in human capital attainment, given that the reduction in *Activity* appears hand-in-hand with a reduction in performance by *Grade*. Negative trends in task engagement are an important determinant of human capital development (Azmat and Iriberry [2010]).

#### *5.4 Limitations and Generalizability*

The field setting for this study is an online education platform that provides advanced training in professional skills. Most respondents to questions about motivation upon enrollment note that they enrolled at least in part due to desires to advance their careers or change their career. The results are therefore most directly applicable to economic settings in which individuals seek to obtain employable human capital. The courses are focused either on proficiency in computer programming and data analysis, or knowledge of laws, justice, and the legal system. Professional training is increasingly delivered through online education, as in this paper's setting, with more than half of fortune 500 companies subscribing to online education services on behalf of their employees (Orbis [2017]). Thus, the results regard a widespread activity in the economy.

The theoretical underpinnings of this study's predictions are, though, not specific to human-capital attainment. These theories relate to expectancy, social comparison, and feedback interventions and help explain behavior in many economic settings that span industries, incentive schemes, organizational hierarchies, and tasks (Messick and Jungeblut [1981], Kluger and DeNisi [1996]). The broad applicability of these theories suggests that the insights from this study will apply beyond human-capital attainment.

In terms of likely limitations on the results' generalizability, the experiment duration is four weeks. The effects are measured as of the end of the course, and so the results represent improvement in a final outcome, but future research could examine effect duration over a longer period. While this study demonstrates significant benefits to constraining the number of ranks shown in RPF, and documents divergent effects for higher and lower initial performers depending on whether the measures included in the constrained are their initially lower ranked measures, it

is possible that the measure selection in reports would need to be updated over time in order to produce similar effects in subsequent periods over a longer time span.

## **6. CONCLUSION**

This paper uses a field experiment in professional education to examine effects of constraining the set of ranks available in feedback to a random subset. This extends the literature on the use of RPF to influence provision and direction of effort (Eyring and Narayanan [2018], Casas-Arce, Lourenço, and Martínez-Jerez [2017], Hannan, Krishnan, and Newman [2008], Tafkov [2013]). There is significant demand from practitioners for insights on how to design feedback (Riegel [2018], Zenger and Folkman [2014]). Moreover, at the same time as the amount of data available in organizations that can be included in feedback proliferates, there is growing evidence that when information outstrips cognitive bounds, this can trigger information overload (Eppler and Mengis [2004], Barber and Odean [2008]). This study addresses the demand for insights on feedback design, drawing on economic models of information processing and exploiting the ability to employ random assignment in a field experiment. The setting involves highly educated participants, the majority of whom have a bachelor's degree, who are engaged in human capital development. The results demonstrate benefits of constraining the ranks included in RPF and shed light on how to best do this for higher and lower initial performers by showing initially higher (lower) performers their own lower (higher) ranks.

Prior research on limited attention has concentrated on methods of providing incremental information—such as reminders, or appeals to altruism—to influence behavior (Calzolari and Nardotto [2017], Sunstein [2021], Bordalo, Gennaioli, and Shleifer [2022]). Yet, research on the usefulness of omitting information is still limited. This study documents that students perform

better when the RPF they receive consists of a constrained set of ranks. Moreover, initially higher performers benefit from receiving feedback with a subset of ranks that are, on average, lower than their average rank by the full set. The results also offer reason for caution when restricting the set of ranks displayed, in the case of lower initial performers. The latter group performs worse and is less active in the online courses when the subset of measures shown have a lower rank, on average, than the recipient's average rank across the full set.

The tests draw on a line of economic theory regarding bounded cognition and information processing that has received increasing attention (DellaVigna [2009], Bordalo, Gennaioli, and Shleifer [2022]) and apply this to RPF, a widespread tool for management control (Hannan, Krishnan, and Newman [2008], Tafkov [2013]). Future research could continue to explore approaches to narrowing an information set in the context of various reports that often include a wide range of measures (Kantor and Streitfeld [2015], Mercer [2019]). Theories from economics, management, and social psychology speak to factors such as resource allocation and salience that could guide further inquiry on how to pare back the provision of information in order to promote goal achievement (Bordalo, Gennaioli, and Shleifer [2015], Calzolari and Nardotto [2017], DellaVigna [2009]).

## ***References***

- ALLCOTT, HUNT, and TODD ROGERS. "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation." *American Economic Review* 104 (2014): 3003–37. Available at <http://www.aeaweb.org/articles?id=10.1257/aer.104.10.3003>.
- ANGRIST, JOSHUA D., and JÖRN-STEFFEN PISCHKE. *Mostly Harmless Econometrics. Mostly Harmless Econometrics*. Princeton University Press, 2009. <https://doi.org/10.1515/9781400829828/HTML>.
- ARROW, KENNETH. "Methodological Individualism and Social Knowledge." *The American Economic Review* 84 (1994): 1–9. Available at [https://www.jstor.org/stable/2117792?casa\\_token=KEYVHSpmq8AAAAA:o70HXII529U](https://www.jstor.org/stable/2117792?casa_token=KEYVHSpmq8AAAAA:o70HXII529U)

- IgBpWkFFOFFWkCfvvAqAunRQ4dlhJ4uAmBwGAZcVqYg17d-  
hAyzV41vJMBJFDLRMOy1GNglCD7EXAFLZB2b1gjFqzVXARsKhhXIMGa9Hfrg.
- ASHRAF, NAVA, ORIANA BANDIERA, and B. KELSEY JACK. “No Margin, No Mission? A Field Experiment on Incentives for Public Service Delivery.” *Journal of Public Economics* 120 (2014): 1–17. <https://doi.org/10.1016/J.JPUBECO.2014.06.014>.
- ASHRAF, NAVA, DEAN KARLAN, and WESLEY YIN. “Tying Odysseus to the Mast: Evidence From a Commitment Savings Product in the Philippines\*.” *The Quarterly Journal of Economics* 121 (2006): 635–72. Available at <http://dx.doi.org/10.1162/qjec.2006.121.2.635>.
- AZMAT, GHAZALA, and NAGORE IRIBERRI. “The Importance of Relative Performance Feedback Information: Evidence from a Natural Experiment Using High School Students.” *Journal of Public Economics* 94 (2010): 435–52. <https://doi.org/10.1016/J.JPUBECO.2010.04.001>.
- BADDELEY, ALAN. “The Magical Number Seven: Still Magic after All These Years?” *Psychological Review* 101 (1994): 353–56. <https://doi.org/10.1037/0033-295X.101.2.353>.
- BANDIERA, ORIANA, IWAN BARANKAY, and IMRAN RASUL. “TEAM INCENTIVES: EVIDENCE FROM A FIRM LEVEL EXPERIMENT.” *Journal of the European Economic Association* 11 (2013): 1079–1114. <https://doi.org/10.1111/jeea.12028>.
- BANDIERA, ORIANA, and IMRAN RASUL. “Social Networks and Technology Adoption in Northern Mozambique.” *The Economic Journal* 116 (2006): 869–902. <https://doi.org/10.1111/J.1468-0297.2006.01115.X>.
- BARBER, BRAD M., and TERRANCE ODEAN. “All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors.” *The Review of Financial Studies* 21 (2008): 785–818. <https://doi.org/10.1093/RFS/HHM079>.
- BAWDEN, DAVID. “Information and Digital Literacies: A Review of Concepts.” *Journal of Documentation* 57 (2001): 218–59. <https://doi.org/10.1108/EUM0000000007083/FULL/HTML>.
- BAYER, PATRICK, RANDI HJALMARSSON, and DAVID POZEN. “Building Criminal Capital behind Bars: Peer Effects in Juvenile Corrections.” *The Quarterly Journal of Economics* 124 (2009): 105–47. <https://doi.org/10.1162/QJEC.2009.124.1.105>.
- BLANES I VIDAL, JORDI, and MAREIKE NOSSOL. “Tournaments Without Prizes: Evidence from Personnel Records.” *Management Science* 57 (2011): 1721–36. <https://doi.org/10.1287/mnsc.1110.1383>.
- BLANKESPOOR, ELIZABETH. “The Impact of Information Processing Costs on Firm Disclosure Choice: Evidence from the XBRL Mandate.” *Journal of Accounting Research* 57 (2019): 919–67. <https://doi.org/10.1111/1475-679X.12268>.
- BLANKESPOOR, ELIZABETH, ED DEHAAN, JOHN WERTZ, and CHRISTINA ZHU. “Why Do Individual Investors Disregard Accounting Information? The Roles of Information Awareness and Acquisition Costs.” *Journal of Accounting Research* 57 (2019): 53–84. <https://doi.org/10.1111/1475-679X.12248>.
- BORDALO, PEDRO, NICOLA GENNAIOLI, and ANDREI SHLEIFER. “Salience Theory of Judicial Decisions.” *Journal of Legal Studies* 44 (2015): S7–33. <https://doi.org/10.1086/676007>.
- . “Salience.” *Annual Review of Economics* 14 (2022): 521–44. <https://doi.org/10.1146/ANNUREV-ECONOMICS-051520-011616>.
- BOUWENS, JAN, and PETER KROOS. “Target Ratcheting and Effort Reduction.” *Journal of Accounting and Economics* 51 (2011): 171–85. <https://doi.org/10.1016/J.JACCECO.2010.07.002>.

- BOYCE, P. R. "Sinus Arrhythmia as a Measure of Mental Load." *Ergonomics* 17 (1974): 177–83. <https://doi.org/10.1080/00140137408931336>.
- BRAUNSTEIN, MYRON L. *Depth Perception Through Motion*. Academic Press Inc., 1976. Available at <https://books.google.com/books?hl=zh-CN&lr=&id=QEi0BQAAQBAJ&oi=fnd&pg=PP1&dq=MYRON+L.+BRAUNSTEIN&ots=5vP4A1bnTH&sig=J-KPBE9x6Eu-TtmKtmmAV8drazU#v=onepage&q=MYRON%20L.%20BRAUNSTEIN&f=false>.
- CALZOLARI, GIACOMO, and MATTIA NARDOTTO. "Effective Reminders." *Management Science* 63 (2017): 2915–32. <https://doi.org/10.1287/mnsc.2016.2499>.
- CAMPBELL, DENNIS. "Choose the Right Measures, Drive the Right Strategy." *Balanced Scorecard Report* 8 (2006): 3–5.
- CAMPBELL, DENNIS, MARC J. EPSTEIN, and F. A. MARTINEZ-JEREZ. "The Learning Effects of Monitoring." *The Accounting Review* 86 (2011): 1909–34. <https://doi.org/10.2308/ACCR-10129>.
- CAMPBELL, DENNIS, MARIA LOUMIOTI, REGINA WITTENBERG-MOERMAN, ETHAN BERNSTEIN, ROBERT LIBBY, JUNG KOO KANG, TRACIE MAJORS, ET AL. "Making Sense of Soft Information: Interpretation Bias and Ex-Post Lending Outcomes We Are Grateful To" (2017).
- CASAS-ARCE, PABLO, SOFIA M. LOURENCO, and F. A. MARTINEZ-JEREZ. "The Performance Effect of Feedback Frequency and Detail: Evidence from a Field Experiment in Customer Satisfaction." *Journal of Accounting Research* 55 (2017): 1051–88. <https://doi.org/10.1111/1475-679X.12184>.
- CHAN, TAT Y., JIA LI, and LAMAR PIERCE. "Learning from Peers: Knowledge Transfer and Sales Force Productivity Growth." *Marketing Science* 33 (2014): 463–84. <https://doi.org/10.1287/MKSC.2013.0831>.
- CHEN, YAN, F MAXWELL HARPER, JOSEPH KONSTAN, and SHERRY XIN LI. "Social Comparisons and Contributions to Online Communities: A Field Experiment on MovieLens." *American Economic Review* 100 (2010): 1358–98. Available at <http://www.aeaweb.org/articles?id=10.1257/aer.100.4.1358>.
- CHEUNG, ALAN C.K., and ROBERT E. SLAVIN. "How Methodological Features Affect Effect Sizes in Education." *Educational Researcher* 45 (2016): 283–92. <https://doi.org/10.3102/0013189X16656615>.
- CHEWNING, EUGENE G., and ADRIAN M. HARRELL. "The Effect of Information Load on Decision Makers' Cue Utilization Levels and Decision Quality in a Financial Distress Decision Task." *Accounting, Organizations and Society* 15 (1990): 527–42. [https://doi.org/10.1016/0361-3682\(90\)90033-Q](https://doi.org/10.1016/0361-3682(90)90033-Q).
- CHOI, JONGWOON (WILLIE), GARY W HECHT, and WILLIAM B TAYLER. "Lost in Translation: The Effects of Incentive Compensation on Strategy Surrogation." *The Accounting Review* 87 (2012): 1135–63. <https://doi.org/10.2308/accr-10273>.
- COSTELLO, ANNA M., ANDREA K. DOWN, and MIHIR N. MEHTA. "Machine + Man: A Field Experiment on the Role of Discretion in Augmenting AI-Based Lending Models." *Journal of Accounting and Economics* 70 (2020): 101360. <https://doi.org/10.1016/J.JACCECO.2020.101360>.
- DELLA VIGNA, STEFANO. "Psychology and Economics: Evidence from the Field." *Journal of Economic Literature* 47 (2009): 315–72. <https://doi.org/10.1257/JEL.47.2.315>.

- DEMSKI, JOEL S, and GERALD A FELTHAM. *Cost Determination: A Conceptual Approach*. Iowa State Pr, 1976.
- DUFLO, ESTHER, MICHAEL KREMER, and JONATHAN ROBINSON. “How High Are Rates of Return to Fertilizer? Evidence from Field Experiments in Kenya.” *American Economic Review* 98 (2008): 482–88. <https://doi.org/10.1257/AER.98.2.482>.
- ELLIOTT, W. BROOKE, JESSEN L. HOBSON, and BRIAN J. WHITE. “Earnings Metrics, Information Processing, and Price Efficiency in Laboratory Markets.” *Journal of Accounting Research* 53 (2015): 555–92. <https://doi.org/10.1111/1475-679X.12080>.
- EMANUEL, JEFFREY P., and ANNE LAMB. “Open, Online, and Blended: Transactional Interactions with MOOC Content by Learners in Three Different Course Formats.” *Online, and Blended: Transactional Interactions with MOOC Content by Learners in Three Different Course Formats* (2015). <https://doi.org/10.24059/olj.v21i2.845>.
- EPPLER, MARTIN, and JEANNE MENGIS. “Side-Effects of the e-Society: The Causes of Information Overload and Possible Countermeasures.” *Proceedings of IADIS International Conference E-Society* (2004), 16–19. Available at [https://www.mgt.ncu.edu.tw/~ckfarn/doc/conference/IADIS%20ES2004\\_Vol2.pdf#page=431](https://www.mgt.ncu.edu.tw/~ckfarn/doc/conference/IADIS%20ES2004_Vol2.pdf#page=431).
- ETTEMA, J. H., and R. L. ZIELHUIS. “Physiological Parameters of Mental Load.” *ERGONOMICS* 14 (1971): 137–44. <https://doi.org/10.1080/00140137108931232>.
- EYRING, HENRY, PATRICK J. FERGUSON, and SEBASTIAN KOPPERS. “Less Information, More Comparison, and Better Performance: Evidence from a Field Experiment.” *Journal of Accounting Research* 59 (2021): 657–711. <https://doi.org/10.1111/1475-679X.12362>.
- EYRING, HENRY, and V. G. NARAYANAN. “Performance Effects of Setting a High Reference Point for Peer-Performance Comparison.” *Journal of Accounting Research* (2018). <https://doi.org/10.1111/1475-679X.12199>.
- FESTINGER, LEON. “A Theory of Social Comparison Processes.” *Human Relations* 7 (1954): 117–40. <https://doi.org/10.1177/001872675400700202>.
- FRYER, ROLAND G., and MEGHAN HOWARD-NOVECK. “High-Dosage Tutoring and Reading Achievement: Evidence from New York City.” *Journal of Labor Economics* 38 (2020): 421–52. [https://doi.org/10.1086/705882/SUPPL\\_FILE/17451DATA.ZIP](https://doi.org/10.1086/705882/SUPPL_FILE/17451DATA.ZIP).
- GIGERENZER, GERD, and DANIEL G. GOLDSTEIN. “Reasoning the Fast and Frugal Way: Models of Bounded Rationality.” *Psychological Review* 103 (1996): 650–69. <https://doi.org/10.1037/0033-295X.103.4.650>.
- GLASGOW, ROBERT, SAM FINLAYSON, and DAVID RAY. “To Enlist Physicians in Reducing Costs, Show Them the Costs.” *Harvard Business Review* (2018).
- GREENLAND, SANDER, STEPHEN J. SENN, KENNETH J. ROTHMAN, JOHN B. CARLIN, CHARLES POOLE, STEVEN N. GOODMAN, and DOUGLAS G. ALTMAN. “Statistical Tests, P Values, Confidence Intervals, and Power: A Guide to Misinterpretations.” *European Journal of Epidemiology* 31 (2016): 337–50. <https://doi.org/10.1007/S10654-016-0149-3/METRICS>.
- GROTE, RICHARD. *Forced Ranking: Making Performance Management Work*. Harvard Business School Press, 2005.
- GUENTHER, CHRISTINA, SEBASTIAN LEHNEN, and RAINER MICHAEL RILKE. “Formalization of Annual Performance Feedback and Employees’ Job Satisfaction in the SME Context.” *Journal of Small Business Management* (2024). <https://doi.org/10.1080/00472778.2024.2326576>.

- HANNAN, R LYNN, RANJANI KRISHNAN, and ANDREW H NEWMAN. “The Effects of Disseminating Relative Performance Feedback in Tournament and Individual Performance Compensation Plans.” *The Accounting Review* 83 (2008): 893–913. <https://doi.org/10.2308/accr.2008.83.4.893>.
- HOLMSTROM, BENGT, and PAUL MILGROM. “Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design.” *JL Econ. & Org.* 7 (1991): 24.
- ITTNER, CHRISTOPHER D, DAVID F LARCKER, and MARSHALL W MEYER. “Subjectivity and the Weighting of Performance Measures: Evidence from a Balanced Scorecard.” *The Accounting Review* 78 (2003): 725–58.
- JACKSON, C. KIRABO, and ELIAS BRUEGMANN. “Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers.” *American Economic Journal: Applied Economics* 1 (2009): 85–108. <https://doi.org/10.1257/APP.1.4.85>.
- KANTOR, JODI, and DAVID STREITFELD. “Inside Amazon: Wrestling Big Ideas in a Bruising Workplace.” *The New York Times*, 2015. Available at <https://www.nytimes.com/2015/08/16/technology/inside-amazon-wrestling-big-ideas-in-a-bruising-workplace.html>.
- KAUR, SUPREET, MICHAEL KREMER, and SENDHIL MULLAINATHAN. “Self-Control and the Development of Work Arrangements.” *American Economic Review* 100 (2010): 624–28. Available at <http://www.aeaweb.org/articles?id=10.1257/aer.100.2.624>.
- KLUGER, AVRAHAM N, and ANGELO DENISI. “The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory.” *Psychological Bulletin* 119 (1996): 254–84. <https://doi.org/10.1037/0033-2909.119.2.254>.
- KOLSTAD, JONATHAN T. “Information and Quality When Motivation Is Intrinsic: Evidence from Surgeon Report Cards.” *American Economic Review* 103 (2013): 2875–2910. <https://doi.org/10.1257/AER.103.7.2875>.
- KRUSCHKE, JOHN, and MARK JOHANSEN. “A Model of Probabilistic Category Learning.” *Journal of Experimental Psychology* 25 (1999): 1083–1119.
- LAWLER, EDWARD E., and J. LLOYD SUTTLE. “Expectancy Theory and Job Behavior.” *Organizational Behavior and Human Performance* 9 (1973): 482–503. [https://doi.org/10.1016/0030-5073\(73\)90066-4](https://doi.org/10.1016/0030-5073(73)90066-4).
- LEONE, ANDREW J., and STEVE ROCK. “Empirical Tests of Budget Ratcheting and Its Effect on Managers’ Discretionary Accrual Choices.” *Journal of Accounting and Economics* 33 (2002): 43–67. [https://doi.org/10.1016/S0165-4101\(01\)00044-1](https://doi.org/10.1016/S0165-4101(01)00044-1).
- MATĚJKA, FILIP, and ALISDAIR MCKAY. “Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model.” *American Economic Review* 105 (2015): 272–98. <https://doi.org/10.1257/AER.20130047>.
- MCKINSEY. “Ahead of the Curve: The Future of Performance Management.” McKinsey Quarterly. 2016. Available at <https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/ahead-of-the-curve-the-future-of-performance-management>.
- MERCER. “Mercer’s 2019 Global Performance Management Survey Creating Performance-Oriented Transformation in the Future of Work.” 2019. Available at <https://www.imercer.com/uploads/common/HTML/LandingPages/AnalyticalHub/june2019-merc-2019-global-performance-management-survey-executive-summary.pdf>.

- MESSICK, SAMUEL, and ANN JUNGBLUT. "Time and Method in Coaching for the SAT." *Psychological Bulletin* 89 (1981): 191–216. <https://doi.org/10.1037/0033-2909.89.2.191>.
- MILLER, GEORGE A. "The Magical Number Seven, plus or Minus Two: Some Limits on Our Capacity for Processing Information." *Psychological Review* 63 (1956): 81–97. <https://doi.org/10.1037/H0043158>.
- MUCHNIK, LEV, SINAN ARAL, and SEAN J. TAYLOR. "Social Influence Bias: A Randomized Experiment." *Science* 341 (2013): 647–51. [https://doi.org/10.1126/SCIENCE.1240466/SUPPL\\_FILE/MUCHNIK.SM.PDF](https://doi.org/10.1126/SCIENCE.1240466/SUPPL_FILE/MUCHNIK.SM.PDF).
- OFFICE OF THE STATE SUPERINTENDENT OF EDUCATION. "High-Dosage Tutoring: A Proven Strategy to Accelerate Student Learning." 2021. 2021. Available at [https://osse.dc.gov/sites/default/files/dc/sites/osse/page\\_content/attachments/HighDosageTutoringGuidance.pdf](https://osse.dc.gov/sites/default/files/dc/sites/osse/page_content/attachments/HighDosageTutoringGuidance.pdf).
- ORBIS. "Global E-Learning Market 2017 to Boom \$275.10 Billion Value by 2022 at a CAGR of 7.5% – Orbis Research." Reuters. 2017. 2017.
- PAYNE, JOHN W. "Task Complexity and Contingent Processing in Decision Making: An Information Search and Protocol Analysis." *Organizational Behavior and Human Performance* 16 (1976): 366–87. [https://doi.org/10.1016/0030-5073\(76\)90022-2](https://doi.org/10.1016/0030-5073(76)90022-2).
- POGONES, SANDY. "Which Quality Measures Should You Report." 2017.
- RIEGEL, DEBORAH. "How to Solicit Negative Feedback When Your Manager Doesn't Want to Give It." *Harvard Business Review* (2018).
- SAATY, T. L., and M. S. OZDEMIR. "Why the Magic Number Seven plus or Minus Two." *Mathematical and Computer Modelling* 38 (2003): 233–44. [https://doi.org/10.1016/S0895-7177\(03\)90083-5](https://doi.org/10.1016/S0895-7177(03)90083-5).
- SAMUELS, RICHARD. "The Magical Number Two, plus or Minus: Dual Process Theory as a Theory of Cognitive Kinds." In *In Two Minds: Dual Processes and Beyond*, 129–46. 1994.
- SAVAGE, LEONARD J. *The Foundations of Statistics*. Vol. 1. John Wiley & Sons, Inc., 1954. <https://doi.org/10.1002/nav.3800010316>.
- SIMON, HERBERT A. "Designing Organizations for an Information-Rich World." *Computers, Communications, and the Public Interest* (1971).
- SIMS, CHRISTOPHER A. "Implications of Rational Inattention." *Journal of Monetary Economics* 50 (2003): 665–90. [https://doi.org/10.1016/S0304-3932\(03\)00029-1](https://doi.org/10.1016/S0304-3932(03)00029-1).
- . "Rational Inattention: Beyond the Linear-Quadratic Case." *American Economic Review* 96 (2006): 158–63. <https://doi.org/10.1257/000282806777212431>.
- SONG, HUMMY, ANITA L TUCKER, KAREN L MURRELL, and DAVID R VINSON. "Closing the Productivity Gap: Improving Worker Productivity Through Public Relative Performance Feedback and Validation of Best Practices." *Management Science* 64 (2017): 2628–49. <https://doi.org/10.1287/mnsc.2017.2745>.
- SUNSTEIN, CASS R. "Hayekian Behavioral Economics." *Behavioural Public Policy* 7 (2021): 170–88. <https://doi.org/10.1017/BPP.2021.3>.
- TAFKOV, IVO D. "Private and Public Relative Performance Information under Different Compensation Contracts." *The Accounting Review* 88 (2013): 327–50. <https://doi.org/10.2308/accr-50292>.
- VISCUSI, W. KIP. "A Bayesian Perspective on Biases in Risk Perception." *Economics Letters* 17 (1985): 59–62. [https://doi.org/10.1016/0165-1765\(85\)90127-2](https://doi.org/10.1016/0165-1765(85)90127-2).
- ZENGER, JACK, and JOSEPH FOLKMAN. "Your Employees Want the Negative Feedback You Hate to Give." *Harvard Business Review* (2014).

## Appendix A: Variable Definitions

---

### Dependent Variables

<i>Δ Grade</i>	The change in the student's <i>Grade</i> (as defined below) from the beginning of the experiment to the end of the course.
<i>Δ Accuracy</i>	The change in the student's <i>Accuracy</i> (as defined below) on problem attempts that allow multiple attempts, from the beginning of the experiment to the end of the course.
<i>Δ Actions</i>	The change in the number of <i>Actions</i> (as defined below) that the student took in the course, from the beginning of the experiment to the end of the course.

### Rank-Based Feedback Variables

<i>Days Active</i>	The number of calendar days on which the student accessed the course.
<i>Actions</i>	All actions in the course that the student took and that are recorded electronically; these include logins, video views, problem attempts, forum posts, other forum actions, and clicks on course material.
<i>Videos</i>	The number of times that the student viewed videos.
<i>Problems</i>	The number of times that the student entered an answer to any problem.
<i>Posts</i>	The number of posts that the student made in discussion forums.
<i>Comments</i>	The number of comments on other students' posts that the student made.
<i>Chapters</i>	The number of course chapters that the student accessed.
<i>Grade</i>	The percentage of the total problems in the course that a student answered correctly, rounded to the nearest integer percentage.
<i>Grade Change Over Prior Week</i>	The change in <i>Grade</i> over the prior week.

### Explanatory Variables

<i>Ranks</i>	An indicator variable equal to one if the student received RPF.
<i>Constrained</i>	An indicator variable equal to one if the student received RPF and the number of measures shown were a randomly selected four from the full set of nine measures listed under the "Rank-Based Feedback Variables" section of this appendix. When <i>Constrained</i> is interacted with <i>Ranks</i> in model 1, and when <i>Ranks</i> is the only other explanatory variable included in the specification, the coefficient on <i>Ranks</i> x <i>Constrained</i> represents the difference in the estimated effects of <i>Ranks</i> <i>Constrained</i> as compared to <i>Ranks</i> <i>Unconstrained</i> . The coefficient thereby

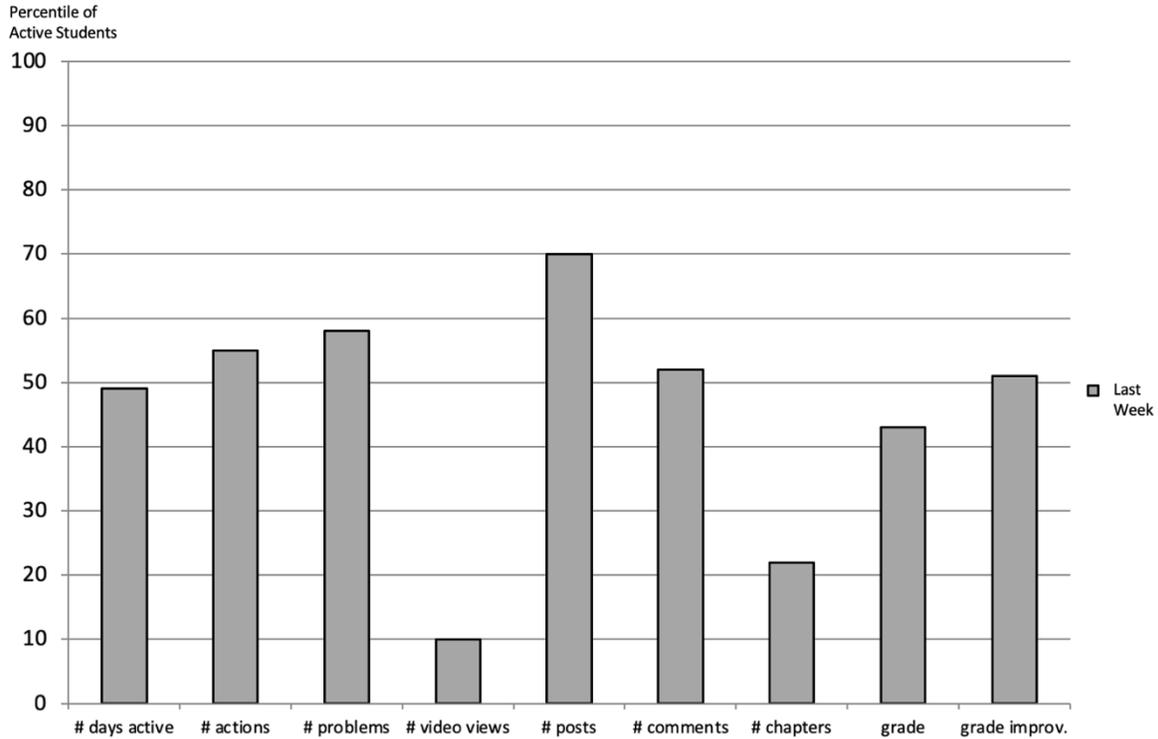
<i>Ranks Constrained</i>	demonstrates whether there is a positive effect of constraining the ranks provided for peer comparison. An indicator variable equal to one if the student received RPF and the number of measures shown were a randomly selected four from the full set of nine measures listed under the “Rank-Based Feedback Variables” section of this appendix.
<i>Ranks Unconstrained</i>	An indicator variable equal to one if the student received RPF and the number of measures shown were the full set of nine measures listed under the “Rank-Based Feedback Variables” section of this appendix.
<i>Lower Rank Focus</i>	An indicator variable equal to one if the student received RPF with a randomly selected four measures and the student’s initial average percentile rank by the subset of four measures shown in his or her feedback was lower than the student’s initial average percentile rank by the full set of nine measures listed under the “Rank-Based Feedback Variables” section of this appendix.
<b>Dependent Variables at Baseline Levels</b>	
<i>Grade Before</i>	The student’s <i>Grade</i> (as defined above) at the start of the experiment prior to the delivery of RPF.
<i>Accuracy Before</i>	The student’s <i>Accuracy</i> (as defined above) at the start of the experiment prior to the delivery of RPF.
<i>Actions Before</i>	The student’s <i>Actions</i> (as defined above) at the start of the experiment prior to the delivery of RPF.
<b>Demographic Variables</b>	
<i>Gender</i>	An indicator variable equal to one if the student chose Male as their gender in the course registration process.
<i>Age</i>	The student’s age reported during course registration, capped at 89 to maintain privacy.
<i>College Educated</i>	An indicator variable equal to one if the student indicated that they had a bachelor’s degree or higher level of education.

---

## Appendix B: Feedback Instruments

**Figure A.B.1: Feedback with Full Set of Nine Measures**

This figure shows an example feedback instrument that the roughly one half of the treatment group sample who received feedback with all nine measures were provided with. The ranks reflected the standing for the given student who received the feedback. Ranks updated each week of the experiment.

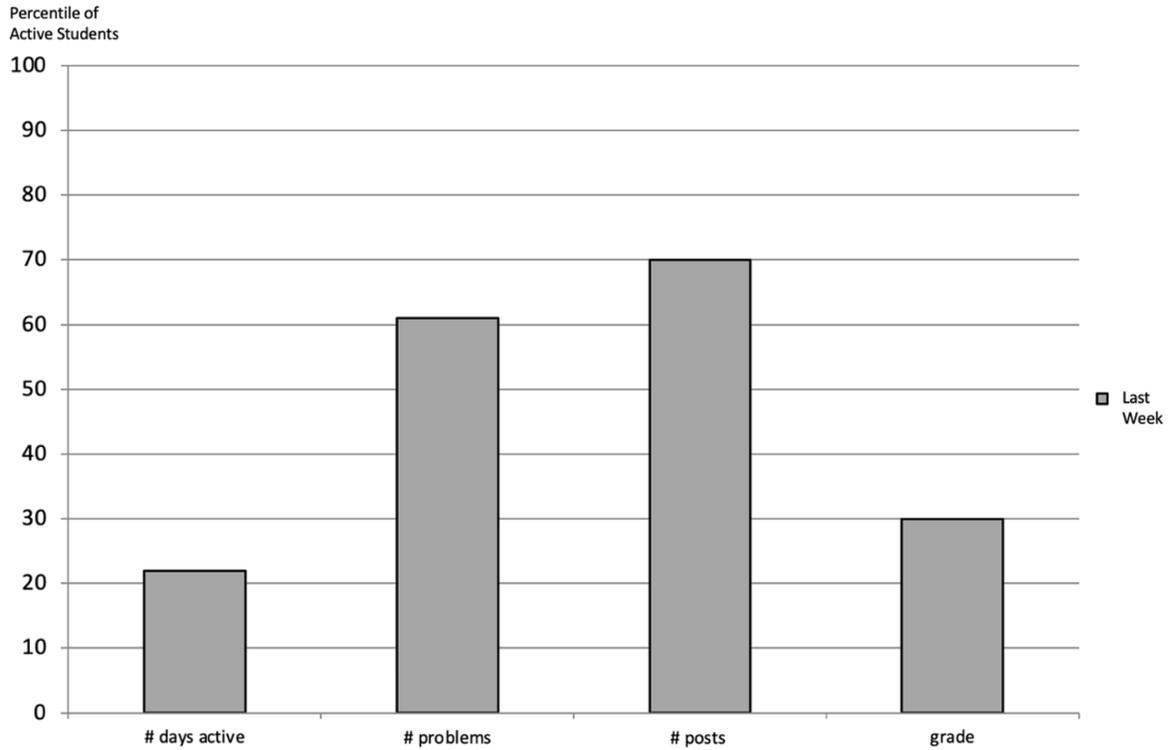


**Measure Key**

# days active	# of days you were active in the course
# actions	# of actions you took in the course
# problems	# of times you attempted any problems
# video views	# of times you viewed any videos
# posts	# of posts you made in the discussion forum
# comments	# of comments you made on others' posts
# chapters	# of chapters you accessed
grade	your grade in the course
grade improv.	your improvement in grade in the course

**Figure A.B.2: Feedback with Subset of Four Randomly Selected Measures**

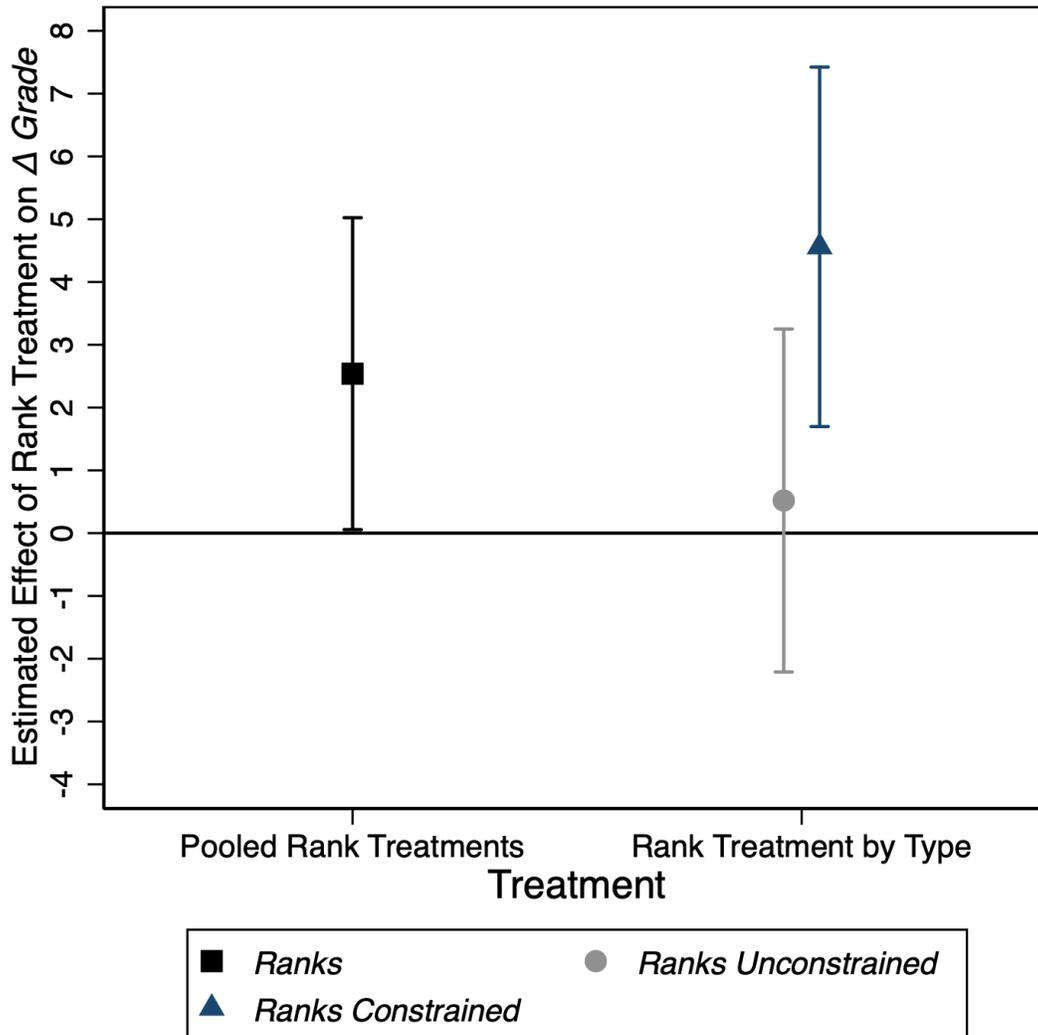
This figure shows an example feedback instrument that the roughly one half of the treatment group sample who received feedback with a randomly selected subset of four measures were provided with. The ranks reflected the standing for the given student who received the feedback. That set was randomly selected at the student-level, such that the same four were not selected for all students in this condition. The set of four selected for a given student stayed constant during the experiment. Ranks on that set were updated each week of the experiment.



**Measure Key**

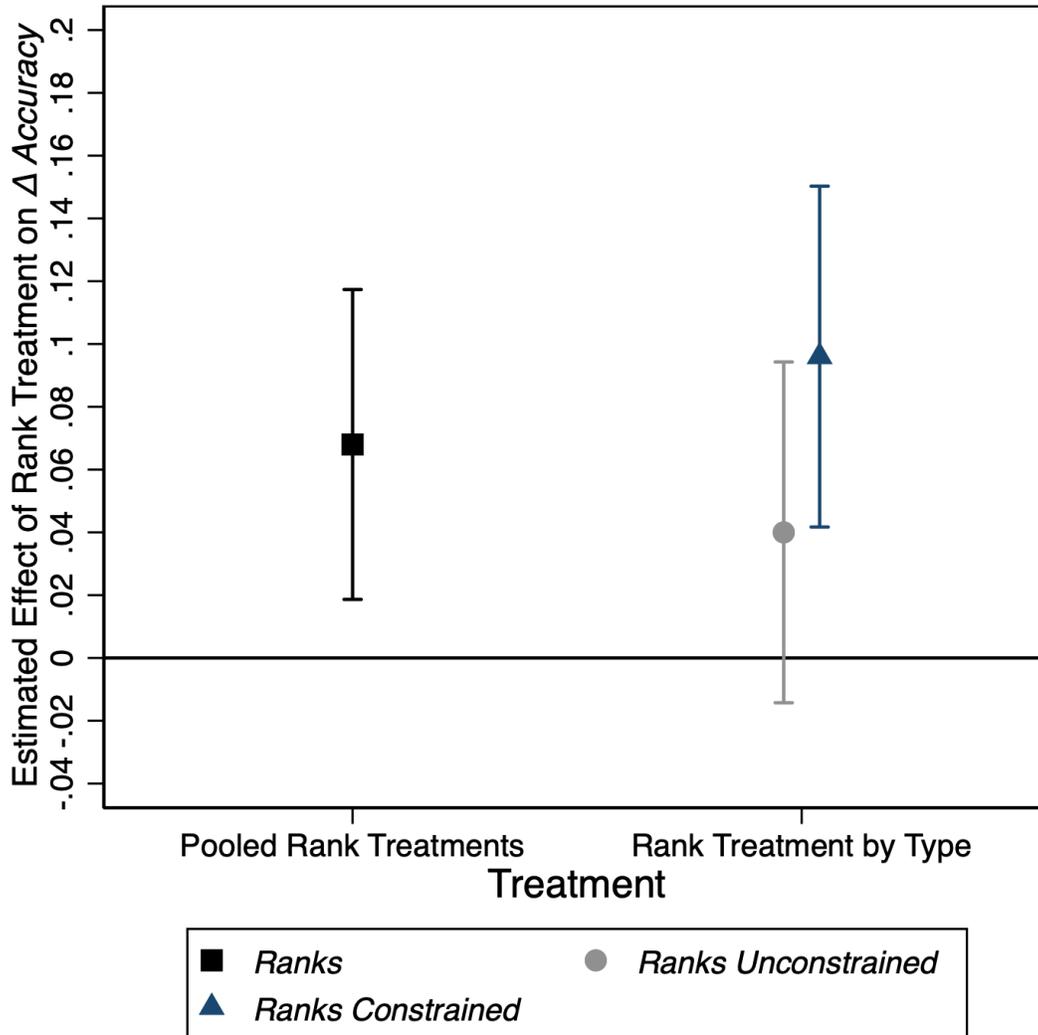
# days active	# of days you were active in the course
# problems	# of times you attempted any problems
# posts	# of posts you made in the discussion forum
grade	your grade in the course

Figure 1: Estimated Effect of Ranks Provided on Change in Grade



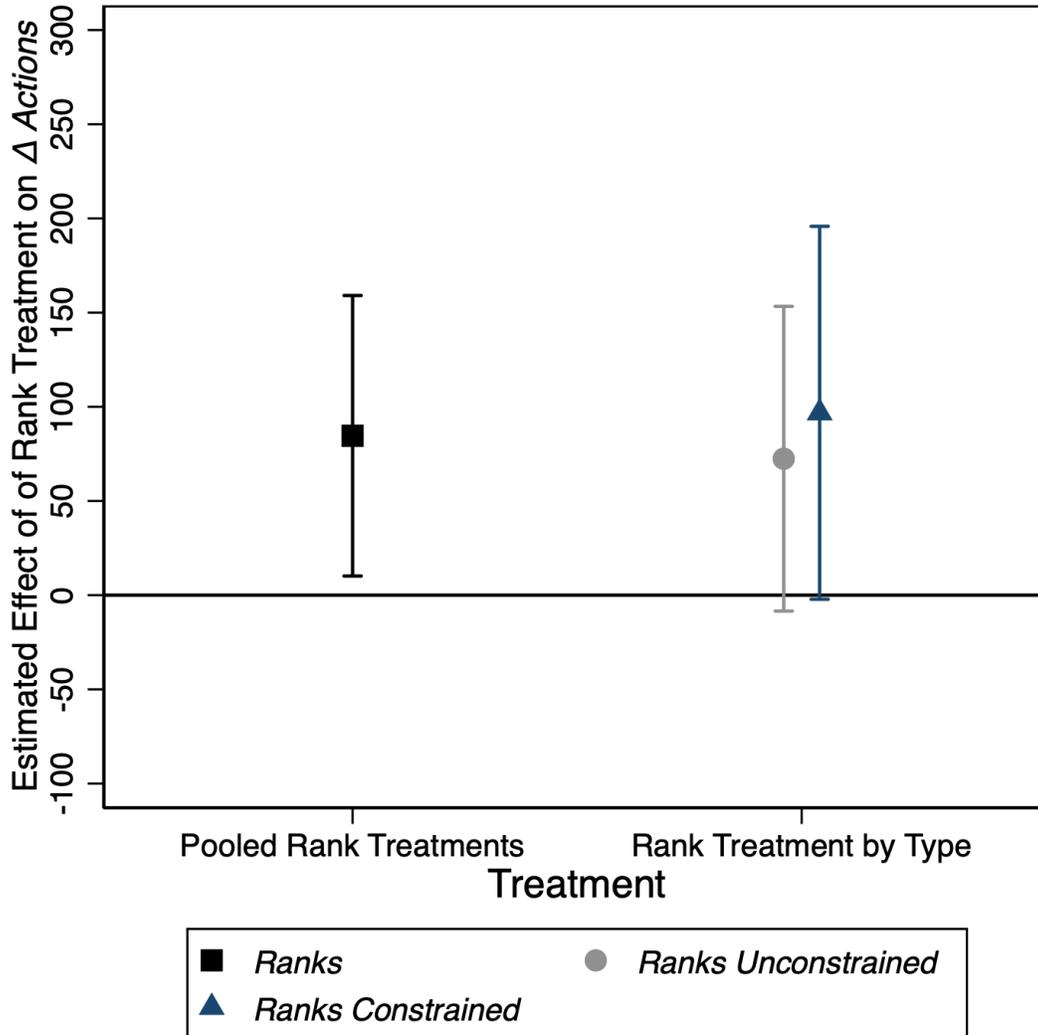
This figure displays the estimated effect on  $\Delta$  Grade of *Ranks*, or assignment to receive RPF, and then with the *Ranks* treatment split into two treatment arms: 1) *Ranks Unconstrained*, where feedback contains the student's ranks on the full set of nine from the list of "Rank-Based Feedback Variables" in the appendix, or 2) *Ranks Constrained*, where feedback contains only a randomly selected subset of four from the full set of nine. The estimated effect of *Ranks* on the left-hand side of the figure is from model 2a with treated as the only explanatory variable. The estimated effects of *Ranks Unconstrained* and *Ranks Constrained* on the right-hand side of the figure are from model 2b with *Ranks Unconstrained* and *Ranks Constrained* included at the same time as explanatory variables. The omitted experiment condition for comparison in both applications of models 2a and 2b are students who did not receive RPF. The estimates are from specifications that include the control variables listed in the main tables: *Gender*, *Age*, and *College Educated*. The sample is all students included in the experiment.  $N=1,397$ . Points represent effect estimates, and vertical lines through the points represent 90% confidence intervals.

Figure 2: Estimated Effect of Ranks Provided on Change in Accuracy



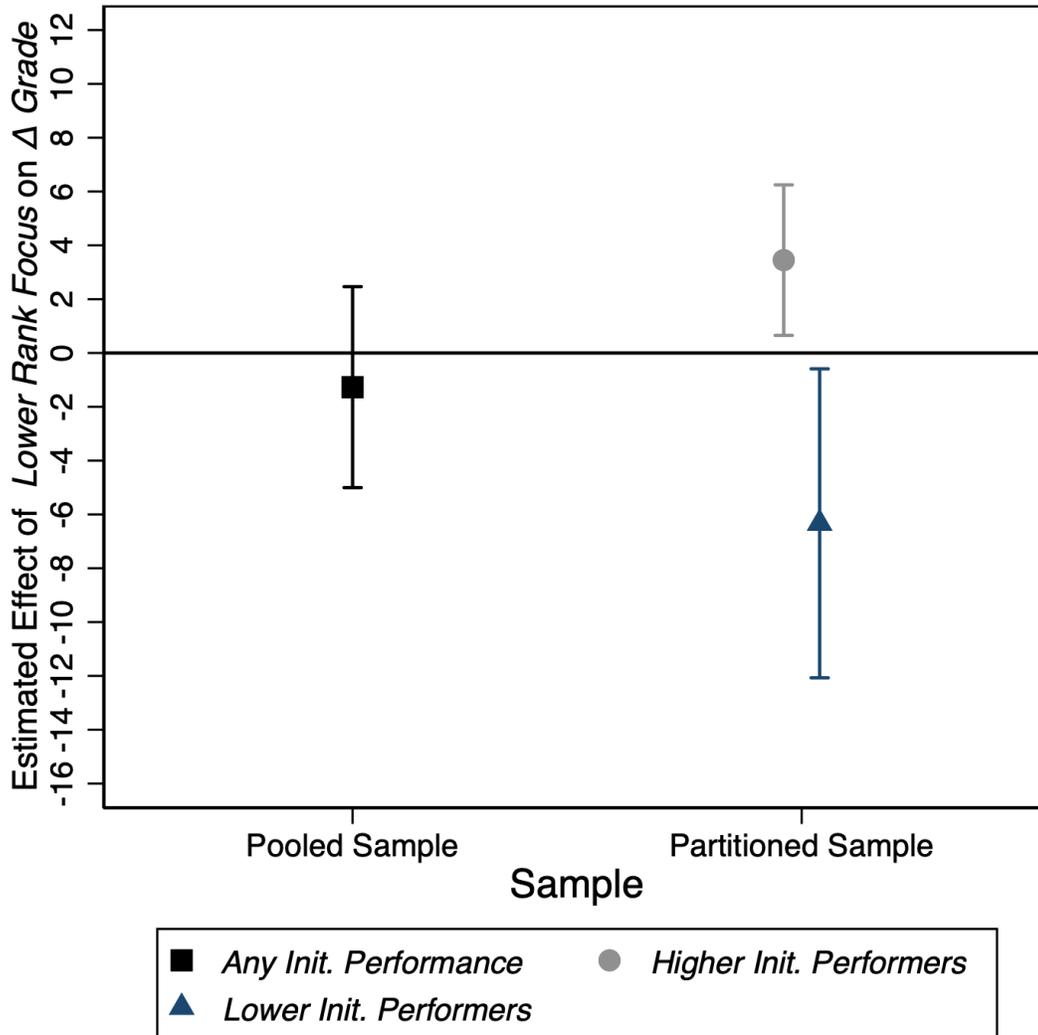
This figure displays the estimated effect on  $\Delta Accuracy$  of *Ranks*, or assignment to receive RPF, and then with the *Ranks* treatment split into two treatment arms: 1) *Ranks Unconstrained*, where feedback contains the student's ranks on the full set of nine from the list of "Rank-Based Feedback Variables" in the appendix, or 2) *Ranks Constrained*, where feedback contains only a randomly selected subset of four from the full set of nine. The estimated effect of *Ranks* on the left-hand side of the figure is from model 2a with treated as the only explanatory variable. The estimated effects of *Ranks Unconstrained* and *Ranks Constrained* on the right-hand side of the figure are from model 2b with *Ranks Unconstrained* and *Ranks Constrained* included at the same time as explanatory variables. The omitted experiment condition for comparison in both applications of models 2a and 2b are students who did not receive RPF. The estimates are from specifications that include the control variables listed in the main tables: *Gender*, *Age*, and *College Educated*. The sample is all students included in the experiment.  $N=1,397$ . Points represent effect estimates, and vertical lines through the points represent 90% confidence intervals.

Figure 3: Estimated Effect of Ranks Provided on Change in Actions



This figure displays the estimated effect on  $\Delta$  Actions of *Ranks*, or assignment to receive RPF, and then with the *Ranks* treatment split into two treatment arms: 1) *Ranks Unconstrained*, where feedback contains the student's ranks on the full set of nine from the list of "Rank-Based Feedback Variables" in the appendix, or 2) *Ranks Constrained*, where feedback contains only a randomly selected subset of four from the full set of nine. The estimated effect of *Ranks* on the left-hand side of the figure is from model 2a with treated as the only explanatory variable. The estimated effects of *Ranks Unconstrained* and *Ranks Constrained* on the right-hand side of the figure are from model 2a with *Ranks Unconstrained* and *Ranks Constrained* included at the same time as explanatory variables. The omitted experiment condition for comparison in the applications of models 2a and 2b are students who did not receive RPF. The estimates are from specifications that include the control variables listed in the main tables: *Gender*, *Age*, and *College Educated*. The sample is all students included in the experiment.  $N=1,397$ . Points represent effect estimates, and vertical lines through the points represent 90% confidence intervals.

Figure 4: Estimated Effect of *Lower Rank Focus* on Change in Grade



This figure displays the estimated effect on  $\Delta$  Grade of *Lower Rank Focus*, which is an indicator equal to one if the student was in the *Ranks Constrained* group, or the group that received ranks for only a randomly selected subset of four measures from the full set of nine from the list of "Rank-Based Feedback Variables" in the appendix, and if the student's average rank on the subset randomly selected to be provided to them was lower than their average rank on the full set. The estimates are shown on the left-hand side of the figure for all students in the *Rank Constrained* arm, and then on the right-hand side as estimated within subsamples of the *Rank Constrained* arm based on whether the student's *Grade Before* was at-or-above median ("Higher Init. Performer") or below median ("Lower Init. Performer"). The omitted experiment condition for comparison to are students in the *Ranks Constrained* arm for whom *Lower Rank Focus* was not equal to one. The estimates are from model 3 with all control variables listed in the tables: *Gender*, *Age*, and *College Educated*.  $N=511$  for the Pooled Sample portion of the figure, and  $N=258$  ( $N=253$ ) for the Partitioned Sample portion for the Higher Init. Performers (Lower Init. Performers) estimate. Points represent effect estimates, and vertical lines through the points represent 90% confidence intervals.

Table 1: Sample Selection and Descriptive Statistics

Panel A: Sample Selection					
All Students Active in Courses One-Week Prior to Experiment					1,819
Exclude Students who did not Participate in Graded Content Prior to Experiment					422
Final Sample					1,397
Panel B: Descriptive Statistics					
	N	Mean	SD	25th Percentile	75th Percentile
<i>Gender</i>	1,397	0.62	0.48	0	1
<i>Age</i>	1,397	45.64	21.62	29	56
<i>College Educated</i>	1,397	0.85	0.35	1	1
<i>Days Active</i>	1,397	11.63	7.22	7	14
<i>Videos</i>	1,397	126.75	213.19	36	130
<i>Problems</i>	1,397	116.40	171.96	48	104
<i>Posts</i>	1,397	2.18	7.80	0	2
<i>Comments</i>	1,397	52.13	283.79	0	28
<i>Chapters</i>	1,397	9.93	4.24	5	13
<i>Grade</i>	1,397	73.00	24.82	65	90
$\Delta$ <i>Grade</i>	1,397	22.45	27.59	6	34
<i>Accuracy</i>	1,397	0.57	0.23	0.44	0.69
$\Delta$ <i>Accuracy</i>	1,397	0.20	0.43	0.08	0.42
<i>Actions</i>	1,397	1,886.82	2,142.71	1,077	2,388
$\Delta$ <i>Actions</i>	1,397	789.75	958.41	280	1,013

This table shows descriptive statistics for demographics, explanatory variables, and dependent variables. The appendix lists variable definitions. Students entered the demographic information during registration. The measures *Days Active*, *Videos*, *Problems*, *Posts*, *Comments*, *Chapters*, and *Grade* are reported as they stood at the end of the online courses.

Table 2: Tests for Pre-Experiment Relationship Between Explanatory Variables and Student Demographics or Pre-Experiment Dependent Variables

	(1)	(2)	(3)	(4)
	<i>Ranks, Ranks x Constrained</i>	<i>Lower Rank Focus</i>	<i>Lower Rank Focus</i>	<i>Lower Rank Focus</i>
	T-stat P-val	T-stat P-val	T-stat P-val	T-stat P-val
<i>Gender</i>	0.870, 0.359	0.489	0.214	0.696
<i>Age</i>	0.893, 0.107	0.304	0.213	0.852
<i>College Educated</i>	0.425, 0.277	0.396	0.535	0.565
<i>Grade Before</i>	0.560, 0.202	0.683	0.489	0.970
<i>Accuracy Before</i>	0.620, 0.610	0.664	0.447	0.164
<i>Actions Before</i>	0.298, 0.802	0.193	0.632	0.214
Sample	All	Ranks Constrained	Ranks Constrained, Higher Init. Performer	Ranks Constrained, Lower Init. Performer
N	1,397	511	258	253

This table presents tests for pre-treatment relationships between explanatory variables and student demographics or pre-experiment course metrics. Column 1 shows P-values associated with T-statistics from OLS regressions using *Ranks* and *Ranks x Constrained* as the explanatory variables, with the given demographic or pre-experiment dependent variable. Columns 2-4 show P-values associated with the T-statistics from OLS regressions using *Lower Rank Focus* as the explanatory variable, with the given demographic or course metric variable of interest as the dependent variable. For Column 2 the sample is the Constrained group, or the group that received ranks for only a randomly selected subset of four measures from the full set of nine from the list of "Rank-Based Feedback Variables" in the appendix. For Columns 3 and 4, the sample is the Constrained group, then further split into subgroups based on whether the student's Grade Before was at-or-above median ("Higher Init. Performer") or below median ("Lower Init. Performer") categories. All OLS regressions used to generate P-values included Course Fixed Effects as do the OLS regressions in all tables that test for treatment effects. These tests demonstrate no statistically significant relationship between student demographics or pre-experiment dependent variables and the explanatory variables at a baseline. \*, \*\*, \*\*\* would denote significance at the .1, .05, and .01 levels respectively in the presence of such a relationship.

Table 3: Estimated Effects of Rank Constraints on  $\Delta$  Grade

	(1)	(2)	(3)	(4)
	$\Delta$ Grade			
<i>Ranks</i>	2.65*	2.55*	0.61	0.53
	[1.75]	[1.68]	[0.37]	[0.32]
<i>Ranks x Constrained</i>			4.07***	4.04***
			[2.64]	[2.62]
<i>Gender</i>		-2.04		-1.90
		[-1.47]		[-1.37]
<i>Age</i>		-0.04		-0.04
		[-1.26]		[-1.14]
<i>College Educated</i>		4.30**		4.39**
		[2.30]		[2.36]
Course FE	Yes	Yes	Yes	Yes
Sample	All	All	All	All
N	1,397	1,397	1,397	1,397

This table presents estimates of the effect on  $\Delta$  Grade of *Ranks*, or assignment to receive RPF, and of *Ranks x Constrained*, or constraining the ranks provided by providing only a randomly selected subset of four from the full set of nine from the list of "Rank-Based Feedback Variables" in the appendix. Note that the specifications does not include *Constrained* separately from its interaction with *Ranks* given that *Constrained* can only take a non-zero value in the presence of RPF. If *Constrained* were entered in an OLS regression along with *Ranks* and *Ranks x Constrained*, the regression automatically drops *Constrained* because it is subsumed. Columns 1 and 3 show estimates without demographic controls, and columns 2 and 4 show estimates with demographic controls. T-statistics are based on robust standard errors and are reported in brackets below each coefficient. \*, \*\*, \*\*\* denote significance at the .1, .05, and .01 levels respectively.

Table 4: Estimated Effects of Rank Constraints on  $\Delta$  Accuracy

	(1)	(2)	(3)	(4)
	$\Delta$ Accuracy			
<i>Ranks</i>	0.07**	0.07**	0.04	0.04
	[2.32]	[2.25]	[1.27]	[1.19]
<i>Ranks x Constrained</i>			0.06**	0.06**
			[1.98]	[1.98]
<i>Gender</i>		-0.04		-0.04
		[-1.48]		[-1.40]
<i>Age</i>		-0.00		-0.00
		[-0.97]		[-0.88]
<i>College Educated</i>		0.11**		0.12**
		[2.52]		[2.55]
Course FE	Yes	Yes	Yes	Yes
Sample	All	All	All	All
N	1,397	1,397	1,397	1,397

This table presents estimates of the effect on  $\Delta$  Accuracy of *Ranks*, or assignment to receive RPF, and of *Ranks x Constrained*, or constraining the ranks provided by providing only a randomly selected subset of four from the full set of nine from the list of "Rank-Based Feedback Variables" in the appendix. Note that the specifications does not include *Constrained* separately from its interaction with *Ranks* given that *Constrained* can only take a non-zero value in the presence of RPF. If *Constrained* were entered in an OLS regression along with *Ranks* and *Ranks x Constrained*, the regression automatically drops *Constrained* because it is subsumed. Columns 1 and 3 show estimates without demographic controls, and columns 2 and 4 show estimates with demographic controls. T-statistics are based on robust standard errors and are reported in brackets below each coefficient. \*, \*\*, \*\*\* denote significance at the .1, .05, and .01 levels respectively.

Table 5: Estimated Effects of Rank Constraints on  $\Delta$  Actions

	(1)	(2)	(3)	(4)
	$\Delta$ Actions			
<i>Ranks</i>	81.97*	84.63*	74.54	72.44
	[1.78]	[1.87]	[1.51]	[1.47]
<i>Ranks x Constrained</i>			14.83	24.38
			[0.23]	[0.39]
<i>Gender</i>		24.92		25.81
		[0.40]		[0.41]
<i>Age</i>		3.13**		3.15**
		[2.05]		[2.11]
<i>College Educated</i>		84.84		85.42
		[1.50]		[1.50]
Course FE	Yes	Yes	Yes	Yes
Sample	All	All	All	All
N	1,397	1,397	1,397	1,397

This table presents estimates of the effect on change in  $\Delta$  Actions of *Ranks*, or assignment to receive RPF, and of *Ranks x Constrained*, or constraining the ranks provided by providing only a randomly selected subset of four from the full set of nine from the list of "Rank-Based Feedback Variables" in the appendix. Note that the specifications does not include *Constrained* separately from its interaction with *Ranks* given that *Constrained* can only take a non-zero value in the presence of RPF. If *Constrained* were entered in an OLS regression along with *Ranks* and *Ranks x Constrained*, the regression automatically drops *Constrained* because it is subsumed. Columns 1 and 3 show estimates without demographic controls, and columns 2 and 4 show estimates with demographic controls. T-statistics are based on robust standard errors and are reported in brackets below each coefficient. \*, \*\*, \*\*\* denote significance at the .1, .05, and .01 levels respectively.

Table 6: Estimated Effects of *Lower Rank Focus* on  $\Delta$  Grade

	(1)	(2)	(3)	(4)	(5)	(6)
	$\Delta$ Grade					
<i>Lower Rank Focus</i>	-1.39	-1.27	3.61**	3.45**	-6.31*	-6.33*
	[-0.61]	[-0.56]	[2.10]	[2.03]	[-1.83]	[-1.81]
<i>Gender</i>		-4.16*		-0.46		0.31
		[-1.75]		[-0.24]		[0.08]
<i>Age</i>		-0.05		0.00		-0.01
		[-0.81]		[0.06]		[-0.12]
<i>College Educated</i>		4.36		6.44***		1.45
		[1.48]		[2.95]		[0.32]
Course FE	Yes	Yes	Yes	Yes	Yes	Yes
			Ranks	Ranks	Ranks	Ranks
Sample	Ranks	Ranks	Constrained,	Constrained,	Constrained,	Constrained,
	Constrained	Constrained	Higher Init.	Higher Init.	Lower Init.	Lower Init.
			Perf.	Perf.	Perf.	Perf.
N	511	511	258	258	253	253

This table presents estimates of the effect on  $\Delta$  Grade of *Lower Rank Focus*. *Lower Rank Focus* is equal to one if the student was in the Ranks Constrained group, or the group that received ranks for only a randomly selected subset of four measures from the full set of nine from the list of "Rank-Based Feedback Variables" in the appendix, and if the student's average rank on the subset randomly selected to be provided to them was lower than their average rank on the full set. For columns 3 through 6, the sample is the Ranks Constrained group, then further split into subgroups based on whether the student's Grade Before was at-or-above median ["Higher Init. Performer"] or below median ["Lower Init. Performer"]. Columns 1, 3, and 5 show estimates without demographic controls, and columns 2, 4, and 6 show estimates with demographic controls. T-statistics are based on robust standard errors and are reported in brackets below each coefficient. \*,\*\*,\*\*\* denote significance at the .1, .05, and .01 levels respectively.

Table 7: Estimated Effects of *Lower Rank Focus* on  $\Delta$  Accuracy

	(1)	(2)	(3)	(4)	(5)	(6)
	$\Delta$ Accuracy					
<i>Lower Rank Focus</i>	0.004 [0.12]	0.008 [0.20]	0.106** [2.38]	0.103** [2.31]	-0.096* [-1.80]	-0.092* [-1.72]
<i>Gender</i>		-0.106** [-2.55]		0.007 [0.13]		-0.116** [-2.03]
<i>Age</i>		-0.000 [-0.19]		0.001 [0.42]		-0.000 [-0.01]
<i>College Educated</i>		0.054 [0.86]		0.052 [0.79]		0.048 [0.48]
Course FE	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Ranks Constrained	Ranks Constrained	Constrained, Higher Init. Perf.	Constrained, Higher Init. Perf.	Constrained, Lower Init. Perf.	Constrained, Lower Init. Perf.
N	511	511	258	258	253	253

This table presents estimates of the effect on  $\Delta$  Accuracy of *Lower Rank Focus*. *Lower Rank Focus* is equal to one if the student was in the Ranks Constrained group, or the group that received ranks for only a randomly selected subset of four measures from the full set of nine from the list of "Rank-Based Feedback Variables" in the appendix, and if the student's average rank on the subset randomly selected to be provided to them was lower than their average rank on the full set. For columns 3 through 6, the sample is the Ranks Constrained group, then further split into subgroups based on whether the student's *Grade Before* was at-or-above median ("Higher Init. Performer") or below median ("Lower Init. Performer"). Columns 1, 3, and 5 show estimates without demographic controls, and columns 2, 4, and 6 show estimates with demographic controls. T-statistics are based on robust standard errors and are reported in brackets below each coefficient. \*, \*\*, \*\*\* denote significance at the .1, .05, and .01 levels respectively.

Table 8: Estimated Effects of *Lower Rank Focus* on  $\Delta$  Actions

	(1)	(2)	(3)	(4)	(5)	(6)
	$\Delta$ Actions					
<i>Lower Rank Focus</i>	-171.78	-176.19*	45.60	34.93	-393.44*	-392.32**
	[-1.62]	[-1.71]	[0.71]	[0.56]	[-1.94]	[-2.00]
<i>Gender</i>		-7.42		115.09*		-93.49
		[-0.05]		[1.70]		[-0.34]
<i>Age</i>		2.09		4.53**		0.46
		[0.60]		[2.58]		[0.07]
<i>College Educated</i>		187.46		70.58		336.71
		[1.45]		[0.86]		[1.28]
Course FE	Yes	Yes	Yes	Yes	Yes	Yes
			Ranks	Ranks	Ranks	Ranks
Sample	Ranks	Ranks	Constrained,	Constrained,	Constrained,	Constrained,
	Constrained	Constrained	Higher Init.	Higher Init.	Lower Init.	Lower Init.
			Perf.	Perf.	Perf.	Perf.
N	511	511	258	258	253	253

This table presents estimates of the effect on  $\Delta$  Actions of *Lower Rank Focus*. *Lower Rank Focus* is equal to one if the student was in the Ranks Constrained group, or the group that received ranks for only a randomly selected subset of four measures from the full set of nine from the list of "Rank-Based Feedback Variables" in the appendix, and if the student's average rank on the subset randomly selected to be provided to them was lower than their average rank on the full set. For columns 3 through 6, the sample is the Ranks Constrained group, then further split into subgroups based on whether the student's *Grade Before* was at-or-above median ("Higher Init. Performer") or below median ("Lower Init. Performer"). Columns 1, 3, and 5 show estimates without demographic controls, and columns 2, 4, and 6 show estimates with demographic controls. T-statistics are based on robust standard errors and are reported in brackets below each coefficient. \*, \*\*, \*\*\* denote significance at the .1, .05, and .01 levels respectively.