Measuring Organizational Capital: A Cross-Language, Cross-Platform Framework*

Kaede Hanazawa† Ritsu Kitagawa‡ Akseli Palomäki§

January 31, 2026

Abstract

Organizational capital is viewed as a key driver of persistent performance differences across firms, but measurement has been largely confined to English-language settings. We construct firm-year measures of organizational capital for Japan using employee reviews from a leading Japanese employee review platform. Closely following the prior work based on Glassdoor employee reviews, we develop a language- and platform-free measurement method and apply it to more than 600,000 Japanese-written employee reviews. The resulting measures are slow-moving, vary meaningfully across firms and CEOs, and positively predict profitability and market valuation, with modest movements around CEO turnover, consistent with the theory. We also build alternative measures using Japanese-specific pre-processing and find substantial agreement with our main measures. These results show that review-based measurement of organizational capital extends to non-English platforms and that our pipeline supports cross-language and cross-platform comparability.

Keywords: Organizational capital; intangible capital; employee reviews

JEL Codes: D23; L25; G30

*We thank OpenWork, Inc. for providing data. We also thank the Center on Japanese Economy and Business (CJEB) at Columbia Business School for providing access to the NEEDS FinancialQUEST database. We thank Hideo Owan, Junichi Yamanoi, Xuanli Zhu, and participants at the Personnel Economics Workshop for helpful feedback. Kitagawa gratefully acknowledges financial support from the Suntory Foundation.

†University of Tokyo.

‡Columbia University.

§Columbia University.

1 Introduction

Large productivity differences persist even among firms that appear observationally similar. For example, Syverson (2011) documents that productivity at the 90th percentile can exceed that at the 10th percentile by roughly 200 percent in U.S. manufacturing. Explaining such dispersion is challenging because many drivers of performance are intangible and therefore only imperfectly observed. Scholars have devoted substantial effort to measuring such intangible assets and their underlying components (Corrado et al., 2022).

We focus on organizational capital, a key component of firms' intangible assets (Crouzet et al., 2022). This concept subsumes various drivers of firm performance. An incomplete list of components that have been studied includes: firm-specific human capital (Prescott and Visscher, 1980), management practices (Bloom and Van Reenen, 2007), relational contracts (Baker et al., 2002), organizational capabilities Teece et al. (1997), and corporate culture (Kreps et al., 1990). These factors governs how the organization converts inputs into output.

A natural place where otherwise latent organizational features are expressed is in employee narratives about how the firm operates. Employee review platforms provide large-scale, high-frequency, and firm-linked text that directly reflects workers' experiences inside organizations. Recent work has shown that such texts can be used to construct firm-year measures of organizational capital and related organizational attributes, and that these measures are empirically meaningful in English-language settings (Cai et al., 2025b; Li et al., 2021, 2025). However, the scope of existing evidence remains narrow: most measurement strategies have been developed and validated in English and on a small number of platforms with specific review formats. As a result, we know far less about organizational capital in the majority of the global economy, where both language and platform design differ substantially.

This paper develops and applies a language- and platform-robust approach to measuring organizational capital from employee reviews, and uses it to provide the first large-scale firm-year measures for Japan. Japan is a particularly informative setting: prior evidence doc-

uments substantial dispersion in management quality and organizational functioning across Japanese firms (Kambayashi et al., 2021), and OpenWork—Japan's largest employee review platform—contains rich, firm-linked narratives that have recently been used to study internal frictions and governance (Iwata, 2025). These features make Japan an attractive context both for extending measurement to a non-English environment and for studying economically relevant organizational heterogeneity within and across firms.

Our empirical setting combines 604,298 OpenWork employee reviews posted between 2015–2025 with financial outcomes for 9,591 firms (listed firms and subsidiaries). We treat the key challenge not as translation per se, but as *harmonization*: mapping heterogeneous, platform-specific review content into a standardized representation suitable for consistent measurement. We implement this harmonization with a large language model (LLM). For each review, the model (i) translates the Japanese text into English without summarizing and (ii) segments the translated content into *pros*, *cons*, and *neutral* statements about the company. This produces a Glassdoor-style representation that supports standard downstream text processing. On the harmonized corpus, we construct organizational-capital measures using established similarity-based approaches (a seed-word measure and a synthetic-review measure) and aggregate review-level signals to the firm-year level (Cai et al., 2025b).

We complement embedding-based measurement with an *LLM-as-rater* approach that directly operationalizes the theoretical definition of organizational capital. Instead of relying on seed vocabularies or embedding-space similarities, we prompt an LLM to read each review and assign an organizational-capital score on a 0–100 scale, using the definition of organizational capital in Dessein and Prat (2022). We implement three rater variants (English-translated text with an English prompt; original Japanese text with an English prompt; and Japanese text with a Japanese prompt) and validate the approach using synthetic reviews designed to represent high versus low organizational capital. The rater method serves two purposes: it provides a conceptually direct measurement benchmark that is not tied to specific dictionaries or embedding choices, and it allows us to assess how sensitive

the extracted signal is to language (Japanese vs. English) and to prompt design. We also construct a parallel Japanese-language version of the embedding-based pipeline (Japanese-specific preprocessing and embeddings) to evaluate robustness without relying on English rendering.

We organize the paper around three questions. First, can organizational capital be measured from employee reviews in a non-English setting with a fundamentally different platform structure? Second, do the resulting measures satisfy the empirical properties implied by theory—that organizational capital is performance-relevant, slow-moving, intangible (informative beyond standard proxies), and produced at least partly through top management (Dessein and Prat, 2022)? Third, once such measures are available at scale, what do they reveal about economically meaningful heterogeneity and frictions within firms?

Our results support five main conclusions. First, organizational capital measured from OpenWork reviews predicts firm performance: higher organizational-capital scores are associated with higher profitability (ROA) and higher market valuation (price-to-book) in firm-year regressions controlling for size, age, and industry-by-year fixed effects, though the magnitudes are weaker than those reported by Cai et al. (2025b). Second, the measures are durable and firm-specific, exhibiting substantial persistence and meaningful firm components, consistent with organizational capital as a slowly evolving intangible (Dessein and Prat, 2022). Third, the measures contain information beyond standard accounting-based proxies, consistent with imperfect observability of organizational capital in financial statements. Fourth, top management matters: CEO fixed effects account for a sizable share of variation in organizational-capital measures, and an event-study around CEO turnover shows modest movements in organizational capital after leadership changes.

Moreover, the core signal is robust across languages and methods: measures produced from English-harmonized text, from Japanese-native preprocessing, and from the LLM-as-rater approach are strongly positively correlated at the firm-year level, while also containing nontrivial method-specific components that help diagnose where language choices or mod-

eling assumptions matter. Given this, our translation and normalization framework offers a general protocol for building comparable organizational-capital measures from heterogeneous employee-review platforms in other countries.

We also show that these measures can be used to study within-firm organizational frictions that are difficult to observe otherwise. Using reviewer gender, we construct gender-specific organizational-capital measures and relate firm performance to both the level and the within-firm gender gap. Firms with larger disagreement between female and male assessments have lower ROA and lower price-to-book ratios, consistent with the interpretation that disagreement reflects organizational frictions—such as unequal workplace experiences or misaligned internal practices—that are economically consequential.

We contribute to several strands of literature. First, we add to the work on organizational capital (Prescott and Visscher, 1980; Atkeson and Kehoe, 2005; Dessein and Prat, 2022).[1] Second, we extend text-based measurement of intangibles using employee reviews (Li et al., 2021; Cai et al., 2025b; Li et al., 2025; Cai et al., 2025c,a) to a non-English setting and introduce an LLM-based approach to harmonizing platform-specific formats. Third, we contribute to the broader agenda of using machine learning and natural language processing methods to measure latent economic concepts from unstructured data, including CEO behavior, amenities, worker rights, and related firm attributes (Bandiera et al., 2020; Cai et al., 2022; Sockin, 2022; Bana, 2022; Dube et al., 2022; He et al., 2024; Clayton et al., 2025).

A closely related line of work studies corporate culture, a component of organizational capital that captures shared norms and values inside the firm (Gorton et al., 2022). Theory tells that culture can shape incentives, coordination, and performance (Kreps et al., 1990; Hermalin, 2012). Empirically, the central challenge has been measurement, and researchers have pursued multiple approaches. Early work proxies for culture using employee surveys and firm communications such as corporate websites (Guiso et al., 2015). Related evidence shows that senior managers view corporate culture as important and often perceive substantial

[1]More generally, we also contribute to the literature on intangible capital, which conceptually subsumes organizational capital (Corrado et al., 2022; Crouzet et al., 2022).

scope for improvement (Graham et al., 2022). More recent work uses machine learning and large language models to measure culture from large-scale text data (Li et al., 2021, 2025). Finally, insider econometric evidence indicates that workplace climate and culture fit can affect worker productivity and related outcomes (Castro et al., 2022; Alan et al., 2023; Cai, 2023). This body of literature motivates measurement strategies that recover otherwise hard-to-observe organizational attributes at scale, and our OpenWork-based approach provides such a tool for corporate Japan.

We also contribute to the literature on corporate Japan by providing new evidence on the sources and economic relevance of organizational heterogeneity in Japanese firms. Kambayashi et al. (2021) administer the management practices survey of Bloom and Van Reenen (2007) in Japan and document substantial dispersion in management quality across Japanese firms. Iwata (2025) uses OpenWork reviews to measure internal organizational frictions and shows that they predict both governance improvements and firm valuation. Our findings suggest that the structure of organizational capital in Japan resembles that in the United States, both qualitatively and quantitatively. We also find that greater within-firm disagreement between female and male employees is associated with worse firm performance. These results contribute to our understanding of the sources of productivity dispersion in Japan.

The remainder of the paper proceeds as follows. Section 2 describes the OpenWork review data and the matched financial outcomes. Section 3 details the LLM-based harmonization step and the construction of our baseline organizational-capital measures. Section 4 evaluates the theoretical validation properties. Section 5 presents applications to CEO turnover and gender disagreement. Section 6 develops alternative measurements, including Japanese-native pipelines and the LLM-as-rater approach. Section 7 concludes.

2 Data

2.1 OpenWork Employee Reviews

Our primary data source is OpenWork (operated by OpenWork, Inc.), a large Japanese platform that hosts anonymous employer reviews posted by current and former employees.[2] Each review is linked to a firm identifier and a posting timestamp and contains a combination of structured ratings and unstructured narrative text. The data provider defines the extract over Japanese firms with at least 50 employees (or corporate groups with at least one affiliate meeting this threshold) and supplies both listed firms and their subsidiaries.

OpenWork is one of the most popular employee review platforms in Japan. A common concern with employee reviews is sample selection. Because users often visit these platforms while searching for jobs, reviewers may be systematically less satisfied with their current employers, which could bias the level of reported sentiment. There are two points that make us less concerned about this issue. First, if the platform attracts relatively dissatisfied workers, this should mainly shift reviews downward on average. As long as there remains meaningful variation across firms, the data can still allow comparisons across firms. Second, OpenWork limits access to others' reviews unless users either post a review or pay a monthly subscription. This "give-to-get" feature encourages broader participation and reduces the risk that the review corpus is dominated by a small group of especially disgruntled employees.

The review-level file contains 604,361 reviews covering 9,591 unique firms over March 2015–October 2025. Aggregating by firm and calendar year of posting yields 63,588 firm-year observations. Table B.6 summarizes sample coverage and reviewer demographics. Gender and employment status at the time of posting are required fields in the review instrument; 70.8% of reviews are posted by men and 54.8% are posted by current employees.

Each review includes an overall employer rating and eight sub-dimension ratings. The dimensions cover compensation satisfaction, employee morale, openness of communication,

[2]OpenWork was previously named Vorkers until May 2019, and the operating company was also renamed from Vorkers, Inc.

teamwork and mutual respect, growth opportunities for younger employees, long-term employee development, compliance orientation, and perceived fairness of evaluation and promotion systems. Each sub-dimension is rated on a 1–5 scale, where 1, 2, 3, 4, and 5 correspond to "Very Bad," "Bad," "Average," "Good," and "Very Good," respectively. The overall rating is the average of the eight sub-ratings. Summary statistics for these ratings are reported in Table B.7.

In addition to ratings, the review form collects free-text responses to multiple open-ended prompts. In our extract, up to nine narrative fields are available, including organizational culture, work-life balance, and pay and evaluation systems. For each review, users are required to write at least 500 characters in total across the nine fields, while there is no requirement to fill any particular field or a minimum number of fields. Due to this relatively flexible response requirement, completion rates and average lengths vary across text fields. Table B.8 reports completion rates and average length for each free-text field.

The review instrument also elicits self-reported workplace conditions and compensation information, including monthly overtime hours (including holiday work) and the approximate usage rate of paid holidays, as well as (optionally) annual salary and its components and an NPS-style recommendation score.

OpenWork supplies a firm master file that includes a stable firm identifier, the firm name, and a four-digit Japanese securities code for listed firms. The firm identifier serves as the primary key linking the firm master file to the review-level file. The data also include a parent–subsidiary relation file that identifies corporate-group links and ownership shares. The mapping is based on an external corporate relationship database (SPEEDA) and provides, for each parent firm, the associated subsidiary identifiers and the ownership rate. We use these identifiers to define the analysis panel at the firm level and, where relevant, to characterize group structure.

For analyses that require accounting and market outcomes, we merge OpenWork firms to external firm-level data using the securities code when available. We conduct empirical

analysis at the firm-year level, where the year index is based on the posting date of the underlying reviews. Any additional sample restrictions induced by the availability of external outcomes are reported alongside the corresponding specifications.

2.2 Other Descriptives of OpenWork Reviews

Since the dataset we use from OpenWork, Inc. is relatively new to the academic community, we document several basic descriptives to introduce it and facilitate broader use in future research. The empirical patterns documented here pertain to our sample of listed firms and their subsidiaries. OpenWork also contains reviews of non-listed firms, many of which are small and medium enterprises with relatively few posted reviews.

2.2.1 Trends Over Time

Figure B.1 reports cumulative firm coverage in OpenWork. For each firm that ever appears in our extract, we compute indicators for whether the firm has reached at least $n \in 1, 5, 10$ reviews per year at least once by year t, and then plot the cumulative share across firms. More than 75% of firms have received at least one review by 2018. In contrast, accumulating a larger stock of reviews is more gradual. By the end of the sample period, roughly 50% of firms have reached at least five reviews and about 25% have reached at least ten reviews.

Figure B.2 reports annual (contemporaneous) coverage, defined as the share of firms that receive at least n reviews in a given year. Coverage increases sharply between 2015 and 2018 and then remains relatively stable thereafter, consistent with rapid adoption of OpenWork followed by a mature phase of platform usage. Even in later years, however, review intensity is uneven across firms: while a majority of firms receive at least one review in a typical year, substantially fewer firms receive five or ten reviews.

The panels in Figure B.3 summarize basic temporal patterns in review activity on OpenWork. Review volume rises sharply in the early part of our sample, increases through the late 2010s, and remains sizable thereafter, with noticeable but moderate year-to-year fluctua-

tions. At a monthly frequency, review posting is also highly persistent, with no evidence that a small number of isolated months drives the data. These patterns suggest that the platform provides a large and continuous stream of employee feedback over our sample window.

The figure also shows clear within-week and within-day regularities. Review activity varies systematically across days of the week and hours of the day. In particular, posting is concentrated in specific parts of the week and day rather than being uniformly distributed, consistent with reviews being written during non-work time or around routine work schedules. Importantly, despite these predictable cycles, activity is present throughout the week, indicating broad participation rather than dependence on a narrow posting window.

Next, we examine annual seasonality. Average review volume varies across calendar months, indicating systematic month-of-year patterns in posting. The variation is not driven by a single month. Instead, activity is relatively stable and moderately decreasing towards the end of the year, with a visible drop in April, which aligns with the timing of Japan's fiscal-year calendar. This seasonality suggests that review posting responds, at least in part, to recurring institutional and organizational cycles, such as annual evaluation and promotion schedules, hiring and mobility seasons, and other regular workplace milestones, rather than reflecting purely idiosyncratic timing.

Figure B.4 reports time trends of the mean overall rating. Panel (a) shows that the mean remains close to three throughout the sample period, with a slight gradual increase. Given the large sample size, this increase is statistically detectable, but the magnitude is economically small. Panels (b)–(f) show that average ratings are stable across year-months, calendar months, days of the week, hours of the day. These results suggest that the composition of reviewed firms and reviewers is broadly stable over our sample period when evaluated at the first moment of the overall rating distribution.

2.2.2 Review Patterns by Reviewer Characteristics

Now we turn to documenting heterogeneity in review content by reviewer characteristics. Although we observe only a limited set of reviewer attributes, examining how ratings and text vary across these groups is informative about the data generating process. The available reviewer characteristics include reviewer gender, employment status (current versus former employee), years of tenure at the firm, and job category.[3] We estimate the following regression equation:

$$Y_r = \beta_1 \left(\text{Female}_r\right) + \beta_2 \left(\text{Tenure}_r\right) + \beta_3 \left(\text{Former Employee}_r\right) + \tau_{t(r)} + \alpha_{k(r)} + \delta_{j(r)} + \eta_r, \quad (1)$$

where Y_r is the rating outcome for review r, $\tau_{t(r)}$ denotes year-month fixed effects (based on the posting date of the review), $\alpha_{k(r)}$ denotes firm fixed effects, and $\delta_{j(r)}$ denotes job-category fixed effects. To aid interpretation, we recode each sub-rating as an indicator equal to 1 if the score is 4 or 5 and 0 otherwise. Under this coding, coefficients can be interpreted as changes in the probability that a review rates the corresponding dimension positively.

Figure B.5 summarizes how ratings vary with observed reviewer characteristics. Each panel reports coefficients from regressions of a rating outcome on indicators for tenure category, gender, and former-employee status, with three specifications that sequentially add firm fixed effects and job-category fixed effects. For the overall rating, longer-tenure reviewers report substantially higher ratings, and the gradient is large even after conditioning on firm and job-category fixed effects. Former employees rate their employers markedly more negatively than current employees, and this gap remains sizable within firm. Female reviewers report slightly higher overall ratings than male reviewers, but the difference attenuates once firm fixed effects are included, consistent with part of the raw gap reflecting differences

[3]These reviewer characteristics are required by OpenWork. Gender is reported as a binary category (male or female) at registration, and employment status, tenure, and job category are required at the time of review submission. This requirement is relatively distinctive to OpenWork. In contrast, Glassdoor typically does not require these attributes at signing up or posting beyond whether the reviewer is a current or former employee.

in the firms reviewed.

The binary outcomes, which indicate whether a given sub-rating is at least four, show a similar structure. Longer tenure is associated with higher probabilities of high ratings for pay and benefits, growth opportunities, compliance, and the evaluation system, while it is associated with lower motivation and long-term development ratings at intermediate tenure levels, before becoming positive at the longest tenure group. Former employees are consistently less likely to assign high ratings across most dimensions, especially compliance and pay. Adding firm fixed effects typically reduces coefficient magnitudes but does not overturn these patterns, indicating that reviewer characteristics predict systematic within-firm differences in reported workplace assessments rather than reflecting only sorting across firms.

Next, for each text field, we construct an indicator variable equal to one if the field contains any non-empty text and zero otherwise. We use these indicators as outcomes and estimate their association with reviewer characteristics using specification (1). Table B.9 shows systematic differences in which text fields reviewers choose to fill, even within the same firm and year-month. Female reviewers are more likely to write in the "environment for women," "work-life balance", and "reason for quiting" field, but are less likely to provide text in most other fields, including "organization structure/culture," "strengths and weaknesses," "suggestions to CEO", and "salary/evaluation." Former employees are substantially more likely to complete the "reason for quitting" field and are modestly less likely to write about "work-life balance" and "salary/evaluation." Tenure also shifts topics. Longer-tenure reviewers are more likely to provide text on "organization structure/culture,", "work fulfillment," and "suggestions to CEO," and less likely to discuss "reason for joining."'

Table B.10 shows systematic differences in how much reviewers write, even comparing reviews posted within the same firm and year-month. The estimates come from specification (1), with outcomes given by the number of non-empty text fields and the total number of characters summed across fields. Female reviewers tend to cover fewer topics and write

shorter total text than male reviewers. Former employees cover a slightly broader set of topics, but their total writing is a bit shorter, consistent with adding brief comments in additional sections. Writing increases strongly with tenure: long-tenure reviewers are much more likely to comment on many different topics and they write substantially longer narratives overall.

We note that these results point to systematic gender differences in review content that may be substantively important. To assess whether they also appear at the firm level, Figure B.6 plots, for each firm-year, the average overall rating among female reviewers against the corresponding average among male reviewers. Panel (a) presents the raw firm-year scatter with the fitted regression line. The relationship is positive but modest, indicating that firms rated more favorably by men also tend to be rated more favorably by women, though far from one-for-one. Panel (b) repeats the exercise after residualizing both series with firm and year fixed effects. The association weakens further, suggesting that much of the comovement in panel (a) reflects persistent firm-level differences and common time effects rather than systematic within-firm, within-year alignment in male versus female rating gaps.

Table B.11 shows that male and female overall ratings covary only weakly at the firm-year level. In the first column, the cross-sectional association is positive but modest. A one-point increase in the male firm-year average is associated with a 0.319-point increase in the female average, and the fit is low ($R^2 = 0.086$). Once we absorb permanent firm heterogeneity with firm fixed effects in the second column, the slope falls to 0.083 and the within-R^2 is essentially zero (0.004). The takeaway is that the limited agreement between male and female ratings comes mainly from persistent differences across firms rather than within-firm fluctuations over time, consistent with gender-specific experiences or reporting that vary even among employees reviewing the same firm in the same period.

These patterns motivate our by-gender measurement exercise in Section 5.2. In particular, we examine whether firms with larger gender disagreement in review-based assessments also differ systematically in performance and other outcomes.

2.3 NEEDS FinancialQUEST

We obtain our primary firm outcomes and characteristics from the NEEDS FinancialQUEST 2.0 database. This database, compiled by Nikkei, Inc., is widely used in research on Japanese firms and provides standardized financial statement data, market valuation metrics, and firm attributes at annual and monthly frequencies. It covers nearly all listed firms in Japan, ensuring broad coverage of our sample.

One caveat of the NEEDS FinancialQUEST database is that it only reports the current CEO and the year in which he or she was inaugurated. As a result, we can identify only the most recent CEO turnover, while any turnover events prior to the current CEO are unobserved. This limitation introduces measurement error in CEO changes, which may bias our variance decomposition in Section 4 toward overstating the firm component and understating the CEO component. Since CEO turnover is relatively rare among listed firms, this concern may be limited, but interpretation of the relative sizes of the firm and CEO components nonetheless requires some caution.

3 Measurement Strategy

In this section, we describe how we transform OpenWork reviews into firm-year measures of organizational capital while closely following the procedure in Cai et al. (2025b). The pipeline has three main steps. First, we translate the Japanese review text into English and separate it into *pros*, *cons*, and *neutral* segments using an LLM (gpt-5-nano). Second, we construct a corpus from the converted review text, apply standard text pre-processing, identify bigrams, and train the word embedding model. Finally, we build two organizational capital measures (a seed-word measure and a synthetic-review measure) and aggregate them to the firm-year level. The main steps in this process are shown in Figure 1.

3.1 Translation and Segmentation

Preparing "Feed" for the LLM. OpenWork reviews are organized into multiple open-ended narrative fields rather than a single pros/cons template. For each review, we therefore construct one review-level raw-text input by combining the narrative text across the free-text fields listed in Table B.8. We do not remove the field identifiers (headers) that appear in original Japanese review text, such as a header like "work-life balance" that precedes the corresponding narrative. We retain these headers because they provide context and can improve the quality of the downstream word-embedding representation. Apart from concatenating fields at the review level, we apply no additional text cleaning before the LLM step.

Translation and pros/cons/neutral segmentation. To render OpenWork reviews comparable to the inputs required by Cai et al. (2025b), we apply an LLM (gpt-5-nano) to the review-level Japanese text and produce an English, Glassdoor-style representation. The LLM performs two tasks jointly. It translates the content without summarization and assigns each translated sentence to one of three mutually exclusive categories about the firm, *pros* (positive statements), *cons* (negative statements), or *neutral* (descriptive statements that are neither positive nor negative). OpenWork reviews respond to multiple open-ended prompts, so a substantial share of the text consists of factual descriptions of the workplace such as business content, internal policies, and day-to-day practices rather than explicit praise or complaints. We therefore introduce a third category, *neutral*, to capture company-related statements that are primarily descriptive rather than evaluative. This design keeps the *pros* and *cons* segments closer to the clearly positive and clearly negative content that underlies Glassdoor-style reviews. Appendix C.1 reports the instruction prompt used for translation and segmentation and provides an example output.

In our downstream construction, organizational capital is measured from the pros–cons difference, as in Cai et al. (2025b). Accordingly, we use neutral text to enrich the corpus for

training word embeddings by adding firm-related terminology and context, but we do not use it directly in the pros–cons scoring formulas.

3.2 Text Pre-Processing

Using the converted English texts, we construct a corpus for training word embeddings. Our text pre-processing consists of two steps. First, we convert each review to lowercase and remove basic punctuation and non-informative symbols. We then apply tokenization, which breaks the text into word-level tokens. Second, to capture organizational concepts that are often expressed as multiword phrases (e.g., human capital), we learn a bigram phrase model on the full corpus using Gensim's Phrases algorithm as in Cai et al. (2025b). We set min_count = 10 and threshold = 10, which limits bigram formation to word pairs that are sufficiently frequent and exhibit strong association in the corpus. When a bigram is detected, we treat the two words as a single token and link them with an underscore.

3.3 Word Embedding

We map each token in our pre-processed corpus into a numeric vector using a Word2Vec model. The key idea is that a token's meaning is reflected in the words that tend to appear around it; accordingly, Word2Vec learns embeddings by predicting nearby tokens within local context windows. We train the model on the pre-processed token sequences constructed in Section 3.2, using an embedding dimension of 300, a context window of 5 words on each side, and 20 training epochs. The resulting embeddings provide a common semantic space that we use to compute similarity-based measures of organizational capital.

3.4 Seed-Word-Based Measure

We construct a seed-word-based measure of organizational capital following Cai et al. (2025b). We use their seed words—human capital, culture, relationship, intangible, and management

—which are intended to capture the core elements of organizational capital.

Let v_s be the embedding vector of seed word s which is obtained in Section 3.3. We define the representative vector for organizational capital as the average of the seed-word vectors:

$$v_R = \frac{1}{5}\sum_{s=1}^{5} v_s$$

Next, let v be the embedding vector of any token in the corpus. For each token, we compute its cosine similarity to v_R, i.e., $\frac{v \cdot v_R}{\|v\|\|v_R\|}$. We then select the top 500 tokens ranked by this cosine similarity and call the resulting set the expanded dictionary, denoted by D. Table C.13 reports the top 50 tokens in D. We index employee reviews by r and let B_r^{pros} and B_r^{cons} denote the sets of tokens appearing in the pros and cons sections of review r, respectively. For each token b in the corpus, we define review-level scores as

$$\text{Score}_r^{\text{pros}} = \sum_{b \in B_r^{\text{pros}}} 1\{b \in D\} \frac{v_b \cdot v_R}{\|v_b\|\|v_R\|}$$

$$\text{Score}_r^{\text{cons}} = \sum_{b \in B_r^{\text{cons}}} 1\{b \in D\} \frac{v_b \cdot v_R}{\|v_b\|\|v_R\|}$$

Finally, let $R_{i,t}$ be the set of employee reviews written about firm i in year t. The firm-year seed-word-based measure is defined by

$$\text{OC}_{i,t}^{\text{Seed Word}} = \frac{1}{|R_{i,t}|} \sum_{r \in R_{i,t}} \left(\text{Score}_r^{\text{pros}} - \text{Score}_r^{\text{cons}}\right). \tag{2}$$

3.5 ChatGPT-Based Measure

We construct a ChatGPT-based measure of organizational capital following the procedure in Cai et al. (2025b). The key idea is to use ChatGPT to generate benchmark texts that represent firms with strong versus weak organizational capital, and then measure how close actual employee reviews are to these benchmarks in the embedding space. Importantly, we use exactly the same organizational capital definition and the same prompt as in Cai

et al. (2025b), including the instruction to generate 100 positive reviews (strong OC) and 100 negative reviews (weak OC). The prompt and examples of the generated outputs are reported in Appendix C.3.

Our construction consists of three steps. First, we use ChatGPT to generate two sets of synthetic reviews and convert them into benchmark vectors in the embedding space. Let R_s and R_w denote the sets of generated synthetic positive and negative reviews describing firms with strong and weak organizational capital, respectively. For any review r, let B_r be the set of tokens appearing in r, and let v_b be the embedding vector of token b. We represent each synthetic review by the average of its token embeddings:

$$v_r = \frac{1}{|B_r|} \sum_{b \in B_r} v_b,$$

and define the strong- and weak-OC benchmark vectors as the averages of these review vectors:

$$v_s = \frac{1}{|R_s|} \sum_{r \in R_s} v_r,$$
$$v_w = \frac{1}{|R_w|} \sum_{r \in R_w} v_r.$$

Second, we score each real employee review by comparing its pros and cons text to the benchmark vectors. Let us define pros- and cons-specific vectors for review r as

$$v_r^{\text{pros}} = \frac{1}{|B_r^{\text{pros}}|} \sum_{b \in B_r^{\text{pros}}} v_b,$$
$$v_r^{\text{cons}} = \frac{1}{|B_r^{\text{cons}}|} \sum_{b \in B_r^{\text{cons}}} v_b.$$

Finally, we aggregate these review-level comparisons to the firm-year level. Let $R_{i,t}$ be the set of employee reviews written about firm i in year t. We define the firm-year ChatGPT-based

organizational capital measure as

$$\mathrm{OC}_{i,t}^{\mathrm{ChatGPT}} = \frac{1}{|R_{i,t}|} \sum_{r \in R_{i,t}} \left(\frac{v_r^{\mathrm{pros}} \cdot v_s}{\|v_r^{\mathrm{pros}}\| \|v_s\|} - \frac{v_r^{\mathrm{cons}} \cdot v_w}{\|v_r^{\mathrm{cons}}\| \|v_w\|} \right) \tag{3}$$

which is high when employees' pros statements resemble the strong organizational capital benchmark and their cons statements resemble it less, or equivalently, are closer to the weak organizational capital benchmark, where resemblance is measured by cosine similarity in the embedding space.

The LLM segmentation described in Section 3.1 may yield an empty pros or cons segment for some reviews. For such reviews, the cosine similarity to the representative vectors v_r^{pros} or v_r^{cons} is undefined because the review vector has zero norm (i.e., $\|v_s\| = 0$ or $\|v_w\| = 0$). In such cases, we impute a value of zero for the cosine similarity instead of discarding the entire review, even when either the pros or cons section is blank. This mirrors the seed-word measure, in which an empty token set mechanically implies a score of zero. Note that if both sections are blank, the resulting measure is pushed toward zero because these reviews add zero to the numerator while increasing the denominator $|R_{i,t}|$. There are 34,979 cases with a missing pro or con section, accounting for 5.79% of all reviews. Only 405 reviews are missing both sections.

3.6 Variants of Measures

We have constructed two text-based measures of organizational capital that capture complementary signals in employee reviews. The seed-word measure is anchored in explicit OC-related vocabulary and its close semantic neighbors in the embedding space. In contrast, the LLM-based (synthetic-review) measure captures broader phrasing patterns that describe strong or weak organizational practices even when canonical organizational capital terms are not used.

To obtain a single, robust proxy for latent organizational capital, we combine the two

measures after standardization, following Cai et al. (2025b). Combining is useful because each measure contains idiosyncratic noise and may perform better in different firm-years; averaging reduces reliance on any one construction. In addition, standardization is required because the raw measures are expressed on different numeric scales: the seed-word score is additive and can vary with the incidence of relevant tokens, whereas the ChatGPT-based score is based on bounded cosine similarities.

Under the empty-segment convention described in Section 3.5, both raw measures are well-defined for every firm-year observation in our review panel. Let $\mathcal{F} = \{(i,t) : |R_{i,t}| > 0\}$ denote the set of firm-year observations with at least one employee review, and let $N = |\mathcal{F}|$. We standardize each raw measure using the mean and standard deviation computed over $\mathcal{F}$:

$$\mu^{\text{Seed Word}} = \frac{1}{N}\sum_{(i,t)\in\mathcal{F}} \text{OC}_{i,t}^{\text{Seed Word}}, \quad \sigma^{\text{Seed Word}} = \sqrt{\frac{1}{N}\sum_{(i,t)\in\mathcal{F}} \left(\text{OC}_{i,t}^{\text{Seed Word}} - \mu^{\text{Seed Word}}\right)^2},$$
$$\mu^{\text{ChatGPT}} = \frac{1}{N}\sum_{(i,t)\in\mathcal{F}} \text{OC}_{i,t}^{\text{ChatGPT}}, \quad \sigma^{\text{ChatGPT}} = \sqrt{\frac{1}{N}\sum_{(i,t)\in\mathcal{F}} \left(\text{OC}_{i,t}^{\text{ChatGPT}} - \mu^{\text{ChatGPT}}\right)^2}.$$

We then construct standardized measures:

$$\widehat{\text{OC}}_{i,t}^{\text{Seed Word}} = \frac{\text{OC}_{i,t}^{\text{Seed Word}} - \mu^{\text{Seed Word}}}{\sigma^{\text{Seed Word}}}, \qquad \widehat{\text{OC}}_{i,t}^{\text{ChatGPT}} = \frac{\text{OC}_{i,t}^{\text{ChatGPT}} - \mu^{\text{ChatGPT}}}{\sigma^{\text{ChatGPT}}}.$$

The combined organizational capital index is the simple average of the standardized measures:

$$\text{OC}_{i,t}^{\text{Average}} = \frac{1}{2}\left(\widehat{\text{OC}}_{i,t}^{\text{Seed Word}} + \widehat{\text{OC}}_{i,t}^{\text{ChatGPT}}\right).$$

In applications where organizational capital is conveyed primarily through explicit OC-related terms, $\widehat{\text{OC}}_{i,t}^{\text{Seed Word}}$ may be especially informative; when organizational capital is expressed more indirectly through broader phrasing, $\widehat{\text{OC}}_{i,t}^{\text{ChatGPT}}$ may be preferred. We therefore use $\text{OC}_{i,t}^{\text{Average}}$ as our baseline and report robustness results using each component measures separately.

We also compute percentile-rank versions for interpretation in relative terms. For each

of $\widehat{\text{OC}}_{i,t}^{\text{Seed Word}}$, $\widehat{\text{OC}}_{i,t}^{\text{ChatGPT}}$, and $\text{OC}_{i,t}^{\text{Average}}$, We compute its rank in the firm-year dataset and rescale it to a 1–100 percentile index.

4 Validation

We validate our measures by examining whether they exhibit empirical patterns that are consistent with the theoretical properties of organizational capital in Dessein and Prat (2022). Dessein and Prat (2022) characterize organizational capital as any intangible asset that satisfies the following four properties: (1) it affects firm performance; (2) it changes slowly over time; (3) being intangible, it is not perfectly observable; and (4) it must be produced at least partly inside the firm with the active participation of the firm's top management. We test each of these properties. Our tests mostly follow the same validation procedure by Cai et al. (2025b). We find that our measures of organizational capital exhibit these properties and estimates are comparable to the ones for the U.S. sample studied by Cai et al. (2025b).

4.1 Firm Performance

The first property posits that organizational capital affects firm performance. Although rigorous causal tests are difficult, we should expect a positive association between our measures and firm performance if they capture organizational capital. To examine whether our organizational capital measures predict firm performance, we estimate the following regression equation:

$$y_{i,t} = \beta_1 \, \text{OC}_{i,t} + \beta_2 \, \ln\left((\text{Total Asset})_{i,t}\right) + \beta_3 \, (\text{Firm Age})_{i,t} + \tau_{I(i),t} + \varepsilon_{i,t}, \tag{4}$$

where $y_{i,t}$ is firm i's return on assets (ROA) or price-to-book (P/B) ratio in year t, $\text{OC}_{i,t}$ is one of our organizational capital measures, and $\tau_{I(i),t}$ denotes industry-by-year fixed effects (with $I(i)$ indexing firm i's industry). We define industries using the two-digit Japan Standard

Industrial Classification (JSIC) code. We cluster standard errors at the firm level.

Table 1 shows the estimation results for ROA. Table 1 reports the estimation results for ROA. Our organizational capital measures are positively and statistically significantly associated with ROA: the coefficient is 0.576 for OC (Seed), 0.586 for OC (ChatGPT), and 0.811 for OC (Avg.), and remains positive when using percentile-scaled measures (0.017–0.028). We note that the magnitudes of these associations are comparable to those reported in Cai et al. (2025b).

Table 2 reports the estimation results for the P/B ratio. Our organizational capital measures are positively associated with P/B and statistically significant: the coefficient is 0.095 for OC (Seed), 0.096 for OC (ChatGPT), and 0.133 for OC (Avg.), and remains positive when using percentile-scaled measures (0.003–0.006). We note that the magnitudes of these associations are more modest than those reported in Cai et al. (2025b), which may reflect the distinct features of Japan's equity market, including persistently low valuations (Kitagawa et al., 2025).

Overall, these results support that our organizational capital measures predict firm performance. The fact that the results hold across alternative constructions of our measure suggests that the relationship is not driven by a particular specification choice. While we do not interpret these estimates causally, the stable positive associations provide validation that our measures capture an economically meaningful component of organizational capital that is reflected in both accounting- and market-based performance outcomes.

4.2 Slow Evolution

The second property posits that organizational capital evolves slowly over time. If our measures capture this slow-moving nature of organizational capital, then, within a finite horizon, most of their variation should be persistent and stable within firms. Accordingly, we should expect firm fixed effects to explain a substantial share of the variation in our organizational capital measures.

To assess how variation in our organizational capital measures is attributed to firm, industry, CEO, and time components, we regress each measure on different combinations of these fixed effects and report the share of variation explained (i.e., the R^2). This variance-decomposition exercise is informative about persistence: if organizational capital is slow-moving and largely firm-specific, then firm fixed effects should account for a large fraction of the variation over our sample window, whereas year effects should play a comparatively smaller role. Including industry and CEO fixed effects further helps distinguish whether our measures primarily capture stable firm-level attributes, common industry-level forces, or CEO-specific components.

Table 3 reports the R^2 from regressions of our organizational capital measures on different sets of fixed effects. Year effects explain almost none of the variation. Year fixed effects account for only 0.2% (Seed Word) and 0.7% (ChatGPT), and adding year fixed effects to industry or firm fixed effects increases the R^2 only slightly. In contrast, firm fixed effects explain a sizable share of the variation, 24.9% (Seed Word) and 23.6% (ChatGPT), indicating that the measures are largely firm-specific over our sample window. CEO fixed effects also explain substantial variation, 32.1% and 31.5%, and including both CEO and firm fixed effects raises the R^2 only a little, to 32.3% and 31.7%. Overall, most variation in our measures comes from persistent differences across firms and CEOs, which is consistent with organizational capital evolving slowly over time.

As we note in Section 2.3, the NEEDS FinancialQUEST data only identify the current CEO and the year of inauguration, so earlier CEO turnovers are not observed. This limitation may attenuate the CEO component and inflate the firm component in the decomposition, so the relative magnitudes should be interpreted with caution. Given that CEO turnover is relatively rare among listed firms, we do not expect this measurement issue to severely bias our estimates. In the worst-case scenario, our estimates of the CEO and firm contributions can be interpreted as a lower bound and an upper bound, respectively.

4.3 Intangible

The third property emphasizes that organizational capital is intangible and therefore not perfectly observable. If our text-based measures capture this latent construct, they should be related to other employee-reported proxies for hard-to-observe workplace attributes. We examine this implication by correlating our organizational capital measures with OpenWork review scores, which summarize employees'evaluations of multiple aspects of the workplace.

Table 4 reports the correlations. All of our organizational capital measures are positively and statistically significantly correlated with the Overall Rating, and the correlations are economically meaningful. The correlation with overall rating is 0.41 for the seed-word-based measure, 0.26 for ChatGPT-based measure, and 0.40 for the average of the two, with similar magnitudes for the percentile-scaled measures. The correlations are generally larger for soft dimensions that reflect internal organization and day-to-day workplace practices, including motivation, openness, teamwork, and opportunities for growth and development, while they are smaller for hard dimensions such as compensation and compliance.

This pattern mirrors the evidence in Cai et al. (2025b), who document that their organizational-capital measures are most strongly related to soft workplace attributes (e.g., culture and management) rather than purely pecuniary benefits. Overall, these correlations provide supporting evidence that our measures capture intangible organizational features that are reflected in employees'assessments of the workplace.

4.4 Top Management's Influence

The fourth property posits that organizational capital is produced, at least in part, through the active engagement of top management. A natural corollary is that organizational capital varies substantially across CEOs. This can be examined in the same variance decomposition exercise reported in Table 3. As can be seen, CEO fixed effects explain a meaningful share of the variation in our measures, with R^2 of 32.1% for Seed Word and 31.5% for ChatGPT, indicating that differences across CEOs account for a substantial portion of organizational

capital variation. These results support the validity of our measures of organizational capital.

5 Applications

In this section, we demonstrate how the organizational capital measures can be deployed to study economically relevant heterogeneity inside firms. We present two applications that leverage the high-frequency, text-based nature of the measures. First, we examine how organizational capital responds to CEO turnover using an event-study design. Second, we study how firm performance relates to within-firm disagreement in employee assessments across reviewer gender.

5.1 CEO Turnover

To further explore top management's engagement with organizational capital, we examine how our organizational capital measures respond to CEO turnover. If organizational capital is produced through the active engagement of top management, then changes in the CEO may be associated with changes in organizational capital. We therefore test whether our measures shift around CEO turnover events. Our estimating equation is

$$\text{OC}_{i,t} = \sum_{k \in K} \beta_k \, \mathbb{1}\{t - t_i^* = k\} + \gamma_1 \ln((\text{Total Assets})_{i,t}) + \gamma_2 (\text{Firm Age})_{i,t} + \varepsilon_{i,t}, \qquad (5)$$

where t_i^* denotes the year of firm i's CEO turnover and $K = \{-3, -2, 0, 1, 2, 3\}$. The indicator $\mathbb{1}\{t - t_i^* = k\}$ equals one if year t is k years relative to the turnover event (and zero otherwise). We normalize one pre-turnover period (e.g., $k = -1$) as the omitted reference category, so each β_k measures the difference in organizational capital in event time k relative to $k = -1$.

Figure 3 plots event-time coefficients around CEO turnover. Overall, we do not observe strong pre-trends: the estimates in the years leading up to turnover are small and generally

close to the omitted year ($t = -1$). After turnover, the ChatGPT-based OC measure shows a modest decline, with point estimates becoming more negative in the first few years following the turnover, although the confidence intervals are wide and often include zero. By contrast, the Seed Word measure exhibits no clear or systematic shift around turnover. The averaged measure largely mirrors the pattern in the ChatGPT-based measure. The percentile-scaled versions deliver similar qualitative patterns. Taken together, the event-study evidence is consistent with a CEO-related component in organizational capital, but the estimated turnover effects are economically small.

The estimates are about half the magnitude of those reported in Cai et al. (2025b) for their U.S. sample. This may indicate that CEO turnover is followed by less disruptive succession in Japan, reflecting more internally planned and coordinated transitions rather than externally driven leadership changes. This interpretation resonates with prior evidence that CEO replacement in Japan is more likely when internal managerial talent is readily available (e.g., Coles and Uchida, 2018; Fujitani et al., 2025).

5.2 Gender Disagreement

As shown in Section 2.2.2, employee assessments differ systematically by reviewer gender, even within the same firm. To capture gender differences in perceived organizational capital, we compute the measures in equations (2) and (3) separately for reviews written by male and female reviewers. We estimate the following equation:

$$\begin{aligned} y_{i,t} = \beta_1 \, \mathrm{OC}_{i,t}^{\mathrm{Female}} + \beta_2 \, \mathrm{OC}_{i,t}^{\mathrm{Male}} + \beta_3 \, \left| \mathrm{OC}_{i,t}^{\mathrm{Female}} - \mathrm{OC}_{i,t}^{\mathrm{Male}} \right| \\ + \beta_2 \, \ln\left((\text{Total Asset})_{i,t} \right) + \beta_3 \, (\text{Firm Age})_{i,t} + \tau_{I(i),t} + \varepsilon_{i,t}, \end{aligned} \tag{6}$$

where $\mathrm{OC}_{i,t}^{\mathrm{Female}}$ and $\mathrm{OC}_{i,t}^{\mathrm{Male}}$ denote the measures of organizational capital computed for female and male reviewers, respectively, and $|\cdot|$ denotes absolute value. We use $\left| \mathrm{OC}_{i,t}^{\mathrm{Female}} - \mathrm{OC}_{i,t}^{\mathrm{Male}} \right|$ as our measure of within-firm gender disagreement, capturing the magnitude of the gap in per-

ceived organizational capital between female and male reviewers within the same firm-year. Our hypothesis is that within-firm gender disagreement reflects frictions such as unequal workplace experiences or misaligned internal practices conceivably due to poor governance, which should be associated with worse outcomes, so we expect $\beta_3 < 0$. For some firm-year observations, we do not observe male or female reviews and exclude those cases, which ends up with a sample smaller than the one in main validation exercise.

Table 5 relates firm performance to gender-specific organizational-capital measures and to the within-firm gender gap. Across both the seed-word- and ChatGPT-based measures, the levels of organizational capital are positively associated with performance, with the male-based measure loading more strongly than the female-based measure in this sample. More importantly, the disagreement term is consistently negative. An increase in the absolute gender gap is associated with lower ROA and a lower P/B ratio. This pattern suggests that greater divergence in employee assessments across gender reflects frictions inside the firm, such as unequal experiences or misaligned practices, that are detrimental to both accounting profitability and market valuation.

Taken together, these applications highlight that the measures provide broadly useful inputs for studying organizations in settings where internal practices and employee experiences are otherwise difficult to observe at scale. We next develop alternative measurement approaches to assess the robustness of our main organizational-capital measures.

6 Alternative Measurements

This section develops alternative constructions of organizational capital to assess the robustness of our baseline measures and to understand how sensitive the signal is to language choices and LLM design.

6.1 Japanese Pipeline

To benchmark our main language-free measurement protocol, we implement a parallel Japanese-language pipeline that is identical in structure to our baseline approach, except that it omits the English translation step. We first obtain structured pros, cons, and neutral fields via batched LLM processing and save the outputs as intermediate files for reproducibility. We then apply Japanese-specific pre-processing—morphological analysis and tokenization via MeCab with UniDic, removal and bigram construction—and estimate Word2Vec embeddings on the resulting corpus using the same parameterization as in the baseline pipeline. Using these embeddings, we construct Japanese analogues of our two main measures: a seed-word measure using Japanese counterparts of the seed words, and a ChatGPT-based measure that scores reviews relative to synthetic high- versus low-organizational-capital reviews generated directly in Japanese by ChatGPT. Finally, we aggregate review-level scores to the firm-year level and compute standardized and percentile-scaled variants.

Table E.16 reports cross-protocol correlations between the organizational-capital measures constructed under our main (language-free) protocol and those constructed directly from Japanese text. The correlations are uniformly positive and statistically significant, with coefficients typically in the 0.30–0.50 range across pairs, indicating substantial agreement between the two pipelines despite differences in language and pre-processing. The Japanese seed-word-based measure is most strongly correlated with the corresponding original seed-word-based and average measures, while the Japanese ChatGPT-based measure aligns particularly closely with the original ChatGPT-based and percentile-scaled variants. At the same time, the correlations are far from one, implying that each protocol retains a nontrivial orthogonal residual component.

To assess whether this residual reflects independent measurement noise or differences in the underlying construct captured by each protocol, we next replicate the downstream validation exercises in Section 4 using the Japanese-constructed measures. If the divergence primarily reflects random noise in each protocol, we should observe similar qualitative

patterns in the validation results. In contrast, if the protocols systematically capture different dimensions of organizational capital, the downstream empirical relationships may differ across measures.

Figure E.13 plots year-by-year averages of the Japanese-pipeline organizational-capital measures. The time-series patterns closely resemble those in Figure 2. The ChatGPT-based series starts above the seed-word-based series in the early years, the two measures converge around zero in the middle of the sample, and the seed-word-based measure drifts modestly upward relative to the ChatGPT-based measure toward the end of the observation period.

Tables E.17 and E.18 report the performance regressions using the Japanese-protocol measures. The results mirror those under the baseline protocol. All Japanese measures are positively and statistically significantly related to both ROA and the P/B ratio, and the percentile-scaled variants yield the same qualitative conclusions. ROA results are very similar to those under the baseline protocol, with only slightly larger coefficients. In contrast, the P/B ratio results differ more sharply. In Table E.18, the Japanese-constructed measures load much more strongly on the P/B ratio than their baseline counterparts. This pattern suggests that working directly in Japanese may recover linguistic signals that are especially informative about the organizational attributes priced by equity markets, so the Japanese-pipeline measures may align more closely with the component of organizational capital that predicts valuation (as opposed to accounting profitability). At the same time, the fact that both protocols deliver the same sign and significance for ROA and P/B ratio indicates a shared underlying construct, while the larger P/B magnitudes point to meaningful protocol-specific content rather than purely random noise.

Table E.19 reports the same variance-decomposition exercise for the Japanese-constructed measures. The results are not only qualitatively similar to the baseline decomposition in Table 3, but also numerically very close. In both pipelines, year fixed effects explain virtually none of the variation, firm fixed effects account for roughly one-quarter, and CEO fixed effects for about one-third, with the joint CEO-firm specification adding little incremen-

tal explanatory power. This near one-to-one correspondence suggests that the slow-moving nature of our organizational-capital measures, as well as the firm- and CEO-level variance components, are robust to whether the text is processed via the baseline protocol or constructed natively in Japanese.

Figure E.10 shows that the CEO-turnover event-study results in the Japanese pipeline. THe estimates are similar to those in the baseline estimates in Figures 3, though less precise. In both sets of results, our organizational capital measures display only modest downward shifts around turnover and the event-time profiles are simila, suggesting that the turnover-response pattern is not sensitive to whether the text is processed via translation or constructed natively in Japanese.

Table E.20 closely mirrors the baseline correlations in Table 4. In both, organizational capital is positively related to overall rating and most subratings, with larger correlations for "soft" workplace dimensions (motivation, openness, teamwork, development) than for compensation or compliance. Magnitudes for the seed-word measure are very similar across protocols, while the Japanese ChatGPT-based measure tends to correlate somewhat more strongly with ratings (e.g., overall rating: 0.35 vs. 0.26). Overall, the employee-rating validation is robust to the protocol, with only modest shifts in strength for the ChatGPT-based series.

The two protocols produce measures that exhibit substantial agreement in levels and deliver closely matching empirical patterns in the downstream validation exercises. The measures remain slow-moving, with variation primarily at the firm- and CEO- levels, and they are positively related to both accounting profitability and market valuation. The correlations with employee ratings are also similar across protocols and are consistently stronger for softer workplace dimensions than for compensation or compliance.

At the same time, the cross-protocol correlations are well below one, and the native-Japanese measures are more predictive of the P/B ratio than the baseline counterparts. These results suggest that translation and language-specific pre-processing affect which tex-

tual signals are emphasized. Our preferred interpretation is that the two protocols largely recover the same underlying construct, as reflected in their broadly comparable validation patterns. Predictive strength can nevertheless vary across downstream outcomes, implying that protocol-specific components may quantitatively matter for particular empirical applications even when the core signal is shared.

6.2 LLM-as-Rater Measure

We also construct more directly-sourced measures of organizational capital by instructing an LLM to evaluate and score each employee review, an approach we term the "LLM-as-Rater" method. Unlike the seed-word and synthetic-review measures, which rely on embedding-space similarities, this method leverages the LLM's ability to directly interpret the semantic content of the review against the theoretical definition of organizational capital. This direct method is depicted in Figure E.12.

We implement three variants of this measure to assess the robustness of our results across languages and prompting strategies. The first variant utilizes the English translation of the full Japanese review and prompts the model in English to rate the review's indication of organizational capital. The second variant involves a mixed-language approach: it feeds the model the original Japanese review text while providing the instructions and definition in English. The third variant utilizes the original Japanese review text and prompts the model entirely in Japanese to perform the same task. For all variants, we use gpt-5-nano as the rater.

The prompt incorporates the exact same definition of organizational capital used in the synthetic-review measure (see Appendix C.3) and instructs the model to assign a score from 0 to 100 representing the strength of organizational capital. The specific instructions used in the English prompt are reported in Appendix D. We define the firm-year OC measure as the average of these review-level scores for each firm-year observation.

We validate the LLM-as-Rater measure using synthetic reviews. Specifically, we apply

the rating pipeline to the same 200 synthetic reviews used to validate the ChatGPT-based approach: 100 reviews depicting firms with strong organizational capital and 100 reviews depicting firms with weak organizational capital. Figure D.8 shows the density distributions of review-level OC scores for each type. Reviews describing strong organizational capital receive substantially higher scores (mean = 64.0) than those describing weak organizational capital (mean = 14.5), with minimal overlap between the two distributions. This clear separation confirms that the LLM-as-rater approach effectively distinguishes textual signals of high versus low organizational capital.

6.3 Cross-Language and Cross-Method Comparisons

We now examine how the different organizational capital measures relate to each other across languages and measurement approaches. Table E.21 reports pairwise correlations among all seven OC measures at both the review level (Panel A) and the firm-year level (Panel B).

Three main patterns emerge. First, measures constructed from the same text but using different languages are positively correlated, indicating that the underlying signal is preserved across the translation process. At the review level, the correlation between the English and Japanese seed-word measures is 0.42, and between the English and Japanese ChatGPT measures is 0.39. These correlations increase substantially at the firm-year level (0.65 and 0.63, respectively), as aggregation reduces idiosyncratic noise in individual reviews. This suggests that while translation introduces some variation at the review level, the firm-level signal is largely preserved.

Second, the three LLM-as-Rater variants—using English-translated text with an English prompt, Japanese text with an English prompt, and Japanese text with a Japanese prompt—are highly correlated with each other. At the review level, pairwise correlations among these three measures range from 0.73 to 0.79; at the firm-year level, they range from 0.92 to 0.95. The high correlation between the English-translated and Japanese-text variants (0.79 at review level; 0.95 at firm-year level) indicates that the LLM extracts similar information

whether it reads translated English or original Japanese. Similarly, the high correlation between Japanese text with English versus Japanese prompts (0.79 at review level; 0.94 at firm-year level) suggests that prompt language has limited influence on the extracted signal once the model processes the same underlying review content.

Third, correlations across different measurement approaches (seed-word, ChatGPT similarity, and LLM-as-Rater) are positive but lower, reflecting methodological differences in how each approach quantifies organizational capital. The moderate cross-method correlations suggest that while all approaches capture related aspects of organizational capital, each method also contains distinct information.

Figure E.11 summarizes the full distribution of English–Japanese divergence for each approach. At the review level, the LLM-as-Rater divergence is tightly centered around zero with visibly thinner tails, while the embedding-based measures exhibit fatter tails, and the ChatGPT-similarity measure is slightly more dispersed than the seed-word measure. At the firm-year level (restricted to firm-years with at least 20 reviews), the distributions compress for all measures, indicating that aggregation reduces cross-language noise. There is still significant divergence between the measures using the Japanese and the English pipeline, which motivates analyzing more in detail what drives higher scores in different languages.

6.3.1 Divergent Reviews Across Measures

To examine why English- and Japanese-based scores can diverge for the same underlying review, we analyze reviews that differ substantially in their relative position in the two language-specific score distributions while remaining near the center of each measure overall. For each measure, we compute English and Japanese percentile ranks and define divergence as the percentile gap (English minus Japanese). We then select reviews from each tail of the divergence distribution, choosing cases whose divergence lies close to ± 2 standard deviations and whose average percentile across English and Japanese is closest to the 50th percentile. For the embedding-based measures, we additionally require that both language versions

contain non-trivial pros and cons text. Tables E.23, E.24, and E.25 report the selected examples, with Panel A showing high-English/low-Japanese cases and Panel B showing the reverse.

For the seed-word measure (Table E.23), divergence reflects how language-specific vocabularies interact with separately trained embeddings and expanded dictionaries. Even when the underlying themes overlap closely, the English and Japanese texts can differ in which terms align most strongly with each language's representative vector. For example, Panel B illustrates that the Japanese text contains phrasing about internal pressure, directives, and overtime that aligns closely with the Japanese expanded dictionary, while the English translation distributes the same content across less tightly aligned terms. As a result, the two languages can assign different percentile ranks even though they refer to comparable organizational features.

For the ChatGPT-based similarity measure (Table E.24), divergence arises from how pros and cons segmentation and translated phrasing affect the score, defined as similarity of pros to the strong-organizational-capital benchmark minus similarity of cons to the weak benchmark. In Panel A, the English version places several workplace positives–such as training quality, controllable overtime, and ease of taking vacations–in the cons segment which decreases the similarity to the weak OC benchmark, whereas the Japanese version places these elements in pros and reserves cons for critiques of organizational direction, seniority-based incentives, and limited originality. Because the measure subtracts similarity of the cons text to the weak benchmark, shifting which content appears in cons materially changes the net score. In Panel B, both language versions discuss team assignment, decision-making frictions, and managerial practices, but the Japanese text includes additional organization-specific detail about internal adoption and team morale, which affects similarity to the benchmark vectors and leads to a higher Japanese score relative to English.

For the LLM-as-rater measure (Table E.25), divergence reflects language-specific reasoning about which elements of the review constitute evidence of organizational capital.

In Panel A, both texts mention transparent evaluation and compensation systems and accommodations such as childcare leave, alongside concerns about power harassment and role demands at higher ranks. The English rationale assigns substantial weight to the existence of formal internal systems, producing a higher score, whereas the Japanese rationale places more emphasis on cultural and leadership aspects highlighted in the same review, resulting in a lower score. In Panel B, the review emphasizes store-level autonomy over merchandising within a structured shift system and relatively good work-life balance, while also noting low pay and limited retention of young full-time workers. The Japanese rationale interprets the structured delegation of responsibility as meaningful organizational capital, tempered by constraints on retention and differentiation, while the English rationale emphasizes the limited evidence of firm-wide capabilities, yielding a lower score.

Taken together, these divergent cases illustrate three channels through which English- and Japanese-based measures can differ at the review level: language-specific embeddings and dictionaries that reward different lexical concentrations, sensitivity of benchmark-similarity scores to how organizational content is allocated across pros and cons and expressed in translation, and language-contingent mappings from textual evidence to scalar ratings in the LLM-as-rater approach. These mechanisms help explain why the measures are positively correlated overall while still displaying meaningful dispersion for individual reviews.

6.4 Validation and Robustness

6.4.1 Bayesian Shrinkage

Firm-year OC measures based on few reviews are noisy. Because each review-level score combines the firm-year latent component with idiosyncratic measurement error, and because firms differ in the number of reviews they receive, the precision of the raw firm-year averages is heterogeneous. We apply an empirical Bayes (EB) shrinkage that pulls imprecise firm-year averages toward the industry-year mean. When industry is not available, we use the year-mean as the target.

Let $\Omega_{j,t} \in \mathbb{R}$ denote latent organizational capital of firm j in year t. Let $R_{j,t}$ be the set of reviews for firm j in year t, with $n_{j,t} \equiv |R_{j,t}|$. Each review $r \in R_{j,t}$ is mapped to a scalar signal $s_r = m(r)$, and we assume

$$s_r = \Omega_{j,t} + \varepsilon_r, \qquad \mathsf{E}[\varepsilon_r \mid j,t] = 0, \qquad \mathsf{Var}(\varepsilon_r \mid j,t) = \sigma_\varepsilon^2, \tag{7}$$

with $\{\varepsilon_r\}$ independent across reviews conditional on (j,t). Define the firm-year mean signal

$$\bar{s}_{j,t} \equiv \frac{1}{n_{j,t}} \sum_{r \in R_{j,t}} s_r, \tag{8}$$

so that $\bar{s}_{j,t} \mid \Omega_{j,t} \sim \mathcal{N}(\Omega_{j,t},\, \sigma_\varepsilon^2/n_{j,t})$.

Let $g(j)$ denote firm j's industry and impose an industry-year prior

$$\Omega_{j,t} \mid g(j), t \sim \mathcal{N}\big(\mu_{g(j),t},\, \sigma^2_{\Omega,g(j),t}\big). \tag{9}$$

The posterior mean is given by

$$\tilde{\Omega}_{j,t} = \mathsf{E}[\Omega_{j,t} \mid \bar{s}_{j,t}, g(j), t] = \lambda_{j,t}\bar{s}_{j,t} + (1-\lambda_{j,t})\mu_{g(j),t}, \qquad \lambda_{j,t} = \frac{\sigma^2_{\Omega,g(j),t}}{\sigma^2_{\Omega,g(j),t} + \sigma_\varepsilon^2/n_{j,t}}. \tag{10}$$

Let $\mathcal{J}_{g,t}$ denote the set of firms observed in industry g and year t, with $J_{g,t} \equiv |\mathcal{J}_{g,t}|$. We estimate the prior mean as the industry-year mean of firm-year averages,

$$\hat{\mu}_{g,t} \equiv \frac{1}{J_{g,t}} \sum_{j \in \mathcal{J}_{g,t}} \bar{s}_{j,t}. \tag{11}$$

Estimate the review-level noise variance using within-firm-year dispersion for $n_{j,t} \geq 2$:

$$\hat{v}_{j,t} \equiv \frac{1}{n_{j,t}-1} \sum_{r \in R_{j,t}} (s_r - \bar{s}_{j,t})^2, \qquad \hat{\sigma}_\varepsilon^2 \equiv \frac{\sum_{j,t:\, n_{j,t} \geq 2} (n_{j,t}-1)\hat{v}_{j,t}}{\sum_{j,t:\, n_{j,t} \geq 2} (n_{j,t}-1)}. \tag{12}$$

Within an industry-year cell, $\mathsf{Var}(\bar{s}_{j,t} \mid g,t) = \sigma^2_{\Omega,g,t} + \sigma_\varepsilon^2/n_{j,t}$, so a method-of-moments

estimator is

$$\widehat{\mathsf{Var}}_{g,t}(\bar{s}) \equiv \frac{1}{J_{g,t}-1}\sum_{j\in\mathcal{J}_{g,t}}(\bar{s}_{j,t}-\hat{\mu}_{g,t})^2, \qquad \hat{\sigma}^2_{\Omega,g,t} \equiv \max\left\{0,\ \widehat{\mathsf{Var}}_{g,t}(\bar{s}) - \hat{\sigma}^2_{\varepsilon}\left(\frac{1}{J_{g,t}}\sum_{j\in\mathcal{J}_{g,t}}\frac{1}{n_{j,t}}\right)\right\}. \tag{13}$$

The implemented EB score is

$$\hat{\Omega}^{EB}_{j,t} \equiv \hat{\lambda}_{j,t}\bar{s}_{j,t} + (1-\hat{\lambda}_{j,t})\hat{\mu}_{g(j),t}, \qquad \hat{\lambda}_{j,t} = \frac{\hat{\sigma}^2_{\Omega,g(j),t}}{\hat{\sigma}^2_{\Omega,g(j),t} + \hat{\sigma}^2_{\varepsilon}/n_{j,t}}. \tag{14}$$

Table E.29 shows that the firm-year review count distribution is highly skewed: 28.1% of firm-years have exactly one review and 43.9% have two or fewer, while the median is 3 (mean 9.5; max 1,032). Under the signal-plus-noise model, the sampling variance of the firm-year mean satisfies $(\bar{s}_{j,t} \mid \Omega_{j,t}) = \sigma^2_{\epsilon}/n_{j,t}$, implying substantial heterogeneity in precision across firm-years. This motivates using Empirical Bayes (EB) shrinkage to reduce the influence of low-n firm-years by shrinking their noisy averages toward the relevant industry-year mean.

Table E.30 reports the EB variance-component estimates and the implied shrinkage weights. Because the OC measures differ mechanically in scale (e.g., the LAR scores are on a 0–100 scale), the levels of $\hat{\sigma}^2_{\epsilon}$ and $\hat{\sigma}^2_{\Omega}$ are not directly comparable across measures; the key summary statistic is the implied weight $\lambda_{j,t} = \hat{\sigma}^2_{\Omega,g,t}/(\hat{\sigma}^2_{\Omega,g,t} + \hat{\sigma}^2_{\epsilon}/n_{j,t})$. The average shrinkage factors are low, with $\bar{\lambda}$ ranging from 0.129 to 0.268 across measures, consistent with a setting in which many firm-years are measured imprecisely. Importantly, $\lambda_{j,t}$ is increasing in $n_{j,t}$, so EB shrinkage is concentrated among the sparse firm-years and becomes modest for firm-years with many reviews.

Table E.31 evaluates the implications for predictive regressions of ROA and the P/B ratio on standardized OC measures. The $n > 10$ restriction is informative about high-precision firm-years, but it also changes the composition of the sample (toward firms that attract many reviews) and yields substantially larger standard errors due to the smaller sample. The EB specifications instead keep the broad sample and implement a precision-weighted

version of within-industry-year variation: because the regressions include industry-by-year fixed effects, shrinking toward the industry-year mean primarily downweights low-precision firm-years by compressing their deviations from the industry-year mean (up to the subsequent standardization). Reassuringly, the EB estimates remain positive and statistically significant across measures and outcomes, and are generally of similar order of magnitude to the baseline. For ROA, EB estimates are close to baseline, indicating that the baseline association is not driven by a small set of noisy low-n firm-years. For P/B, EB often strengthens the association, consistent with measurement error in raw firm-year averages attenuating the baseline relationship. Finally, the "EB (Industry Available)" rows yield qualitatively similar results, suggesting that the handling of firm-years without industry codes is not driving the findings.

6.4.2 Company-Name Bias

A general concern with LLM-based measurement is that the model may use information beyond the intended signal in the text. Employee reviews can contain direct cues about the firm (e.g., references to market success, profitability, or other outcomes), which could mechanically raise the inferred organizational-capital (OC) score even if the underlying organizational content is weak. More subtly, even when the review text is held fixed, mentioning a firm name can activate the LLM's pre-trained beliefs about that firm. Because large language models are trained on broad internet corpora, they may encode firm-specific reputational priors and implicitly "recognize" well-known companies; the resulting OC score may then reflect a combination of review content and an identity-conditioned prior, rather than review content alone. This is a form of rater bias (or lack of measurement invariance) that can generate non-classical measurement error if firm reputation is correlated with outcomes such as profitability.

We assess the magnitude of this firm-identity channel using a controlled prompt-perturbation test. Starting from the same synthetic high-OC and low-OC reviews used in the LLM-as-

Rater validation exercise, we re-run the rater while prepending each review with the name of a company that is publicly ranked as a strong workplace.[4] Because the underlying review text is identical across conditions, any systematic change in the assigned OC score isolates the effect of firm-name priming, i.e., an identity-driven reputational prior.

Figure D.8 reports the resulting score distributions. Two patterns emerge. First, the rater continues to clearly differentiate high-OC from low-OC synthetic reviews even when company names are added, indicating that semantic content remains the primary driver of scores. Second, mentioning a "best workplace" company name shifts the low-OC distribution meaningfully upward (a +5.9 point increase in the mean), while the high-OC distribution shifts modestly downward (a -4.3 point change in the mean). This response is consistent with a reputational "halo" effect: when the firm name carries a prior, the model partially discounts negative organizational signals in the text, compressing the distance between low and high OC scores (attenuation) even though separation remains substantial. The attenuation of high OC-scores could reflect the fact the LLM's the prior about these companies is not as high as what is reflected in the review text.

These findings suggest that, while the LLM-as-Rater measure is largely content-driven, firm identifiers can induce systematic shifts in levels. In applications where reviews frequently name the employer (or contain enough context to infer it), the estimated firm-year OC may partly reflect firm reputation rather than organizational practices described in the text.

6.5 Limitations

Our use of LLMs to measure a latent construct like organizational capital comes with significant caveats. The classification and scoring outcomes can vary significantly with changes in how the LLM is used. Small differences in prompt phrasing or the choice of LLM model (and even using Japanese vs. English text) may lead to inconsistent results. Recent work shows that without a grounding truth or calibration, seemingly innocuous choices like which model

[4]We use 15 large firms from the "Best Workplaces in Japan 2025" list (Hatarakigai, 2025)

or prompt is used can produce dramatically different outputs and downstream estimates (Ludwig et al., 2025). In our context, this means the measured organizational capital is not perfectly stable and it could shift if we altered the prompt wording, used a different LLM, or relied on one language over the other.

A second concern is look-ahead bias and memorization. Large language models are trained on vast corpora and may implicitly "know" far more than they should for a given task. If an LLM has seen data or outcomes related to our sample during training, it could baselessly inflate its performance by regurgitating that knowledge. Ludwig et al. (2025) stress that any overlap between the LLM's training data and the researcher's sample violates the no-leakage condition for valid inference. In our setting, we cannot rule out that a closed-source model was trained on text that gives away information about certain firms. Absent transparency into the training set, there is a risk that the model's ratings are contaminated by prior knowledge of the firm's identity or success. For instance, an LLM might recognize a company's name (or deduce it from context) and recall that firm's general reputation or performance, thus biasing the "organizational capital" score upward or downward. This undermines the idea that the score comes solely from the content of the review. In extreme cases, LLMs have been shown to reconstruct masked information from minimal textual cues (Lopez-Lira et al., 2025).

Levy (2024) finds that much of the apparent "superhuman" performance of LLMs in finance stems from such modeling artifacts rather than true economic insight. In our context, one worry is that the LLM-as-rater method might partly be picking up spurious cues or even outcome hints in the text (e.g. an employee review explicitly mentioning strong profits or growth) instead of genuine organizational traits. If some reviews directly reference the firm' s success, the model's score would trivially correlate with performance – a form of label leakage that inflates the perceived validity of the measure. These issues imply that our LLM-derived organizational capital metrics, especially those using a black-box model, may be overly optimistic indicators unless we rigorously guard against information leakage.

One difficulty in measuring an intangible concept such as organizational capital is that there is no established "gold standard" against which we can validate the LLM-based measures. Unlike a supervised classification task with known labels or a numeric prediction with observed outcomes, here the true value of a firm's organizational capital (as inferred from text) is inherently unobservable. We chose not to construct a human-coded validation set because it's unclear that human judges would agree on these classifications or produce a more valid measure than LLMs. However, the absence of any human benchmark or external criterion makes it difficult to assess the accuracy or bias of our LLM outputs. Ludwig et al. (2025) argue that without a small set of verified labels to calibrate against, one cannot quantify the error in LLM-generated measures, and as a result even minor design choices can yield materially different conclusions. In essence, we are treating the LLM's own output as the object of interest – defining organizational capital in terms of what the model identifies in the text. This approach can be useful when a concept has no clear definition or ground truth, but it carries interpretational risk (Ludwig et al., 2025). The LLM's score may capture some latent aspect of firm culture or employee sentiment, but without an external point of reference, we must be cautious in claiming what that score truly represents.

Finally, we outline one potential robustness strategy as a future step to address these concerns. One simple guard against firm-specific bias is to remove identifiable firm information from the input texts. By using only reviews that do not mention the company name (or redacting names), we can force the LLM to rely solely on the content of the review rather than prior knowledge about the firm. If the organizational capital scores remain stable without company identifiers, it would increase confidence that the model is measuring intrinsic qualities rather than reputation. On the other hand, if scores drop or change significantly, that would reveal the extent of identity-driven bias in the original ratings.

In summary, while our study demonstrates the promise of LLMs in quantifying organizational capital at scale, these models'known limitations warrant a careful and nuanced approach. By refining the pipeline by excluding potentially biasing information and verify-

ing stability across conditions, future research can bolster confidence that the LLM-based measures reflect true organizational intangibles rather than artifacts.

7 Conclusion

This paper develops firm-year measures of organizational capital for Japan using employee reviews from OpenWork. The Japanese setting poses two practical hurdles for implementing existing review-based measurement approaches: the reviews are written in a linguistically distant language and the platform's narrative format differs from the pros/cons template used in prior work. We address these hurdles with an LLM-based conversion step that translates and restructures OpenWork narratives into standardized pros, cons, and neutral fields, and we then apply a text-embedding procedure to construct two complementary measures of organizational capital, a seed-word measure and a synthetic-review (ChatGPT-based) measure, along with standardized and percentile-scaled variants.

The resulting measures exhibit empirical properties consistent with organizational capital as defined in Dessein and Prat (2022). They are positively related to both accounting performance and market valuation in firm-year regressions with rich controls. They are slow-moving in the sense that a large share of variation is explained by firm and CEO components, while time effects contribute little. They are also meaningfully correlated with employee-reported workplace ratings, with stronger associations for softer organizational dimensions than for compensation or compliance. Event-study evidence around CEO turnover indicates modest shifts in organizational capital following leadership changes, consistent with a CEO-linked component, although the magnitudes are economically small.

A central contribution is to show that these measurement results are not an artifact of translating Japanese reviews into English. We implement a parallel native-Japanese pipeline that preserves the structure of the baseline approach but omits translation and instead applies Japanese-specific tokenization and preprocessing. The native-Japanese measures are

positively correlated with the translation-based measures and replicate the main validation patterns. At the same time, the correlations are well below one and some downstream relationships differ in magnitude, most notably for market valuation. This combination of similarity and divergence suggests that both protocols recover a common core construct while placing different weights on linguistic signals that may be more salient in Japanese text or in translated English renderings.

Our findings have two implications. Substantively, they provide new evidence that organizational capital is an economically meaningful and persistent source of heterogeneity among Japanese firms, with variation that is associated with firm performance and is reflected in employee assessments and, to some extent, in top-management changes. Methodologically, they illustrate a practical framework for exporting review-based measurement of intangible assets to settings where language and platform formats differ, using LLMs to harmonize inputs while retaining transparency in the downstream construction. Future work can apply this translation-and-normalization approach to other countries and platforms and conduct cross-country comparison.

One limitation of our study is that we cannot cleanly disentangle cross-country differences in underlying organizational fundamentals from differences driven by measurement error in text-based proxies. When our estimates differ from those in other countries or platforms, it is difficult to determine whether the gap reflects true variation in organizational capital or artifacts of language, platform design, review composition, and model-based processing. Even when downstream validation exercises produce similar qualitative patterns, concluding that the underlying economic relationships are truly comparable still requires assumptions. In principle, distinct structural parameters could generate similar reduced-form evidence if measurement errors differ across settings and happen to offset in a way that preserves the downstream correlations. Developing a statistical framework to formalize this measurement problem is left for future work.

References

Alan, S., Corekcioglu, G., and Sutter, M. (2023). Improving workplace climate in large corporations: A clustered randomized intervention. *The Quarterly Journal of Economics*, 138(1):151–203.

Atkeson, A. and Kehoe, P. J. (2005). Modeling and measuring organization capital. *Journal of Political Economy*, 113(5):1026–1053.

Baker, G., Gibbons, R., and Murphy, K. J. (2002). Relational contracts and the theory of the firm. *The Quarterly Journal of Economics*, 117(1):39–84.

Bana, S. H. (2022). Work2vec: using language models to understand wage premia.

Bandiera, O., Prat, A., Hansen, S., and Sadun, R. (2020). Ceo behavior and firm performance. *Journal of Political Economy*, 128(4):1325–1369.

Bloom, N. and Van Reenen, J. (2007). Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics*, 122(4):1351–1408.

Cai, W. (2023). Formalizing the Informal: Adopting a Formal Culture-Fit Measurement System in the Employee-Selection Process. *The Accounting Review*, 98(3):47–70.

Cai, W., Campbell, D., Chen, Y., Chen, Y., and Prat, A. (2025a). Measuring quality. *Available at SSRN 5840182*.

Cai, W., Prat, A., and Yu, J. (2025b). Measuring organizational capital. CEPR Discussion Paper 20229, Centre for Economic Policy Research.

Cai, W., Prat, A., and Yu, J. (2025c). Mergers and Non-contractible Benefits: The Employees' Perspective.

Cai, W., Rouen, E., and Zou, Y. (2022). Passing the mic: Career and firm outcomes of executive interactions. *Available at SSRN 3942049*.

Castro, S., Englmaier, F., and Guadalupe, M. (2022). Fostering psychological safety in teams: Evidence from an rct. *Available at SSRN 4141538*.

Clayton, C., Coppola, A., Maggiori, M., and Schreger, J. (2025). Geoeconomic pressure. *Columbia Business School Research Paper Forthcoming, Stanford University Graduate School of Business Research Paper Forthcoming*.

Coles, J. L. and Uchida, K. (2018). Power of tournament incentives: Evidence from japanese corporate boards. *SSRN Electronic Journal*.

Corrado, C., Haskel, J., Jona-Lasinio, C., and Iommi, M. (2022). Intangible capital and modern economies. *Journal of Economic Perspectives*, 36(3):3–28.

Crouzet, N., Eberly, J. C., Eisfeldt, A. L., and Papanikolaou, D. (2022). The Economics of Intangible Capital. *Journal of Economic Perspectives*, 36(3):29–52.

Dessein, W. and Prat, A. (2022). Organizational Capital, Corporate Leadership, and Firm Dynamics. *Journal of Political Economy*, 130(6):1477–1536. Publisher: The University of Chicago Press.

Dube, A., Naidu, S., and Reich, A. D. (2022). Power and dignity in the low-wage labor market: Theory and evidence from wal-mart workers. Technical report, National Bureau of Economic Research.

Fujitani, R., Kusaka, Y., and Ning, D. (2025). Internal candidate pool and ceo turnover in japan: A pre-registered report. *Pacific-Basin Finance Journal*, page 102788.

Gorton, G. B., Grennan, J., and Zentefis, A. K. (2022). Corporate culture. *Annual Review of Financial Economics*, 14(1):535–561.

Graham, J. R., Grennan, J., Harvey, C. R., and Rajgopal, S. (2022). Corporate culture: Evidence from the field. *Journal of Financial Economics*, 146(2):552–593.

Guiso, L., Sapienza, P., and Zingales, L. (2015). The value of corporate culture. *Journal of Financial Economics*, 117(1):60–76.

Hatarakigai (2025). Best workplaces in japan 2025. `https://hatarakigai.info/ranking/japan/2025.html`. Accessed: YYYY-MM-DD.

He, B., Mostrom, L. I., and Sufi, A. (2024). Investing in customer capital. Technical report, National Bureau of Economic Research.

Hermalin, B. E. (2012). Leadership and corporate culture. *Handbook of Organizational Economics*, pages 432–78.

Iwata, K. (2025). The dark side of dual loyalty: How internal organizational frictions stifle corporate governance reform in japan. *Available at SSRN 5882524*.

Kambayashi, R., Ohyama, A., and Hori, N. (2021). Management practices and productivity in japan: Evidence from six industries in jp mops. *Journal of the Japanese and International Economies*, 61:101152.

Kitagawa, R., Liu, W., and Nagata, R. (2025). Stock Exchanges as Governance Institutions: Evidence from the Tokyo Stock Exchange. *Available at SSRN 5866042*.

Kreps, D. M. et al. (1990). Corporate culture and economic theory. *Perspectives on Positive Political Economy*, 90(109-110):8.

Levy, B. L. (2024). Caution ahead: Numerical reasoning and look-ahead bias in ai models. *Fama-Miller Working Paper, Chicago Booth Research Paper*.

Li, K., Mai, F., Shen, R., and Yan, X. (2021). Measuring Corporate Culture Using Machine Learning. *The Review of Financial Studies*, 34(7):3265–3315.

Li, K., Mai, F., Shen, R., Yang, C., and Zhang, T. (2025). Dissecting Corporate Culture Using Generative AI. *The Review of Financial Studies*, page hhaf081.

Lopez-Lira, A., Tang, Y., and Zhu, M. (2025). The memorization problem: Can we trust llms' economic forecasts? *arXiv preprint arXiv:2504.14765*.

Ludwig, J., Mullainathan, S., and Rambachan, A. (2025). Large language models: An applied econometric framework. Working Paper 33344, National Bureau of Economic Research.

Prescott, E. C. and Visscher, M. (1980). Organization Capital. *Journal of Political Economy*, 88(3):446–461. Publisher: The University of Chicago Press.

Sockin, J. (2022). Show me the amenity: Are higher-paying firms better all around?

Syverson, C. (2011). What Determines Productivity? *Journal of Economic Literature*, 49(2):326–365.

Teece, D. J., Pisano, G., and Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, 18(7):509–533.

Tables

Table 1: Organizational Capital and Firm Performance (ROA)

	Dependent variable: ROA (%)					
OC (Seed)	0.576***					
	(0.100)					
OC (ChatGPT)		0.586***				
		(0.099)				
OC (Avg.)			0.811***			
			(0.119)			
OC (Seed, %)				0.017***		
				(0.003)		
OC (ChatGPT, %)					0.022***	
					(0.004)	
OC (Avg., %)						0.028***
						(0.004)
log(Total Assets)	0.711***	0.709***	0.703***	0.708***	0.701***	0.695***
	(0.116)	(0.116)	(0.116)	(0.116)	(0.116)	(0.115)
Firm Age	−0.045***	−0.045***	−0.044***	−0.045***	−0.044***	−0.044***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Observations	23,918	23,918	23,918	23,918	23,918	23,918
R^2	0.216	0.216	0.217	0.215	0.216	0.217
Industry-Year FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is return on assets (ROA is in percent for readability). Each column reports a separate regression of ROA on one organizational capital (OC) measure, controlling for the logarithm of total assets and firm age, with industry-by-year fixed effects. OC (Seed), OC (ChatGPT), and OC (Avg.) use standardized (z-score) firm-year measures; OC (Avg.) is the simple average of the two standardized measures. The percentile measures ("%") are firm-year percentile ranks (1–100). Standard errors are clustered at the firm level and reported in parentheses. $p < 0.01$, $p < 0.05$, $p < 0.10$.

Table 2: Organizational Capital and Firm Performance (P/B Ratio)

	Dependent variable: P/B Ratio					
OC (Seed)	0.095*					
	(0.055)					
OC (ChatGPT)		0.096*				
		(0.050)				
OC (Avg.)			0.133**			
			(0.065)			
OC (Seed, %)				0.005**		
				(0.002)		
OC (ChatGPT, %)					0.003*	
					(0.002)	
OC (Avg., %)						0.006**
						(0.002)
log(Total Assets)	−0.174***	−0.174***	−0.175***	−0.177***	−0.175***	−0.178***
	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)
Firm Age	−0.037***	−0.037***	−0.037***	−0.037***	−0.037***	−0.037***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	24,063	24,063	24,063	24,063	24,063	24,063
R^2	0.188	0.188	0.188	0.189	0.188	0.189
Industry-Year FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is the price-to-book (P/B) ratio. Each column reports a separate regression of P/B on one organizational capital (OC) measure, controlling for the logarithm of total assets and firm age, with industry-by-year fixed effects. OC (Seed), OC (ChatGPT), and OC (Avg.) use standardized (z-score) firm-year measures; OC (Avg.) is the simple average of the two standardized measures. The percentile measures ("%") are firm-year percentile ranks (1–100). Standard errors are clustered at the firm level and reported in parentheses. $p < 0.01$, $p < 0.05$, $p < 0.10$.

Table 3: Variance Decomposition of Organizational Capital

No.	Fixed Effects	OC (Seed Word)	OC (ChatGPT)
1	Year Fixed Effects	0.2%	0.7%
2	Industry Fixed Effects	4.1%	4.2%
3	Industry Fixed Effects + Year Fixed Effects	4.2%	5.0%
4	Industry × Year Fixed Effects	21.1%	21.1%
5	Firm Fixed Effects	24.9%	23.6%
6	Firm Fixed Effects + Year Fixed Effects	25.0%	24.3%
7	CEO Fixed Effects	32.1%	31.5%
8	CEO Fixed Effects + Year Fixed Effects	32.2%	32.0%
9	CEO–Firm Fixed Effects	32.3%	31.7%
10	CEO–Firm Fixed Effects + Year Fixed Effects	32.4%	32.2%

Notes: Each entry reports the overall R^2 from a regression of firm-year organizational capital on the listed fixed effects only (no covariates). The Seed Word and ChatGPT measures are standardized (z-scores). Values are shown as percentages of total variation explained by the fixed-effects structure.

Table 4: Correlations between Organizational Capital Measures and Review Scores

	OC (Seed)	OC (ChatGPT)	OC (Avg.)	OC (Seed, %)	OC (ChatGPT, %)	OC (Avg., %)	Overall Rating	Compensation Rating	Motivation Rating	Openness Rating	Teamwork Rating	Chance to Grow	Long-term Development	Compliance Rating	Employee Evaluation System
OC (Seed)	1.00														
OC (ChatGPT)	0.44***	1.00													
OC (Avg.)	0.85***	0.84***	1.00												
OC (Seed, %)	0.91***	0.41***	0.78***	1.00											
OC (ChatGPT, %)	0.41***	0.94***	0.79***	0.42***	1.00										
OC (Avg., %)	0.78***	0.80***	0.94***	0.84***	0.84***	1.00									
Overall Rating	0.41***	0.26***	0.40***	0.42***	0.26***	0.40***	1.00								
Compensation Rating	0.22***	0.07***	0.17***	0.23***	0.06***	0.17***	0.64***	1.00							
Motivation Rating	0.31***	0.25***	0.33***	0.32***	0.25***	0.34***	0.74***	0.36***	1.00						
Openness Rating	0.34***	0.20***	0.32***	0.36***	0.20***	0.33***	0.68***	0.30***	0.45***	1.00					
Teamwork Rating	0.29***	0.22***	0.30***	0.29***	0.22***	0.30***	0.67***	0.24***	0.51***	0.50***	1.00				
Chance to Grow	0.28***	0.21***	0.29***	0.29***	0.21***	0.30***	0.68***	0.27***	0.54***	0.43***	0.38***	1.00			
Long-term Development	0.30***	0.21***	0.30***	0.30***	0.21***	0.30***	0.74***	0.45***	0.45***	0.39***	0.42***	0.45***	1.00		
Compliance Rating	0.19***	0.08***	0.16***	0.20***	0.08***	0.17***	0.53***	0.36***	0.22***	0.25***	0.25***	0.11***	0.38***	1.00	
Employee Evaluation System	0.25***	0.17***	0.25***	0.25***	0.16***	0.24***	0.64***	0.39***	0.48***	0.32***	0.30***	0.44***	0.39***	0.13***	1.00

Notes: Entries report pairwise correlations. The diagonal elements are 1.00 by construction and the upper triangle is omitted. OC (Seed), OC (ChatGPT), and OC (Avg.) use standardized (z-score) firm-year measures; OC (Avg.) is the simple average of the two standardized measures. The percentile measures ("%") are firm-year percentile ranks (1–100). Significance levels are based on two-sided tests of the null hypothesis of zero correlation. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 5: Organizational Capital by Reviewer Gender and Firm Performance

	ROA (%)	P/B Ratio	ROA (%)	P/B Ratio
Female OC (Seed)	0.141	0.160***		
	(0.124)	(0.052)		
Male OC (Seed)	0.554***	0.206*		
	(0.173)	(0.112)		
Gender Disagreement (Seed)	−0.317**	−0.148*		
	(0.157)	(0.089)		
Female OC (ChatGPT)			0.238*	−0.006
			(0.122)	(0.066)
Male OC (ChatGPT)			0.685***	0.237**
			(0.181)	(0.105)
Gender Disagreement (ChatGPT)			−0.439***	−0.191**
			(0.166)	(0.074)
log(Total Assets)	0.439***	−0.187***	0.409***	−0.194***
	(0.154)	(0.072)	(0.154)	(0.070)
Firm Age	−0.041***	−0.033***	−0.040***	−0.033***
	(0.008)	(0.005)	(0.008)	(0.005)
Observations	12,824	12,824	12,824	12,824
R^2	0.250	0.171	0.252	0.171
Industry-Year FEs	Yes	Yes	Yes	Yes

Notes: The unit of observation is a firm-year. Columns (1)–(2) use the seed-word-based organizational capital measure; columns (3)–(4) use the ChatGPT-based organizational capital measure. We include the measures based on female and male reviewers, separately, and the absolute value of the difference between the two as our proxy for disagreement between genders. The measures are standardized using a pooled mean and standard deviation across female and male series. All specifications control for log(total assets) and firm age and include industry-by-year fixed effects. Standard errors are clustered at the firm level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Figures

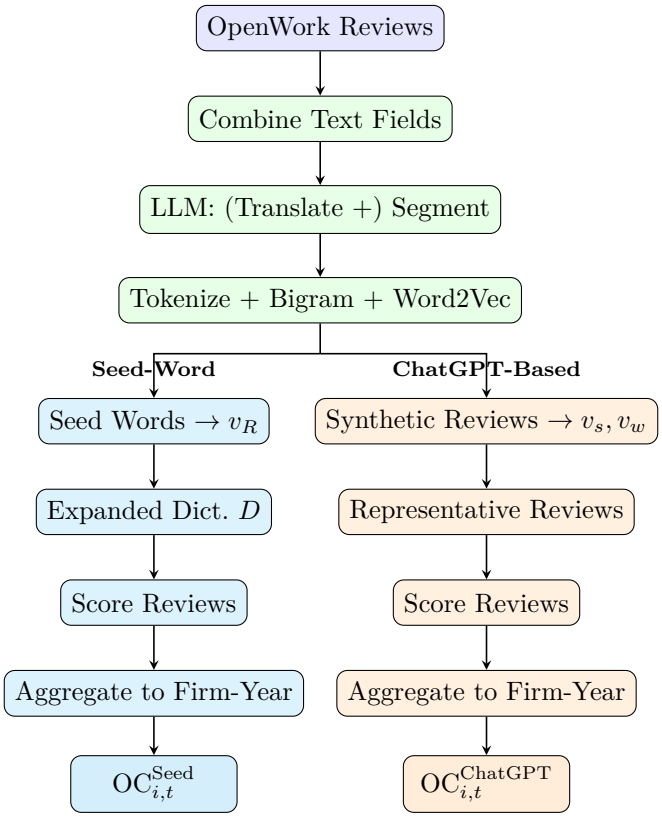


Figure 1: Pipeline for Constructing Organizational Capital Measures. Multiple narrative fields are combined and segmented into pros/cons/neutral via LLM. The main pipeline translates to English before segmentation; the Japanese pipeline (Section 5.1) segments directly and uses MeCab tokenization. The seed-word method builds a representative vector v_R, expands the dictionary D, and scores reviews based on similarity. The ChatGPT-based method generates synthetic benchmark vectors v_s (strong OC) and v_w (weak OC), computes review vectors, and scores by similarity. Both methods compute pros minus cons differences, aggregated separately to firm-year measures.

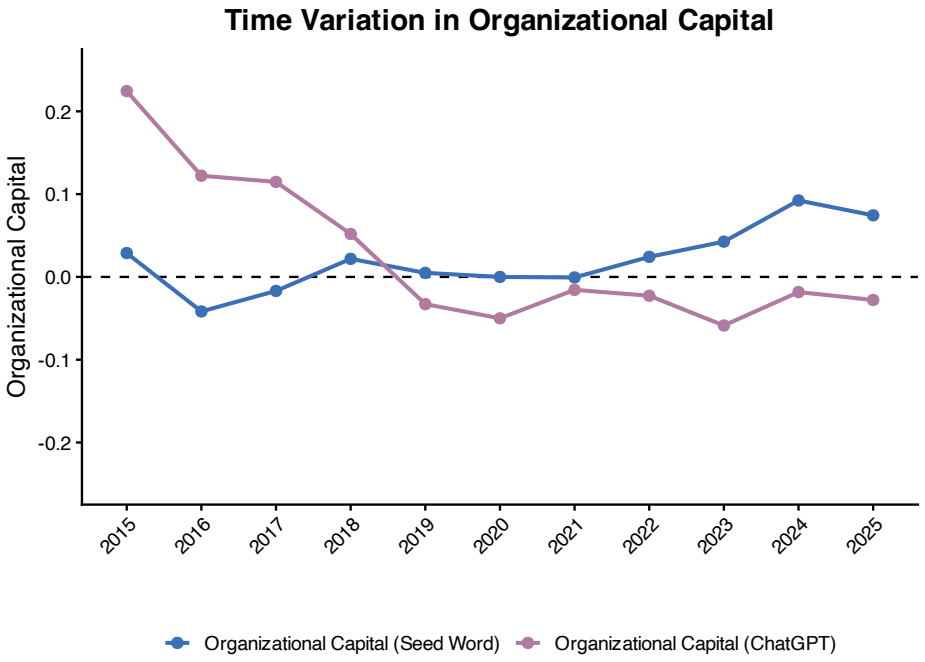
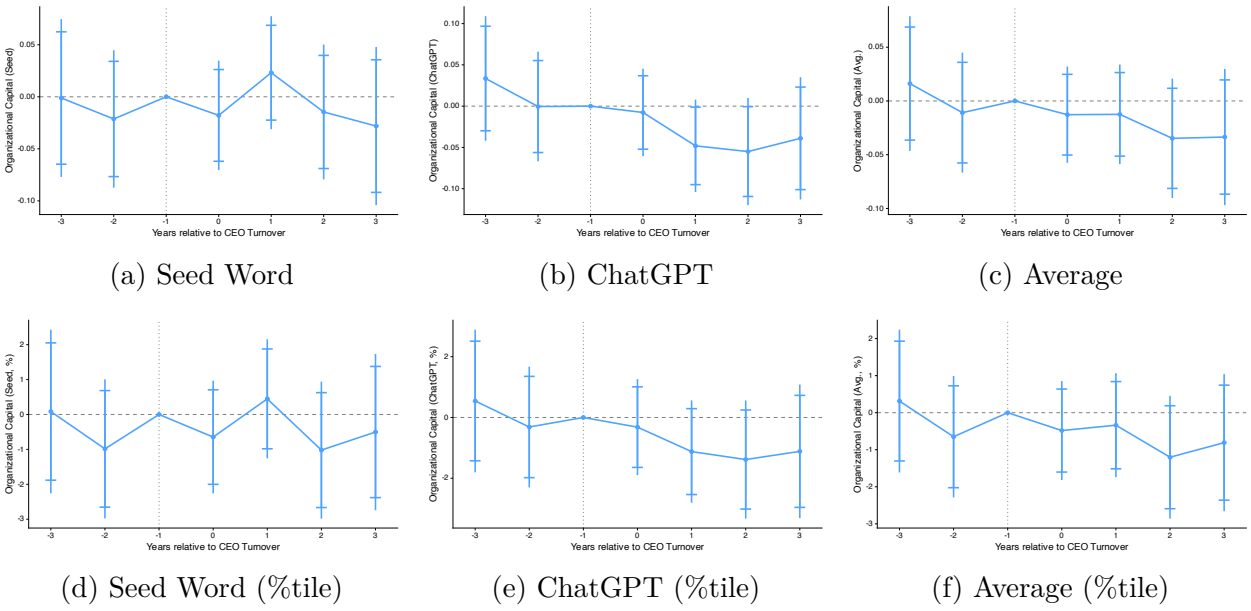


Figure 2: Time Variation in Organizational Capital

Notes: The figure plots yearly averages of firm-year organizational capital measures constructed from employee reviews. The seed-word and ChatGPT measures are standardized (z-scores) and averaged within year; the horizontal dashed line denotes zero.



(a) Seed Word (b) ChatGPT (c) Average

(d) Seed Word (%tile) (e) ChatGPT (%tile) (f) Average (%tile)

Figure 3: CEO Turnover and Organizational Capital

Notes: The figure reports event-study estimates around CEO turnover. The horizontal axis is event time in years relative to the CEO turnover ($t = 0$), with the year before turnover ($t = -1$) omitted as the reference period. Coefficients are from firm- and year-fixed-effects regressions of the indicated organizational capital (OC) measure on event-time indicators, controlling for the logarithm of total assets and firm age. Points show coefficient estimates; vertical bars show 95% confidence intervals and the horizontal bars on each end show 90% confidence intervals. The dashed horizontal line denotes zero and the dotted vertical line marks the reference year ($t = -1$). Standard errors are clustered at the firm level.

Appendices

A OpenWork Data

The OpenWork extract includes three files. The firm master file contains firm identifiers, names, and securities codes. The corporate-group file contains parent–subsidiary links and ownership rates. The review-level file contains the review posting timestamp, reviewer characteristics, quantitative ratings, narrative text fields, and additional self-reported workplace and compensation items. Reviewer characteristics include job category, tenure at the employer, employment status at the time of posting, and gender.

Table A.1: Coding of Key Reviewer Characteristics

Variable	**Values**	**Definition**
`sex`	1, 2	1: male; 2: female.
`status_of_employment`	1, 2	1: current employee; 2: former employee.
`job_tenure`	1, 3, 5, 10, 15, 20	Approximate tenure bracket (years): 1: < 3; 3: 3–5; 5: 5–10; 10: 10–15; 15: 15–20; 20: ≥ 20.
`year_of_retiring`	YYYY	Year of retirement (observed only for former employees; optional field).

Table A.2 and Table A.3 report the OpenWork question labels along with an English translation, and Table A.4 report other self-reported items. All rating items are recorded on a 1–5 scale. Free-text fields are optional open responses.

Table A.2: Rating Items (1–5 Scale): Japanese Labels and English Translation

Japanese label	**Japanese description**	**English translation**
総合評価	個別スコア（待遇面の満足度〜人事評価の適正感）を総合したスコア	Overall employer rating.
待遇面の満足度	給料や賞与に対する満足度	Satisfaction with salaries and benefits.
社員の士気	組織内の士気の高さ	Employee morale.
風通しの良さ	部署や役職を越えて議論しやすいか	Openness of communication and discussion.
社員の相互尊重	社員同士が尊重し合える文化とチームワークの高さ	Mutual respect and teamwork.
20 代成長環境	20 代の社員が成長できる仕組みや環境があるか	Growth opportunities for younger employees.
人材の長期育成	採用した社員を 30 代〜50 代にいたるまで長期に育成する環境にあるか	Long-term employee development.
法令遵守意識	法令を遵守しようとする組織内の意識の高さ	Compliance orientation.
人事評価の適正感	実力に応じて評価、昇進がされているか	Fairness of evaluation and promotion.

Table A.3: Free-Text Prompts: Japanese Labels and English Translation

Japanese label	**Japanese description**	**English translation**
組織体制・企業文化	企業の組織体制、企業文化の特徴への自由回答	Organizational structure and corporate culture.
入社理由と入社後ギャップ	入社を決めた理由とその妥当性、入社前に認識すべき事柄への自由回答	Reason for joining and any gap after joining.
働きがい・成長	どのような点で働きがいが得られるか、また自身の成長やキャリア開発への影響に関する自由回答	Work fulfillment and growth opportunities.
女性の働きやすさ	女性がやりがいを持って働き続けやすい環境が整っているかに対する自由回答	Working environment for women.
ワークライフバランス	プライベートと業務のバランスを調整しやすいか、に対する自由回答	Work–life balance.
退職検討理由	転職を考えた理由やきっかけへの自由回答	Reason for quitting or considering leaving.
企業分析（強み・弱み・展望）	事業の強みや弱みの分析と、今後の事業展望への予測に関する自由回答	Strengths, weaknesses, and outlook of the company.
経営者への提言	経営者に向けた企業改善につながる提言	Suggestions to top management.
給与制度・評価制度	給与や評価制度への自由回答	Salary and evaluation systems.

Table A.4: Additional Self-Reported Quantitative Items

Japanese label	**Japanese description**	**English translation**
残業時間	休日出勤も含めた残業時間	Monthly overtime hours including holiday work.
有給消化率	有給休暇のおおよその消化率	Approximate usage rate of paid holidays.
年収	年収（単位：万円）	Annual salary (units of 10,000 JPY).
年収内訳：基本給（月）	月の基本給（単位：万円）	Monthly base salary (10,000 JPY units).
年収内訳：残業代（月）	月の残業代（単位：万円）	Monthly overtime pay (10,000 JPY units).
年収内訳：賞与（年）	年間賞与（単位：万円）	Annual bonus (10,000 JPY units).
年収内訳：その他の手当など（年）	その他の手当など（単位：万円）	Other annual allowances (10,000 JPY units).
NPS	この企業に転職・就職することを家族や友人にどれくらいおすすめするか。	Recommendation likelihood for working at the firm.

Table A.5 reports annual trends in the number of reviews and covered firms, together with the mean overall rating by year. The 2025 counts reflect reviews posted through October.

Table A.5: Temporal Trends in OpenWork Reviews

Year	**Avg Rating**	**N Reviews**	**N Companies**
2015	2.95	17,561	3,187
2016	2.96	34,072	4,280
2017	2.98	52,396	5,385
2018	2.99	70,942	6,227
2019	2.99	65,963	6,144
2020	3.03	56,132	6,014
2021	3.05	58,033	6,123
2022	3.04	63,088	6,459
2023	3.07	62,886	6,606
2024	3.11	68,382	6,827
2025	3.13	54,906	6,336

B Empirical Facts of OpenWork Reviews

Table B.6: OpenWork Dataset Overview

Panel (a): Sample Coverage		
Total reviews	604,361	
Unique companies	9,591	
Sample period	2015-03 to 2025-10	
Company-year observations	63,588	
Panel (b): Reviewer Demographics	**N**	**%**
Male	427,850	70.8
Female	176,511	29.2
Current employee	331,236	54.8
Former employee	273,125	45.2

Notes: Panel (a) summarizes the OpenWork review extract used in the analysis. "Total reviews" counts all review postings in the raw review-level file. "Unique companies" counts distinct firm identifiers in the OpenWork master file (listed firms and subsidiaries). "Sample period" is based on the review posting timestamp (year-month). "Company-year observations" counts unique firm-year observations with at least one posted review in that year, where the year is defined by the calendar year of the posting date. Panel (b) reports reviewer characteristics. Gender and employment status (current versus former employee) are required fields in the OpenWork review instrument; percentages are computed within each demographic item and sum to 100 within the male/female and current/former categories.

Table B.7: Rating Dimensions: Summary Statistics

Dimension	N	Mean	SD	P25	P75
Overall rating	604,361	3.04	0.66	2.50	3.50
Salary/benefits satisfaction	604,361	2.93	1.02	2.00	4.00
Employee motivation	604,361	2.88	0.92	2.00	3.00
Openness	604,361	3.09	0.94	2.00	4.00
Teamwork	604,361	3.12	0.97	2.00	4.00
Growth opportunities (young)	604,361	3.00	1.07	2.00	4.00
Long-term development	604,361	2.66	1.04	2.00	3.00
Compliance	604,361	3.71	1.07	3.00	5.00
Evaluation system	604,361	2.88	1.04	2.00	4.00

Notes: The table reports summary statistics for the OpenWork rating items at the review level. Each rating is recorded on a 1–5 scale (higher values indicate a more favorable evaluation) except "Overall rating." "Overall rating" is the platform's overall employer score, defined as the average of the eight sub-dimension ratings listed below. N is the number of reviews with a non-missing value for the item. P25 and P75 denote the 25th and 75th percentiles.

Table B.8: Text Field Completion Rates and Average Length

Text Field	Overall %	Male %	Female %	Average # of Characters
Organization structure/culture	71.3	75.3	61.4	112
Reason for joining	64.4	66.1	60.2	141
Work fulfillment	76.0	77.8	71.7	149
Environment for women	65.3	58.8	81.1	78
Work-life balance	74.1	73.5	75.6	92
Reason for quitting	57.2	56.2	59.8	98
Strength/weakness	47.7	52.4	36.2	120
Suggestions to CEO	20.8	23.6	14.2	84
Salary/evaluation	68.2	70.6	62.4	118

Notes: The table reports completion rates and average length for each OpenWork free-text prompt. "Overall %" is the share of reviews that contain a non-empty response in the corresponding text field; "Male %" and "Female %" report the same completion rate within reviewer gender. "Average # of Characters" is the mean number of characters in the field among completed responses (that is, conditional on the field being non-empty). Percentages are computed at the review level.

Table B.9: Reviewer Characteristics and Whether Text Fields Are Non-Empty

	Organization Structure/Culture	Reason for Joining	Work Fulfillment	Environment for Women	Work-Life Balance	Reason for Quitting	Strength/ Weakness	Suggestions to CEO	Salary/ Evaluation
Female	−0.105***	−0.055***	−0.054***	0.235***	0.037***	0.005**	−0.136***	−0.076***	−0.066***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)
Former Employee	0.014***	0.023***	0.017***	−0.004***	−0.047***	0.227***	−0.024***	−0.003**	−0.058***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.002)
Tenure: 20+ Years	0.120***	−0.001	0.067***	0.165***	0.002	−0.053***	0.138***	0.149***	0.039***
	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.005)	(0.004)	(0.003)	(0.003)
Tenure: 15–20 Years	0.078***	−0.054***	0.034***	0.111***	−0.006*	−0.027***	0.079***	0.086***	0.041***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Tenure: 10–15 Years	0.048***	−0.061***	0.027***	0.079***	−0.012***	−0.005*	0.039***	0.050***	0.043***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
Tenure: 5–10 Years	0.022***	−0.054***	0.015***	0.038***	−0.010***	0.009***	0.008***	0.018***	0.046***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Tenure: 3–5 Years	0.004**	−0.039***	0.008***	0.016***	−0.004**	0.012***	−0.005**	−0.000	0.036***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	604,359	604,359	604,359	604,359	604,359	604,359	604,359	604,359	604,359
R^2	0.070	0.033	0.035	0.080	0.027	0.090	0.058	0.052	0.047
Within R^2	0.015	0.005	0.005	0.047	0.004	0.052	0.020	0.015	0.010
Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The unit of observation is a review. Each column reports a separate regression where the dependent variable is an indicator equal to 1 if the corresponding text field contains any text and 0 otherwise. All specifications include firm and year-month fixed effects. Standard errors are clustered at the firm level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table B.10: Reviewer Characteristics and Review Text Completeness and Length

	Number of Non-Empty Fields	Total Characters (Sum Across Fields)
Female	−0.216***	−32.865***
	(0.008)	(0.806)
Former Employee	0.145***	−5.694***
	(0.006)	(0.643)
Tenure: 20+ Years	0.627***	25.858***
	(0.018)	(1.459)
Tenure: 15–20 Years	0.343***	19.644***
	(0.015)	(1.575)
Tenure: 10–15 Years	0.207***	19.459***
	(0.011)	(1.149)
Tenure: 5–10 Years	0.093***	16.429***
	(0.008)	(0.785)
Tenure: 3–5 Years	0.029***	9.109***
	(0.008)	(0.770)
Observations	604,359	604,359
R^2	0.046	0.040
Within R^2	0.008	0.006
Year-Month FE	Yes	Yes
Firm FE	Yes	Yes

Notes: The unit of observation is a review. Column (1) uses as the dependent variable the number of text fields that contain any text (across the eight review text fields). Column (2) uses the total number of characters summed across these fields. Both specifications include firm and year-month fixed effects. Standard errors are clustered at the firm level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table B.11: Male and Female Overall Ratings at the Firm-Year Level

	Female Overall Rating	Female Overall Rating
Male Overall Rating	0.319***	0.083***
	(0.009)	(0.010)
Observations	30,281	30,281
R^2	0.086	0.417
Within R^2	0.078	0.004
Year FE	Yes	Yes
Firm FE	No	Yes

Notes: The unit of observation is firm-year. The dependent variable is the firm-year average overall rating among female reviewers, and the main regressor is the firm-year average overall rating among male reviewers. Column (1) includes year fixed effects; column (2) includes year and firm fixed effects. Standard errors are clustered at the firm level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

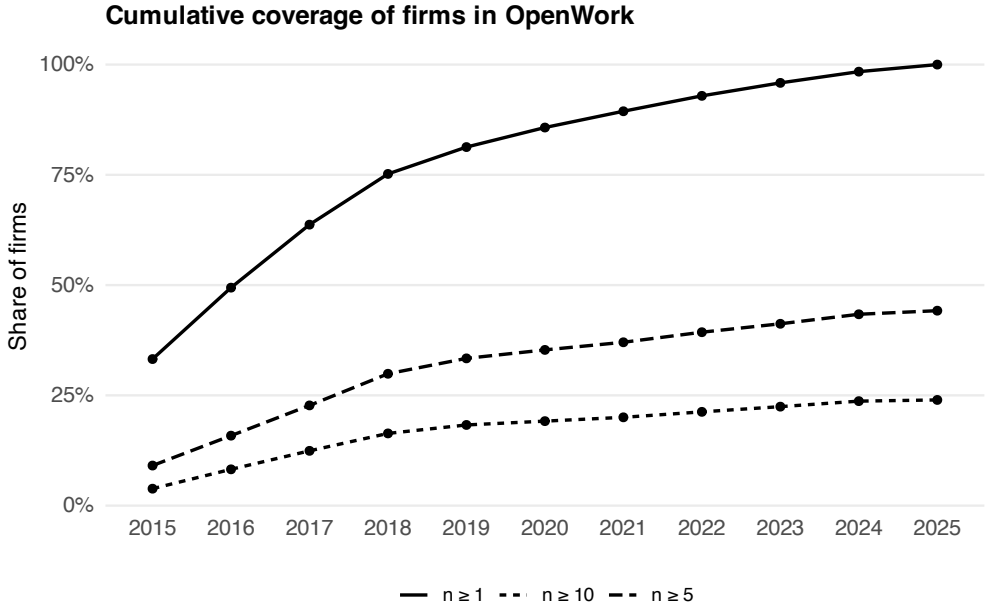


Figure B.1: Cumulative Coverage of Firms in OpenWork

Notes: The figure plots the cumulative share of firms that have reached at least n reviews in a given year, for $n \in \{1, 5, 10\}$. A firm is counted in year t if it has at least n reviews posted in year t or in any earlier year, and it remains counted thereafter. The sample includes listed firms and subsidiaries observed in our OpenWork extract.

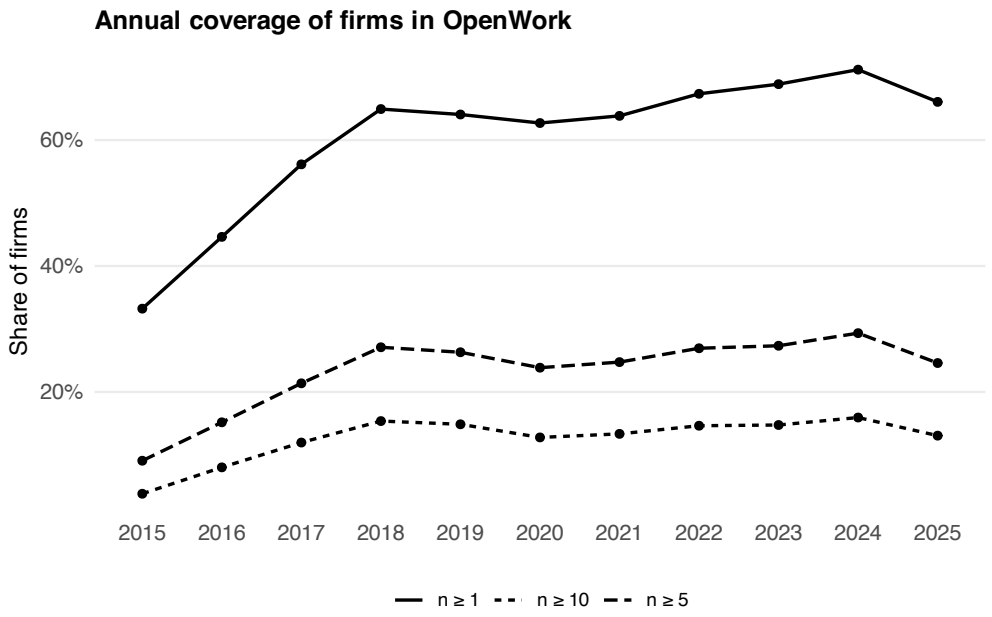
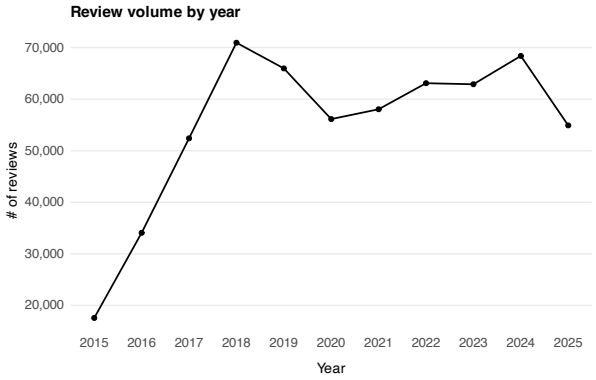
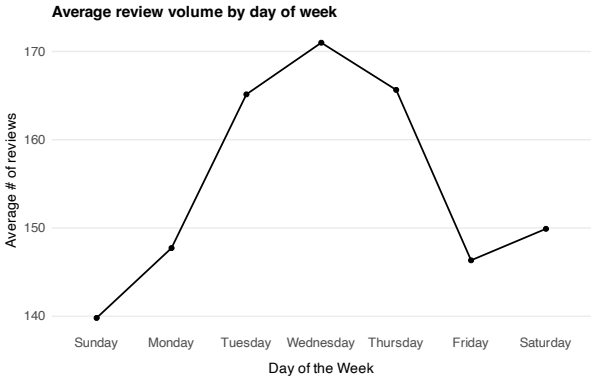


Figure B.2: Annual Coverage of Firms in OpenWork

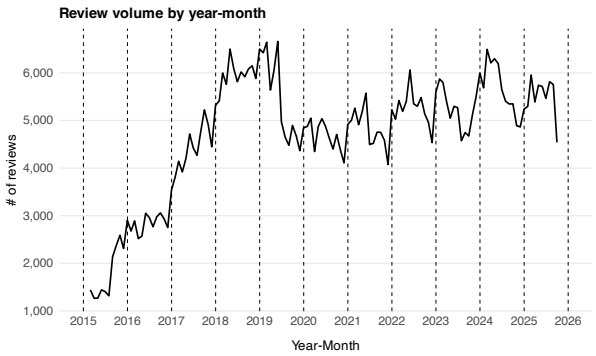
Notes: The figure plots, for each year, the share of firms with at least n reviews posted in that year, for $n \in \{1, 5, 10\}$. The sample includes listed firms and subsidiaries observed in our OpenWork extract.



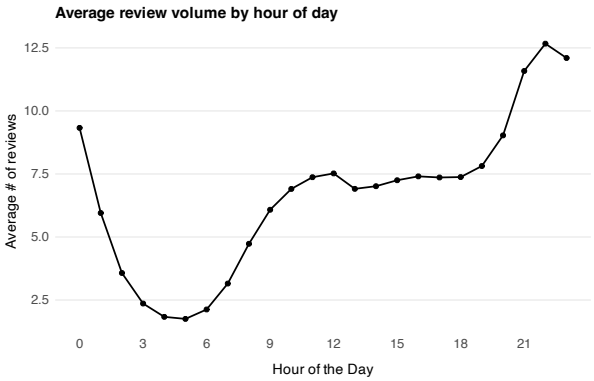
(a) Annual review volume



(b) Average daily volume by day of week



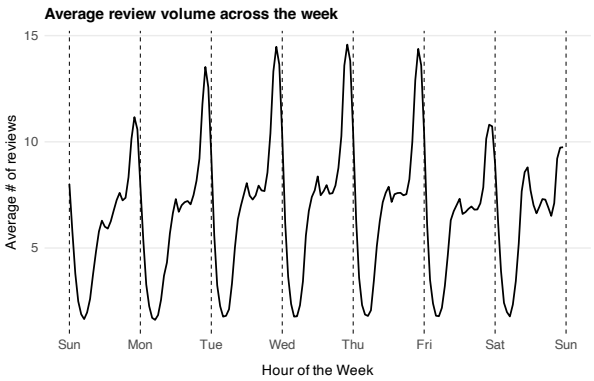
(c) Monthly review volume



(d) Average hourly volume by hour of day



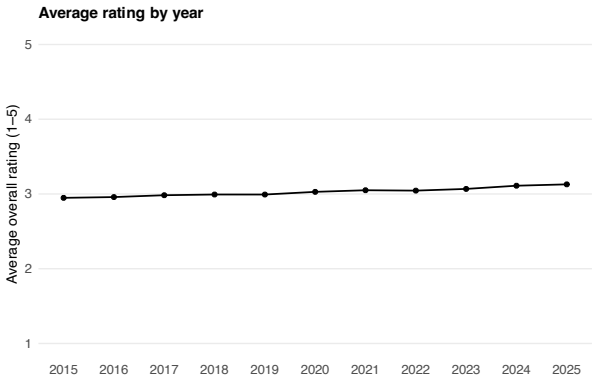
(e) Seasonality by calendar month



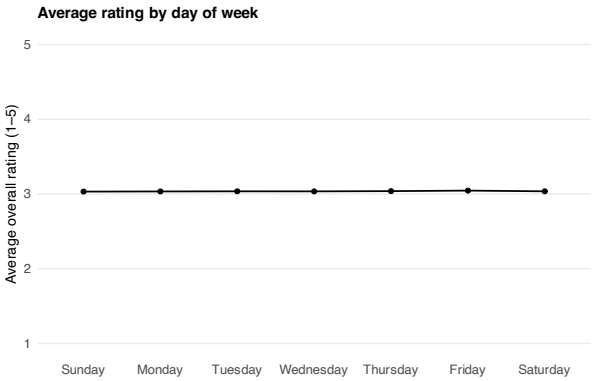
(f) Average hourly volume across the week

Figure B.3: Temporal Patterns in OpenWork Review Volume

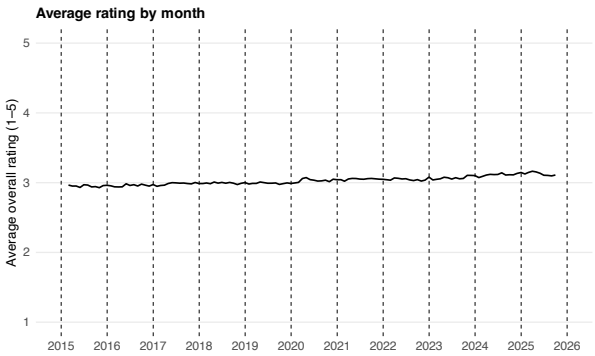
Notes: Panel (a) plots the number of reviews posted each year. Panel (b) reports the average number of reviews posted on a given calendar day, by day of week. Panel (c) plots the number of reviews posted each month; vertical dashed lines mark January of each year. Panel (d) reports the average number of reviews posted in a given hour of the day (averaged across calendar days). Panel (e) reports average monthly review volume by calendar month (January–December), averaged across years. Panel (f) reports the average number of reviews posted in each hour of a "typical week" (Sunday 0:00–Saturday 23:00), averaged across calendar days; vertical dashed lines mark 0:00 at the start of each day.



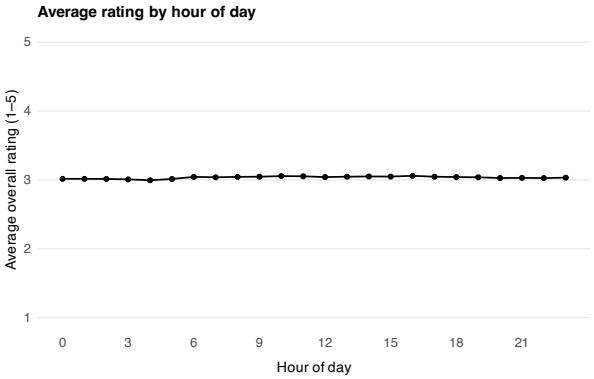
(a) Average overall rating by year



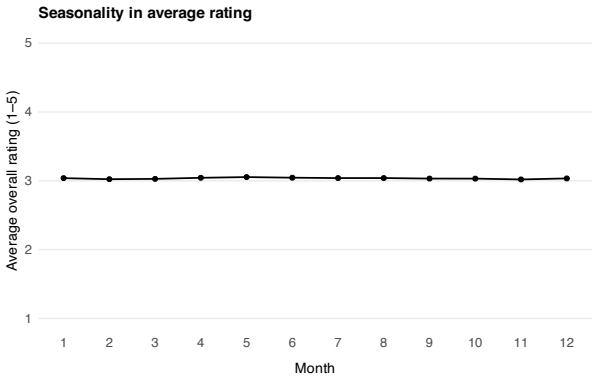
(b) Average overall rating by day of week



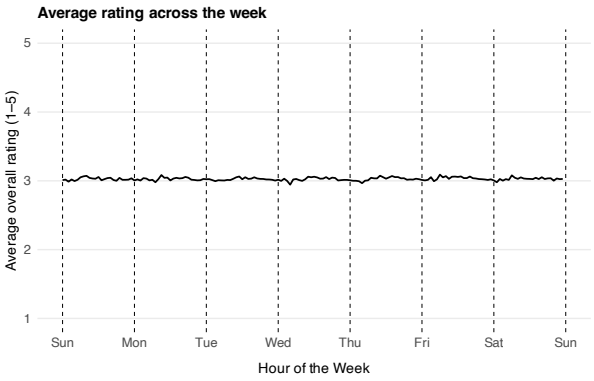
(c) Average overall rating by month



(d) Average overall rating by hour of day



(e) Seasonality in average rating



(f) Average overall rating across the week

Figure B.4: Temporal Patterns in OpenWork Overall Ratings

Notes: The figure reports averages of the overall employer rating (1–5). Panel (a) plots the average rating by year. Panel (b) plots the average rating by day of week. Panel (c) plots the average rating by year-month; vertical dashed lines mark January of each year. Panel (d) plots the average rating by hour of day (averaged across calendar days). Panel (e) reports average rating by calendar month (January–December), averaged across years. Panel (f) reports the average rating in each hour of a "typical week" (Sunday 0:00–Saturday 23:00), averaged across calendar days; vertical dashed lines mark 0:00 at the start of each day.

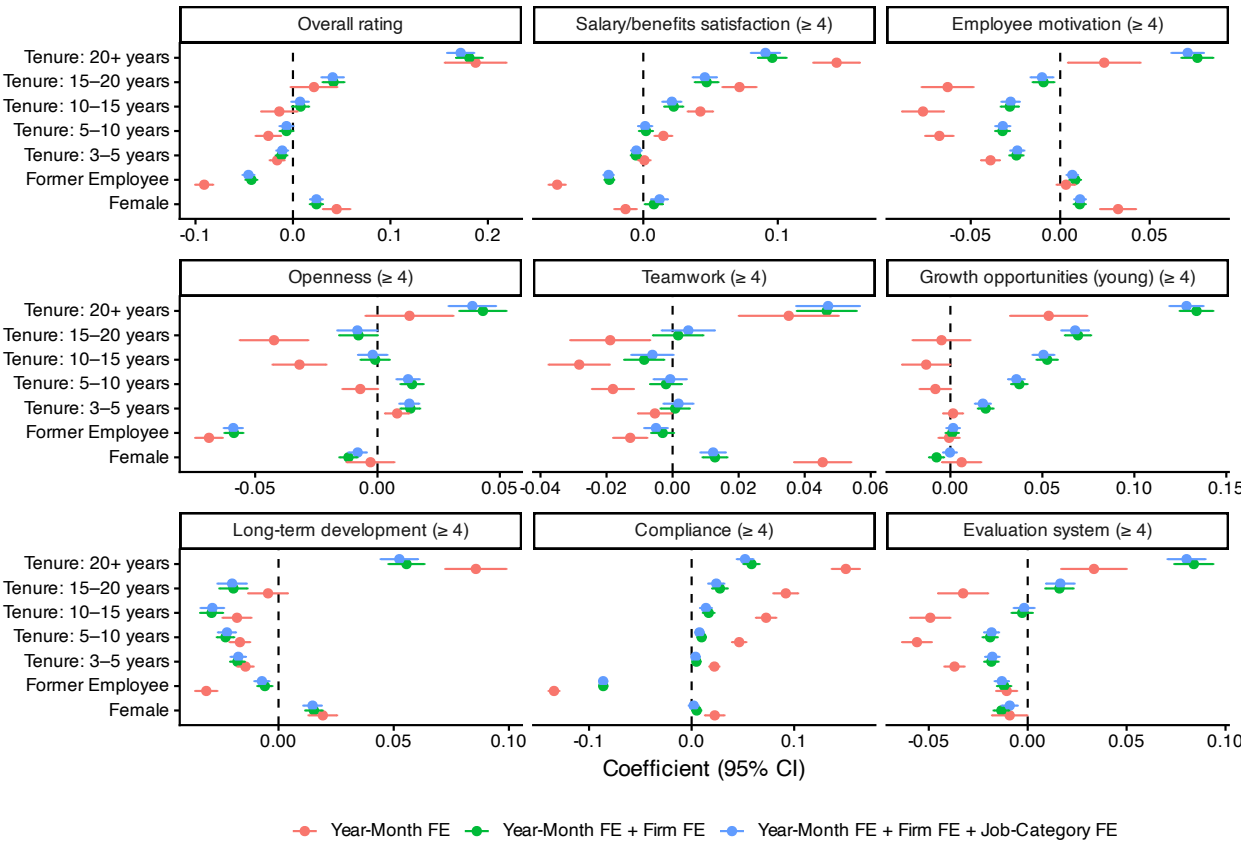
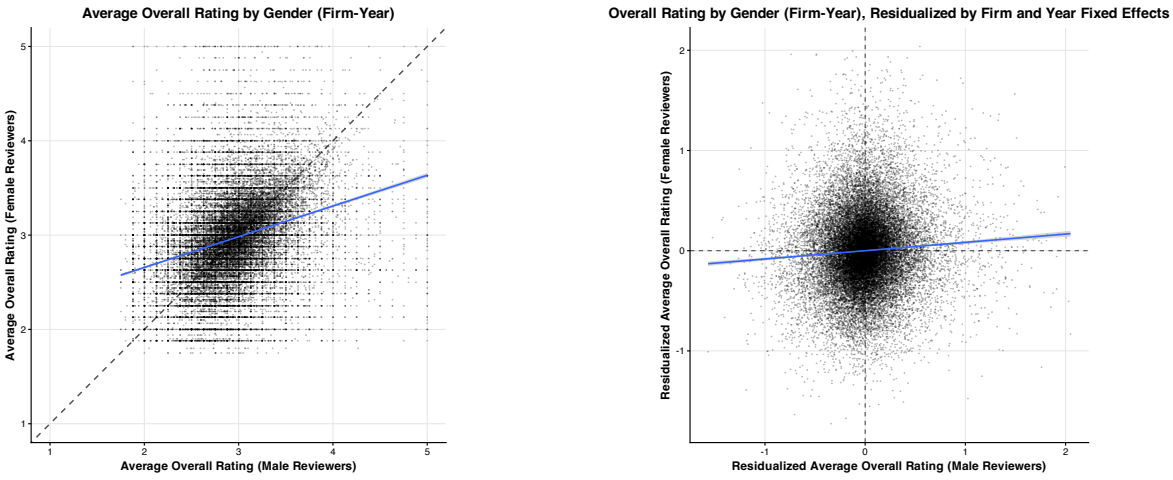


Figure B.5: Reviewer Characteristics and Employee Ratings

Notes: The figure plots coefficient estimates from review-level regressions of employee rating outcomes on reviewer characteristics. Outcomes include the overall rating (in levels) and indicator variables equal to one if the corresponding sub-rating is at least 4 (out of 5), i.e., positive. The key regressors are an indicator for female reviewers, an indicator for former employees, and job-tenure categories (with the omitted reference group given by the excluded tenure category in the specification). Points show coefficient estimates and horizontal bars show 95% confidence intervals. Each outcome is estimated under multiple fixed-effects specifications (as indicated in the legend), including year-month fixed effects and, in some specifications, firm and job-category fixed effects. Standard errors are clustered at the firm level.



(a) Raw firm-year averages

(b) Residualized by firm and year fixed effects

Figure B.6: Firm-Year Average Overall Ratings by Reviewer Gender

Notes: The unit of observation is a firm-year. The horizontal axis reports the mean overall rating (1–5) among male reviewers in a given firm-year, and the vertical axis reports the mean overall rating among female reviewers in the same firm-year. Panel (a) plots raw firm-year averages. Panel (b) plots residuals from regressions of each gender-specific firm-year average on firm and year fixed effects (i.e., within-firm, within-year deviations from the corresponding means). The dashed 45-degree line in Panel (a) indicates equality between male and female averages. Solid blue lines show best linear fits with 95% confidence intervals.

C Details of the Measurement Protocol

C.1 LLM Conversion Prompt and Example Output Placeholders

This appendix reports the exact instruction prompt used for the translation and segmentation step.

C.1.1 Prompt

```
You are an information extraction system.

TASK:
Translate the following employee review written in Japanese into English WITHOUT summarizing.
Classify the translated content into three categories ABOUT THE COMPANY:

1. Pros: Positive statements about the company.
2. Cons: Negative statements about the company.
3. Neutral: Descriptive or factual statements about the company that are neither positive nor negative.

REQUIREMENTS:
- Return ONLY valid JSON.
- JSON keys must be exactly: "pros", "cons", and "neutral".
- Values must be plain strings (not lists, not nested).
- Translate faithfully; do NOT omit details.
- If the review contains no content for a category, return an empty string.
- All content must be about the company, not the reviewer's personal situation unless explicitly tied to the company.
- Do NOT add labels, comments, headings, backticks, or any markdown (no ```json, no code fences).
- Do NOT include trailing commas in the JSON.
- Escape quotes properly so the JSON parses.

OUTPUT FORMAT (copy exactly):
{
  "pros": "...",
  "cons": "...",
  "neutral": "..."
}

REVIEW TEXT:
{review_text}
```

C.1.2 Output Example

Table C.12: Example Output from `gpt-5-nano` for LLM-Based Segmentation (Pros/Cons/Neutral)

Field	Output
Pros	I was greatly influenced by Sompo Japan, for better or worse. But I feel they should recognize that there aren't as many employees as Sompo Japan. Reason for joining: I wanted to do work that supports people in trouble. Benefits were well established.
Cons	The company may require a higher staffing level; in my department we were understaffed. Some departments may have issues with balancing sales and damage service staff. Senior colleagues seemed dissatisfied with their salaries. People who stay for a long time may find it a fit, but it didn't suit me. Overtime was paid properly, and the salary was decent, but there were complaints about pay among seniors.
Neutral	The job involves dealing with emotionally charged policyholders; overall, the company is aiming for growth with improved remuneration and work-life balance for those who are adaptable; if growth continues, significant hiring would be required. The company's reward structure paid for overtime and salary was reasonably provided.

Notes: This table reports a representative example of the structured JSON output returned by `gpt-5-nano` when segmenting a single employee review into `pros`, `cons`, and `neutral` fields.

C.2 Expanded Dictionary

Table C.13: Top 50 Words in the Expanded Dictionary

Rank	Word	Similarity	Rank	Word	Similarity
1	culture	0.728	26	spirit	0.438
2	relationship	0.612	27	relationships	0.437
3	atmosphere	0.568	28	ethos	0.431
4	hierarchy	0.565	29	approach	0.426
5	climate	0.534	30	environment	0.420
6	organizational_structure	0.529	31	customer-first_mindset	0.419
7	temperament	0.524	32	relations	0.418
8	mindset	0.520	33	hierarchies	0.409
9	organization	0.518	34	strict_hierarchy	0.407
10	management	0.517	35	management_layer	0.406
11	mentality	0.514	36	absolute_authority	0.400
12	governance	0.500	37	rapport	0.399
13	attitude	0.484	38	posture	0.398
14	pyramid_structure	0.481	39	bottom-up_approach	0.395
15	stance	0.473	40	logic	0.394
16	vibe	0.471	41	senior-junior_hierarchy	0.393
17	structure	0.470	42	upper_management	0.393
18	vertical_hierarchy	0.463	43	customer-centric_approach	0.391
19	rigid_hierarchy	0.455	44	alignment	0.391
20	tradition	0.455	45	discipline	0.389
21	top-down_approach	0.454	46	autocratic_leadership	0.389
22	leadership	0.450	47	parent-subsidiary_relationship	0.388
23	communication	0.445	48	collaboration	0.386
24	image	0.442	49	family-like_atmosphere	0.384
25	character	0.438	50	job-for-life	0.384

Notes: The table lists the top 50 tokens in the expanded dictionary for the seed-word organizational-capital measure. Similarity is the cosine similarity between each token's Word2Vec embedding and the seed representative vector v_R, defined as the average of the embeddings of the five seed words. Tokens are ranked in descending order of similarity. Bigrams detected using Gensim's `Phrases` algorithm in the corpus are shown with underscores.

C.3 Synthetic-review Construction Prompt and Output

In this Appendix, we report the exact prompt used to generate the synthetic reviews. This prompt is identical to that in Cai et al. (2025b): it instructs the model to produce 100 positive reviews (strong OC) and 100 negative reviews (weak OC). We also provide illustrative examples of the generated outputs.

C.3.1 Prompt

This is the definition of organizational capital.
Organizational capital encompasses any intangible firm asset with four properties: (1) it affects firm performance; (2) it changes slowly over time; (3) being intangible, it is not perfectly observable; and (4) it must be produced at least partly inside the firm with the active participation of the firm's top management.
What could organizational capital be in practice? A leading example is the management practices analyzed by Bloom and Van Reenen (2007), which arguably affect firm performance (as in condition 1) and are slow-moving (as in condition 2). In support of the imperfect observability condition (condition 3), note that management practices are difficult to measure precisely (Bloom et al. [2019] estimate that 45% of the observed variance in management scores is due to measurement error). For condition 4, Simons (1994b) uses 10 years of observational data collected in over 50 US businesses to document how top managers use control systems—namely, mechanisms for influencing human endeavor within the company —to maintain or alter patterns in organizational activities; in particular, new CEOs use the first 18 months of their tenure to define and measure critical performance variables (Simons 1994a). There are other possible examples of organizational capital. Our definition includes at least partial constructs, such as relational contracts (Baker, Gibbons, and Murphy 2002), corporate culture (Schein 2010), firm-specific human capital (Prescott and Visscher 1980), or firm capabilities (Teece, Pisano, and Shuen 1997).
Based on the above definition, please give me 100 employee reviews of their company which indicate that their company has strong organizational capital, and 100 employee reviews of their company which indicate that their company does not have strong organizational capital. Each review is one or two sentences long. The language should be in the style of an employee's review of their company.

C.3.2 Example outputs

Table C.14: Ten Sample Synthetic Reviews (Strong Organizational Capital)

Leadership provides a clear vision and follows through, and everyone knows how their work contributes to long-term goals.
Management practices are consistent and transparent, and decisions are explained thoughtfully.
The company invests heavily in internal training and long-term capability building.
We have strong cross-department communication, so projects run smoothly without confusion.
Management metrics are stable and well-defined, which helps us focus on what matters most.
Cultural norms here are clear, supportive, and widely shared across teams.
The company has a deliberate onboarding process that sets up employees for long-term success.
Leadership listens carefully and acts on employee feedback rather than ignoring it.
Projects are planned with realistic expectations, and leadership sticks with commitments.
Internal systems and processes make it easy to do great work without wasted time.

Table C.15: Ten Sample Synthetic Reviews (Weak Organizational Capital)

Leadership changes direction every week, and we never know what matters.
There are no clear goals and priorities constantly shift.
We don't have formal processes—everything is improvised last minute.
Management rarely communicates and decisions seem random.
The culture is chaotic and nobody is aligned on strategy.
We reinvent the wheel constantly because nothing is documented.
Leadership doesn't listen and feedback goes into a black hole.
Metrics change constantly and feel meaningless.
Processes are inconsistent across teams, leading to constant confusion.
Promotion criteria aren't clear and seem political rather than merit-based.

D Direct LLM Rater Details and Validation

The prompt used in the direct LLM-based measure (English version) begins with the definition of organizational capital, which is identical to the one reported in Appendix C.3, followed by the task instructions below.

D.1 Prompt (Latter Part)

TASK: Based ONLY on the employee review text provided below, rate how much the review indicates the company has strong organizational capital on a scale of 0 to 100.

EMPLOYEE REVIEW: {review_text}

OUTPUT REQUIREMENTS: Return ONLY valid JSON with exactly these fields:

- "oc_score": integer from 0 to 100
- "reasoning": brief explanation (1-2 sentences) of your rating

D.2 LLM-as-Rater Validation

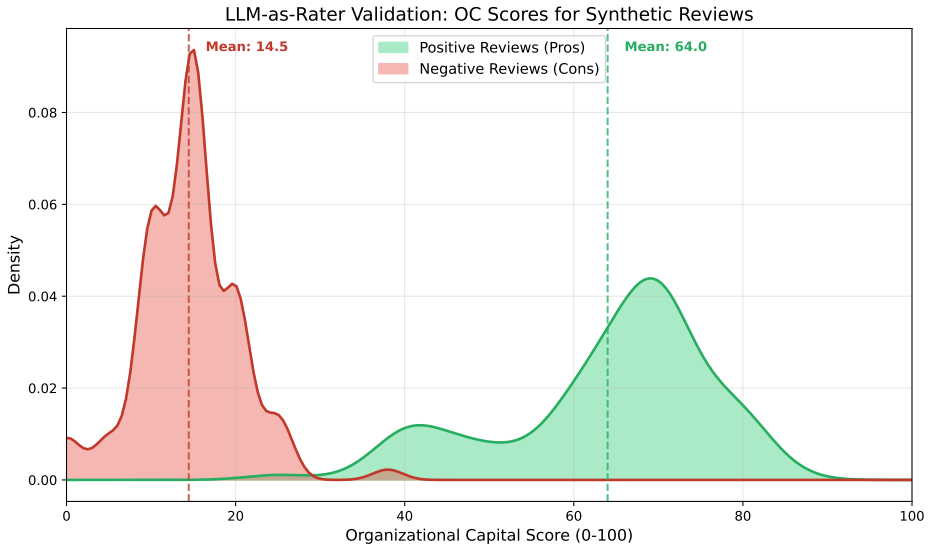


Figure D.7: LLM-as-Rater Validation: OC Scores for Synthetic Reviews

Notes: Density distributions of OC scores for 200 synthetic reviews (100 positive, 100 negative) generated by ChatGPT and rated by gpt-5-nano using the same prompt as the main analysis.

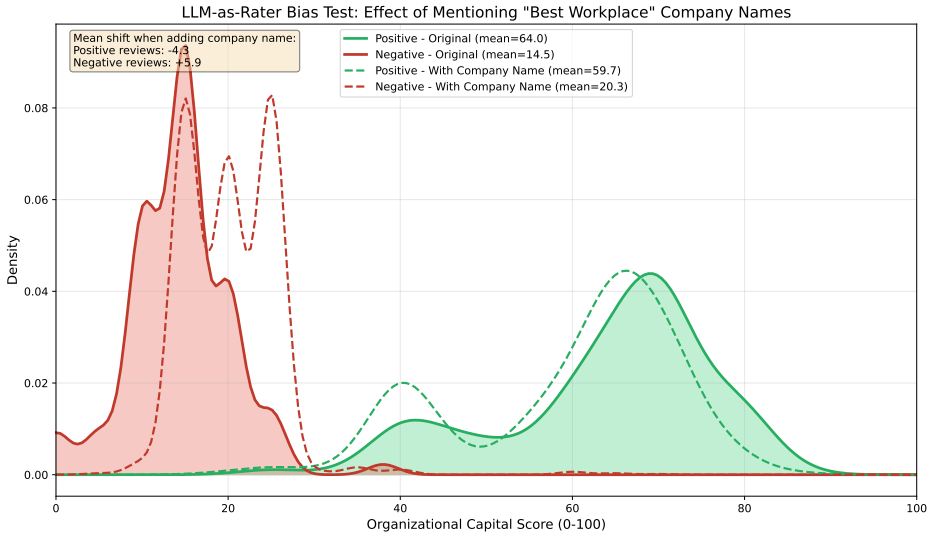


Figure D.8: LLM-as-Rater Validation: Company Name Effect

Notes: This figure tests whether mentioning a company name influences LLM-assigned OC scores. The solid lines show the original density distributions from Figure D.8. The dashed lines show distributions when each of 100 "Best Workplaces in Japan 2025" large company names is prepended to the same reviews (3,000 ratings: 15 large companies with 200 reviews).

E Alternative Measurements

Table E.16: Cross-Protocol Correlations: Japanese Pipeline vs. Original Protocol

	OC (Seed)	OC (ChatGPT)	OC (Avg.)	OC (Seed, %)	OC (ChatGPT, %)	OC (Avg., %)
OC (Seed) [JPN]	0.44***	0.32***	0.45***	0.40***	0.30***	0.41***
OC (ChatGPT) [JPN]	0.32***	0.41***	0.44***	0.31***	0.39***	0.41***
OC (Avg.) [JPN]	0.44***	0.42***	0.51***	0.41***	0.39***	0.47***
OC (Seed, %) [JPN]	0.41***	0.30***	0.42***	0.42***	0.31***	0.43***
OC (ChatGPT, %) [JPN]	0.30***	0.39***	0.41***	0.31***	0.39***	0.41***
OC (Avg., %) [JPN]	0.40***	0.40***	0.47***	0.41***	0.40***	0.48***

Notes: Entries report Pearson correlations between the Japanese-pipeline measures (rows) and the original-protocol measures (columns). Significance levels are based on two-sided tests of the null hypothesis of zero correlation. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table E.17: Organizational Capital and Firm Performance (ROA) (Japanese Pipeline)

	Dependent variable: ROA (%)					
OC (Seed)	0.622***					
	(0.096)					
OC (ChatGPT)		0.809***				
		(0.095)				
OC (Avg.)			0.929***			
			(0.114)			
OC (Seed, %)				0.022***		
				(0.003)		
OC (ChatGPT, %)					0.031***	
					(0.003)	
OC (Avg., %)						0.035***
						(0.004)
log(Total Assets)	0.711***	0.709***	0.704***	0.705***	0.700***	0.694***
	(0.116)	(0.116)	(0.116)	(0.116)	(0.115)	(0.115)
Firm Age	−0.044***	−0.042***	−0.042***	−0.044***	−0.041***	−0.042***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Observations	23,907	23,907	23,907	23,907	23,907	23,907
R^2	0.217	0.218	0.219	0.217	0.220	0.220
Industry-Year FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is return on assets (ROA, percent). Each column reports a separate regression of ROA on one organizational capital (OC) measure constructed under the Japanese pipeline, controlling for the logarithm of total assets and firm age, with industry-by-year fixed effects. OC (Seed), OC (ChatGPT), and OC (Avg.) are standardized (z-score) firm-year measures; OC (Avg.) is the simple average of the two standardized measures. The percentile measures ("%") are firm-year percentile ranks (1–100). Standard errors are clustered at the firm level and reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table E.18: Organizational Capital and Firm Performance (P/B Ratio) (Japanese Pipeline)

	Dependent variable: P/B Ratio					
OC (Seed)	0.190***					
	(0.050)					
OC (ChatGPT)		0.252***				
		(0.062)				
OC (Avg.)			0.286***			
			(0.063)			
OC (Seed, %)				0.007***		
				(0.002)		
OC (ChatGPT, %)					0.009***	
					(0.002)	
OC (Avg., %)						0.011***
						(0.002)
log(Total Assets)	−0.178***	−0.178***	−0.180***	−0.180***	−0.181***	−0.183***
	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)
Firm Age	−0.037***	−0.036***	−0.036***	−0.037***	−0.036***	−0.036***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	24,052	24,052	24,052	24,052	24,052	24,052
R^2	0.190	0.190	0.190	0.190	0.190	0.190
Industry-Year FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is the price-to-book (P/B) ratio. Each column reports a separate regression of P/B on one organizational capital (OC) measure constructed under the Japanese pipeline, controlling for the logarithm of total assets and firm age, with industry-by-year fixed effects. OC (Seed), OC (ChatGPT), and OC (Avg.) use standardized (z-score) firm-year measures; OC (Avg.) is the simple average of the two standardized measures. The percentile measures ("%") are firm-year percentile ranks (1–100). Standard errors are clustered at the firm level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table E.19: Variance Decomposition of Organizational Capital (Japanese Pipeline)

No.	Fixed Effects	OC (Seed Word)	OC (ChatGPT)
1	Year Fixed Effects	0.1%	0.1%
2	Industry Fixed Effects	4.6%	5.5%
3	Industry Fixed Effects + Year Fixed Effects	4.7%	5.7%
4	Industry × Year Fixed Effects	20.5%	20.9%
5	Firm Fixed Effects	25.5%	26.8%
6	Firm Fixed Effects + Year Fixed Effects	25.5%	27.0%
7	CEO Fixed Effects	32.2%	34.3%
8	CEO Fixed Effects + Year Fixed Effects	32.2%	34.4%
9	CEO–Firm Fixed Effects	32.4%	34.5%
10	CEO–Firm Fixed Effects + Year Fixed Effects	32.4%	34.6%

Notes: Each entry reports the overall R^2 from a regression of firm-year organizational capital constructed under the Japanese pipeline on the listed fixed effects only (no covariates). The Seed Word and ChatGPT measures are standardized (z-scores). Values are shown as percentages of total variation explained by the fixed-effects structure. Observations dated before the CEO inauguration year are assigned a firm-specific placeholder CEO.

Table E.20: Correlations between Organizational Capital Measures and Review Scores (Japanese Pipeline)

	OC (Seed)	OC (ChatGPT)	OC (Avg.)	OC (Seed, %)	OC (ChatGPT, %)	OC (Avg., %)	Overall Rating	Compensation Rating	Motivation Rating	Openness Rating	Teamwork Rating	Chance to Grow	Long-term Development	Compliance Rating	Employee Evaluation System
OC (Seed)	1.00														
OC (ChatGPT)	0.53***	1.00													
OC (Avg.)	0.88***	0.87***	1.00												
OC (Seed, %)	0.87***	0.51***	0.79***	1.00											
OC (ChatGPT, %)	0.49***	0.94***	0.81***	0.52***	1.00										
OC (Avg., %)	0.78***	0.83***	0.92***	0.88***	0.87***	1.00									
Overall Rating	0.37***	0.35***	0.41***	0.38***	0.34***	0.41***	1.00								
Compensation Rating	0.19***	0.14***	0.19***	0.18***	0.12***	0.17***	0.64***	1.00							
Motivation Rating	0.31***	0.33***	0.37***	0.32***	0.34***	0.38***	0.74***	0.36***	1.00						
Openness Rating	0.28***	0.24***	0.30***	0.29***	0.22***	0.30***	0.68***	0.30***	0.45***	1.00					
Teamwork Rating	0.25***	0.25***	0.29***	0.25***	0.25***	0.29***	0.67***	0.24***	0.51***	0.50***	1.00				
Chance to Grow	0.28***	0.29***	0.33***	0.29***	0.29***	0.33***	0.68***	0.27***	0.54***	0.43***	0.38***	1.00			
Long-term Development	0.28***	0.27***	0.31***	0.28***	0.26***	0.31***	0.74***	0.45***	0.45***	0.39***	0.42***	0.45***	1.00		
Compliance Rating	0.15***	0.09***	0.13***	0.14***	0.07***	0.12***	0.53***	0.36***	0.22***	0.25***	0.25***	0.11***	0.38***	1.00	
Employee Evaluation System	0.26***	0.28***	0.31***	0.26***	0.28***	0.31***	0.64***	0.39***	0.48***	0.32***	0.30***	0.44***	0.39***	0.13***	1.00

Notes: Entries report pairwise correlations. The diagonal elements are 1.00 by construction and the upper triangle is omitted. OC (Seed), OC (ChatGPT), and OC (Avg.) use standardized (z-score) firm-year measures; OC (Avg.) is the simple average of the two standardized measures. The percentile measures ("%") are firm-year percentile ranks (1–100). Significance levels are based on two-sided tests of the null hypothesis of zero correlation. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. [(Japanese Pipeline)]

Table E.21: Correlation of Organizational Capital Measures

	Seed (Eng) (1)	Seed (Jpn) (2)	ChatGPT (Eng) (3)	ChatGPT (Jpn) (4)	LLM E→E (5)	LLM J→E (6)	LLM J→J (7)
Panel A: Review Level ($N = 604,360$)							
(1) Seed (Eng)	1.000						
(2) Seed (Jpn)	0.419	1.000					
(3) ChatGPT (Eng)	0.405	0.296	1.000				
(4) ChatGPT (Jpn)	0.309	0.511	0.386	1.000			
(5) LLM Eng→Eng	0.433	0.401	0.305	0.400	1.000		
(6) LLM Jpn→Eng	0.449	0.410	0.300	0.413	0.793	1.000	
(7) LLM Jpn→Jpn	0.410	0.373	0.266	0.370	0.731	0.789	1.000
Panel B: Firm-Year Level ($N = 6,451$; $\geq$ 20 reviews)							
(1) Seed (Eng)	1.000						
(2) Seed (Jpn)	0.650	1.000					
(3) ChatGPT (Eng)	0.572	0.574	1.000				
(4) ChatGPT (Jpn)	0.524	0.724	0.626	1.000			
(5) LLM Eng→Eng	0.660	0.586	0.487	0.486	1.000		
(6) LLM Jpn→Eng	0.677	0.597	0.483	0.494	0.952	1.000	
(7) LLM Jpn→Jpn	0.666	0.603	0.472	0.509	0.922	0.940	1.000

Notes: Pairwise Pearson correlations. Panel A reports correlations at the review level. Panel B reports correlations at the year level, restricted to firm-years with at least 20 reviews. Seed = CPY seed-word dictionary measure; ChatGPT = synt review similarity; LLM = direct LLM rating (0–100 scale). Eng/Jpn indicates input text language; →Eng/→Jpn indicates pr language.

Table E.22: Top 50 words in the Expanded Dictionary (Japanese)

Rank	Word	Similarity	Rank	Word	Similarity
1	性	0.602	26	一体 _ 感	0.364
2	経営	0.547	27	姿勢	0.364
3	文化	0.516	28	インテグリティ	0.363
4	カルチャー	0.495	29	理念	0.363
5	風土	0.489	30	ガバナンス	0.362
6	思想	0.451	31	戦略	0.360
7	ポリシー	0.447	32	性質	0.358
8	リレーション	0.434	33	リレーションシップ	0.356
9	哲学	0.430	34	エコシステム	0.356
10	アイデンティティ	0.412	35	構造	0.354
11	マインド	0.405	36	価値 _ 観	0.352
12	社風	0.400	37	サプライ _ チェーン	0.351
13	収益 _ 構造	0.400	38	M_A	0.351
14	体質	0.398	39	資源 _ 配分	0.351
15	ビジネス _ モデル	0.393	40	ブランディング	0.351
16	統治	0.390	41	オーナー	0.351
17	求心 _ 力	0.381	42	共創	0.350
18	発想	0.375	43	DNA	0.349
19	気風	0.374	44	マインド _ セット	0.349
20	パートナーシップ	0.372	45	政策	0.347
21	ビジョン	0.372	46	再生	0.347
22	倫理 _ 観	0.369	47	ｂｔｏｂ	0.347
23	信念	0.369	48	ＭＶＶ	0.345
24	思考	0.366	49	創業	0.344
25	コネクション	0.366	50	原理 _ 原則	0.344

Notes: The expanded dictionary is constructed from Word2Vec embeddings trained on the Japanese review corpus after Japanese-specific pre-processing. Text is morphologically anazlyzed and tokenized using MeCab with the UniDic dictionary. Similarity is the cosine similarity between each token's embedding and the Japanese seed-word representative vector. Bigrams detected using Gensim's `Phrases` algorithm in the corpus are shown with underscores.

Table E.23: Divergent Reviews: Seed-Word Measure

Panel A: High English Score (81%) / Low Japanese Score (19%)	
English Pros/Cons	Japanese Pros/Cons
Pros: Stable by old Japan standards. For better or worse, it has not changed for a long time. Reason for joining: it was a company with a high average salary that even beginners could enter. Strong points: since the company is registered with Udemy, you can receive a certain amount of training for free. They also cover the costs for qualifications and obtaining them. It is 充実している. Because there are many women, I think it is an easy workplace for women. In the department I worked in, coordination was easy and I could take time off freely. Strengths: they have a track record of selling their own packaged products to municipalities with a stable demand, so sales are stable. **Cons:** There are parts that are ブラック体質 in some departments depending on the department and season. Job is not much due to seniority-based system, and there are many older people, so you don't feel motivated. Weaknesses: because municipalities are the main customers, they are slow to adopt new technologies. Many people feel that the pay does not match the work. A fundamental reform of the salary system seems necessary. Salary system: almost uniform raises every year. Raises vary with effort, but within the margin of error. Whether you work hard or not, the amount is almost the same. Evaluation system: seems proper, but unless you achieve significant accomplishments, you only rise within a margin of error.	**Pros:** 平均給与は確かに高いが、給与体系は毎年ほぼ一律で働きぶりによって昇給具合は変わるが誤差の範囲。資格補助金や資格の取得に必要な費用も出してくれる。Ｕｄｅｍｙに会社で登録しているのである程度の研修は無料で受けることができる。充実している。女性が多いので、働きやすい職場ではあると思う。私が働いていた部署は調整が効きやすく、自由に休むことができた。強み：安定志向の自治体に対して自社パッケージ製品を売り込んでいる実績があるので、売り上げが安定している。**Cons:** 平均給与は高いが、平均年齢も高く、若手のうちの給料は安いし上がらない。部署や時期によるところが大きく部署によってはブラック体質な箇所が一部存在する。年功序列制で仕事はあまりできないが年齢が高い人が多く、働きがいが感じられない。自治体がメイン顧客なのも相まって新しい技術の導入に疎い。給料のわりに働きが見合っていない人が多すぎる。給与制度：頑張っても頑張らなくてもほぼ同じ。評価制度：よほどの功績を残さない限り誤差の範囲でしか上がらない。
Panel B: High Japanese Score (83%) / Low English Score (17%)	
English Pros/Cons	Japanese Pros/Cons
Pros: After the first year, the position changes to a candidate for store manager, and everyone aims to become a store manager. It is rewarding to think for yourself about how to run the store and plan the sales floor as you learn store operations. There are many female employees who actively work; in cosmetics and customer service, I think women are better suited. Part-time workers are mostly women, and men feel somewhat out of place by contrast. Most stores close at 21:00, so if the work is finished, you can go home by 21:00 even if late. Basic shifts are arranged as morning and evening shifts, but if the store is understaffed, you may work straight through. **Cons:** The company pressures employees to make as many store managers as possible, and those who do not wish to become managers or lack confidence quit due to the pressure. There are many company directives and many mistakes in the Product Department, causing unfinished tasks and overtime without pay, leading to employee dissatisfaction that is hard to resolve. The company tells employees not to do unpaid overtime, but because work cannot be finished, there is no choice but to do it. When I asked my supervisor to allow overtime, it was denied, so I clock in and finish the work before leaving. The workload keeps increasing, and unpaid overtime seems likely to rise, so I want to quit.	**Pros:** 1 年目が終わると店長候補という扱いに変わり、全員店長を目指すようになります。店舗運営を学んでいくので自分で考え店を運営したり売場を考えていくのはやりがいがありますが、女性社員も多く活躍しています。化粧品などの接客では女性の方が向いていると思います。パートさんも女性がほとんどで、逆に男性の方が肩身が狭いです。営業時間が 21 時までの店がほとんどなのでやるべきことが終わっていれば遅くても 21 時には帰れます。基本は早番と遅番でシフトを組んでいますが人がいない店だと通しで出ることになります。**Cons:** 会社が店長を増やしたいとの事で店長にできるだけ早くしようと迫ってきます。店長になるのを望んでない者や自信が無い人はそれがプレッシャーになって辞めていきます。会社からの指示も多く、商品部のミスも多い為、作業が終わらずサービス残業したり、従業員からの不満も出るため解消していくのが大変です。残業させてほしいと上司に相談したが却下されたのでタイムカード押してから作業を終わらせて帰ります。作業はどんどん増えていく感じがしてサービス残業がもっと増えそうなので辞めたいです。

Notes: Reviews with divergence $\approx \pm 2$ standard deviations and average language percentile ≈ 50. English is LLM-translated.

Table E.24: Divergent Reviews: ChatGPT Similarity Measure

Panel A: High English Score (82%) / Low Japanese Score (18%)	
English Pros/Cons	Japanese Pros/Cons
Pros: Compared to other domestic banks, there is more freedom here, but it still has the feel of a large Japanese company where carrying out tasks without errors is demanded above all else. Reason for joining: for career advancement from sales. Also, since talent outflow from the sell-side to the buy-side is becoming common, and will likely increase in the future, I decided to move early. **Cons:** You cannot suddenly do fund manager duties; if you want to become a fund manager right away, you should build a track record at another company before moving. The job offers strong learning in asset management basics. Training system is solid. Most employees do not leave even after getting married. Vacations are easy to take and overtime can be adjusted, which is very good. In two years you can grasp most things, so if you are aiming for higher compensation, it is better to move outside. If you stay for ten years, it becomes difficult to move outside, so it is better to leave while you are young.	**Pros:** 入社を決めた理由：営業職からのキャリアアップのため。働きがい：アセットマネジメントの基礎が学べる。成長・キャリア開発：研修制度がしっかりしている。おそらく良い。結婚してもやめない社員がほとんど。休暇はとりやすく、残業も調整できるのでとてもよい。二年もあれば大抵のことは掴めるので、更なる報酬を目指すなら外に出たほうが良い。十年選手になってしまうとなかなか外には出にくくなってしまうため、若いうちに出たほうが良い。強み：預かり資産が国内トップクラス。またグループの証券会社の個人向けの販売力が強い。**Cons:** いきなりファンドマネージャー業務などは当然やれないので、いきなりファンドマネージャーをやりたい人は他社でトラックレコードを作ってから移るべき。弱み：証券会社からの影響を受けやすい。オリジナリティがない。受け身体制の仕事になりがち。運用商品がコモディティ化しているのと親会社からの影響を受けるので、リーディングカンパニーとして新しいことに挑戦して欲しい給与制度：基本的には年功序列
Panel B: High Japanese Score (82%) / Low English Score (18%)	
English Pros/Cons	Japanese Pros/Cons
Pros: Being a large company, the team you are assigned to varies greatly, but overall there are many people with high motivation. I was in a relatively foundation-oriented team, so I was able to focus on technical matters. **Cons:** Depending on the department, there were times when it was hard to make decisions like stopping something or halting system development, and there were cases where I worked on systems whose users were unknown. I left because I did not get along with the manager; it was a management style that did almost no management, the evaluations were sloppy and there was no sense of satisfaction.	**Pros:** 大企業なので配属チームによって大きく変わりますが総じてモチベーションが高めの人が多い印象があります。比較的基盤側のチームにいたので技術的なことに集中できていました。入社を決めた理由：ソフトウェアエンジニアとして技術的なことに集中できそうだと感じたので入社しました。**Cons:** 部署次第ですが何かをやめる・システム開発を中止するみたいな決断がなかなかできず誰に使われるかわからないシステムの開発もすることがありました。基盤開発をするチームに入りましたが社内でほとんど使われていませんでした。マネージャーと合わなくて退職しました。マネージメントをほとんどしないタイプの人で評価などが雑で納得感が全くなかったので退職しました。チームで開発したものが社内でほとんど使われていなかったため、チームの雰囲気が悪く同僚たちのモチベーションがかなり低くなっていたのも退職した理由のひとつです。

Notes: Reviews with divergence $\approx \pm 2$ standard deviations and average language percentile ≈ 50. English is LLM-translated.

Table E.25: Divergent Reviews: LLM Direct Rating

Panel A: High English Score (75%) / Low Japanese Score (32%)			
English Review	Japanese Review	English Reasoning	Japanese Reasoning
There is still a tendency toward power harassment. Motivation / Job satisfaction: I think there is a sense of purpose. If it's a task the company is actively addressing, it's fine. However, once you become a director, balancing private life and work may be difficult. As you rank up, there are people who must be reachable at any time. There is consideration for women, and it seems that they can properly take maternity leave and childcare leave. Compensation system: Transparent and good. Evaluation system: Transparent and good.	パワハラ気質がまだ残ってる働きがい：やりがいはあると思う。社として取り組んでいる事務ならば大丈夫。しかしディレクターになるとプライベートとの調節は難しいかもしれません。役職が上がると何時でも連絡を受けなければならない人もいます。女性には配慮があり、産休や育児休暇などはちゃんと取得できるようです。給与制度：明朗で良い評価制度：明朗で良い	**Score: 52** The review notes transparent compensation and evaluation systems and some consideration for work-life balance, which reflect partial internal systems; however, concerns about power harassment and the implication that top leadership influence on practices is not clearly evidenced, suggesting moderate but not strong organizational capital.	**Score: 35** レビューは組織資本の一部要素（評価制度の明朗性、育児休暇の取得容易さ、組織としての取り組み）があることを示すが、パワハラ気質の残存や上位職での連絡頻度の要求は組織文化・リーダーシップの健全性に欠け、観察困難性を含む複雑性を減じる要因が不足している。
Panel B: High Japanese Score (84%) / Low English Score (41%)			
English Review	Japanese Review	English Reasoning	Japanese Reasoning
Motivation to work: Because I was entrusted with parts of the store's sales floors, I was able to create a sales floor in line with my own ideas. There is responsibility that my own methods affect sales, but of course I'm happy when sales increase year over year. Depending on the store, there are many women, and I think there is an understanding toward women. In the store I worked at, there were many women, and heavy items were basically carried by men. Because it was on a shift system, I had one fixed weekday off, and there were about 8–10 days off per month. There wasn't much overtime, and it was easy to take paid leave. Since shorter working hours were possible, I think the work-life balance is pretty good. Strengths: As with the entire 100-yen/100-yen shop industry, stores located in tourist cities and major cities'centers due to inbound tourism are performing well, and overall business is stable. Weaknesses: Compared with other companies in the same industry, the recognition of what the company's strengths are is 薄く (thin/weak), and the annual income is also low. Salary system: Because I wasn't a full-time employee, the pay was low. The salary isn't much different from a full-time part-time job. However, even for full-time employees, the salary is low, and perhaps because of that, many 20-something full-time employees do not stay long and leave. One good thing is that overtime pay is calculated to the minute.	働きがい：店舗の一部の売り場を任されるため、自分の思うような売り場づくりをすることができた。自分のやり方によって売上に影響が出る責任感もあるが、売上を前年比増にできるとやはりうれしい。店舗にもよるが、女性が多く、女性に対する理解はあると思う。自分のいた店舗では女性が多く、重い商品は基本的には男性が運んでくれていた。シフト制だったので平日１日を固定休として、月では 8〜10 程が休日となっていた。残業はあまりなく、有給も取りやすかった。時短勤務もできるのでワークライフバランスは良い方だと思う。強み：100 均業界全体に言えることであるが、インバウンドにより観光都市や大都市の中心に立地する店舗は売上が好調であり、業績が安定していること。弱み：同業他社と比べて知名度や強みが何であるかというイメージも薄く、年収も低い。給与制度：正社員ではないので安い。フルタイムのアルバイトと大して変わらない給料。ただ、正社員であっても給料は低く、それもあってか 20 代の正社員は長続きせず退職される方が多い気がした。残業代は１分単位で計算される点は良かった。	**Score: 35** The review mentions empowerment in some local store decisions and some work-life balance, but overall there is little evidence of systematic management practices, durable relational contracts, or firm-wide capabilities; observed control is limited and performance effects are modest, resulting in a low-to-moderate organizational capital signal.	**Score: 62** 従業員は自店舗の売り場運用や顧客対応に対する影響力を認識しており、時間をかけて変化する組織的手法（責任範囲・シフト制・L&W/Balanced)を示唆。ただし、知名度・給与の低さ、同業他社との差異の認識、正社員の低い定着が強い組織資本の推進を阻む要因となっており、内発的資本生成は限定的な点がある。

Table E.26: Organizational Capital Measures and Firm Performance (LAR)

	LAR E→E	LAR J→E	LAR J→J	LAR (Avg.)	LAR E→E (%)	LAR J→E (%)	LAR J→J (%)	LAR (Avg., %)
Panel (a)	0.867***	0.866***	0.877***	1.000***	0.030***	0.029***	0.029***	0.032***
ROA (%)	(0.114)	(0.108)	(0.104)	(0.120)	(0.004)	(0.004)	(0.004)	(0.004)
Panel (b)	0.290***	0.345***	0.394***	0.394***	0.011***	0.013***	0.014***	0.013***
P/B Ratio	(0.066)	(0.065)	(0.068)	(0.073)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	24,006	24,006	24,006	24,006	24,006	24,006	24,006	24,006
log(Total Assets)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry-Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Panel (a) reports firm-year regressions with ROA (in percent) as the dependent variable; Panel (b) reports analogous regressions with the price-to-book (P/B) ratio as the dependent variable. Each column reports a separate regression on one LLM-as-Rater (LAR) organizational-capital measure constructed from employee reviews. In the LAR approach, an LLM assigns each review a score from 0 to 100 based on the extent to which the review reflects organizational capital; firm-year measures are the average of review-level scores within a firm-year. Column labels indicate the language of the review text and the prompt: E→E uses English-translated review text with an English prompt; J→E uses Japanese review text with an English prompt; J→J uses Japanese review text with a Japanese prompt; (Avg.) is the simple average of the three corresponding z-score measures. Columns labeled (%) use firm-year percentile ranks (1–100) of the underlying LAR score (constructed within year), and (Avg., %) is the simple average of the three percentile-rank measures. All specifications control for log(total assets) and firm age and include industry-by-year fixed effects. Standard errors are clustered at the firm level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table E.27: Variance Decomposition of Organizational Capital (LAR)

No.	Fixed Effects	OC (LAR) E→E	OC (LAR) J→E	OC (LAR) J→J
1	Year Fixed Effects	0.9%	0.9%	1.7%
2	Industry Fixed Effects	7.0%	7.4%	7.2%
3	Industry Fixed Effects + Year Fixed Effects	7.8%	8.2%	8.8%
4	Industry × Year Fixed Effects	21.5%	22.0%	22.7%
5	Firm Fixed Effects	36.7%	36.9%	34.2%
6	Firm Fixed Effects + Year Fixed Effects	37.4%	37.5%	35.5%
7	CEO Fixed Effects	43.2%	43.5%	41.3%
8	CEO Fixed Effects + Year Fixed Effects	43.7%	43.9%	42.1%
9	CEO–Firm Fixed Effects	43.3%	43.6%	41.5%
10	CEO–Firm Fixed Effects + Year Fixed Effects	43.9%	44.0%	42.2%

Notes: Each entry reports the overall R^2 from a regression of the firm-year LLM-as-Rater (LAR) organizational-capital measure on the listed fixed effects only (no covariates). E→E uses English-translated review text with an English prompt; J→E uses Japanese review text with an English prompt; J→J uses Japanese review text with a Japanese prompt. All three measures are standardized (z-scores). Values are shown as percentages of total variation explained by the fixed-effects structure. Observations dated before the CEO inauguration year are assigned a firm-specific placeholder CEO.

Table E.28: Correlations between OC Measures and Review Scores (LAR)

	LAR E→E	LAR J→E	LAR J→J	LAR (Avg.)	LAR E→E (%)	LAR J→E (%)	LAR J→J (%)	LAR (Avg., %)	Overall Rating	Compensation Rating	Motivation Rating	Openness Rating	Teamwork Rating	Chance to Grow	Long-term Development	Compliance Rating	Employee Evaluation System
LAR E→E	1.00																
LAR J→E	0.84***	1.00															
LAR J→J	0.78***	0.82***	1.00														
LAR (Avg.)	0.93***	0.95***	0.93***	1.00													
LAR E→E (%)	0.96***	0.80***	0.75***	0.90***	1.00												
LAR J→E (%)	0.80***	0.96***	0.79***	0.91***	0.82***	1.00											
LAR J→J (%)	0.75***	0.79***	0.96***	0.89***	0.77***	0.81***	1.00										
LAR (Avg., %)	0.89***	0.91***	0.89***	0.96***	0.92***	0.94***	0.92***	1.00									
Overall Rating	0.64***	0.65***	0.62***	0.68***	0.62***	0.64***	0.60***	0.66***	1.00								
Compensation Rating	0.42***	0.44***	0.40***	0.45***	0.42***	0.43***	0.39***	0.44***	0.64***	1.00							
Motivation Rating	0.47***	0.48***	0.47***	0.50***	0.45***	0.46***	0.45***	0.48***	0.74***	0.35***	1.00						
Openness Rating	0.45***	0.47***	0.45***	0.49***	0.44***	0.46***	0.44***	0.47***	0.68***	0.30***	0.45***	1.00					
Teamwork Rating	0.40***	0.40***	0.39***	0.42***	0.38***	0.38***	0.38***	0.41***	0.67***	0.24***	0.51***	0.50***	1.00				
Chance to Grow	0.45***	0.46***	0.46***	0.49***	0.43***	0.45***	0.45***	0.47***	0.68***	0.27***	0.54***	0.43***	0.38***	1.00			
Long-term Development	0.47***	0.47***	0.43***	0.49***	0.46***	0.46***	0.42***	0.47***	0.74***	0.45***	0.45***	0.39***	0.42***	0.44***	1.00		
Compliance Rating	0.33***	0.34***	0.29***	0.35***	0.34***	0.35***	0.30***	0.35***	0.53***	0.36***	0.21***	0.25***	0.25***	0.11***	0.38***	1.00	
Employee Evaluation System	0.39***	0.41***	0.40***	0.43***	0.37***	0.39***	0.38***	0.40***	0.65***	0.39***	0.47***	0.32***	0.30***	0.44***	0.40***	0.13***	1.00

Notes: Entries report pairwise Pearson correlations. The diagonal elements are 1.00 by construction and the upper triangle is omitted. LAR denotes LLM-as-Rater (LAR) organizational-capital measures. E→E uses English-translated review text with an English prompt; J→E uses Japanese review text with an English prompt; J→J uses Japanese review text with a Japanese prompt. are firm-year averages of review-level rater scores. For the z-score LAR measures (top four LAR columns/rows), each firm-year measure is standardized (z-score) within the analysis sample. The percentile measures ("%") are firm-year percentile ranks (1–100). Significance levels are based on two-sided tests of the null hypothesis of zero correlation. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

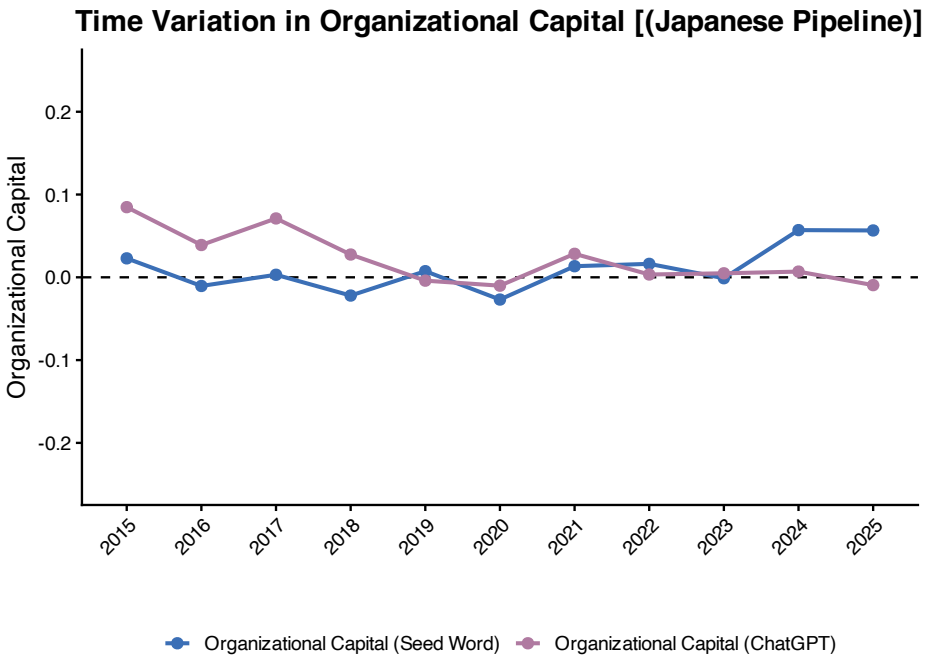
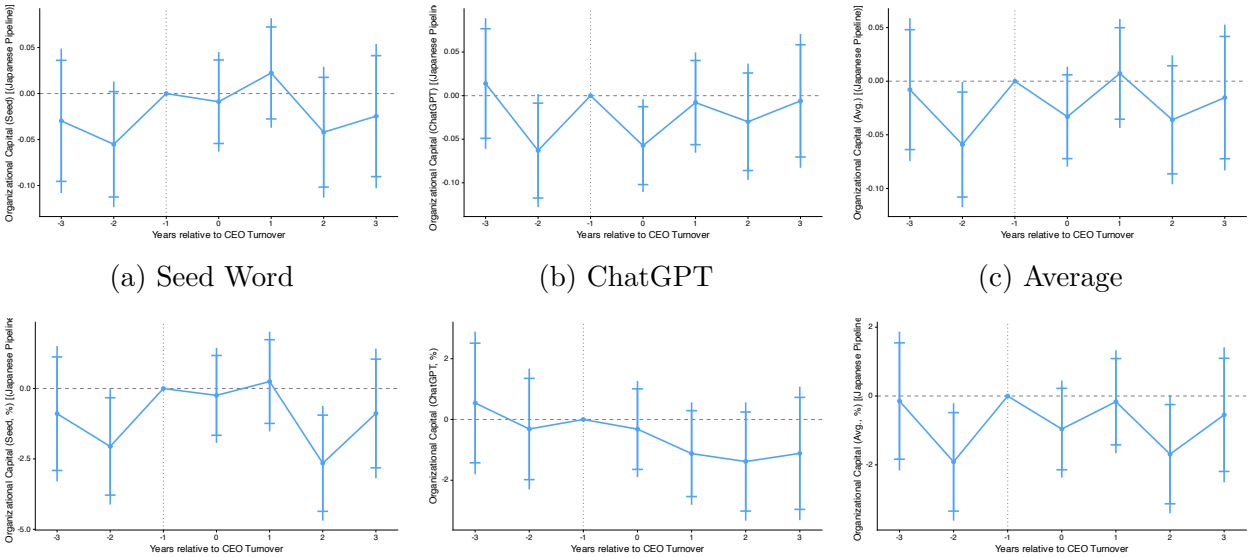


Figure E.9: Time Variation in Organizational Capital (Japanese Pipeline)

Notes: The figure plots yearly averages of firm-year organizational capital measures constructed from employee reviews. The seed-word and ChatGPT measures are standardized (z-scores) and averaged within year; the horizontal dashed line denotes zero.

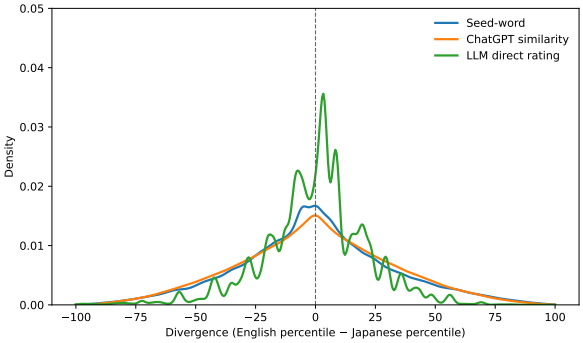


(a) Seed Word (b) ChatGPT (c) Average

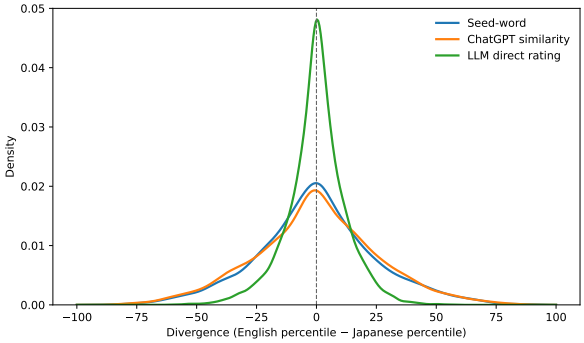
(d) Seed Word (%tile) (e) ChatGPT (%tile) (f) Average (%tile)

Figure E.10: CEO Turnover and Organizational Capital (Japanese Pipeline)

Notes: The figure reports event-study estimates around CEO turnover. The horizontal axis is event time in years relative to the CEO turnover ($t = 0$), with the year before turnover ($t = -1$) omitted as the reference period. Coefficients are from firm- and year-fixed-effects regressions of the indicated organizational capital (OC) measure on event-time indicators, controlling for the logarithm of total assets and firm age. Points show coefficient estimates; vertical bars show 95% confidence intervals and the horizontal bars on each end show 90% confidence intervals. The dashed horizontal line denotes zero and the dotted vertical line marks the reference year ($t = -1$). Standard errors are clustered at the firm level.



(a) Review-level



(b) Firm-year level ($n \geq 20$ reviews)

Figure E.11: Distribution of English–Japanese Divergence Across Measures

Notes: Each line plots the kernel density of the divergence metric, defined as the percentile score in English minus the percentile score in Japanese, for the seed-word, ChatGPT-similarity, and LLM-as-Rater measures. The review-level panel uses all reviews; the firm-year panel uses firm-years with at least 20 reviews.

Table E.29: Distribution of Reviews per Firm-Year

N	Mean	Median	p75	p99	Max
63,579	9.5	3	8	112	1,032

Notes: 28.1% of firm-years have only one review; 66.6% have five or fewer; 19.0% have more than ten.

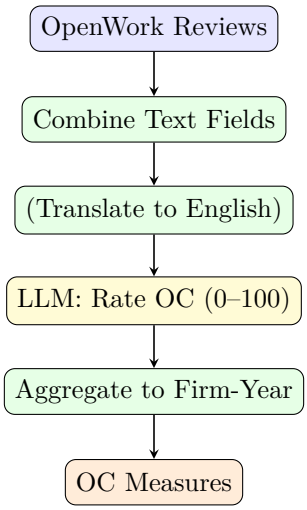


Figure E.12: Pipeline for LLM-as-Rater Organizational Capital Measure. Each review is directly scored by an LLM on a 0–100 scale based on how strongly it indicates the company has organizational capital. The LLM prompt includes the theoretical definition of organizational capital. Translation is optional: we implement variants using English translations, Japanese text with English prompt, and Japanese text with Japanese prompt.

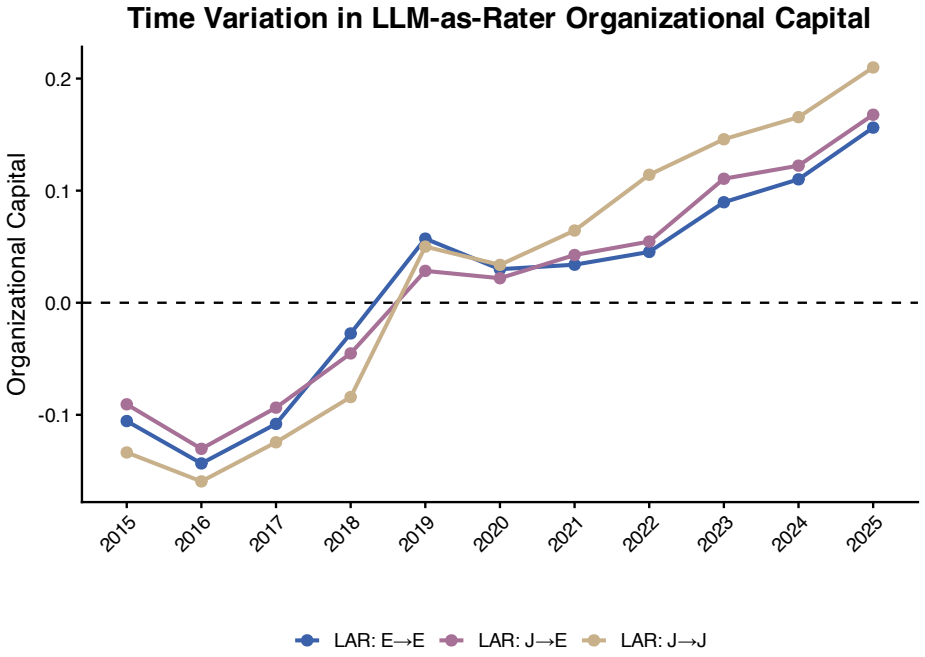
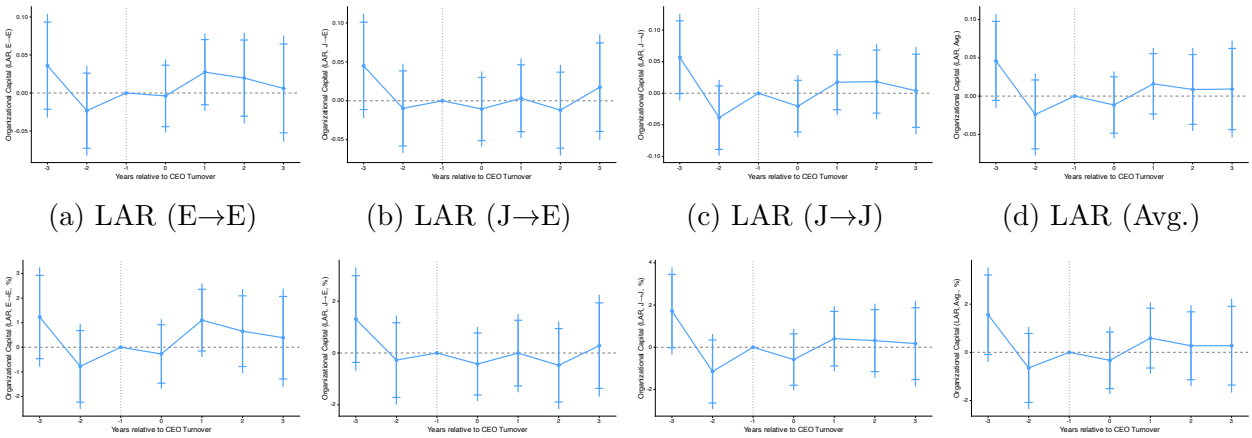


Figure E.13: Time Variation in Organizational Capital (LAR)

Notes: The figure plots yearly averages of firm-year organizational capital measures constructed from employee reviews using the LLM-as-Rater (LAR) method. The three series correspond to the English-translated review rated with an English prompt (E→E), the original Japanese review rated with an English prompt (J→E), and the original Japanese review rated with a Japanese prompt (J→J). The LAR measures are standardized (z-scores) and averaged within year; the horizontal dashed line denotes zero.



(a) LAR (E→E) (b) LAR (J→E) (c) LAR (J→J) (d) LAR (Avg.)

(e) LAR (E→E, %) (f) LAR (J→E, %) (g) LAR (J→J, %) (h) LAR (Avg., %)

Figure E.14: CEO Turnover and Organizational Capital (LAR)

Notes: The figure reports event-study estimates around CEO turnover for the organizational capital measures constructed from employee reviews using the LLM-as-Rater (LAR) method. The top four subplots correspond to the English-translated review rated with an English prompt (E→E), the original Japanese review rated with an English prompt (J→E), the original Japanese review rated with a Japanese prompt (J→J), and the average of the three measures. The bottom subplots correspond to firm-year percentile ranks (1–100) of each measure (constructed within year), and the average of the three percentile-rank measures. The horizontal axis is event time in years relative to the CEO turnover ($t = 0$), with the year before turnover ($t = -1$) omitted as the reference period. Coefficients are from firm- and year-fixed-effects regressions of the indicated LAR measure on event-time indicators, controlling for the logarithm of total assets and firm age. Points show coefficient estimates; vertical bars show 95% confidence intervals and the horizontal bars show 90% confidence intervals. The dashed horizontal line denotes zero and the dotted vertical line marks the reference year ($t = -1$). Standard errors are clustered at the firm level.

Table E.30: Empirical Bayes Estimation Parameters

OC Measure	$\hat{\sigma}^2_\epsilon$	$\bar{\hat{\sigma}}^2_\Omega$	$\bar{\lambda}$	Firm-Years	Reviews	Ind-Years
Seed (English)	2.0324	0.2355	0.230	63,579	604,057	4,892
ChatGPT (English)	0.0476	0.0045	0.163	63,579	604,057	4,892
Seed (Japanese)	0.9411	0.1109	0.129	63,547	602,885	4,893
ChatGPT (Japanese)	0.0522	0.0051	0.166	63,547	602,885	4,893
LAR E→E	190.0413	18.8246	0.244	63,588	604,313	4,893
LAR J→E	206.5127	21.4043	0.268	63,588	604,341	4,893
LAR J→J	224.2493	21.9802	0.249	63,588	604,340	4,893

Notes: This table reports the estimated parameters from the empirical Bayes shrinkage procedure. $\hat{\sigma}^2_\epsilon$ is the estimated review-level noise variance (within-firm-year dispersion). $\bar{\hat{\sigma}}^2_\Omega$ is the mean of estimated prior variance across industry-years. $\bar{\lambda}$ is the mean shrinkage factor, where $\lambda_{j,t} = \hat{\sigma}^2_{\Omega,g,t}/(\hat{\sigma}^2_{\Omega,g,t} + \hat{\sigma}^2_\epsilon/n_{j,t})$. Higher λ means less shrinkage toward the industry-year mean. LAR = LLM-as-Rater. E→E: English text, English prompt; J→E: Japanese text, English prompt; J→J: Japanese text, Japanese prompt.

Table E.31: Organizational Capital and Firm Performance: Robustness

	Seed (E)	ChatGPT (E)	Seed (J)	ChatGPT (J)	LAR E→E	LAR J→E	LAR J→J
Panel A: ROA (%)							
Baseline							
OC Measure	0.576***	0.586***	0.622***	0.809***	0.862***	0.861***	0.876***
	(0.089)	(0.089)	(0.086)	(0.085)	(0.102)	(0.096)	(0.093)
Observations	23,918	23,918	23,907	23,907	23,921	23,921	23,921
$n > 10$ *reviews*							
OC Measure	1.416***	2.133***	2.178***	2.930***	1.688***	1.551***	1.726***
	(0.479)	(0.445)	(0.489)	(0.428)	(0.485)	(0.473)	(0.484)
Observations	5,763	5,763	5,763	5,763	5,763	5,763	5,763
Empirical Bayes							
OC Measure	0.547***	0.542***	0.392***	0.742***	0.741***	0.738***	0.812***
	(0.089)	(0.081)	(0.089)	(0.086)	(0.114)	(0.114)	(0.115)
Observations	23,918	23,918	23,907	23,907	23,921	23,921	23,921
EB (Industry Available)							
OC Measure	0.462***	0.376***	0.318***	0.547***	0.606***	0.593***	0.635***
	(0.092)	(0.077)	(0.091)	(0.087)	(0.119)	(0.119)	(0.115)
Observations	13,814	13,814	13,806	13,806	13,814	13,814	13,814
Panel B: P/B Ratio							
Baseline							
OC Measure	0.095*	0.096**	0.190***	0.252***	0.290***	0.344***	0.393***
	(0.049)	(0.045)	(0.045)	(0.056)	(0.059)	(0.058)	(0.061)
Observations	24,063	24,063	24,052	24,052	24,066	24,066	24,066
$n > 10$ *reviews*							
OC Measure	0.407	0.432	0.768*	0.577	1.160***	1.166***	1.242***
	(0.511)	(0.488)	(0.452)	(0.467)	(0.197)	(0.209)	(0.207)
Observations	5,774	5,774	5,774	5,774	5,774	5,774	5,774
Empirical Bayes							
OC Measure	0.124***	0.180***	0.166***	0.257***	0.384***	0.438***	0.514***
	(0.047)	(0.044)	(0.048)	(0.060)	(0.069)	(0.065)	(0.065)
Observations	24,063	24,063	24,052	24,052	24,066	24,066	24,066
EB (Industry Available)							
OC Measure	0.084**	0.146***	0.153***	0.217***	0.287***	0.344***	0.422***
	(0.042)	(0.036)	(0.044)	(0.058)	(0.084)	(0.078)	(0.075)
Observations	13,940	13,940	13,932	13,932	13,940	13,940	13,940
log(Total Assets)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry-Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports firm-year regressions of firm performance on organizational capital (OC) measures. Panel A uses return on assets (ROA, in percent) as the dependent variable; Panel B uses the price-to-book (P/B) ratio. Seed and ChatGPT columns use dictionary-based measures; (E) denotes English-translated reviews, (J) denotes Japanese reviews. LAR (LLM-as-Rater) columns use an LLM to score reviews: E→E uses English text with English prompt; J→E uses Japanese text with English prompt; J→J uses Japanese text with Japanese prompt. All OC measures are standardized (z-scores). Robustness check filters to firm-years with > 10 reviews. Empirical Bayes (EB) specifications shrink firm-year means toward industry-year means; 'EB (Industry Available)' restricts to firms with industry codes available in the EB computation. All specifications include log(total assets) and firm age as controls, with industry-by-year fixed effects. Standard errors clustered at firm level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.