Does Feedback Sentiment Affect Entrepreneurial Performance: A Quasi-Experimental Approach

Sandy Yu
UC Berkeley
sandyyu@berkeley.edu

Alicia Robb
Kauffman Foundation
UC Berkeley
aliciarobb@berkeley.edu

This draft: December 6, 2016

Preliminary Draft– Please Do Not Circulate

Abstract

Does feedback affect start-up firm performance? Feedback can be beneficial and particularly actionable for early-stage firms—it may motivate firms to experiment and improve on their ideas, or help firms speed up the failure process for lower quality ideas. In this paper, we use a proprietary database of business plan competition participants and judges combined with text analysis of feedback to investigate whether firms incorporate feedback and change over time. Specifically, we leverage the feedback sentiment of randomly assigned judges as an instrumental variable to estimate causal effects of feedback on short-term and long-term performance. We find that firms improve within-competition performance after receiving more negative feedback that is mapped to specific areas of improvement and received in the initial round. Interestingly, there are heterogeneous treatment effects based on firm quality for post-competition performance. Specifically, after receiving negative feedback, higher quality firms are more likely to remain operational, while lower quality firms are more likely to shut down.

Keywords: entrepreneurship, feedback, pitch competitions, performance, failure

* We are grateful to Howard Aldrich, Aaron Chatterji, Laura Huang, and participants of the Kauffman Emerging Scholars Workshop, Duke Strategy Conference, West Coast Research Symposium, and Social Enterprise @ Goizueta for their helpful comments. We also thank Adam Rentschler for access to data and Trevor McCoy for excellent research assistance. Financial support from the Kauffman Foundation is gratefully acknowledged.

1 Introduction

Timely feedback and experimentation are important to innovation and the entrepreneurial process (Kerr, Nanda, and Rhodes-Kropf, 2014; Manso, 2011). Feedback can be particularly beneficial and actionable for early-stage firms since it can motivate firms to experiment and improve on their idea, or help firms fail faster. In fact, many business plan competitions, accelerators, and entrepreneurship courses emphasize feedback and mentorship as core components. Yu (2016) finds that in an accelerator setting, accelerator participants receive more frequent feedback than non-accelerator counterparts, which enables founders to fail faster if the quality of the idea is low. However, our understanding of what type of feedback is most effective and actionable for entrepreneurs is quite limited. The main research question we seek to answer is: how does feedback sentiment contribute to short-term and long-term performance for early-stage ventures? In particular, how do entrepreneurs who advance through business plan competition stages incorporate feedback? Using a sample of firms who participated in various stages of a large-scale business plan competition, we apply sentiment analysis and text analysis techniques to investigate how the sentiment of judges' comments affect the performance of firms within a business plan competition as well as post-competition.

While feedback may play a role in various stages of firm development, we are particularly interested in feedback provided to entrepreneurs during the nascent stage of the firm. In addition, founders can be exposed to feedback from peers, mentors, and teachers through both informal and formal channels. If we focus on feedback in an entrepreneurship setting, the results in existing studies vary depending on the stage of the firm and source of feedback. In particular, Lerner and Malmendier (2013) leverage random assignment of entrepreneurs in MBA sections to analyze the impact of entrepreneurial peers on new ventures. They find that sections with more entrepreneurial peers have lower startup rates, likely due to feedback given by peers that might discourage founding new ventures based on lower quality ideas. In a different setting, Wagner (2016) uses a randomized controlled trial to test the effect of written feedback on business plans and whether it impacts the subsequent survival of these ventures. All of the firms in that sample participate in an incubator program and receive funding. After four years, the entrepreneurs who received feedback were more likely to list their ventures in specific databases. In both papers, the analyses are framed around entrepreneurs receiving any feedback compared to no feedback at all. However, in our study, all entrepreneurs receive feedback from randomly assigned judges and the variation comes from different sentiment of the feedback.

To our knowledge, this is the first paper to utilize sentiment analysis in an entrepreneurship context and investigate specific attributes of that feedback. The sentiment of feedback comments gives us insight into more subtle aspects of the judges' evaluations that is not captured in the quantitative scores. In the competition press releases, it is highlighted that companies "will receive constructive feedback generated from the evaluation process," and in this context, constructive feedback tends to be negative feedback. In other words, a firm may receive both high scores and written feedback that is negative in sentiment. Research in psychology suggests there is a "negativity effect" where negative stimuli have greater impact and higher informational value than equally intense positive stimuli (Peeters and Czapinski, 1990). Furthermore, credible negative feedback motivates higher goals and higher performance (Podsakoff and Farh, 1989). Given that firms receive feedback from a panel of credible judges (judges are often industry experts or investors) and negative feedback may contain higher informational value, one would expect founders to act upon negative feedback and reassess their strategy, which may improve both short-term and long-term performance.

The data includes a sample of 1,459 firms that participated in a large-scale state-sponsored business plan competition. All of the applicants receive structured feedback from three or four judges in the first round. Firms that were selected to advance to the semifinal round were then asked to submit updated and more detailed business plans. All of the semifinalists receive a second round of structured feedback and finalists are then chosen for a live presentation to a panel of judges. Based on the live pitch, 6 winners are chosen per competition and awarded up to $250,000 in prize money. Through the electronic submission platform used by both applicants and judges, we are able to collect firm information, judges' scores for individual firms, and the full text of comments that firms receive from the judges. Further information is hand-collected to track individual firm history, including funding, exit events, and closures.

Our empirical strategy consists of a quasi-experimental approach where we leverage the random assignment of judges to firms as a source of variation in feedback sentiment. This methodology builds upon the labor economics literature where judge "incarceration propensity" is used to establish causal effects of incarceration length (Aizer and Dolye, Jr., 2015; Di Tella and Schargrodksy, 2013; Kling 2006). More recently in the innovation literature, patent examiner leniency has also been used to study which patent applications are granted (Farre-Mensa, Hegde, Ljungqvist, 2016; Sampat and Williams, 2015). In this research context, judges have different propensities to give more negative or positive feedback, and the random judge assignment allows us to investigate the causal effect of feedback sentiment on performance.

There are three main results. First, while negative comments seem to help firms improve, less negative comments on average are positively associated with winner status. Second, firms improve their within-competition performance after receiving more negative feedback in the initial round, but only if the comments are specific to particular areas of improvement. Third, firms of heterogeneous quality respond differently to feedback sentiment with respect to long-term performance. In particular, high quality and low quality firms respond in opposite ways following negative feedback. Higher quality firms are more likely to remain operational, while lower quality firms are more likely to shut down. Overall, these findings suggest that the sentiment and specificity of feedback contribute to how well a firm performs in subsequent rounds of the competition as well as in the longer term post-competition.

While business plan competitions may be a potential source of financing and certification (Howell 2016), our research provides evidence that entrepreneurs benefit from judges' feedback, even if they do not ultimately win the competition. There are many ways for entrepreneurs to acquire feedback and business plan competitions are a popular source of feedback that can guide entrepreneurs as they launch their ventures. In addition to understanding whether feedback helps entrepreneurs, we take a further step and investigate what types and sentiment of feedback can be more effective in helping entrepreneurs improve firm performance. A better understanding of the impact of sentiment can inform the implementation of more effective feedback structures that can benefit entrepreneurs as well as inform judges.

2 Institutional Context

State-Sponsored Business Plan Competition

The context of this paper is a large-scale state-sponsored business plan competition. Applicants to this competition must be aiming for commercialization of an innovation technology or scientific solution in high-technology industries, so the companies that apply generally have sophisticated technologies and applications for patents. The business plan competition is held twice a year, and winners are awarded up to $250,000 in non-dilutive grant funding. In the first round of the competition, all firms submit business plans and receive online feedback from the judges. Then, a subset of firms are chosen for the semifinalist round and asked to submit an updated and more detailed business plan. Based on the semifinalist business plans, finalists are then chosen to pitch in front of a panel of judges. The entire competition cycle is a multi-month process that results in selection of the winners, who are awarded grant funds to commercialize their product over the following year. Judges for the competition possess expertise in relevant industries and they volunteer their time to judge various rounds of the competition. Primary job functions of the judges include venture capitalists, angel investors, entrepreneurs, industry experts, startup and industry executives, and start-up mentors.

Unique features of this competition allow us identify the effect of feedback sentiment and other feedback attributes. First, the same scoring rubric is used for all competitions and judges are randomly assigned. Second, the selection criteria for semifinalists is consistent across the competitions. Specifically, firms are ranked according to overall judges' scores and companies above certain cutoff ranks are guaranteed or likely to advance. Participants in the sample are also tracked over time after the end of the competition in order to capture various longer term performance measures.

Overall scores and dimension scores

After each round of the competition, founders have access to a dashboard containing a summary of the feedback and specific text comments from the judges. Figure 1 provides example of what the founders see. There is a quantitative summary where founders see not only their overall score and score in each dimension, but also where they stand relative to other firms in the competition and the distribution of scores. In addition, founders also receive a heat map for each scoring dimension where the darker shade of the color of the box reflects how much consensus there was between the judges for a particular evaluation.

Judges and Comments

In addition to overall scores for each dimension, founders also receive comments from judges who choose to give direct feedback. In fact, almost all judges give comments, and very few firms receive no comments at all. All feedback is anonymous, so founders only know if comments were given by different judges. The length of the comments range from one sentence to full paragraphs, and the tone ranges from sarcastic to very encouraging. For example, "Premise is very very flawed…this business is going to need all the talent and advisors it can get" or "Huge market opportunity and really interesting technology." Each firm is presented with 3-4 comments from different judges, and founders must also process all this information and decide whether and how to revise the business plan. We take advantage of the random assignment of judges and the corresponding full text of their feedback, and use text analysis to investigate different aspects of the comments, most prominently, the sentiment of the comments.

3 Data and Descriptive Statistics

3.1 Data Sources

The main data source of this paper is Valid Eval, which is a proprietary online evaluation platform created with the goal of helping organizations make defensible decisions. It is used by business plan competitions, grant administrators, and organizations such as accelerators to record and distribute team applications and evaluations. Through Valid Eval, we have access to data on all firms and judges who participated in the competition from years 2012 through 2015. This results in a full sample of 1,459 firms from two competitions each year. Summary statistics for the competition events are presented in Table 1. We see that there are 8 competitions, and each event receives 182 applicants on average, of which 31 applicants move onto the semi-final round, 10 pitch in the final round, culminating in 6 eventual winners. For each event, there is a large pool of around 48 judges, and it is worth pointing out that judges for the competition follow a strict scoring rubric, and all scores and comments are logged by the system. Thus, we have access to the qualitative and quantitative feedback each firm receives from each judge. Table 2 contains summary statistics of firm-level data. We can see that average firm scores improve between the first and semifinal rounds. Interestingly, the number of comments increases across rounds, too, but more importantly, firms are receiving around three comments, which indicates comments from three different judges. The length of the comments almost doubles in the semifinal round, indicating that judge comments may be more elaborate in later stages of the competition. If we break down the firm scores based on applicant status, we see in Figure 2 that firms that advance to later rounds in the competition score much higher than rejected firms in the first round. However, semifinal round scores are fairly close between semifinalists, finalists, and eventual winner. However, within the same status group, scores actually decrease. One possibility is that the judges have higher standards in the semifinals, so given the same quality of firms, the average scores are lower.

In addition to within-competition outcomes, further data is collected across several sources to track firm performance over time. Even though public information on funding rounds and operational status is very sparse and inconsistent, we were able to collect and cross check additional details using SEC Form D's, Crunchbase, Angellist, LinkedIn, SBIR, company websites, news articles, and even YouTube videos created by founders. The main sample thus includes Valid Eval data combined with additional firm details including founding date, funding rounds, operational status, and exit dates.

3.2 Text Analysis

Sentiment Analysis

The first type of text analysis is sentiment analysis, or opinion mining. This technique has been widely used to analyze news articles, online reviews, and tweets. In this context, we use it to determine whether judges' comments were positive or negative. We leverage the Stanford CoreNLP natural language analysis tools, which uses a deep learning model to compute the sentiment based on how words compose the meaning of longer phrases. Consequently, it can learn that a sentence is negative overall even if it contain positive words (Socher et al, 2014). Using this tool, we gain insight into the polarity of each judges' comments. The polarity of a sentence takes on the following values: very negative, negative, neutral, positive, and very positive. For example, "This is a services company" is neutral, "The financial projections include gross margins of 95%...I'm unaware of any business with that gross margin" is very negative, and "The management team and advisory board are top-notch" is very positive. Since some judges give multi-sentence comments, each sentence is scored and then aggregated to produce the overall sentiment of the comment. Descriptive statistics of the sentiment of judge comments is presented in Figure 3. Surprisingly, the mean sentiment of comments is slightly negative, regardless of competition status. Across stages of the competition, the average sentiment becomes more negative, and there is wider dispersion of sentiment in later rounds. This is consistent with judges having higher standards in the semifinal round, while providing feedback that comes across as more extreme.

Dimension-Mapping

The second type of text analysis involves using keywords to map judge comments to specific dimensions of evaluation. As seen in Figure 1, dimensions of evaluation include 1) Market validation & analysis, 2) Industry attractiveness, 3) Product/Solution, 4) Business model, 5) Risk vs. Talent, and 6) Presentation quality. For example, the comment "Founder very strong, but needs a team….Need better, more detailed market analysis and segmentation" is mapped to the "Market validation and analysis" and "Risk vs. Talent" dimensions. Mapping comments to dimensions allows us to measure the specificity of a comment, but also whether there is mismatch between the quantitative and qualitative feedback founders receive.

3.3 Variables

Outcome Variables

There are three sets of outcomes variables. The first two capture within-competition performance and the third captures post-competition performance.

Within-competition performance – advancement in competition. This includes the semifinal score, whether a firm becomes a finalist, and whether a firm becomes a winner. *Semifinal Score* is the overall score received based on the semifinal business plan. This means the sample for analysis only includes firms that advanced past the semifinal round. *Finalist* is a binary variable, with 1 indicating that a firm was selected as a finalist, and 0 indicating otherwise. Similarly, *Winner* is a binary variable with 1 indicating that a firm was selected as a final winner.

Within-competition performance –improvement across competition rounds. In order to capture direct improvements as a result of feedback, we create three binary variables to indicate whether a firm improved between Round 1 and Round 2, which is the semifinalist round. These variables track improvement in specific dimensions as a result of comments-mapping, *DimensionScoreImproved*, whether the overall score improved, *ScoreImproved*, and whether the ranking of a firm improved, *RankImproved*.

Post-competition performance and sentiment. Post-competition performance is captured both by the funding activity of the firm and operational status. *Funding Total* measures the total amount of funding, in millions, a firm receives after the competition. The operational status of the firm can take on the following values: operating, IPO, acquired, and closed. In particular, we create binary variables, *Acquired* and *Closed* to indicate that a firm has exited the market. For acquired companies, the date of acquisition is also recorded if available. The inherent challenge to determining whether a firm has shut down is that some firms become "zombie firms" and the operational status becomes unclear. In order to verify that a firm has gone out of business, we use social media activity as an indicator for whether or not a firm is still active. If the last tweet or Facebook post is older than December 2015, the firm is marked as closed. For companies with no social media presence, we check the company website or blog for a copyright 2016 mark or blogpost later than December 2015. If either of these do not exist, the company is marked as closed. If a firm has no social media or web presence, the founders are contacted to verify the operational status directly. In these cases, all but a single firm was closed, reaffirming the usability of an online presence as a measurement of whether a firm has shut down. The post-competition measures are current as of August 2016.

Explanatory Variables

We are interested in how different attributes of feedback affect performance outcomes, so the explanatory variables include *Sentiment Mean, Sentiment Max*, *Sentiment Min*, which are aggregated measures of judge sentiments across all comments for the focal firm. Dimension-mapping measures include *Comments Mapped*, which is a binary variable, with 1 indicating that the comment mapped to at least one dimension; and *Number Dimensions Mapped*, which is the number of dimensions a comment is mapped to.

Control Variables

Given the variation in the competition cycle and firm industries, we include competition fixed effects and industry fixed effects as control variables in the analysis. We also include year fixed effects to control for macro-economic trends.

4 Empirical Framework

The main analysis uses an instrumental variables approach to examine how the polarity of judges' comments affect performance such as improvements in score across competition rounds, fundraising and closures, and within-competition performance including improvements in score and ranking across competition rounds.

Instrumental Variables Setup and Calculation

Here we exploit the random assignment algorithm for judges to tease out how the polarity of judges' feedback affects future performance. Specifically, we use the tendency of a randomly assigned judge to give more positive or negative feedback as an instrument for the overall sentiment of comments that a firm receives. Then, we compare within-competition and post-competition performance outcomes between firms assigned to judges that have different propensities to give harsher or more encouraging feedback. Due to the instrumental variables approach, we can interpret differences in performance as a causal effect of the change in feedback sentiment as a consequence of differences in these propensities.

Assignment of Judges to Firms

The assignment of judges in this competition is automated and relies on a confidential and proprietary algorithm. At a high level, each judge is automatically assigned to multiple firms during a given competition round and matched based on industry if possible. Even though there is a higher likelihood of a judge being assigned to a firm in the same industry, this is orthogonal to their sentiment propensity, and for our purposes, judges need to be randomly assigned relative to their tendency to give either more negative or positive feedback. There is no evidence that judges with certain technology backgrounds always give more positive or negative comments, or that judges matched to firms outside their technology type would have a propensity to fall on either side of the sentiment scale. Judges tend to vary in feedback sentiment on an individual level. This gives us confidence that the assignment of judges creates randomization, resulting in a valid instrument for analysis.

IV Calculation

Each firm is assigned several judges for evaluation during a given competition round. For each firm we create an instrument that represents an aggregate measure of the "sentiment propensity" of judges assigned to evaluate the firm. The instrument, *Z1*, is an aggregate of leave-out means for each firm i assigned to n_i judges $j_1, j_2, \ldots, j_n$:

$$Z1_i = \min_{n_i}\{\left(\frac{1}{N_{j_1}-1}\right)\sum_{m\neq i}^{N_{j_1}-1} P_m, \dots, \left(\frac{1}{N_{j_n}-1}\right)\sum_{m\neq i}^{N_{j_n}-1} P_m\} \quad (1)$$

N_{j_n} is the total number of firms within a competition evaluated by judge *j*, *m* indexes the firms evaluated by judge *j* where *P* is the polarity of the focal judge's feedback based on sentiment analysis of feedback comments. In order to focus on the polarity between positive and negative sentiments, comments with sentiment score = 0 are excluded. The firm-level leave-out means consists of the minimum sentiment score, across *n* judges for firm *i*.
s are aggregated on the firm level as an average across *n* judges. For a second instrument, *Z2*, the leave-out means are aggregated on the firm level as an average across *n* judges for firm *i*:

$$Z2_i = \left(\frac{1}{n_i}\right)\sum_{1}^{n_i}\left[\left(\frac{1}{N_{j_n}-1}\right)\sum_{m\neq i}^{N_{j_n}-1} P_m\right] \quad (2)$$

In both the first and second stages of the IV regressions, we also include year and industry fixed effects. In Figure 4, we show the distribution of the leave-out means of the sentiment scores to demonstrate the variation in the instruments. There is substantial variation across judges present in both instruments.

We estimate the effect of feedback sentiment on within-competition outcomes according to Equation (6) below. A logit model is used to regress *DimensionScoreImproved*, *ScoreImproved*, and *RankImproved* on feedback sentiment received by firm *i*, and include competition fixed effects. To address endogeneity concerns that lower quality of firms receive more negative feedback, or that judges are selective about types of feedback they give, we instrument feedback sentiment using firm-level leave-out means specified by *Z1* and *Z2*.

$$P(Improved_i = 1) = P(\alpha + \theta FeedbackSentiment_i + Competition_i + w_i \geq 0) \quad (3)$$

To estimate the effect of feedback sentiment on post-competition performance outcomes, we use logit models specified in Equations (7) and (8) to regress whether a firm goes out of business or becomes acquired on the feedback sentiment received by firm *i*, controlling for round 1 scores, year fixed effects, and industry fixed effects. Similarly, we use ordinary least squares to regress the log of total funding raised post-competition on feedback sentiment with the same set of controls. Feedback sentiment is instrumented by firm-level leave-out means specified by *Z1* and *Z2*.

$$P(Closed_i = 1) = P(\alpha + \theta FeedbackSentiment_i + ScoreRd1_i + Year_i + Industry_i + w_i \geq 0) \quad (4)$$

$$P(Acquired_i = 1) = P(\alpha + \theta FeedbackSentiment_i + ScoreRd1_i + Year_i + Industry_i + \xi_i \geq 0) \quad (5)$$

$$\log(FundingTotal_i) = \alpha + \theta FeedbackSentiment_i + ScoreRd1_i + Year_i + Industry_i + \epsilon_i \quad (6)$$

5 Results

5.1 Feedback Attributes and Performance Outcomes

The regression results of how different sentiment measures influence performance outcomes are reported in Table 3.[1] Consistent with the descriptive statistics, the coefficient on average sentiment is positive and statistically significant in both rounds of the competition, whereas the min and max are not statistically significant. It is worth noting that very negative sentiments are associated with lower semifinal scores, which is what we would expect.

The results in Table 4 address how the specificity of comments may affect semifinal round scores and the probability of advancing to finalist or winner status. In columns 1 and 3, we see that the fact that comments are mapped at all seem important for the semifinal score and becoming a winner. However, in columns 2 and 3, the number of dimensions mapped have mixed effects during different rounds of the competition. One explanation is that when finalists pitch their products in the final round a focused pitch is better, whereas in a written business plan, it is better to address all dimensions. Overall, the results here are less clear.

5.2. Main Results: Feedback Sentiment and Performance Outcomes

Within-competition performance. The results in Table 5 show that the feedback sentiment is predictive of whether a firm improves over the course of the competition, specifically for overall score and rank improvements. In column 1 we see the first stage results of instrument *Z1* with

[1] In addition to sentiment, other attributes of feedback, such as the number of sentences in the comments, also contribute to performance. Regression results of these quantitative attributes on performance outcomes are reported in the Appendix Table A1.

additional controls. The coefficient is statistically significant, giving us confidence of the validity of the instrument. However, in column 2, feedback sentiment does not have a statistically significant effect on dimension score improvements. In columns 3 through 5, we run conditional regressions for firms where *CommentsMapped*=1, meaning the feedback comments mapped to at least one dimension for improvement. Column 3 shows first stage results without controls for *CommentsMapped*, and the coefficient is again statistically significant. Interestingly, feedback sentiment contributes negatively to score and rank improvement in columns 4 and 5. This means that as feedback sentiment becomes more positive, firms are less likely to have higher scores and improvements in ranking in the semifinalist round.

Post-competition performance. In Table 6, we run ordinary least squares for the first stage regression on both instruments, *Z1* and *Z2*. In column 1 where the instrument is the minimum value of judge's leave-out mean sentiment scores for the focal firm, the coefficient is statistically significant with an F-statistic of 22.32. In column 2, the instrument is the mean value of judge's leave-out mean sentiment scores, and the coefficient is also statistically significant but with an F-statistic of 11.6. Although both instruments appear valid and pass the rule of thumb for weak instruments (Stock, Wright, and Yogo, 2002), we focus on *Z1* for the remainder of the analysis. In columns 3 through 5, we see that feedback sentiment during the competition does not appear to affect the probability of closure, acquisition, or the total funding amount post-competition. Instead, a higher score in Round 1 decreases the probably that a firm will shut down and increases the funding raised, which suggests that the initial quality of the firm, as proxied by Round 1 scores, may be a better predictor of post-competition performance than the type of feedback the firm received. However, there may be heterogeneous treatment effects of feedback sentiment based on the quality of the firm. In Table 7, we present regression results based on quartiles of the Round 1 scores using *Z1* as the instrument. In column 1, the coefficient on feedback sentiment is positive and statistically significant. Given that feedback sentiment ranges from -2 to 2, this means that 1st quartile and higher quality firms that receive more negative feedback are less likely to go out of business. We see the opposite result in column 4, where the 4th quartile and lower quality firms that receive more negative feedback are more likely to shut down. In addition, firms of medium quality in the 2nd and 3rd quartile seem unaffected by the feedback sentiment. One interpretation of these results is that high quality firms are better-positioned to incorporate feedback, which may help them improve and stay operational. Conversely, low quality firms may interpret negative feedback as a signal to close down, or be forced to close down before feedback is completely implemented.

6 Experimental Design/Survey

<This section to be completed>

In order to identify the mechanisms underlying the nonlinear effect of negative feedback, we complement the observational study with a survey distributed during one of the competitions.

7 Discussion and Conclusion

The goal of this paper is to investigate whether and how feedback plays a role in the early stages of venture formation. Specifically, if different sentiments of feedback can help entrepreneurs improve performance over time. To answer this question, we use comprehensive data from a state-sponsored business plan competition to analyze how firm performance changes in response to feedback from judges within the competition and after the competition. By leveraging the random assignment of judges throughout the competition, we establish causal linkages between feedback sentiment and various performance outcomes. In addition to tracking the frequency and diversity of comments, the full text of feedback was analyzed using sentiment analysis and other text analysis tools to extract the sentiment and specificity of all judge comments.

The analysis produces three main results. First, negative comments appear to be more helpful for improving performance, but less negative comments on average are positively associated with winner status. Second, firms improve subsequent performance in the competition after receiving more negative feedback in the initial round, but only if the comments are mapped to specific dimensions to improve. Third, in terms of long-term performance, firms of heterogeneous quality respond differently to feedback sentiment. In particular, higher quality firms are more likely to remain operational, while lower quality firms are more likely to shut down. These results indicate that both sentiment and specificity contribute to how well a firm performs in subsequent rounds of the competition and longer term post-competition. One limitation to this study is we take firm performance as reflective of changes in the business plan rather than analyzing the full text of the business plan. However, we believe the judges' subsequent comments combined with firm performance serve as reasonable proxies for how the business plan changes over time.

This paper contributes to a better understanding of how specific attributes of feedback can benefit entrepreneurs, especially in early stages of the firm. Future work includes investigating other long term effects of feedback outside the business plan competition setting, including firm growth and employment.

References

Aizer, Anna, and Doyle Jr., Joseph J. "Juvenile Incarceration, Human Capital and Future Crime: Evidence from randomly Assigned Judges." Quarterly Journal of Economics.

Di Tella, Rafael, and Ernesto Schargrodsky. "Criminal recidivism after prison and electronic monitoring." Journal of Political Economy 121.1 (2013): 28-73

Farre-Mensa, Joan and Hegde, Deepak and Ljungqvist, Alexander. "The Bright Side of Patents." (January 26, 2016). USPTO Economic Working Paper 2015-5.

Howell, Sabrina T. 2016. "Very Early Venture Finance: How Competitions Reduce Financial Frictions." Working paper.

Kerr, William R. and Nanda, Ramana and Rhodes-Kropf, Matthew. "Entrepreneurship as Experimentation." (July 28, 2014). Journal of Economic Perspectives (Forthcoming); Harvard Business School Entrepreneurial Management Working Paper No. 15-005.

Kling, Jeffrey R. 2006. "Incarceration Length, Employment, and Earnings." American Economic Review, 96(3): 863-876.

Lerner, Josh and Malmendier, Ulrike. 2013. "With a Little Help from My (Random) Friends: Success and Failure in Post-Business School Entrepreneurship." Review of Financial Studies, 26 (10), pp. 2411-52.

Manning, Christopher D., Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55-60.

Manso, Gustavo. 2010. "Motivating Innovation." Journal of Finance, vol. 66, pp. 1823-1869, 2011; AFA 2007 Chicago Meetings Paper. Available at SSRN: http://ssrn.com/abstract=891514 or http://dx.doi.org/10.2139/ssrn.891514

Peeters, G., & Czapinski, J. (1990). "*Positive-negative asymmetry in evaluations: The distinction between affective and informational negativity effects.*" In W. Stroebe & M. Hewstone (Eds.), European review of social psychology (Vol. 1, pp. 33-60). Chichester, England: Wiley.

Podsakoff, P.M., and Farh, J.L. (1989). "*Effects of feedback sign and credibility on goal setting and task performance.*" Organizational Behavior and Human Decision Processes, 44, 45-67.

Sampat, Bhaven N. and Williams, Heidi L. "How Do Patents Affect Follow-On Innovation? Evidence from the Human Genome" (October 2015). NBER Working Paper No. w21666.

Socher, Richard, Alex Perelygin, Jean Wu, Jason Chuang, Christopher Manning, Andrew Ng and Christopher Potts. 2013. “Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank.” Conference on Empirical Methods in Natural Language Processing

Stock, James, Jonathan Wright, and Motohiro Yogo, “A survey of weak instruments and weak identification in generalized method of moments,” Journal of Business and Economic Statistics, 2002, 20 (4), 518–529.

Wagner, Rodrigo Andres. 2016. “Does Feedback to Business-Plans Impact New Ventures? Evidence from a Field Experiment.” Available at SSRN: http://ssrn.com/abstract=2766566

Yu, Sandy. 2016. “How Do Accelerators Impact the Performance of High-Technology Ventures?” Working paper.

Figure 1. Sample dashboard of feedback for firm founders

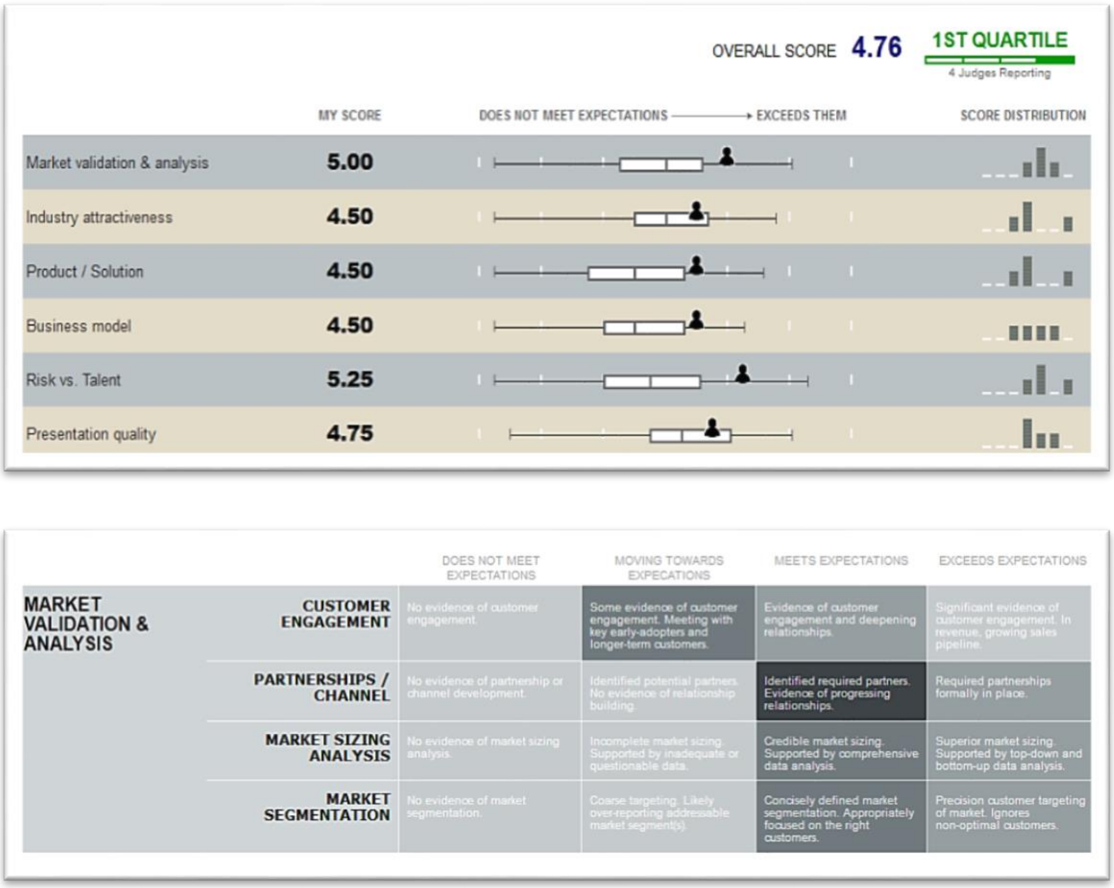


Figure 2. Average firm scores across competition rounds

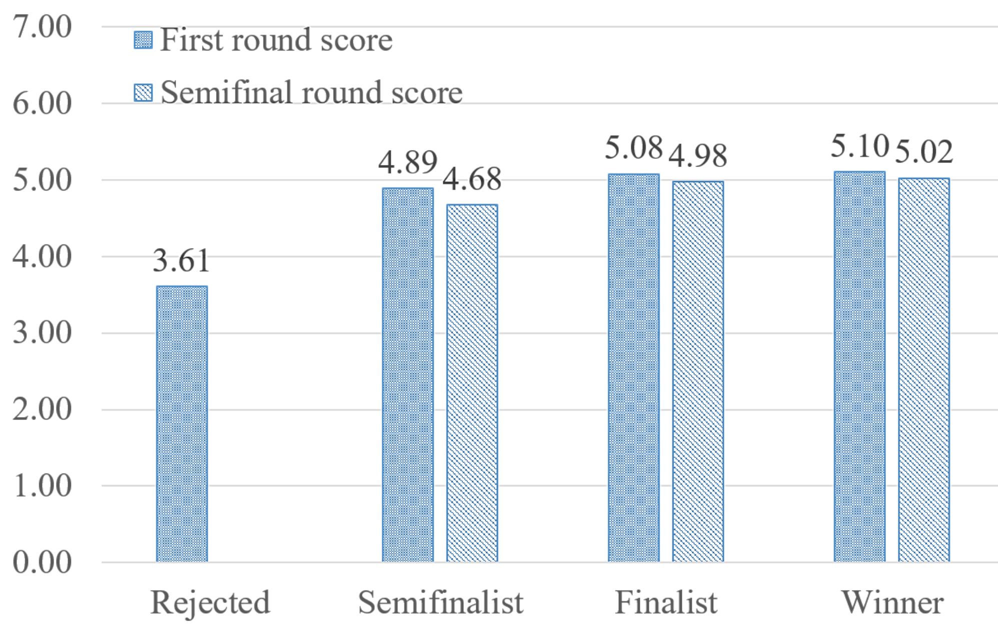


Figure 3. Sentiment of judge comments across competition rounds

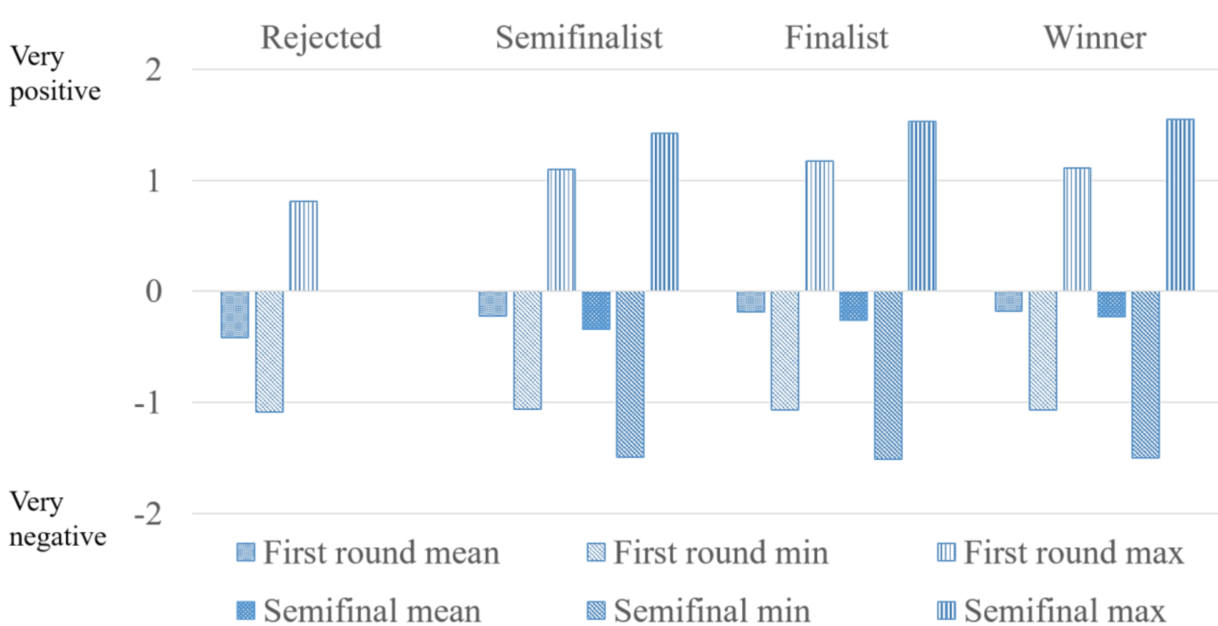
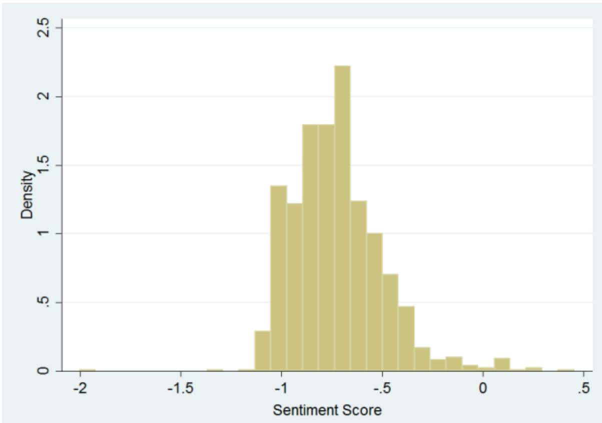
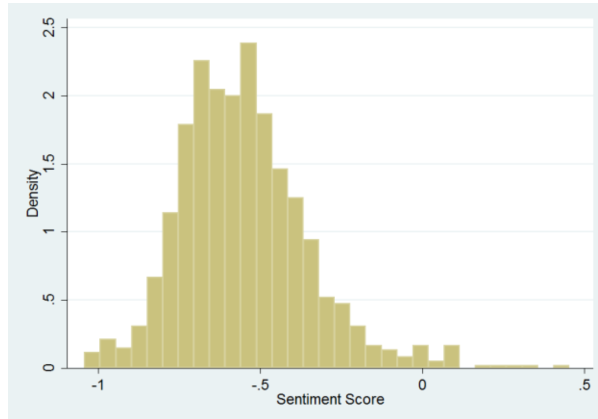


Figure 4. Distribution of Instruments: Polarity of Judge's Feedback



Z1: Min of judges' leave-out means



Z2: Average of judges' leave-out means

Table 1. Summary statistics at the competition event level

	N	Mean	Std. Dev.	Min	Max
Num applicants	8	182.4	61.0	132	314
Num semifinalists	8	30.9	3.6	25	35
Num winners	8	6	0	6	6
Num judges	8	47.5	10.4	30	64

Table 2. Summary statistics at the firm level

	N	Mean	Std. Dev.	Min	Max
First round scores (scale 1-7)	1,361	3.81	0.88	1	6.25
Semifinal scores (scale 1-7)	242	4.63	0.47	3.28	5.86
No. comments in first round	1,452	2.54	1.39	0	7
No. comments in semifinals	245	3.72	2.69	0	8
No. sentences in first round	1452	13.41	11.19	0	63
No. sentences in semifinals	245	23.95	19.65	0	73

Table 3. Regression Results: Sentiment of Feedback on Performance

Variables	(1) OLS Semifinal Score	(2) Logit Finalist	(3) Logit Winner
Sentiment Mean first round	0.142	1.614***	
	(0.109)	(0.445)	
Sentiment Max first round	-0.0251	0.475	
	(0.0593)	(0.306)	
Sentiment Min first round	-0.186**	-0.325	
	(0.0864)	(0.338)	
Sentiment Mean semifinal round			4.182***
			(1.163)
Sentiment Max semifinal round			0.254
			(0.481)
Sentiment Min semifinal round			-0.202
			(0.449)
Event fixed effects	Y	Y	Y
Industry fixed effects	Y	Y	Y
Constant	5.047***	-2.839***	-0.0807
	(0.150)	(0.602)	(1.236)
Observations	194	1,321	163
R-squared	0.296		

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 4. Regression Results: Dimension-Mapping on Performance

	(1)	(2)	(3)
	OLS	Logit	Logit
	Semifinal	Finalist	Winner
Variables	Score		
Comments Mapped first round	0.148**	0.345	
	(0.0695)	(0.328)	
No. Dimensions first round	-0.00892	0.326**	
	(0.0247)	(0.131)	
Comments Mapped semifinal round			1.234*
			(0.677)
Event fixed effects	Y	Y	Y
Industry fixed effects	Y	Y	Y
Constant	5.152***	-4.300***	1.261
	(0.152)	(0.803)	(1.420)
Observations	194	1,321	163
R-squared	0.296		

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 5. Regression Results: Feedback Sentiment on Within-Competition Performance Improvements

VARIABLES	(1) OLS First Stage	(2) Logit Dimension Score Improved	(3) OLS First Stage	(4) Logit Score Improved	(5) Logit Rank improved
Feedback Sentiment		-0.161		-2.172***	-1.215*
		(0.924)		(0.557)	(0.623)
Z1: Min (judge's leave-out mean sentiment)	0.483***		0.509***		
	(0.114)		(0.106)		
Comments Mapped	0.0899	0.130			
	(0.105)	(0.317)			
Num Dimensions Mapped	-0.0683	-0.0714	-0.0569	-0.0630	0.0206
	(0.0429)	(0.151)	(0.0374)	(0.120)	(0.104)
Competition FE	Y	Y	Y	Y	Y
Constant	0.373	-0.222	0.468**	-0.486	-0.331
	(0.230)	(0.656)	(0.235)	(0.736)	(0.671)
Observations	173	173	172	172	172

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6. Regression Results: Feedback Sentiment on Post-Competition Performance

VARIABLES	(1) OLS First Stage	(2) OLS First Stage	(3) Logit Closed	(4) Logit Acquired	(5) 2SLS Log(Funding)
Feedback Sentiment			0.0983	0.647	0.423
			(0.540)	(1.034)	(0.340)
Round 1 Score	0.165***	0.163***	-0.472***	0.309	0.277***
	(0.0127)	(0.0121)	(0.0843)	(0.230)	(0.0716)
Z1: Min (judge's leave-out mean sentiment)	0.297***				
	(0.0856)				
Z2: Mean (judge's leave-out mean sentiment)		0.389***			
		(0.0712)			
Year FE	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y
Constant	-0.957***	-0.950***	0.285	-3.683**	-0.490
	(0.111)	(0.0853)	(0.614)	(1.804)	(0.408)
Observations	1,275	1,275	1,275	1,275	421
F-statistic	22.32	11.6			
R-squared					0.185

Robust standard errors clustered by competition in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 7. Regression Results: Feedback Sentiment on Post-Competition Performance by Quartiles. Feedback sentiment is instrumented by the Min (judge's leave-out mean sentiment) for focal firm.

VARIABLES	(1) Logit Closed 1st quartile	(2) Logit Closed 2nd quartile	(3) Logit Closed 3rd quartile	(4) Logit Closed 4th quartile
Feedback Sentiment	1.182***	0.320	-0.0918	-2.715***
	(0.398)	(3.341)	(1.689)	(0.326)
Round 1 Score	-0.263	-1.431***	-0.998***	0.288**
	(0.180)	(0.511)	(0.345)	(0.143)
Year FE	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y
Constant	-0.300	3.928	1.815	-2.692***
	(1.210)	(2.494)	(1.813)	(0.420)
Observations	333	306	321	296

Robust standard errors clustered by competition in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Appendix

Table A1. Regression Results: Quantitative Feedback Attributes on Performance

Variables	(1) OLS Semifinal Score	(2) Logit Finalist	(3) Logit Winner	(4) OLS Semifinal Score	(5) Logit Finalist	(6) Logit Winner
Num. sentences in first round	-0.000536	0.0192*				
	(0.00296)	(0.0110)				
Num. sentences in semifinal round			0.0236***			
			(0.00855)			
Num. of judges who gave comments in first round				-0.0266	0.221**	
				(0.0310)	(0.107)	
Num. of judges who gave comments in semifinal round						0.154***
						(0.0564)
Event fixed effects	Y	Y	Y	Y	Y	Y
Industry fixed effects	Y	Y	Y	Y	Y	Y
Constant	4.380***	-2.797***	-2.174***	4.749***	-3.138***	-1.989***
	(0.128)	(0.455)	(0.634)	(0.223)	(0.539)	(0.576)
Observations	198	1,452	245	198	1,452	208
R-squared	0.283			0.286		

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1