

The Activity Inside the Expense: Measuring Internally Generated Intangible Investment

Aneel Iqbal

Arizona State University

aneel.iqbal@thunderbird.asu.edu

October 2026

ABSTRACT

I develop a firm-level measure of internally generated organization-capital investment from what employees do rather than from capitalization parameters inferred from accounting outcomes or market prices. Using 384 million employment records and O*NET occupational task data, I identify capital-building labor and translate its cost into an investment fraction of SG&A. The approach separates two measurement questions that expenditure-based methods combine: how much capital-building work a firm performs and how that work maps into reported expense. Capital-building workforce structure is highly firm-specific. Most variation occurs among firms within the same industry-year, firm differences persist over time, and firms visibly accumulate this structure around listing. Translating activity into an SG&A fraction requires scaling by SG&A's labor share. Decomposing the resulting fraction reveals that the labor share varies more across firms than capital-building workforce structure and is negatively related to that structure, explaining why activity-based and expenditure-based measures rank firms differently. Independent accounting evidence supports both the workforce-based construction and its translation into reported expense. By making investment activity within reported expenses separately observable, the paper shows why the measurement challenge logically precedes the question of recognition.

Keywords: intangible capital; organization capital; capitalization; human capital; selling, general, and administrative expense; workforce data

JEL Classifications: M41; J24; E22.

1. INTRODUCTION

Most internally generated intangible investment is expensed under U.S. GAAP, so financial statements do not directly reveal how much firms invest in intangible capital. A large literature addresses this measurement problem by capitalizing portions of research and development (R&D) and selling, general, and administrative (SG&A) expenditures.¹ The resulting estimates depend critically on an unobserved parameter, the fraction of expenditure that creates long-lived assets rather than supporting current operations. Prior work imposes a uniform thirty-percent investment share (Peters and Taylor 2017)², estimates industry-level parameters from accounting outcomes (Enache and Srivastava 2018; Iqbal, Rajgopal, Srivastava, and Zhao 2025), or infers them from market prices at firm exits (Ewens, Peters, and Wang 2025). In each case, investment is inferred from reported expenditures, accounting outcomes, or market prices rather than from direct evidence about the underlying capital-building activity. As a result, these approaches cannot reveal whether firms assigned the same capitalization parameter are actually performing the same amount of investment activity.

This paper starts from a different place, the activity performed inside the firm rather than the expenditure reported for it. Firms use labor both to run current operations and to build longer-lived capabilities, such as organizational design, trained teams, and customer relationships. Occupational task data reveal which employees perform which tasks. The measurement therefore proceeds in a fixed order. It observes capital-building activity, expresses it as an investment flow, and only then translates that flow into accounting units. The translation is not neutral. Capital-

¹ These studies generally reconstruct intangible assets using the perpetual-inventory method. A fraction of annual expenditure is treated as investment and accumulated into a capital stock at an assumed depreciation rate. This paper focuses on the investment fraction. The depreciation rate of the resulting capital is examined separately in Iqbal (2026).

² The thirty-percent convention is widely used in firm-level finance applications of intangible capital and is embedded in the total-q dataset distributed through Wharton Research Data Services.

building occupations make up roughly half of the SG&A labor at a major beverage company and about a quarter at a large semiconductor manufacturer. However, labor constitutes less than a quarter of the beverage company's total SG&A and nearly all of the manufacturer's SG&A. The implied investment fraction is therefore approximately 2 percent of SG&A for the beverage company and 5 percent for the manufacturer, reversing the ranking (Figure 2). Thus, observing capital-building activity and estimating the investment content of an expense are different measurement problems, because the latter also depends on what else the expense contains.

I implement this idea using 384 million employment records linked to public firms from Revelio Labs³ and O*NET occupational task data.⁴ I classify organization-capital occupations using the task-based criteria of Squicciarini and Le Mou  l (2012)⁵ and keep formal research outside the measure by classifying research occupations separately under the Frascati Manual definition.⁶ The central quantity is capital-building workforce structure, defined as the share of a firm's SG&A labor employed in capital-building occupations. It shows where capital-building labor resides. Weighting occupations by the intensity with which they perform capital-building tasks yields task-weighted capital-building activity, and the payroll cost of that activity is own-account labor investment. None of the three values the activity's output. Employees developing customer relationships count toward them, but the measures do not value the customer relationships or brand capital that may result. The construction has two stages (Figure 1). The first

³ Revelio Labs aggregates public online professional profiles into standardized employment histories containing employer, job title, occupation, seniority, estimated salary, and position start and end dates.

⁴ The Occupational Information Network (O*NET  ), sponsored by the U.S. Department of Labor, provides standardized information on occupational tasks and work activities. I use O*NET 25.1 (November 2020), the first release coded to the O*NET-SOC 2019 taxonomy used by the workforce data (National Center for O*NET Development 2020).

⁵ O*NET describes occupational task content using standardized work activities. Squicciarini and Le Mou  l (2012) use these activities to identify occupations associated with building organization capital.

⁶ The OECD's Frascati Manual defines R&D personnel as researchers, technicians and equivalent staff, and other supporting staff (OECD 2015). I use occupations to identify the first two groups. I do not classify general support occupations as research because occupation alone does not reveal whether those employees support R&D.

identifies who performs capital-building work and how intensively, using only occupations and externally measured O*NET task content and frequency. No firm accounting outcome, capitalization parameter, or market price enters this stage. The second stage uses the labor share of SG&A to express the cost of that activity as a fraction of reported expense. Accordingly, the firm-level tests of capital-building activity use workforce structure, while the labor-core fraction is used when the question concerns own-account labor investment as a share of SG&A. Accounting outcomes can therefore evaluate the measurement without determining which firms are classified as doing more capital building.

The first finding is that capital-building workforce structure is primarily a firm characteristic rather than an industry characteristic. Across 75,440 firm-years, 88 percent of the variation in workforce structure occurs among firms within the same industry-year.⁷ These differences are highly persistent. The annual autocorrelation of the within-industry component is 0.90, and five years later, 69 percent of firms initially in the top within-industry quintile remain there. A leave-one-out industry-year average explains only 10 percent of the firm-level measure. Therefore, the principal variation in capital-building activity is precisely the variation that an industry-level capitalization parameter cannot observe.

Persistence alone may reflect stable business models rather than capital building. The second finding is that the structure is visibly built. Employment histories begin before firms list, so the workforce can be followed through the years just before listing as well as after. Relative to the same firm's mature years, the capital-building share of SG&A labor is already elevated three years before listing, remains at that level through the listing year, and declines steadily afterward. Net hiring into capital-building occupations peaks three years before listing and falls by more than

⁷ The decomposition separates variation across industry-year cells from variation among firms within the same industry-year and is applied identically to workforce structure and the labor-core fraction.

three quarters within two years after it. Managerial and professional occupations outside the capital-building set are not similarly elevated before listing, making differential representation of professional employees in early employment histories unlikely to explain the pattern. Listing therefore occurs late in the organizational build-out rather than marking its start. This timing application shows what observing activity adds, because a measure built from public financial statements begins only when the firm enters the accounting data. The build-out is concentrated in the activities the classification targets rather than in the reporting and compliance staffing a newly public company must add. Excluding finance and compliance occupations leaves 90 percent of the listing-year increase. The persistence result therefore does not simply identify a stable type of firm, because the same workforce structure is visibly accumulated during a period in which firms are building organizational capacity.

The third finding concerns the accounting translation. Translating task-weighted activity into an investment fraction requires scaling it by the labor share of SG&A. The labor share is more dispersed than workforce structure, and the two components are negatively related. Expense composition therefore dominates the resulting fraction. Its Spearman correlation is 0.872 with the labor share but only 0.129 with workforce structure. The fraction is thus not a pure measure of capital-building activity. It measures how that activity sits inside the reported expense base. This distinction matters whenever a capitalization fraction is interpreted as a ranking of firms' underlying investment activity. A classification application shows the consequence. Within industry-year, only 22 to 36 percent of the firm-years in the top quintile of workforce structure are also in the top quintile of SG&A intensity or a published capitalization measure, compared with 20 percent if the rankings were unrelated (Section 7.3 and Table 8). In levels, the baseline own-account labor estimate averages 4.8 percent of SG&A net of R&D. Calculations that broaden the

activity and cost scope move the figure toward the 30 percent convention and the parameters estimated in prior work (Section 7.2).

Because the construction is independent of accounting outcomes, accounting data can evaluate three separate links in the measurement. Reported staff expense evaluates modeled payroll. Subsequent R&D initiation evaluates whether the occupational classification separates organization building from formal research. Disclosed software capitalization evaluates the second-stage translation of workforce payroll into a share of reported expense. A mapping estimated in an initial period and then frozen predicts disclosed capitalization in a later holdout period, with a correlation of 0.44 between predicted and disclosed intensities. A component diagnostic shows that this predictive content comes from the labor share of reported expense rather than workforce structure. The software test therefore supports the labor-share scaling used in the second-stage expense translation. It does not validate the identification of capital-building activity. Together, the three settings support the payroll input, the research boundary, and the labor-share scaling without using any of these outcomes to define capital-building activity.

This paper makes two contributions. First, it provides a measure of capital-building activity that is independent of the outcomes and prices used to evaluate capitalization measures. Existing firm-level measures impose an investment fraction (Peters and Taylor 2017) or estimate it from accounting outcomes or market prices (Lev and Sougiannis 1996; Enache and Srivastava 2018; Ewens et al. 2025; Iqbal et al. 2025), so tests against related outcomes partly evaluate information used to build the measure. Growth-accounting research identifies capital-building occupations but applies them at the national or industry level (Corrado, Hulten, and Sichel 2009; Squicciarini and Le Mou  l 2012). Recent workforce-based studies measure the organizational capital a firm has accumulated (Cai, Prat, and Yu 2026), the amount and functional location of its labor costs (Chen,

Jain, Li, and Rajgopal 2026), or organization-capital-building work within AI employment (Babina, He, and Jiang 2026). I measure the activity through which employees build organization capital across the occupational distribution and at the firm level. The resulting workforce measure varies with observed occupational choices rather than parameters inferred from outcomes it is later used to explain. Translating the cost of that activity into an accounting investment flow is a separate second-stage calculation.

The second contribution is evidence that measuring investment activity and estimating an accounting investment fraction are distinct problems. Capital-building workforce structure varies substantially and persistently among firms in the same industry-year, precisely the variation that industry-level parameters average away. A capitalization fraction combines this activity with the composition of the expense base, and expense composition dominates the resulting ranking in this setting. Firms with similar capital-building activity can therefore have different fractions, while similar fractions can arise from different workforces. These findings bear on the debate over internally generated intangibles (Lev and Zarowin 1999; Kanodia, Sapra, and Venugopalan 2004; Barker, Lennard, Penman, and Teixeira 2022). Financial statements identify the expense category but not the activity within it that creates future benefits. The evidence shows that this missing activity is observable at the firm level. The measurement challenge therefore precedes the recognition decision, because the investment activity within reported expense must be identified before one can decide how to recognize it.

Section 2 reviews related research and the measurement setting. Section 3 develops the measurement framework, and Section 4 describes the data and construction. Section 5 evaluates the payroll input, the research boundary, and the labor-share scaling in the expense translation. Section 6 documents firm-level capital-building workforce structure, including its variation,

persistence, and the timing of its build-out around listing. Section 7 separates activity from expense composition, relates the measures to broader investment estimates and existing capitalization measures, and shows how the choice of measure changes which firms are classified as capital builders. Section 8 concludes.

2. RELATED LITERATURE AND MEASUREMENT SETTING

2.1. Existing Approaches to Measuring Intangible Investment

The intangible-capital literature typically reconstructs expensed investment using the perpetual-inventory method, which requires two parameters, the fraction of expenditure treated as investment and the rate at which the resulting capital depreciates. Early work estimated these parameters from accounting outcomes. Lev and Sougiannis (1996) estimated industry-specific amortization schedules for R&D from earnings, while Lev and Zarowin (1999) and Kothari, Laguerre, and Leone (2002) documented the consequences and uncertainty associated with expensing investments whose benefits extend beyond the current period. Corrado et al. (2009) instead specified category-level investment fractions and depreciation rates within an expanded national-accounts framework.

These approaches subsequently shaped firm-level measures of intangible capital. Eisfeldt and Papanikolaou (2013) use SG&A to construct organization capital, while Peters and Taylor (2017) treat 30 percent of SG&A as investment and depreciate the resulting capital at 20 percent annually. More recent work estimates rather than assumes the investment fraction. Enache and Srivastava (2018) estimate the investment component of SG&A using industry-year regressions, Iqbal et al. (2025) estimate investment and depreciation parameters from the relation between expenditures and future revenues, and Ewens et al. (2025) infer intangible-capital parameters from market prices at firm exits.

These methods differ in how they obtain the capitalization parameter, but they confront the same underlying measurement problem. The investment activity contained within reported expenditure is not directly observed. The resulting parameters must therefore be imposed or inferred from accounting outcomes or market prices and are generally applied to expenditures at the category or industry level. Occupational data provide a different source of information. They reveal the capital-building activity performed inside individual firms before that activity is translated into an accounting expenditure fraction. This distinction matters when firms within the same industry devote different shares of their workforces to capital building, variation that a common industry parameter necessarily averages away. Internally generated intangible capabilities can also be highly firm-specific and may have limited recoverable value outside the firm (Kermani and Ma 2023; Payne-Mann and Sloan 2026), further motivating measurement at the firm level.

The distinction matters for applications as well. Capitalizing intangible expenditures affects measured earnings quality, the interpretation of losses, and assessments of firm performance (Srivastava 2014, 2023; Rajgopal, Srivastava, and Zhao 2023; Rajgopal, Basirian, Iqbal, and Srivastava 2026). Each application ultimately depends on the same first-stage measurement question of which portion of reported expenditure represents investment rather than current-period operations. The workforce approach provides evidence on that question from the underlying activity rather than from the accounting outcomes that capitalization subsequently alters.

2.2. Workforce-Based Measurement of Intangible Investment

The idea that labor embodies intangible investment is longstanding, but the data needed to measure that activity at individual firms have only recently become available. The growth-accounting literature measures own-account organization-capital investment using occupations.

Squicciarini and Le Mouël (2012) identify capital-building occupations from O*NET task content and treat 20 percent of their labor costs as investment, extending to a broader set of occupations the share that Corrado, Hulten, and Sichel (2005, 2009) apply to executives.⁸ Le Mouël and Squicciarini (2015) extend the occupation-based approach across countries, while Piekkola (2016) uses occupational payroll from Finnish linked employer-employee data to construct intangible investment. These studies establish the link from occupational activity to own-account investment but apply it primarily at the national or industry level. Workforce microdata make it possible to apply this approach at the firm level and observe differences in capital-building activity among firms within the same industry.

Workforce data are increasingly used to measure firm characteristics that financial statements do not reveal (e.g., Li, Lourie, Nekrasov, and Shevlin 2022; Fedyk and Hodson 2023; Baker, Larcker, McClure, Saraph, and Watts 2024; Hampole, Papanikolaou, Schmidt, and Seegmiller 2025). Closest in spirit, He, Mostrom, and Sufi (2024) use sales-and-marketing payroll, among other measures, to measure investment in customer capital, with evidence focusing on differences across industries and over time. I extend the occupation-based approach to the broader set of activities through which firms build organization capital and measure that investment among firms within the same industry-year.

The closest studies using workforce data to measure internally generated intangibles capture different points in the investment process. Cai, Prat, and Yu (2026) use employee reviews to measure the organizational capital a firm has accumulated, the stock produced by past investment. Occupational task data instead identify where capital-building labor is currently deployed, the flow that builds that stock. Consistent with the firm specificity documented below, Cai et al. find that

⁸ The 20 percent convention treats 20 cents of each dollar of compensation in a capital-building occupation as investment and the remaining 80 cents as current operating cost.

industry-year effects explain less than 8 percent of the variation in their measure. Chen, Jain, Li, and Rajgopal (2026) use Revelio data to estimate labor costs by job function and show that this disaggregation is informative, motivated in part by the FASB's expense-disaggregation standard (FASB 2024). Even within a functional category such as SG&A, however, employees perform economically distinct activities, so identifying where labor costs are incurred does not identify their investment content. Babina, He, and Jiang (2026) classify job descriptions to separate AI jobs that build durable firm-specific systems from other AI jobs and show that this organization-capital channel helps explain the productivity effects of AI investment. Their accounting proxy capitalizes all SG&A rather than estimating the share of SG&A that represents investment.

I measure a different object, capital-building activity, across the broader occupational distribution, and the resulting firm-level own-account investment flow, expressed in accounting units. To my knowledge, prior work does not make this translation from workforce activity to a firm-level investment fraction across occupations.

3. MEASUREMENT FRAMEWORK

3.1. From Workforce Activity to Investment Flows

The measurement framework starts from a simple distinction. Firms use labor both to support current operations and to build intangible capital. Occupations consist of different tasks, some directed toward current operations and others toward activities whose benefits extend beyond the current period. Capital-building tasks include developing strategy, organizing work, building and training teams, and establishing customer relationships.

The framework distinguishes workforce structure, capital-building activity, and investment. Let o index occupations and i,t index firm-years. Define s_{oit} as occupation o 's share of the firm's

SG&A labor and b_o as an indicator for whether occupation o is classified as capital building. Capital-building workforce structure is

$$Structure_{it} = \sum_o s_{oit} b_o. \quad (1)$$

This quantity measures the share of the firm's SG&A labor employed in capital-building occupations. It depends on who the firm employs, not on how much of each employee's time is devoted to capital building.

Employees in capital-building occupations do not necessarily spend all of their time building intangible capital. I therefore use O*NET task information to estimate an occupation-specific capital-building task share, τ_o . Weighting each occupation by this share gives the share of the firm's SG&A labor devoted to capital-building activity:

$$Activity_{it} = \sum_o s_{oit} b_o \tau_o. \quad (2)$$

The framework therefore distinguishes three quantities, and the paper uses each term in one sense only. Capital-building workforce structure, equation (1), records where capital-building labor resides in the firm's SG&A workforce. Task-weighted capital-building activity, equation (2), estimates how intensively that labor engages in capital-building tasks. Own-account labor investment, equation (3) below, is the payroll cost of that activity.

Workforce structure is the paper's firm-level measure because it captures the margin firms choose, the occupational composition of their workforce. The task shares in equation (2) are occupation-level conversion weights taken from O*NET. They are the same for every firm employing an occupation and are not separately observed firm choices, so they vary across firms only through that composition. They therefore enter where they are needed, in translating workforce structure into investment dollars in equation (3), rather than in the firm-level analysis.

The corresponding own-account labor investment is the labor cost of that activity. Let $Payroll_{it}^{SG\&A}$ denote payroll allocated to SG&A. Then

$$LaborInvestment_{it} = Payroll_{it}^{SG\&A} \times Activity_{it}. \quad (3)$$

Equation (3) is an investment flow, the estimated labor resources devoted during year t to activities that build organization capital. To express that flow relative to the reported expense base, define the firm's labor share of SG&A as

$$LaborShare_{it} = \frac{Payroll_{it}^{SG\&A}}{SG\&A_{it} - R\&D_{it}}. \quad (4)$$

Substituting equation (4) into equation (3) gives the labor-core investment fraction:

$$Fraction_{it} = LaborShare_{it} \times \sum_o s_{oit} b_o \tau_o. \quad (5)$$

$$Fraction_{it} = (\sum_o s_{oit} b_o \tau_o) \times \frac{Payroll_{it}^{SG\&A}}{SG\&A_{it} - R\&D_{it}}.$$

In the expanded form, the term in parentheses measures what SG&A labor is doing, and the payroll ratio measures how much of SG&A is labor. Equation (5) makes the two stages of the measurement explicit. The first stage uses occupational composition and task content to identify where the firm's capital-building labor resides and how intensively it builds capital. The second uses the labor share of SG&A to translate that activity into a fraction of reported expenditure. The labor-core fraction therefore measures own-account labor investment as a share of SG&A net of R&D. It does not measure purchased or other non-labor intangible investment.

A final measurement choice concerns τ_o , the share of work in a capital-building occupation treated as investment. Rather than impose a common share, the baseline specification uses O*NET task-frequency ratings to construct occupation-specific capital-building shares (Section 4.5). Their payroll-weighted average is 17 percent, close to the 20 percent convention of Corrado, Hulten, and

Sichel (2005, 2009), but obtained independently from task data. I retain the uniform 20 percent convention as a literature-comparable benchmark.⁹

3.2. Scope of the Measure and the Boundary with Research

The measure captures own-account labor investment, the labor cost of capital-building activity performed by the firm's own employees. This is narrower than total intangible investment. It excludes non-labor resources that support those employees, such as purchased services, materials, occupancy, and overhead, as well as investment produced outside the firm's workforce through contract research, advertising agencies, consulting, contractors, and acquisitions. It also excludes intangible investment created through learning by doing within production occupations. These distinctions define the scope of the measure rather than whether the excluded activities create capital. In particular, outsourcing can reduce measured own-account labor investment without reducing a firm's total intangible investment. The three concepts are therefore nested:

Own-account labor investment $\subset$ Own-account investment $\subset$ Total intangible investment

The workforce measure estimates the first from employee activity and its labor cost. Financial statements do not separately identify the investment content of the additional inputs within SG&A, so I do not impute those costs firm by firm. Section 7.2 instead shows how broadening the activity and cost scope under explicit assumptions moves the labor-based estimate toward capitalization parameters designed to capture broader investment. Keeping these quantities separate makes clear which part of intangible investment is observed from workforce activity and which part would require additional assumptions.

⁹ Corrado, Hulten, and Sichel (2005, 2009) assume that executives devote 20 percent of working time to building organization capital and value own-account organization investment accordingly. Squicciarini and Le Mouél (2012) extend this convention to their broader set of capital-building occupations. Section 4.5 describes the task-frequency construction.

The measure focuses on organization capital rather than knowledge capital because the accounting measurement problem differs across the two. R&D has a separately reported expense line when disclosed, whereas organization-capital investment is commingled with current operations inside SG&A. The problem addressed here is therefore particularly acute for organization capital, because the financial statements report the expenditure but do not identify the investment activity within it. Whether such investment should ultimately be reported separately from operating expense is a long-standing question in accounting theory (Kanodia, Sapra, and Venugopalan 2004). The prior measurement question is whether the underlying investment activity can be distinguished from current operations at all.

Research nevertheless creates an occupational boundary problem. Some technical occupations satisfy both the organization-capital criteria and the Frascati Manual definition of research personnel, and the same occupation can perform formal R&D or internal technology work, such as systems implementation and maintenance. I therefore classify research occupations separately and exclude formal research from the organization-capital measure. For the baseline measure, occupations that fall into both classifications are assigned to research when the firm reports R&D expense and to organization capital otherwise (Section 4.4). Section 5.2 validates the distinction using mutually exclusive occupation sets whose construction does not depend on reported R&D status.

3.3. Measurement Assumptions and Sources of Error

Applying the framework to workforce data requires three assumptions, each with a corresponding source of measurement error. First, O*NET assigns a common task profile to each occupation, so firm-level variation reflects differences in occupational composition rather than differences in how the same occupation is performed across firms. For example, a marketing

manager at a consumer-products firm may devote a different share of time to capital building than one at a pharmaceutical firm. Such within-occupation differences introduce measurement error when the O*NET profile does not capture the common capital-building content of the occupation. Job-description text could provide greater firm-level granularity but is proprietary. Using occupations instead makes the classification transparent and reconstructible from public O*NET files.

Second, positions in the workforce data may be assigned imperfectly to occupations. Third, online professional profiles may represent some parts of a firm's workforce better than others, even after applying Revelio's occupation-level sampling weights. These potential errors affect different parts of the measurement and can therefore be evaluated separately. I compare modeled payroll with independently reported staff expense (Section 5.1), test the occupational classification against reported R&D and an independently re-estimated occupation set (Sections 5.2 and 5.4), and evaluate workforce representation against reported employment, high-coverage subsamples, alternative employee-count weights, and the listing-period professional-workforce placebo (Sections 4.3, 5.4, and 6.2). The empirical tests therefore evaluate the principal inputs on which the workforce measure depends rather than assuming that the underlying data are measured without error.

3.4. Two Measures and the Validation Logic

The framework yields two measures that answer different questions. The capital-building workforce structure, equation (1), measures the share of the firm's SG&A labor employed in capital-building occupations. It depends only on the firm's occupational composition. The labor-core fraction, equation (5), goes further by expressing capital-building activity as a share of reported SG&A. It therefore also depends on the share of SG&A that is labor. One measures

capital-building activity in the workforce, whereas the other expresses that activity relative to the accounting expense. Because the second measure incorporates the composition of SG&A, the two can rank firms differently, as Section 7 shows.

The validation logic follows from what a capital-building workforce structure should look like. If it captures capital-building activity rather than merely a static occupational characteristic, it should persist across firms yet change when firms build organizational capacity. Section 6 tests these implications by examining persistence within industry-year and the organizational build-out before and after listing. Appendix Table A10 reports a supplementary test based on subsequent capital-building hiring. Together, these tests ask whether the measured workforce structure exhibits the durability and formation expected of organization-building activity. They are not causal tests. Workforce structure is an endogenous firm choice reflecting strategy, business model, growth opportunities, and other firm characteristics, so I interpret the estimates as descriptive and predictive evidence about the behavior of the measure rather than as effects of changing workforce structure.

[See Figure 1]

4. DATA AND CONSTRUCTION

4.1. Workforce Data

My workforce data come from Revelio Labs, which constructs employment histories from online professional profiles.¹⁰ Each position record identifies the employer, job title, occupation, seniority, modeled annual salary, sampling weight, and position start and end dates. Occupations

¹⁰ The Revelio Labs data are proprietary and cannot be redistributed. The accompanying public package provides the occupation classifications, crosswalk lineage, and code that constructs the O*NET inputs and converts licensed position records into firm-year measures. Code for the remaining analyses is available to editors and referees. Reconstructing the firm-year measures requires licensed Revelio Labs data.

follow the 2019 O*NET taxonomy, and position dates are recorded with monthly precision. Revelio provides occupation-country sampling weights to adjust for differences in online-profile coverage.¹¹ Company reference files provide market identifiers and parent-subsidiary relationships, which I use to link positions to public firms and aggregate subsidiary employment to the ultimate parent.

Two features of the data shape the research design. First, historical coverage is weaker in earlier years because employment histories are reconstructed from profiles in the database. I therefore begin the panel in 2005, the earliest year with broad workforce coverage, and end it in fiscal year 2024, the last year with complete fiscal-year employment histories. I require at least 5 percent workforce coverage before grossing observed payroll to the full workforce (Section 4.5). Section 4.3 evaluates coverage against independently reported employment, and later tests assess whether the principal results are sensitive to coverage.

Second, salaries are modeled rather than observed. Because payroll enters directly into the investment measure, this input is independently testable, and Section 5.1 compares modeled payroll with staff expense reported in firms' financial statements. I also reconstruct workforce structure using employee counts rather than salary weights to assess whether the firm-level results depend on modeled compensation.

4.2. Firm Linkage and Sample Construction

Table 1 reports the analysis samples and descriptive statistics, while Appendix Table A2 provides details on the company linkage and workforce-record extraction.

[See Table 1]

¹¹ A position active for six months contributes 0.5 full-time equivalents (FTEs) before weighting. With an occupation-country sampling weight of 1.5, it contributes 0.75 weighted FTEs. A position active for the full year contributes one FTE before weighting.

I link Revelio employers to Compustat firms, aggregate subsidiaries to their ultimate parents, and supplement Revelio's links with manual matches. The resulting data contain 384 million usable positions at linked public firms. I allocate each position to firm-years according to the portion of the year it is active and apply Revelio's sampling weights to adjust for differences in online-profile coverage across occupations and countries. Positions without usable start dates, 7.5 percent of the total, are excluded. After merging the workforce and accounting data and excluding regulated utilities and financial firms,¹² the U.S.-headquartered sample contains 75,893 firm-years for 7,969 firms from 2005 through 2024. Capital-building workforce structure is measurable for 75,440 firm-years, and the labor-core investment fraction can be constructed for 54,291 firm-years at 6,258 firms across 43 industries. I define industries using the 48 Fama and French (1997) groups and industry-year cells using fiscal year. Firm age is measured from the firm's first appearance in Compustat. All continuous variables entering regressions, correlations, variance decompositions, or reported summary statistics are winsorized at the 1st and 99th percentiles within fiscal year on the relevant estimation sample.

Table 1, Panel B describes firm-years for which the labor-core fraction can be constructed. The sample spans a wide range of public firms. The median firm has \$339 million in total assets and approximately 760 employees, while the corresponding means are substantially larger. On average, capital-building occupations account for 48.7 percent of SG&A labor and labor accounts for 59.2 percent of SG&A net of R&D, and the labor-core fraction averages 4.8 percent (Table 2). The dispersion in these underlying components is important because the paper's firm-level tests exploit differences in workforce structure and expense composition rather than differences in firm

¹² I exclude regulated utilities and financial firms because their reporting formats and expense structures make their SG&A-based capitalization measures difficult to compare with those of other firms.

scale alone. The median workforce coverage ratio is 1.01, and Section 4.3 examines its cross-sectional and time-series variation.

4.3. Workforce Coverage

Weighted Revelio full-time-equivalent employment tracks independently reported employment. The median ratio of Revelio to Compustat employment rises from 0.46 in 2005 to 1.25 in 2024, and the correlation between the two measures in logs is 0.85 (Appendix Table A2). The coverage ratio should not be interpreted as the percentage of a firm's employees observed in Revelio. Revelio measures sampling-weighted positions active during the fiscal year, whereas Compustat reports employment at a point in time, so the ratio can exceed one.

For the workforce measure, the central concern is not the level of coverage itself but whether uneven coverage systematically distorts occupational composition. Revelio's sampling weights vary by occupation and country to address differential representation across occupations. Section 5.4 therefore tests the measurement directly under alternative coverage conditions and shows that the principal results are similar among high-coverage firm-years and in later years, when coverage is greater. Thus, incomplete coverage affects how much of the workforce is observed, but the evidence does not indicate that it drives the firm-level occupational differences used in the analysis.

4.4. Occupation Classification

The measurement begins with an occupational classification determined independently of the accounting outcomes used to evaluate it. Appendix Table A3 summarizes the classification and its alternatives.

For organization capital, I use the task-based classification of Squicciarini and Le Mouél (2012). Their procedure identifies 84 occupations from 11 O*NET work activities associated with

organization building, including developing strategy, organizing and prioritizing work, building and training teams, supervising activities, and communicating across groups. I carry their published occupation set forward to the O*NET-SOC 2019 taxonomy used in the workforce data through official occupational crosswalks.¹³ The resulting baseline classification contains 250 detailed O*NET codes with usable task ratings. Using the published set fixes the occupational boundary outside this study and before any firm-level accounting outcome is observed.

As a sensitivity test, I separately reapply the Squicciarini and Le Mouél selection procedure to contemporary O*NET task ratings, using their original distributional and clustering criteria rather than carrying forward their selected occupations. This re-estimation produces a substantially narrower set of 51 detailed occupations, 35 of which also appear in the baseline classification (Appendix Table A3). The exercise therefore changes the occupational boundary rather than simply remapping the published classification. Section 5.4 examines how the firm-level measures and results change under this independently re-estimated set.

Research occupations are identified separately using the Frascati Manual definition of R&D personnel (OECD 2015). Forty-eight occupations satisfy both the organization-capital and research classifications. For the baseline workforce structure and investment measure, I assign these overlapping occupations to research when the firm reports R&D expense and to organization capital otherwise, as described in Section 3.2. Because this overlap rule uses reported R&D status, it is not used to validate the classification itself. The validation in Section 5.2 instead uses mutually exclusive occupation sets that remove the 48 overlapping occupations from both groups, so

¹³ The original classification uses the 2000 Standard Occupational Classification. I map it to the O*NET-SOC 2019 taxonomy using the Bureau of Labor Statistics 2000-to-2010 and 2010-to-2018 crosswalks, following occupation splits, merges, and group changes across revisions rather than relying on occupation titles.

reported R&D status enters the construction of neither validation measure. Section 5.4 examines alternative treatments of the overlap.

4.5. From Workforce Structure to Investment Fractions

I define organization-capital payroll for each firm-year as the wage bill of employees in the 250 organization-capital occupations, measured in full-time-equivalent terms and adjusted using Revelio's sampling weights. For occupations classified as both organization capital and research, I apply the assignment rule described in Section 3.2. Capital-building workforce structure is the share of SG&A payroll attributable to organization-capital occupations.

Translating workforce activity into an investment fraction requires aligning the workforce measure with the expense reported in the financial statements. SG&A contains both labor and non-labor costs.¹⁴ I adjust observed payroll for workforce coverage and allocate it across expense lines using employees' functional roles, assigning research and development roles to R&D, production roles to cost of sales in goods-producing industries, and the remaining roles to SG&A.¹⁵ I do not construct the labor-core fraction when workforce coverage is below 5 percent. Appendix B states the allocation rules.

The numerator and denominator must refer to the same expense base. I therefore classify each position along two dimensions, namely whether its occupation performs capital-building work and which expense line its function maps to. Capital-building payroll enters the numerator only when

¹⁴ SG&A net of R&D includes advertising, rent, purchased services, and other non-labor costs. I estimate its labor share separately for each firm-year.

¹⁵ Observed payroll is adjusted for workforce coverage so that modeled payroll corresponds to the reported workforce. The adjustment applies symmetrically when modeled employment is below or above reported employment and is subject to fixed bounds. It assumes that observed and unobserved employees have similar average compensation after applying Revelio's occupation-level sampling weights. Section 5.4 examines a construction based on employee counts rather than modeled salaries.

the position is assigned to SG&A. This joint classification separates the economic activity an employee performs from where the associated labor cost is recorded in the financial statements.

The distinction matters empirically. In aggregate, 13 percent of organization-capital payroll is assigned outside SG&A, approximately 10 percentage points to R&D and 3 percentage points to cost of sales in goods-producing industries. The measure therefore captures the capital-building labor contained within SG&A rather than all capital-building labor in the firm. These allocations use common functional rules rather than firms' internal cost accounting, and capital-building labor mapped to R&D or cost of sales lies outside the measure by construction.

I estimate the labor share of SG&A as allocated SG&A payroll divided by SG&A net of R&D, capped at one, as defined in equation (4):

$$LaborShare_{it} = \frac{Payroll_{it}^{SG\&A}}{SG\&A_{it} - R\&D_{it}}.$$

The remaining input is the occupation-specific capital-building task share, τ_o . For each organization-capital occupation, I use O*NET task-frequency ratings to estimate the share of expected task frequency devoted to the eleven organization-building work activities. Occupations without their own frequency ratings, representing 17 percent of organization-capital payroll, receive the mean task share of the nearest rated occupations. Appendix B describes the task mapping, the frequency conversion, and the imputation, which use only O*NET information.

Task frequency serves a different purpose from the Importance-by-Level ratings used to classify occupations. Importance-by-Level identifies which occupations perform organization-building work, whereas task frequency estimates how intensively those occupations perform it. The two O*NET constructs are only weakly correlated. Because O*NET reports frequency categories rather than hours, I interpret the resulting task share as a proxy for the share of working time devoted to capital-building activity rather than as observed time.

Combining workforce composition, task intensity, and the labor share of SG&A gives the baseline labor-core investment fraction:

$$Fraction_{it} = LaborShare_{it} \times \sum_o s_{oit} b_o \tau_o. \quad (6)$$

The summation term measures task-weighted capital-building activity within SG&A labor, and multiplying it by the firm's labor share expresses that activity as a fraction of the reported expense base. The payroll-weighted capital-building task share averages 17 percent across firm-years. Equation (6) therefore makes explicit that the accounting fraction combines measured capital-building activity within the workforce with the composition of the reported expense.

For comparison, I also construct a uniform 20 percent benchmark that assigns the same capital-building share to every organization-capital occupation:

$$Fraction_{it}^{20\%} = 0.20 \times Structure_{it} \times LaborShare_{it}. \quad (7)$$

I refer to the resulting measure as the labor-core investment fraction, defined as the estimated own-account labor investment in organization capital as a fraction of SG&A net of R&D. Under the uniform benchmark, changing the common task share scales the fraction proportionally but does not change firm rankings. Appendix Table A1 provides the variable definitions.

4.6. Estimated Investment Fractions

Under the baseline task-frequency specification, the labor-core fraction averages 4.8 percent of SG&A net of R&D, with a median of 4.4 percent and an SG&A-weighted aggregate of 4.1 percent (not tabulated). Under the uniform 20 percent benchmark, the mean, median, and SG&A-weighted aggregate are 5.5, 5.0, and 4.9 percent. Table 2 reports the mean components (Panel A) and the mean fractions of both specifications (Panel B).

[See Table 2]

The task-frequency calibration affects the level of the estimate much more than the relative ordering of firms. The payroll-weighted capital-building task share averages 17.5 percent (Table 2, Panel A), close to the uniform 20 percent benchmark. The baseline and uniform specifications produce nearly identical firm rankings, with a Spearman correlation of 0.97. The firm-level variation used in the paper is therefore not driven by whether capital-building time is estimated from O*NET task frequencies or imposed at the conventional 20 percent rate.

Figure 2 illustrates why the resulting investment fraction should nevertheless be distinguished from the underlying capital-building activity. Among sixty large U.S. firms, some have a large share of SG&A labor in capital-building occupations but a relatively low labor share of SG&A, while others have the opposite combination. The dashed curves connect combinations that imply approximately the same labor-core fraction at a common task intensity. Firms can therefore arrive at similar investment fractions through very different combinations of capital-building workforce structure and expense composition. Section 7 examines this distinction formally.

[See Figure 2]

5. VALIDATING THE MEASUREMENT

Because accounting outcomes do not enter the construction of the workforce measure, they provide independent evidence with which to evaluate it. No single outcome benchmarks the full construction, so I evaluate three links separately. They are the measurement of labor cost, the boundary between organization building and formal research, and the translation of workforce payroll into a share of reported expense. Section 5.1 compares modeled payroll with independently reported staff expense. Section 5.2 tests whether the occupational classification distinguishes organization building from formal research. Section 5.3 evaluates the expense translation against disclosed software capitalization by estimating the mapping in an initial period, freezing it, and

evaluating it on subsequent disclosures that enter neither the workforce construction nor the mapping's estimation. Section 5.4 examines sensitivity to alternative construction choices, including the occupational classification itself.

5.1. Validating Payroll Measurement

The investment measure relies on model-predicted salaries and workforce coverage, so the first validation asks whether the resulting payroll estimates correspond to labor costs reported independently in firms' financial statements. I compare modeled payroll with reported staff expense (Compustat XLR) for 5,830 firm-years at 861 firms that disclose both measures. Because most U.S. filers do not separately report staff expense, these disclosures provide an independent accounting benchmark for the payroll input to the investment measure.

[See Table 3]

Table 3 shows close correspondence between modeled payroll and reported staff expense. The slope of reported staff expense on modeled payroll is 0.951 overall and 0.967 with industry-year fixed effects. The latter comparison is particularly relevant because the paper's principal firm-level analyses exploit differences among firms within the same industry-year. The level and rank correspondence are also high, with a Pearson correlation in logs of 0.957 and a Spearman correlation of 0.961. Thus, the workforce data recover not only the ordering of firms by labor cost but also approximately its accounting magnitude.

Chen et al. (2026) similarly use Revelio data to estimate firm-level labor costs, but their construction anchors total employment to Compustat-reported headcount while using Revelio to estimate compensation and workforce composition. I instead retain Revelio-based employment so that aggregate payroll is constructed from the same occupation-level workforce used to measure capital-building activity. Both compensation and employment coverage are therefore estimated

from the workforce data rather than mechanically anchored to reported headcount. The close correspondence with independently reported staff expense consequently evaluates both inputs underlying modeled payroll. Section 4.3 separately examines workforce coverage, and Section 5.4 evaluates the measure in high-coverage subsamples.

5.2. Distinguishing Organization Capital from Research Activity

The classification must distinguish organization-building activity from formal research because some technical occupations satisfy both definitions (Section 3.2), while the organization-capital measure excludes research by design. Reported R&D provides an independent accounting benchmark. ASC 730 requires R&D costs to be expensed as incurred and the amount charged to expense to be disclosed. If the occupational classification distinguishes the underlying activities, research-intensive workforce composition should be more strongly associated with the subsequent onset of reported R&D than organization-building workforce composition.

The sample consists of firm-years from 2005 through 2021 without reported R&D expense whose three following fiscal years are all observed, and the outcome indicates whether the firm reports positive R&D expense in any of those three years. Requiring the full window prevents firm-years near the end of the sample, whose later years are unobserved, from being counted as non-initiators. For this validation only, I use mutually exclusive occupational groups. The 48 occupations that satisfy both definitions are excluded from both groups, so no occupation enters both measures, and reported R&D status is not used in constructing either. I further decompose the remaining organization-capital payroll into three activity classes (training and development, management and organization, and customer and external relations) using the underlying O*NET work activities. Each occupation's payroll is allocated across the classes in proportion to its activity scores. The research workforce share and each organization-capital activity class enter separate

regressions, standardized within industry-year, with industry-year fixed effects and controls for firm size and sales growth.

[See Table 4]

Table 4 shows clear separation between research and organization-building activity. A one-standard-deviation increase in the research workforce share is associated with a 1.18-percentage-point higher probability of initiating reported R&D within three years ($t=4.35$). The organization-capital coefficients are substantially lower. Management and customer-related activity are statistically indistinguishable from zero, while training-and-development activity is negatively associated with R&D initiation ($t=-2.11$). The validation prediction concerns the distinction between research and organization-building activity rather than zero coefficients across all organization-capital classes. Consistent with that prediction, the research coefficient exceeds each of the three organization-capital coefficients, with difference-test t -statistics ranging from 2.76 to 3.79 and a joint p -value below 0.001. Thus, occupational composition distinguishes the research activity the measure excludes from the organization-building activity it is designed to capture, rather than simply identifying firms that are generally intangible-intensive. The ordering is the same when base years are restricted to 2005 through 2021 without the window requirement and when base years through 2024 are included with unobserved years coded as non-initiation (Table 4 note). Including the 48 overlapping occupations in both groups also produces the same ordering and coefficient differences (Appendix Table A4).

5.3. Validating the Expense Translation

Software-development disclosures provide an independent benchmark for the step that translates workforce payroll into a share of reported expenditure. Firms independently disclose the amount of qualifying software-development spending they capitalize, making the corresponding

accounting investment fraction observable (Aboody and Lev 1998).¹⁶ The test serves a different purpose from the R&D test in Section 5.2. It does not validate the organization-capital classification. It asks whether the expense translation is informative when the capitalized amount is independently observed.

[See Table 5]

I identify 109 sample firms with usable software-development disclosures, contributing 574 firm-years with disclosed capitalization intensity, defined as capitalized software-development additions divided by development spending. Of these, 534 firm-years meet the requirements for the training-and-holdout validation reported in Table 5. I estimate the expense translation in an initial period, freeze the estimated relation, and evaluate it in a subsequent holdout period. The relation remains informative on disclosures that played no role in its estimation. In Panel A, the holdout-period slope is 6.09 ($t=2.89$), predicted and disclosed capitalization intensities have a correlation of 0.44, and the fitted model has lower prediction error than the training-period mean benchmark.

Panel B identifies which component carries this result. On the common sample, the labor share of reported SG&A alone produces a holdout correlation of 0.51 and a root-mean-square error of 0.175, compared with 0.44 and 0.179 for the labor-core fraction. Capital-building workforce structure and task-weighted activity have holdout correlations of 0.05 and -0.02 and essentially match the prediction error of the training-period mean. The software test therefore supports the

¹⁶ The disclosures cover software developed for sale under ASC 985-20 and internal-use and hosted software under ASC 350-40. ASC 985-20 requires capitalization of qualifying development costs after technological feasibility, while ASC 350-40 requires capitalization during the application-development stage. The development-spending base in the software sample is separated from SG&A in the empirical construction because ASC 985-20 filers typically expense qualifying development within R&D. The software tests therefore use the fraction built on the reported SG&A line rather than SG&A net of R&D.

labor-share scaling used in the second-stage expense translation. It does not validate the identification of capital-building activity, which the R&D test in Section 5.2 evaluates.

An international setting provides a separate check on the connection between workforce payroll and capitalized development cost. IAS 38 requires qualifying development expenditure to be capitalized once its recognition criteria are met (IASB 2004). Linking annual reports from listed firms in twelve European countries to the workforce data, modeled payroll closely tracks audited employee benefits expense reported under IAS 19 (IASB 2011), and development-occupation payroll predicts capitalized development additions in both a primary sample and a holdout sample of countries (Appendix Table A6). The relation therefore appears in a different accounting regime, set of firms, and capitalization setting rather than being specific to the U.S. software sample.

5.4. Measurement Robustness

I examine sensitivity to the principal construction choices. Replacing salary weights with employee counts, changing the treatment of occupations that overlap the organization-capital and research classifications, and replacing occupation-specific task shares with the uniform 20 percent convention produce similar firm rankings and principal results. The main conclusions therefore do not depend on modeled compensation, the research-overlap rule, or the task-intensity specification.

The results are also robust to re-estimating the occupational classification itself. Reapplying the Squicciarini and Le Mouél procedure to contemporary O*NET task data produces a substantially narrower set of 51 occupations, compared with 250 in the mapped published classification (Appendix Table A3). Despite the different occupational boundary, the central evidence on firm-specific variation, persistence, and the listing-period build-out remains similar.¹⁷

¹⁷ Under the re-estimated classification, 80 percent of workforce-structure variation remains within industry-year, the within-industry component has an annual autocorrelation of 0.91, and both workforce structure and net capital-building hiring increase significantly in the listing year ($t=3.66$ and 13.23). The relation between workforce structure

I also examine sensitivity to workforce coverage. The principal results are similar in high-coverage firm-years and in later years, when coverage is greater. Coverage changes do not account for the software-validation result. In Appendix Table A5, changes in coverage contain no incremental information for disclosed software capitalization once the labor-core fraction is included, while the fraction remains predictive. Results are also similar after excluding firms with potentially ambiguous parent assignments. Taken together, the robustness tests indicate that the evidence on firm specificity, persistence, and organizational build-out is not driven by compensation weights, research-overlap rules, task-intensity assumptions, workforce coverage, parent assignment, or the particular occupational boundary used in the baseline classification.

6. FIRM-LEVEL CAPITAL-BUILDING WORKFORCE STRUCTURE

The validation shows that the payroll input, the research boundary, and the labor-share scaling track independent evidence. This section asks what workforce structure reveals about capital-building activity at the firm level. If the measure captures economically meaningful organizational structure rather than transitory workforce noise, differences among firms in the same industry should persist. If that structure reflects capital building rather than simply a fixed firm type, it should also change when firms build organizational capacity. Sections 6.1 and 6.2 examine these implications.

6.1. Firm-Specific Variation and Persistence

Figure 3 decomposes variation in capital-building workforce structure and the resulting labor-core fraction into differences across industry-year cells and differences among firms within the same industry-year. Eighty-eight percent of the variation in workforce structure occurs among

and subsequent net capital-building hiring does not carry over to the re-estimated classification (Appendix Table A10, Panel B).

firms within the same industry-year. Translating workforce structure into the labor-core fraction through each firm's labor share of SG&A preserves substantial firm-level variation, with 75 percent of the resulting fraction varying within industry-year. The dominant variation in both measures therefore occurs among firms that an industry-level capitalization approach would assign to the same industry-year environment.

[See Figure 3]

These firm-specific differences are also highly persistent. Table 6 shows that the annual autocorrelation of the within-industry component is 0.92 in the matched sample and 0.90 in the full sample. Five years later, 69 percent of firms initially in the top within-industry quintile remain there, and 74 percent of firms initially in the bottom quintile remain in the bottom, compared with 20 percent under random assignment. Firm means account for 80 percent of the variance of the within-industry component in the matched sample. The within-industry variation is therefore predominantly a durable firm characteristic rather than transitory measurement noise.

[See Table 6]

This persistence also shows why an industry average poorly represents individual firms. Replacing each firm's workforce structure with the leave-one-out mean of the other firms in its industry-year yields a proxy that is approximately proportional to the firm-level measure, with a slope of 0.92, but explains only 10 percent of its variation in the full sample and 14 percent in the investment-fraction sample (Table 6, Panel C). The industry proxy therefore discards most of the differences among firms within an industry-year, which is the variation that dominates the firm-level measure.

Section 6.2 therefore examines whether workforce structure changes during an identifiable period of organizational build-out.

6.2. Application: The Timing of the Organizational Build-Out

Persistence alone does not establish that workforce structure reflects capital building rather than a stable firm type. The years around listing provide a setting in which to observe the buildup of organizational capacity. Because employment histories extend to years before a firm's first Compustat observation, I can follow the workforce both before and after listing. Figure 4 traces workforce structure and net capital-building hiring from three years before listing through five years after, relative to the same firm's years six or more after listing. The estimates use calendar-year workforce cells with firm and calendar-year fixed effects (Table 7).

[See Figure 4]

[See Table 7]

Relative to the same firm's mature years, the capital-building share of SG&A labor is 2.8, 2.6, and 3.1 percentage points higher three, two, and one years before listing ($t=3.23$, 3.27 , and 4.22), remains similarly elevated in the listing year (2.8 points), and declines steadily afterward. Net capital-building hiring is highest three years before listing, at 27.1 percentage points of lagged payroll ($t=18.61$), exceeds its listing-year level of 22.5 points ($t=3.25$ on the difference), and falls to 5.2 points two years after listing. Capital-building hiring before departures follows the same pattern, at 21.4 percentage points three years before listing and 19.3 points in the listing year ($t=14.78$ and 17.97). The workforce structure is therefore already being built well before listing, and listing occurs late in the organizational build-out rather than initiating it.

The pre-listing estimates rely on employment histories reconstructed from current profiles, which may disproportionately represent managerial and professional employees in earlier years. The data do not show that pattern overall. The payroll share of managerial and professional occupations that are neither capital-building nor research is 1.3 percentage points below its mature

level in each of the three pre-listing years (*t-stat* between -5.04 and -4.18), while the capital-building share of total payroll is 2.1 to 2.8 points above it, and the differences are significant in each year (*t-stat* between 4.13 and 6.05). The pre-listing elevation also persists when workforce composition is measured by headcount rather than payroll (*t-stat* between 2.99 and 4.79, Appendix Table A7). The pre-listing pattern is therefore specific to capital-building occupations rather than a general feature of reconstructed professional-workforce histories.

A complementary fiscal-year analysis begins at listing, when reported employment is available and workforce coverage can be measured. The listing-year increase in workforce structure is 2.3 percentage points and remains after controlling for coverage, restricting the sample to firms with stable coverage, and measuring workforce structure by headcount rather than payroll (Appendix Table A8).

The build-out also does not simply reflect the reporting, internal-control, and compliance staff associated with becoming a public company. Excluding finance and compliance occupations leaves 90 percent of the listing-year increase in the capital-building share of payroll. The increase appears separately in training and development, management and organization, and customer and external relations, and most of the increase in capital-building hiring also occurs outside finance and compliance occupations (Appendix Table A9). Together, the event-time evidence shows that the persistent workforce structure documented in Section 6.1 is not merely a fixed firm characteristic. It is visibly accumulated during a period of organizational expansion and subsequently declines as firms mature. Appendix Table A10 examines whether workforce structure predicts subsequent capital-building hiring. The relation appears under the baseline classification but not the independently re-estimated occupation set, so I do not treat it as central evidence.

The listing setting illustrates why observing investment activity separately from reported expenditure matters. A measure constructed from public financial statements begins when the firm enters the accounting data and therefore cannot identify when the organizational build-out occurred. The workforce histories show that net capital-building hiring peaks three years before listing and that capital-building workforce structure is already elevated at that point. The measurement approach therefore changes the timing conclusion. Listing marks a late stage of the organizational build-out rather than its beginning. Studies that begin at listing would therefore misdate the formation of organization capital if they inferred it only from post-listing expenditures.

7. FROM ACTIVITY TO ACCOUNTING INVESTMENT

Translating workforce activity into an investment fraction introduces a second source of firm-level variation. Workforce structure records where capital-building labor resides, whereas the labor share measures how much of SG&A is labor. The distinction matters because an accounting investment fraction combines these two dimensions rather than measuring capital-building activity alone. This section first separates the two components and then examines the scope of the labor-based measure relative to broader investment estimates and existing capitalization measures.

7.1. Activity and Expense Composition

The labor-core fraction combines task-weighted capital-building activity within SG&A labor with the labor share of SG&A. Because the task-frequency adjustment leaves firm rankings almost unchanged (Spearman correlation 0.97, Section 4.6), I use workforce structure to represent differences in occupational composition and the labor share to represent differences in the composition of SG&A. As developed in Section 4.5,

$$Fraction_{it} = \left(\sum_o s_{oit} b_o \tau_o \right) \times LaborShare_{it}.$$

The fraction's close relation to the labor share follows from one mechanical fact and one economic fact. Mechanically, the labor share varies more across firms than workforce structure, with coefficients of variation of 0.52 and 0.30, so it contributes more to cross-firm variation in their product. Economically, the components move in opposite directions. Their Spearman correlation is -0.247 , indicating that firms with a larger capital-building share of SG&A labor tend to have less labor-intensive SG&A. That negative relation offsets the activity signal. Together, the two facts explain why the labor-core fraction has a Spearman correlation of 0.872 with labor share but only 0.129 with workforce structure.

The accounting translation therefore changes the economic object, not merely its units. Workforce structure and task-weighted activity answer how much capital-building work the firm's SG&A workforce performs. The labor-core fraction answers how much of reported SG&A is own-account labor investment. Firms with similar activity can have different fractions because their expense bases differ, while similar fractions can arise from different combinations of activity and labor intensity. Workforce data make those components separately observable.

The appropriate measure depends on the economic question. Workforce structure and task-weighted activity are appropriate when the question concerns differences in capital-building work across firms or over time because they are not mechanically affected by the labor share of SG&A. The labor-core fraction is appropriate when the question concerns the amount of reported SG&A net of R&D attributable to internally performed capital-building labor. Its 4.8 percent average should therefore be interpreted as the own-account labor component of a potential expense adjustment, not as total intangible investment. Broader expenditure-based measures remain appropriate when the question concerns the full resource cost or stock of intangible capital,

including purchased services and other non-labor inputs. The measures are therefore complements, and the appropriate choice follows from the scope of the question.

This distinction also disciplines comparisons with uniform or industry-level capitalization fractions. Such parameters treat a dollar of SG&A alike regardless of what it buys, whereas the labor-core fraction varies systematically with labor intensity. These differences should not be read as direct errors in measures of total intangible investment because the labor-core fraction has a narrower scope. To the extent that non-labor SG&A contains complementary investment, the labor-core measure will be least complete where non-labor SG&A is largest. The data establish the negative relation between the components, but they do not distinguish complementary non-labor spending from differences in firms' business models.

7.2. Scope and Relation to Existing Capitalization Measures

The level of the labor-based measure must be interpreted in light of its scope. Own-account labor investment is the narrowest of the three concepts defined in Section 3.2. It captures labor directly attributable to capital-building activity but excludes supporting non-labor and purchased inputs. Some capital-building labor also falls outside SG&A, primarily in R&D and, for goods-producing firms, in cost of sales (Section 4.5). Existing capitalization parameters generally target a broader investment concept.

Across the 54,291 firm-years with a labor-core fraction, the baseline measure averages 4.8 percent of SG&A net of R&D, compared with the 30 percent convention of Peters and Taylor (2017) and the 52.4 percent average parameter of Iqbal et al. (2025). Table 2, Panel B shows how differences in scope account for much of this gap. Treating all SG&A payroll of employees in capital-building occupations as investment raises the mean to 27.5 percent. Extending the calculation to associated non-labor SG&A, using each firm's ratio of non-labor to labor SG&A as

in the national-accounts treatment of own-account investment (BEA 2024), raises it to 48.8 percent. These calculations are not alternative estimates. They show how progressively broader definitions of investment move the workforce-based measure toward capitalization parameters designed to capture broader resources.

Separating the two components also clarifies how the workforce measures relate to published capitalization measures. Four of five published measures are strongly positively related to SG&A intensity (Appendix Table A11, Panel A). Capital-building workforce structure is weakly positively related to SG&A intensity and to each of these four measures, with rank correlations of 0.15 and of 0.08 to 0.17. The labor-core fraction has the opposite sign, with rank correlations of -0.19 and of -0.18 to -0.27 , and it remains negatively related to each published measure after holding SG&A intensity fixed. The fraction differs from workforce structure mainly through the labor share of SG&A, because the task-frequency weighting barely changes firm rankings (Section 4.6). The reversal therefore reflects expense composition rather than capital-building activity. Measured capital-building activity is broadly consistent with what expenditure-based measures capture, whereas the composition of the expense base that a capitalization fraction embeds is not. Appendix Table A11, Panel B further compares the workforce measures with the industry-year capitalization parameters of Iqbal et al. (2025).

The difference in levels therefore reflects a difference in economic objects rather than a mere disagreement over the appropriate capitalization rate. Expenditure-based measures capture a broader set of resources devoted to intangible investment, whereas the workforce measure isolates internally performed labor and reveals firm-level differences in capital-building activity that industry-level expenditure parameters cannot capture.

7.3. Application: Classifying Firms by Capital-Building Workforce Intensity

The distinction between activity and expenditure matters when researchers use these measures to identify intensive organization builders. Table 8 compares firms' within-industry-year rankings by capital-building workforce structure with rankings by the labor-core fraction, by SG&A intensity, which corresponds to a uniform capitalization rate, and by five published capitalization measures. First, the classifications overlap only modestly. Of the firm-years in the top quintile of workforce structure, only 22 to 36 percent are also in the top quintile of SG&A intensity or a published measure, compared with 20 percent under unrelated rankings (Panel A). Second, task-weighted capital-building activity closely reproduces the workforce-structure ranking, with a top-quintile overlap of 73 percent, so the disagreement does not reflect an unstable activity measure. Third, scaling own-account labor investment by lagged assets, the denominator of the published measures, raises its overlap with SG&A intensity and four of the five published measures to 55 to 65 percent (Panel B). The baseline rankings therefore differ both because they capture different workforce composition and because expenditure-based measures weight the scale of resources.

[See Table 8]

The discordant firms also differ systematically (Appendix Table A12). Firms with high workforce structure but low SG&A intensity are of typical size and are approximately at break-even. Firms with high SG&A intensity but low workforce structure are smaller, have less labor-intensive SG&A, and report losses. These results do not indicate that expenditure-based rankings are incorrect. Rather, the measures answer different questions. Expenditure-based measures identify firms with high spending relative to assets, whereas workforce structure identifies firms whose SG&A labor is concentrated in capital-building occupations. The appropriate ranking therefore depends on whether the research question concerns resource intensity or workforce

activity. Because most variation in workforce structure occurs within industry-year, an industry-level capitalization parameter cannot recover the activity-based classification.

8. CONCLUSION

This paper develops a firm-level measure of capital-building activity from tasks performed by a firm's own employees and translates the cost of that activity into own-account labor investment. Occupational task data distinguish labor devoted to building longer-lived organizational capabilities from labor supporting current operations, allowing activity to be observed before it is expressed in accounting units. The approach therefore begins with the underlying activity and then connects its labor cost to the expense in which that cost is reported.

Capital-building workforce structure is primarily a persistent firm characteristic rather than an industry characteristic, yet it is not fixed. Its formation is visible as firms build organizational capacity around listing. Three accounting settings evaluate different links in the measurement. Reported staff expense supports the modeled payroll input. R&D initiation supports the boundary between organization building and formal research. Disclosed software capitalization supports the labor-share scaling used in the second-stage expense translation. Its predictive content comes from the labor share rather than workforce structure, so it does not validate the identification of capital-building activity. Together, these results support using workforce information to identify firm-level investment activity that conventional expense classifications do not separately reveal.

The usefulness of the workforce measures lies in questions for which reported expenditures do not reveal the underlying activity. Workforce structure identifies which firms perform more capital-building work and allows that activity to be observed before it appears in public financial statements. The listing setting illustrates this advantage. Net capital-building hiring peaks three years before listing, and capital-building workforce structure is already elevated at that point. A

measure based on public-company expenditures would see only the later part of this process and could mistake listing for the start of the build-out rather than a late stage in it. The choice of measure also changes which firms are classified as intensive capital builders. Within industry-year, most firms in the top quintile of workforce structure are not in the top quintile of expenditure-based measures. This divergence does not arise from instability in the activity measure, because task-weighted activity produces a similar ranking. Rankings move toward expenditure-based measures only when own-account labor investment is scaled by assets, showing that workforce composition and resource scale are distinct dimensions of investment intensity.

The labor-core fraction serves a different purpose. It translates observed capital-building activity into the portion of reported SG&A attributable to internally performed capital-building labor. It is not a complete measure of intangible investment. Firms with similar workforce structures can therefore have different labor-core fractions when the labor share of SG&A differs. Instead, the fraction supplies the own-account labor component needed when the question concerns the composition of reported expense. Keeping these uses separate prevents the fraction from being read as a ranking of firms' investment activity. Broader expenditure-based measures remain appropriate when the question concerns the full resource cost or stock of intangible capital.

These distinctions matter for financial reporting. For public business entities that present SG&A as a relevant expense caption, ASU 2024-03 will make the employee compensation included in that caption observable (FASB 2024). That disclosure reveals the labor share needed for the expense translation, but not how much of that labor performs capital-building activity. Workforce information supplies the missing activity component.

The broader implication is that addressing the measurement challenge should precede the question of recognition. Before asking whether internally generated intangibles should be

recognized as assets, one must identify the activities that constitute investment and distinguish them from current operations. Conventional expense classifications show where costs are reported, while workforce information reveals which economic activity produced them. Separating activity from expense composition makes the investment activity within reported expenses observable before recognition is considered.

REFERENCES

- Aboody, D., and B. Lev. 1998. The value relevance of intangibles: The case of software capitalization. *Journal of Accounting Research* 36 (Supplement): 161–191.
- Babina, T., A. X. He, and R. Jiang. 2026. Canaries in the gold mine: Early productivity gains from artificial intelligence creating organization capital. NBER Working Paper No. 35684.
- Baker, A., D. F. Larcker, C. McClure, D. Saraph, and E. M. Watts. 2024. Diversity washing. *Journal of Accounting Research* 62 (5): 1661–1709.
- Barker, R., A. Lennard, S. Penman, and A. Teixeira. 2022. Accounting for intangible assets: Suggested solutions. *Accounting and Business Research* 52 (6): 601–630.
- Bureau of Economic Analysis (BEA). 2024. *Concepts and Methods of the U.S. National Income and Product Accounts*. Washington, DC: U.S. Department of Commerce.
- Cai, W., A. Prat, and J. Yu. 2026. Measuring organizational capital. *Journal of Accounting Research*. <https://doi.org/10.1111/1475-679x.70073>
- Chen, Y., K. Jain, N. Li, and S. Rajgopal. 2026. Decision usefulness of disaggregated labor costs: Ex-ante evidence and implications for the FASB’s expense disclosure mandate. Columbia Business School Research Paper No. 5050805.
- Corrado, C., C. Hulten, and D. Sichel. 2005. Measuring capital and technology: An expanded framework. In *Measuring Capital in the New Economy*, edited by C. Corrado, J. Haltiwanger, and D. Sichel, 11–46. Chicago: University of Chicago Press.
- Corrado, C., C. Hulten, and D. Sichel. 2009. Intangible capital and U.S. economic growth. *Review of Income and Wealth* 55 (3): 661–685.
- Eisfeldt, A. L., and D. Papanikolaou. 2013. Organization capital and the cross-section of expected returns. *Journal of Finance* 68 (4): 1365–1406.
- Enache, L., and A. Srivastava. 2018. Should intangible investments be reported separately or commingled with operating expenses? New evidence. *Management Science* 64 (7): 3446–3468.
- Ewens, M., R. H. Peters, and S. Wang. 2025. Measuring intangible capital with market prices. *Management Science* 71 (1): 407–427.
- Fama, E. F., and K. R. French. 1997. Industry costs of equity. *Journal of Financial Economics* 43 (2): 153–193.
- Fedyk, A., and J. Hodson. 2023. Trading on talent: Human capital and firm performance. *Review of Finance* 27 (5): 1659–1698.
- Financial Accounting Standards Board (FASB). 2024. Accounting Standards Update No. 2024-03, Income Statement, Reporting Comprehensive Income, Expense Disaggregation Disclosures (Subtopic 220-40): Disaggregation of Income Statement Expenses. Norwalk, CT: FASB.
- Hampole, M., D. Papanikolaou, L. D. W. Schmidt, and B. Seegmiller. 2025. Artificial intelligence and the labor market. NBER Working Paper No. 33509.
- He, B., L. I. Mostrom, and A. Sufi. 2024. Investing in customer capital. NBER Working Paper No. 33171.

- International Accounting Standards Board (IASB). 2004. IAS 38 Intangible Assets. London: IFRS Foundation.
- International Accounting Standards Board (IASB). 2011. IAS 19 Employee Benefits. London: IFRS Foundation.
- Iqbal, A. 2026. When capital walks out the door: Measuring the depreciation of organization capital from workforce microdata. Working paper, Arizona State University. Available at SSRN: <https://ssrn.com/abstract=7400239>.
- Iqbal, A., S. Rajgopal, A. Srivastava, and R. Zhao. 2025. A better estimate of internally generated intangible capital. *Management Science* 71 (1): 731–752.
- Kanodia, C., H. Sapra, and R. Venugopalan. 2004. Should intangibles be measured: What are the economic trade-offs? *Journal of Accounting Research* 42 (1): 89–120.
- Kermani, A., and Y. Ma. 2023. Asset specificity of nonfinancial firms. *Quarterly Journal of Economics* 138 (1): 205–264.
- Kothari, S. P., T. E. Laguerre, and A. J. Leone. 2002. Capitalization versus expensing: Evidence on the uncertainty of future earnings from capital expenditures versus R&D outlays. *Review of Accounting Studies* 7 (4): 355–382.
- Le Mouél, M., and M. Squicciarini. 2015. Cross-country estimates of employment and investment in organisational capital: A task-based methodology using PIAAC data. *OECD Science, Technology and Industry Working Papers* 2015/08. <https://doi.org/10.1787/5jrs3smfgcjb-en>
- Lev, B., and T. Sougiannis. 1996. The capitalization, amortization, and value-relevance of R&D. *Journal of Accounting and Economics* 21 (1): 107–138.
- Lev, B., and P. Zarowin. 1999. The boundaries of financial reporting and how to extend them. *Journal of Accounting Research* 37 (2): 353–385.
- Li, Q., B. Lourie, A. Nekrasov, and T. J. Shevlin. 2022. Employee turnover and firm performance: Large-sample archival evidence. *Management Science* 68 (8): 5667–5683.
- National Center for O*NET Development. 2020. *O*NET 25.1 Database*. Raleigh, NC: U.S. Department of Labor, Employment and Training Administration. <https://www.onetcenter.org/database.html>
- OECD. 2015. *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264239012-en>
- Payne-Mann, C., and R. G. Sloan. 2026. Intangible asset specificity. Working paper, University of Southern California (SSRN 5737322).
- Peters, R. H., and L. A. Taylor. 2017. Intangible capital and the investment-q relation. *Journal of Financial Economics* 123 (2): 251–272.
- Piekkola, H. 2016. Intangible investment and market valuation. *Review of Income and Wealth* 62 (1): 28–51.
- Rajgopal, S., E. Basirian, A. Iqbal, and A. Srivastava. 2026. Reassessed earnings with capitalized intangibles. Working paper, Arizona State University, University of Calgary, and Columbia University.
- Rajgopal, S., A. Srivastava, and R. Zhao. 2023. Do digital technology firms earn excess profits? Alternative perspectives. *The Accounting Review* 98 (4): 321–344.

- Squicciarini, M., and M. Le Mouél. 2012. Defining and measuring investment in organisational capital: Using US microdata to develop a task-based approach. *OECD Science, Technology and Industry Working Papers* 2012/05. <https://doi.org/10.1787/5k92n2t3045b-en>
- Srivastava, A. 2014. Why have measures of earnings quality changed over time? *Journal of Accounting and Economics* 57 (2): 196–217.
- Srivastava, A. 2023. Trivialization of the bottom line and losing relevance of losses. *Review of Accounting Studies* 28 (3): 1190–1208.

TABLES AND FIGURES

Table 1. Sample Construction and Descriptive Statistics

This table reports the analysis samples and descriptive statistics. The sample consists of U.S.-headquartered public firms whose Revelio workforce records are linked to Compustat, for fiscal years 2005 through 2024. Regulated utilities and financial firms (Fama-French 48 industries 31 and 44 through 47) are excluded. Panel A reports the sample at each construction step, and Panel B reports descriptive statistics for firm-years with a labor-core investment fraction. Dollar amounts are in millions of U.S. dollars and employees are in thousands. R&D intensity is R&D expense divided by total assets, with missing R&D set to zero. Revenue growth is sales divided by prior-year sales, minus one. Firm age is the number of years since the firm's first Compustat fiscal year. The labor-core investment fraction is estimated own-account labor investment in organization capital divided by SG&A net of R&D. The research workforce share is payroll in Frascati Manual research occupations divided by total payroll. Workforce coverage is weighted Revelio full-time-equivalent employment divided by reported employment and can exceed one because the two employment measures are constructed differently. Continuous variables are winsorized at the 1st and 99th percentiles within fiscal year.

Panel A: U.S.-headquartered analysis sample

Sample	Firms	Firm-years
U.S.-headquartered panel with workforce and accounting data	7,969	75,893
With measurable workforce structure	7,930	75,440
With a labor-core investment fraction	6,258	54,291

Panel B: Descriptive statistics for the estimation sample

Variable	N	Mean	SD	P25	Median	P75
<i>Firm characteristics</i>						
Total assets	54,291	3,433	10,657	49	339	1,801
Revenue	54,291	2,801	9,032	27	261	1,476
Market capitalization	50,236	5,146	18,393	58	411	2,228
Employees (thousands)	50,894	6.74	18.56	0.11	0.76	4.30
SG&A net of R&D	54,291	472	1,349	14	67	297
R&D intensity	54,268	0.106	0.252	0.000	0.014	0.104
Revenue growth	45,364	0.189	0.750	-0.044	0.066	0.216
Firm age	54,291	16.6	13.1	5.0	14.0	26.0
<i>Workforce-based measures</i>						
Labor-core investment fraction	54,291	0.048	0.029	0.026	0.044	0.067
Research workforce share of payroll	54,291	0.289	0.180	0.138	0.267	0.423
Workforce coverage ratio	54,291	1.360	1.521	0.580	1.012	1.526

Table 2. Components and Scope of the Workforce-Based Investment Measure

This table reports the components and scope of the labor-core investment fraction for the 54,291 firm-years with a nonmissing fraction from fiscal 2005 through 2024. Panel A reports firm-year means of the components. The capital-building share of SG&A labor is organization-capital payroll allocated to SG&A divided by total SG&A payroll. The labor share of SG&A expense is coverage-adjusted SG&A payroll divided by SG&A net of R&D, capped at one. The capital-building task share is the payroll-weighted O*NET task-frequency share of the firm's organization-capital occupations. Memorandum rows report the share of organization-capital payroll assigned outside SG&A and its allocation to research and cost of sales. Panel B reports mean investment fractions under progressively broader scope. The baseline uses occupation-specific O*NET task-frequency shares, and the uniform benchmark assigns a 20 percent capital-building share to every organization-capital occupation. Treating all payroll in capital-building occupations as investment removes the task-intensity adjustment, and the own-account full-cost calculation also assigns non-labor SG&A to capital building in proportion to labor. These broader calculations illustrate scope and are not alternative estimates. The Peters and Taylor (2017) convention and the average of 1,927 replicated Iqbal et al. (2025) industry-year parameters are shown for comparison.

Panel A: Components (percent)

Component	Mean
Capital-building share of SG&A labor	48.7
Labor share of SG&A expense	59.2
Capital-building task share (O*NET task frequency)	17.5
<i>Memorandum (aggregate share of organization-capital payroll)</i>	
Organization-capital payroll assigned outside SG&A	13.2
Organization-capital payroll assigned to research	10.2
Organization-capital payroll assigned to cost of sales	2.9

Panel B: Investment fraction under alternative scope

Measure	Percent of SG&A net of R&D
<i>Workforce-based estimates</i>	
Baseline own-account labor investment (O*NET task-frequency estimate)	4.8
Uniform 20 percent benchmark	5.5
All capital-building SG&A payroll treated as investment, labor cost only	27.5
Own-account full-cost calculation under proportional non-labor allocation	48.8
<i>Prior-research benchmarks</i>	
Peters and Taylor (2017) convention	30.0
Iqbal et al. (2025), replicated average parameter	52.4

Table 3. Validation Against Reported Staff Expense

This table compares modeled payroll with independently reported staff expense (Compustat XLR) for 5,830 firm-years at 861 firms with both measures available from fiscal 2005 through 2024. Modeled payroll is adjusted for workforce coverage before comparison with firm-wide staff expense. Panel A reports the Pearson correlation in logs and the Spearman rank correlation. Panel B reports regressions of log reported staff expense on log modeled payroll, estimated pooled and with industry-year fixed effects, where an industry-year is a Fama-French 48 industry in a fiscal year. Both logs are winsorized at the 1st and 99th percentiles within fiscal year. Standard errors are clustered by firm. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively, using two-tailed tests.

Panel A: Association between modeled payroll and reported staff expense

Measure or specification	Estimate
Pearson correlation, logs	0.957
Spearman correlation	0.961

Panel B: Reported staff expense on modeled payroll

Specification	Estimate	S.E.
Pooled	0.951***	0.010
Within industry-year	0.967***	0.013

Table 4. Research and Organization-Capital Occupations and R&D Initiation

This table reports associations between mutually exclusive workforce activity classes and the start of reported R&D expense (ASC 730). The sample is firm-years from fiscal 2005 through 2021 without reported R&D expense whose three following fiscal years are all observed in the panel, so the outcome is never coded from an unobserved year. The outcome equals one if the firm reports positive R&D expense in any of fiscal years $t+1$ through $t+3$, which occurs in 4.1 percent of the sample. The 48 occupations that belong to both the organization-capital set and the Frascati research set are excluded from both sides, so no occupation counts toward both measures and reported R&D status enters the construction of neither. In aggregate they carry 14.0 percent of research payroll and 9.9 percent of organization-capital payroll. The remaining organization-capital payroll is allocated across three activity classes in proportion to each occupation's O*NET activity scores, and each class is expressed as a share of total payroll. The research workforce share is the remaining research payroll divided by total payroll. Each coefficient is from a separate linear probability regression on one workforce measure, standardized within industry-year, with industry-year fixed effects and controls for firm size and sales growth. An industry-year is a Fama-French 48 industry in a fiscal year. Coefficients are changes in outcome probability per standard deviation, and multiplying by 100 gives percentage-point effects. Continuous variables are winsorized at the 1st and 99th percentiles within fiscal year on each estimation sample. Standard errors are clustered by firm, and t -statistics are in parentheses. The lower rows report the difference between the research share and each organization-capital class, estimated jointly across the four regressions with firm-clustered standard errors, and a joint test of equality of the four coefficients. Two alternative samples give the same ordering. Restricting base years to 2005 through 2021 without requiring the following years to be observed gives a research coefficient of +0.0100 ($t=4.56$) and difference t -statistics of 2.98 to 3.92. Including base years through 2024 and coding unobserved following years as non-initiation gives +0.0093 ($t=4.71$) and 3.05 to 4.04. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively, using two-tailed tests.

Workforce class	R&D initiation within three years
Training and development	-0.0050** (-2.11)
Management and organization	-0.0020 (-0.79)
Customer and external relations	-0.0008 (-0.30)
Research workforce share	+0.0118*** (4.35)
Observations	15,810
Research share minus class (t)	
Training and development	3.79***
Management and organization	3.04***
Customer and external relations	2.76***
Joint equality, p -value	<0.001

Table 5. Out-of-Sample Validation of the Expense Translation Against Disclosed Software Capitalization

This table evaluates the translation of workforce payroll into a share of reported expense against disclosed software capitalization. Disclosed capitalization intensity is capitalized software-development additions divided by the sum of those additions and R&D expense. Additions are taken from the CapitalizedComputerSoftwareAdditions element of annual Form 10-K XBRL filings, which report software-development costs capitalized under U.S. GAAP. A predictor measured in fiscal year t is evaluated against disclosed intensity in fiscal year $t+1$. The training period is fiscal 2005 through 2014, and the holdout period is fiscal 2015 through 2024. Panel A estimates the relation between the labor-core fraction and disclosed intensity on 166 training-period firm-years, freezes it, and evaluates it on 368 holdout-period firm-years whose disclosures play no role in estimation. Panel B is a component diagnostic on the same 534 firm-years, in which each predictor enters separately. Workforce structure is organization-capital payroll allocated to SG&A divided by SG&A payroll, task-weighted activity weights each organization-capital occupation in that ratio by its O*NET capital-building task share, and the labor share is SG&A payroll divided by reported SG&A. The fraction and the labor share use reported SG&A (Compustat XSGA, which includes R&D), because firms that capitalize software development expense the remaining development cost within R&D. The benchmark predicts every holdout firm-year with the training-period mean of disclosed intensity, and the training-period mean benchmark is re-estimated on the sample used in each panel. Significance of the slopes uses standard errors clustered by firm, and significance of the correlations tests whether each correlation differs from zero. Panel B reports no significance tests. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively, using two-tailed tests.

<i>Panel A: Baseline expense translation</i>		
Validation Statistic	Estimate	
Training-period slope	3.09**	
Training-period correlation	0.26***	
Holdout-period slope	6.09***	
Holdout correlation, predicted against disclosed	0.44***	
Holdout root-mean-square error, fitted model	0.179	
Holdout root-mean-square error, training-period mean	0.197	
<i>Panel B: Component diagnostic on the common sample</i>		
Predictor	Holdout correlation	Holdout RMSE
Labor-core fraction	0.44	0.179
Labor share of reported SG&A	0.51	0.175
Capital-building workforce structure	0.05	0.196
Task-weighted capital-building activity	−0.02	0.197
Training-period mean benchmark		0.196

Table 6. Firm-Specific Capital-Building Workforce Structure

This table reports firm-level variation, persistence, and industry-level imputation of capital-building workforce structure, the share of SG&A payroll in organization-capital occupations, for fiscal 2005 through 2024. The matched sample contains firm-years with a labor-core fraction, and the full sample contains all firm-years with measurable workforce structure. An industry-year is a Fama-French 48 industry in a fiscal year. Panel A reports the share of total variation occurring among firms within the same industry-year and, separately, the share occurring within firms over time, each measured as one minus the R-squared from a regression on industry-year or firm fixed effects. The within-industry component is the firm's measure minus its contemporaneous industry-year mean. Panel B reports persistence among firms with at least five years of data. The persistent firm component is the variance of firm means of the within-industry component divided by its total variance, an upper bound because firm means are estimated with noise. The annual autocorrelation is the slope from regressing the within-industry component on its prior-year value. Five-year transitions use quintiles of the within-industry component formed within industry-year cells with at least ten firm-years. Panel C regresses firm-level workforce structure on the leave-one-out mean of the other firms in its industry-year. Workforce structure is winsorized at the 1st and 99th percentiles within fiscal year. Parentheses report firm-clustered standard errors for the autocorrelation and industry-proxy regressions. Variance shares, persistent-component shares, and transition rates are reported in percent. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively, using two-tailed tests of whether each coefficient differs from zero.

Panel A: Firm-level variation (shares in percent)

Statistic	Matched sample	Full sample
Firm-years	54,291	75,440
Variance share within industry-year, workforce structure	83.7	87.8
Variance share within industry-year, investment fraction	75.1	
Variance share within firm, workforce structure	15.5	19.6

Panel B: Persistence

Persistent firm component of within-industry variation (percent)	80.1	75.6
Firms with at least five years	4,262	5,778
Annual autocorrelation of the within component	0.922*** (0.003)	0.905*** (0.003)
Five-year transitions observed	25,477	39,799
Top within-industry quintile stays (percent)	68.8	66.9
Bottom within-industry quintile stays (percent)	73.8	71.5

Panel C: Industry-level imputation

Industry proxy: slope on the leave-one-out mean	0.919*** (0.031)	0.920*** (0.030)
Industry proxy: R-squared	0.138	0.103

Table 7. The Organizational Build-Out Before and After Listing

This table reports workforce changes around listing for the 2,271 firms that list during the sample period. A listing firm is one whose first Compustat fiscal year is no more than one year after the calendar year of its Compustat IPO date, and that calendar year is the listing year. The sample contains calendar-year workforce observations for 2005 through 2024, which extend to years before a firm's first Compustat year. It includes the three years before listing, the listing year, and the five years after listing, together with years six and later after listing, which form the omitted reference period. Years more than three before listing are excluded because employment histories are reconstructed from fewer profiles the further back they reach. Workforce structure is SG&A-allocated organization-capital payroll over SG&A payroll. Net capital-building hiring is the payroll of positions beginning in capital-building occupations minus the payroll of positions ending in them, scaled by lagged total payroll. Regressions include event-time indicators and firm and calendar-year fixed effects, and standard errors are clustered by firm. Reported employment is unavailable before listing, so workforce coverage is not controlled. Outcomes are winsorized at the 1st and 99th percentiles within calendar year on each estimation sample. Coefficients are in percentage points, with *t*-statistics in parentheses. Panel B reports the difference between each pre-listing coefficient and the listing-year coefficient. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively, using two-tailed tests.

Panel A: Event-time coefficients

Years relative to listing	Capital-building share of SG&A labor	Net capital-building hiring
−3	2.8*** (3.23)	27.1*** (18.61)
−2	2.6*** (3.27)	23.5*** (18.07)
−1	3.1*** (4.22)	22.5*** (17.74)
Listing year	2.8*** (4.31)	22.5*** (20.73)
+1	2.1*** (3.78)	11.9*** (14.03)
+2	1.6*** (3.11)	5.2*** (7.00)
+3	1.4*** (3.15)	2.0*** (3.13)
+4	1.2*** (3.33)	1.2** (2.13)
+5	0.5** (2.04)	1.5** (2.54)
Observations	28,597	26,802

Panel B: Pre-listing years against the listing year

Difference from the listing year	Capital-building share of SG&A labor	Net capital-building hiring
−1 minus listing year	0.3 (1.26)	0.0 (0.00)
−2 minus listing year	−0.1 (−0.40)	1.0 (0.82)
−3 minus listing year	0.0 (0.06)	4.6*** (3.25)

Table 8. Alternative Classifications of Capital-Building Intensity

This table compares within-industry-year rankings of capital-building intensity. An industry-year is a Fama-French 48 industry in a fiscal year. For each comparison, quintiles and statistics are computed on the pairwise-complete sample of firm-years with both measures, within industry-year cells with at least ten such firm-years, so the number of firm-years varies across rows. Rank correlations are firm-year-weighted means of within-cell Spearman correlations. Top-quintile overlap is the share of firm-years in the top quintile of the reference measure that are also in the top quintile of the comparison measure. Under unrelated rankings, the overlap would be 20 percent and 48 percent of firm-years would be two or more quintiles apart. The reference measure is capital-building workforce structure in Panel A and own-account labor investment over lagged total assets in Panel B. Own-account labor investment over lagged total assets is the labor-core fraction multiplied by SG&A net of R&D, divided by lagged total assets. SG&A intensity is SG&A net of R&D divided by lagged total assets, and the published measures are also scaled by lagged total assets. The sample is firm-years with a labor-core fraction from 2005 through 2024.

Panel A: Rankings compared with capital-building workforce structure

Comparison measure	Rank correlation	Top-quintile overlap (percent)	Two or more quintiles apart (percent)	Firm-years
Task-weighted activity (positive control)	0.86	73	7	53,336
Labor-core investment fraction	0.21	34	41	53,336
SG&A intensity	0.21	36	39	50,810
Peters and Taylor (2017)	0.12	34	43	50,810
Eisfeldt and Papanikolaou (2013)	0.17	35	41	50,810
Iqbal et al. (2025)	0.16	35	41	49,921
Enache and Srivastava (2018)	−0.09	22	49	46,912
Ewens, Peters, and Wang (2025)	0.15	35	41	33,733

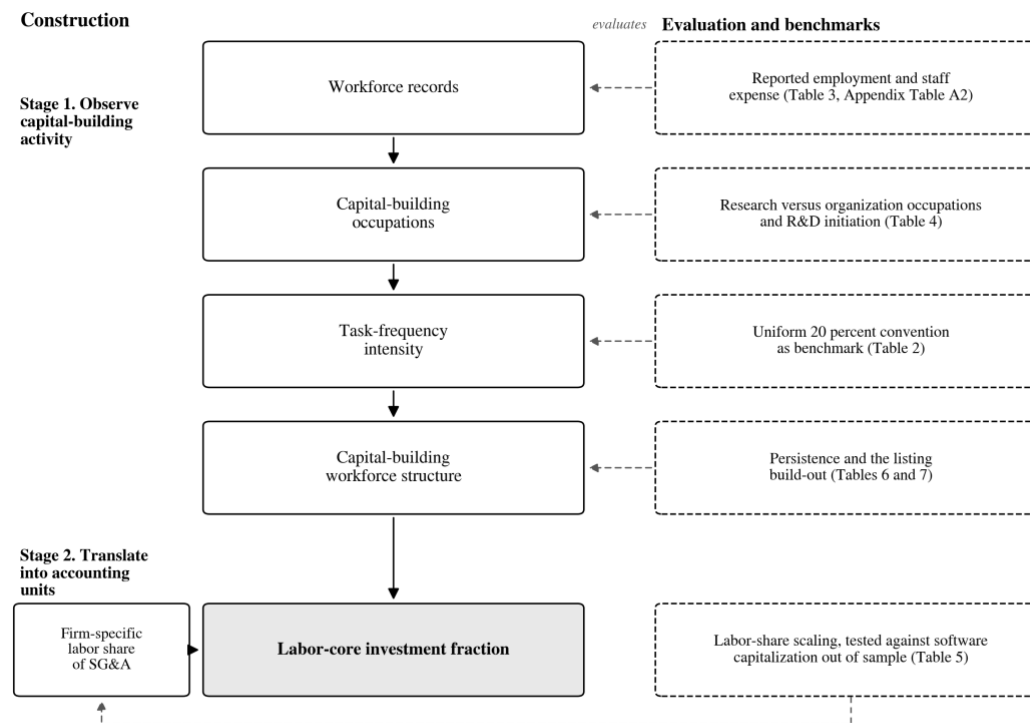
Panel B: Rankings compared with own-account labor investment over lagged assets

Comparison measure	Rank correlation	Top-quintile overlap (percent)	Two or more quintiles apart (percent)	Firm-years
Capital-building workforce structure	0.32	40	35	50,810
SG&A intensity	0.73	65	15	50,810
Peters and Taylor (2017)	0.55	55	25	50,810
Eisfeldt and Papanikolaou (2013)	0.64	60	20	50,810
Iqbal et al. (2025)	0.60	57	22	49,921
Enache and Srivastava (2018)	0.05	34	43	46,912
Ewens, Peters, and Wang (2025)	0.60	58	23	33,733

FIGURES

Figure 1. From Workforce Activity to the Labor-Core Investment Fraction

This figure summarizes the two-stage measurement framework and the evaluation evidence and benchmarks for its components. Stage 1 combines occupational composition with O*NET task information to measure capital-building workforce structure. Stage 2 combines this activity with the firm's labor share of SG&A to construct the labor-core investment fraction. Dashed boxes identify the evaluation evidence or benchmark for the corresponding measurement inputs and report the tables containing those tests. The software capitalization test supports the labor-share scaling used in the Stage 2 translation (Table 5, Panel B). It does not validate the identification of capital-building activity. Evaluation outcomes do not enter the construction of the workforce measures.



Workforce structure records where capital-building labor resides. Labor share measures how much of SG&A is labor.

Figure 2. Two Components of the Labor-Core Investment Fraction

This figure plots the capital-building share of SG&A labor against the labor share of SG&A expense for the sixty largest U.S.-headquartered firms with at least three valid observations over the most recent five fiscal years. Each firm is measured using its average over those years. Dashed curves show combinations of the two components that imply approximately equal labor-core fractions at the median capital-building task intensity of the plotted firms. Six firms are highlighted by business description to illustrate different combinations of the two components. Firm identities are not reported because the measures are constructed from proprietary workforce data.

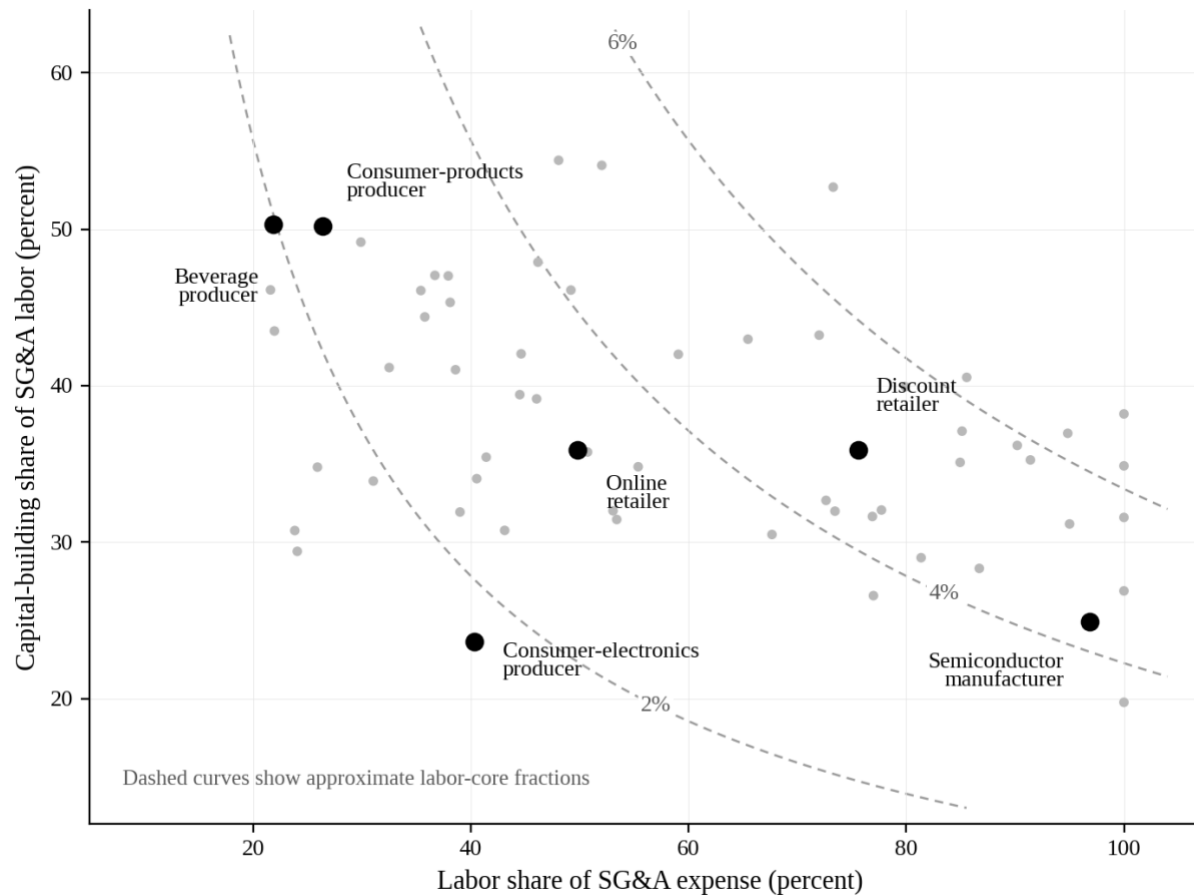


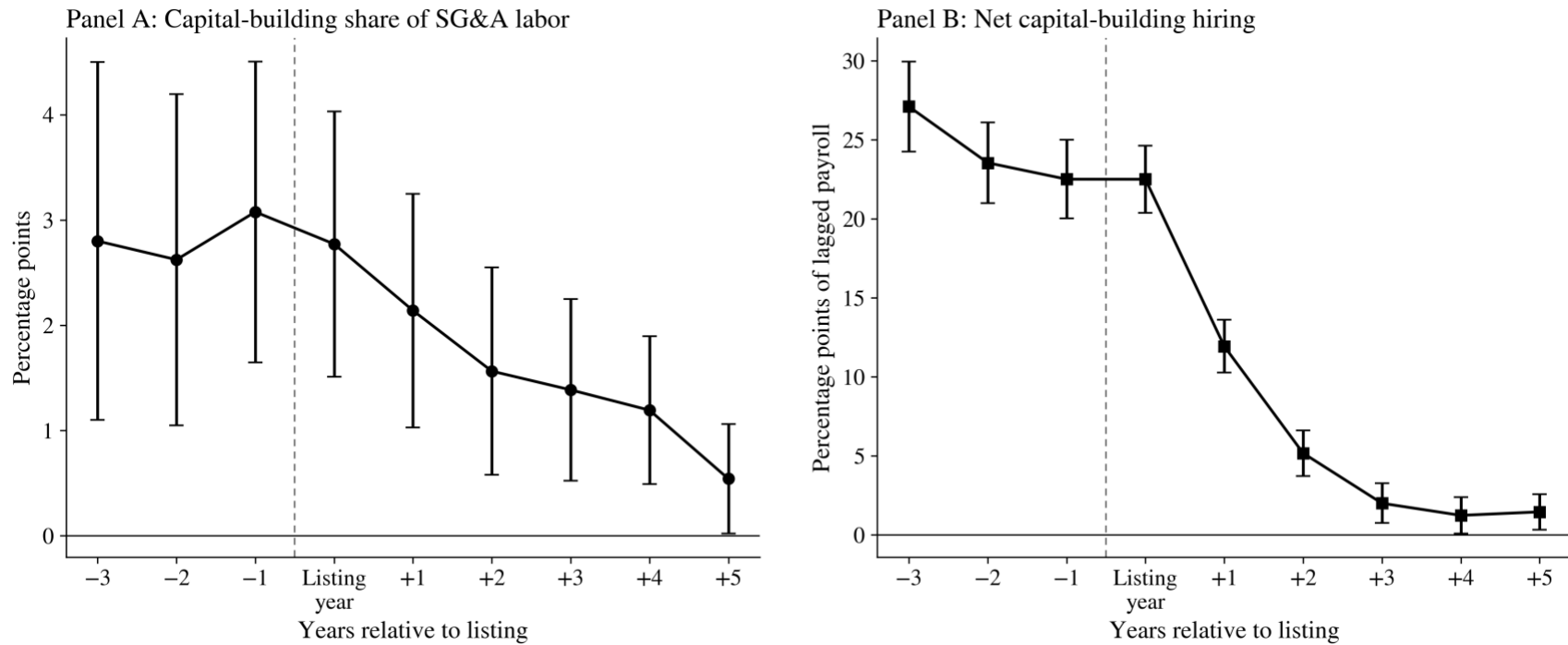
Figure 3. Firm-Level Variation Within Industry-Year Cells

This figure decomposes variation in capital-building workforce structure and the labor-core fraction into variation across industry-year cells and among firms within the same industry-year. Workforce structure is measured for 75,440 firm-years and the labor-core fraction for 54,291 firm-years. Within-industry-year variation accounts for 88 percent of total variation in workforce structure and 75 percent in the labor-core fraction. Table 6 reports the underlying variance decomposition.



Figure 4. The Organizational Build-Out Before and After Listing

This figure reports workforce changes in the years before and after listing, relative to the same firm's years six or more after listing. Panel A plots the capital-building share of SG&A labor, and Panel B plots net capital-building hiring in percentage points of lagged payroll. The estimates are those of Table 7, whose note describes the sample and specification. Vertical bars show 95 percent confidence intervals.



APPENDIX A. ADDITIONAL TABLES

Table A1. Variable Definitions

This table defines the principal variables used in the analysis. Payroll-based measures use Revelio sampling weights and full-time-equivalent employment based on the portion of the year each position is active. Salaries are model-predicted annual base pay. Compustat mnemonics appear in parentheses where applicable.

Variable	Definition	Source
Activity-class shares	Organization-capital payroll allocated across training and development, management and organization, and customer and external relations according to each occupation's relative O*NET activity-class scores. Each share is expressed relative to total payroll and standardized within industry-year.	Revelio Labs; O*NET 25.1
Annual revenue growth (<i>Growth</i>)	Sales in year t divided by sales in year $t-1$, minus one (SALE), for consecutive fiscal years, bounded at -1 and 3 .	Compustat
Capital-building hiring (<i>Hiring</i>)	Payroll of positions beginning in capital-building occupations during the year, divided by lagged total payroll.	Revelio Labs
Capital-building workforce structure (<i>Structure</i>)	Organization-capital payroll assigned to SG&A divided by total SG&A payroll; in firm-years with reported R&D expense, the 48 occupations also classified as research are excluded. Standardized within industry-year where indicated.	Revelio Labs
Coverage ratio (<i>Coverage</i>)	Weighted Revelio full-time-equivalent employment divided by Compustat-reported employment (EMP). Δ Coverage is the annual change.	Revelio Labs; Compustat
Disclosed software capitalization intensity (<i>DisclosedCap</i>)	Capitalized software-development additions divided by the sum of those additions and R&D expense (XRD), from the CapitalizedComputerSoftwareAdditions element of Form 10-K XBRL filings.	SEC filings
Firm age (<i>Age</i>)	Number of years since the firm's first fiscal-year observation in Compustat.	Compustat
Firm size (<i>Size</i>)	Natural logarithm of total assets (AT).	Compustat
Full-time-equivalent employment	Sum across positions of the Revelio sampling weight multiplied by the fraction of the fiscal year the position is active.	Revelio Labs
Labor share of SG&A (<i>LaborShare</i>)	Coverage-adjusted payroll allocated to SG&A divided by SG&A net of R&D; capped at one. SG&A is reconstructed from operating expense less cost of goods sold when XSGA is unavailable.	Revelio Labs; Compustat
Labor-core investment fraction (<i>Fraction</i>)	Labor share of SG&A multiplied by the sum across organization-capital occupations of each occupation's share of SG&A labor times its O*NET capital-building task-frequency share. The uniform specification replaces occupation-specific task shares with 0.20.	Revelio Labs; O*NET; Compustat
Net capital-building hiring (<i>NetHiring</i>)	Capital-building hiring minus the payroll of positions ending in capital-building occupations during the year, divided by lagged total payroll.	Revelio Labs

Table A1 (continued)

Variable	Definition	Source
Organization-capital payroll	Payroll in the 250 O*NET-SOC codes mapped from the Squicciarini and Le Mou��l (2012) organization-capital classification.	Revelio Labs; O*NET
Payroll	Full-time-equivalent employment multiplied by model-predicted annual base salary and aggregated across positions.	Revelio Labs
R&D intensity (<i>RD_Int</i>)	R&D expense divided by total assets (XRD/AT); missing XRD is set to zero.	Compustat
Research payroll	Payroll in occupations classified as R&D under the Frascati Manual definition, whatever the firm's R&D status.	Revelio Labs; O*NET; OECD
Research workforce share (KCShare)	Research payroll divided by total payroll. In mutually exclusive validation specifications, the 48 occupations belonging to both classifications are removed from research payroll and from the organization-capital activity classes.	Revelio Labs
SG&A intensity (<i>SGA</i>)	SG&A expense net of R&D divided by lagged total assets (AT).	Compustat

Table A2. Revelio-Compustat Linkage and Workforce Coverage

This table reports the Revelio-Compustat linkage and compares weighted Revelio full-time-equivalent employment with Compustat-reported employment (EMP). Panel A reports the linkage universe and the scale of the workforce data from which the analysis sample is drawn. Panels B and C use the 68,881 firm-years of U.S.-headquartered public firms from fiscal 2005 through 2024 that report employment, excluding regulated utilities and financial firms. Panel B reports the association between Revelio and reported employment, and Panel C reports the median coverage ratio by year. The coverage ratio is weighted Revelio employment divided by reported employment. It can exceed one because Revelio measures sampling-weighted positions active during the fiscal year, whereas Compustat reports employment at a point in time. Correlations between Revelio and reported employment are computed in logs. Coverage ratios are medians of firm-year ratios and use occupation-country sampling weights and full-time-equivalent employment.

Panel A: Revelio-Compustat linkage universe

Linkage statistic	Count
Revelio company records linked to Compustat	334,564
Matched directly by a Compustat identifier	65,283
Reached through an ultimate parent	268,264
Resolved by hand	1,017
Compustat firms represented by linked records	65,720
Compustat firms with employee position records	61,152
<i>Workforce-data scale</i>	
Position records in the delivery	1.819 billion
Position records at linked firms	383.7 million
Share of position records with usable start dates	92.5 percent
Firm-year-occupation cells in the workforce panel	84,603,594

Panel B: Association between Revelio-based and reported employment

Statistic	Value
Correlation between Revelio and reported employment (logs, pooled)	0.85
Slope: log Revelio employment on log reported employment	0.81
R-squared of that association	0.73

Table A2 (continued)*Panel C: Coverage of reported employment by year*

Fiscal year	Median coverage ratio	Firm-years
2005	0.46	3,877
2006	0.51	3,782
2007	0.56	3,738
2008	0.61	3,535
2009	0.67	3,470
2010	0.70	3,398
2011	0.74	3,366
2012	0.79	3,388
2013	0.86	3,520
2014	0.90	3,502
2015	0.96	3,401
2016	0.99	3,300
2017	1.05	3,262
2018	1.08	3,271
2019	1.14	3,210
2020	1.20	3,402
2021	1.17	3,582
2022	1.21	3,520
2023	1.25	3,289
2024	1.25	3,068

Table A3. Classification of Occupations

This table summarizes the occupation classifications. The baseline maps the organization-capital occupations of Squicciarini and Le Mouél (2012) to the O*NET-SOC 2019 codes used in the workforce data. The re-estimated classification reapplies their task-based selection procedure to O*NET 25.1 task data rather than carrying forward the published occupation set. Rank correlations are pooled Spearman correlations for fiscal 2005 through 2024, over the 71,609 firm-years with both workforce-structure measures and the 54,291 firm-years with both fractions. Research occupations follow the Frascati Manual definition (OECD 2015). Occupations that qualify as both organization capital and research are assigned to research when the firm reports R&D expense and to organization capital otherwise.

Classification statistic	Value
Original occupations in Squicciarini and Le Mouél (2012)	84
Baseline: O*NET-SOC 2019 codes after the crosswalk	250
Re-estimated: occupations selected from O*NET 25.1 task data	51
Occupations in both classifications	35
Rank correlation with the baseline: workforce structure	0.23
Rank correlation with the baseline: labor-core fraction	0.74
Occupations classified as both organization capital and research	48

Table A4. R&D Initiation with Overlapping Occupation Sets

This table repeats the R&D-initiation test with the 48 occupations that belong to both the organization-capital set and the Frascati research set included in both groups. The research workforce share counts every Frascati research occupation, whatever the firm's R&D status, and the three activity classes allocate all baseline organization-capital payroll. The sample is firm-years from fiscal 2005 through 2021 without reported R&D expense whose three following fiscal years are all observed in the panel, so the outcome is never coded from an unobserved year. The outcome equals one if the firm reports positive R&D expense in any of fiscal years $t+1$ through $t+3$, which occurs in 4.1 percent of the sample. Each workforce measure is standardized within Fama-French 48 industry-year, and each coefficient is from a separate linear probability regression with industry-year fixed effects and controls for firm size and sales growth. Continuous variables are winsorized at the 1st and 99th percentiles within fiscal year. Standard errors are clustered by firm. Coefficients are changes in outcome probability per standard deviation, with t -statistics in parentheses. The lower rows report the difference between the research share and each organization-capital class, estimated jointly across the four regressions, and a joint test of equality of the four coefficients. Including base years through 2024 and coding unobserved following years as non-initiation gives a research coefficient of +0.0096 ($t=5.01$). ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively, using two-tailed tests.

Workforce class	R&D initiation within three years
Training and development	-0.0028 (-1.21)
Management and organization	+0.0002 (0.07)
Customer and external relations	+0.0009 (0.36)
Research workforce share	+0.0120*** (4.53)
Observations	15,810
Research share minus class (t)	
Training and development	3.62***
Management and organization	2.82***
Customer and external relations	2.57**
Joint equality, p -value	<0.001

Table A5. Workforce Coverage and the Software Validation

This table tests whether changes in workforce coverage account for the software validation. Workforce coverage is weighted Revelio full-time-equivalent employment divided by Compustat-reported employment. The first row is the correlation between annual changes in workforce coverage and annual changes in the labor-core fraction over 44,362 firm-year changes at 5,592 firms from fiscal 2006 through 2024. Its significance is assessed with firm-clustered standard errors from a regression of the standardized change in the fraction on the standardized change in coverage, whose slope equals the correlation ($t=-31.54$). The remaining rows report one regression of disclosed software capitalization intensity in fiscal year $t+1$ on the labor-core fraction and the change in workforce coverage in year t . Disclosed intensity is capitalized software-development additions, from Form 10-K XBRL filings, divided by the sum of those additions and R&D expense. The regression uses all 438 firm-years at 98 firms from fiscal 2008 through 2023 with both predictors and a disclosure in the following year. It pools the training and holdout periods of the out-of-sample validation (141 and 297 firm-years) and includes no fixed effects or other controls. The fraction is measured on reported SG&A, which includes R&D. Variables are winsorized at the 1st and 99th percentiles within fiscal year, and standard errors are clustered by firm. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively, using two-tailed tests.

Coverage Test	Estimate
Correlation between changes in the fraction and in coverage	-0.30***
Change in workforce coverage, conditional on the labor-core fraction	+0.015
Labor-core fraction, conditional on the change in workforce coverage	+5.15***

Table A6. International Validation of Workforce Payroll and Development Capitalization

This table presents international validation using European Single Electronic Format (ESEF) annual reports for fiscal years 2020 through 2025, linked to workforce data by legal entity identifier. The primary sample covers firms that are their own ultimate parent in the United Kingdom, the Netherlands, Sweden, Denmark, Norway, Finland, and France, and the holdout covers Poland, Italy, Spain, Belgium, and Austria. Development occupations are engineers, architects, engineering technicians, software developers and testers, and architectural and engineering managers. Research occupations are life, physical, and social scientists, science technicians, mathematical and data scientists, computer and information research scientists, and natural sciences managers. Panel A compares modeled payroll with IAS 19 employee benefits expense, and its correlation is computed on log values after removing country-year means. Panels B and C use positive IAS 38 development additions. Zero or undisclosed additions are excluded, so these panels describe amounts capitalized, not the decision to capitalize. In a blind rereading of 10 percent of filings, capitalization-policy agreement exceeds 97 percent and extracted amounts agree within 5 percent in at least 97.8 percent of matched readings. Panel B ratios are upper bounds on the capitalized labor share because additions may include nonlabor costs. Panel C jointly regresses log capitalized additions on log development- and research-occupation payroll, with log assets, return on assets, log total payroll, country-year and two-digit NAICS fixed effects, and firm-clustered standard errors. The development-minus-research row is the difference between the two coefficients in that regression, tested with the firm-clustered covariance matrix. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively, using two-tailed tests.

Panel A: Payroll validation

Statistic	Firm-years	Estimate
Within-country-year log correlation between modeled payroll and employee benefits expense	2,247	0.89
Median ratio of modeled payroll to employee benefits expense	2,247	0.94

Panel B: Capitalized development relative to development-occupation payroll

Sample	Firm-years	Median	Interquartile range
Primary	687	0.26	0.07 to 0.61
Country holdout	238	0.22	0.07 to 0.65

Panel C: Capitalized development and occupation payroll

Sample and payroll measure	Firm-years	Coefficient	t-statistic
Primary: development occupations	658	0.523***	5.17
Primary: research occupations	658	0.115	1.53
Primary: development – research	658	0.408***	3.05
Country holdout: development occupations	224	0.808***	2.73
Country holdout: research occupations	224	0.324*	1.95
Country holdout: development – research	224	0.484	1.35

Table A7. Professional-Workforce Placebo and Headcount Around Listing

This table checks whether the pre-listing elevation in capital-building workforce structure reflects how early employment histories are reconstructed. If reconstructed histories represented managerial and professional employees more fully than others, every managerial and professional share would be elevated before listing. The sample contains calendar-year workforce observations for 2005 through 2024 for the 2,271 firms that list during the sample period. A listing firm is one whose first Compustat fiscal year is no more than one year after the calendar year of its Compustat IPO date, and that calendar year is the listing year. Observations run from three years before listing through five years after listing, with years six and later after listing as the omitted reference period. Regressions include event-time indicators and firm and calendar-year fixed effects, and standard errors are clustered by firm. All three outcomes share the denominator of total payroll or total headcount, namely the capital-building share of total payroll, the payroll share of other managerial and professional occupations (O*NET-SOC major groups 11 through 29 outside the capital-building and research sets), and the capital-building share of total headcount. Their mature-period means are 46.8, 8.3, and 39.6 percent. The lower rows report the difference between the first two, estimated jointly with outcome-specific fixed effects. Outcomes are winsorized at the 1st and 99th percentiles within calendar year on each estimation sample. Coefficients are in percentage points, with *t*-statistics in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively, using two-tailed tests.

Years relative to listing	Capital- building share of total payroll	Other managerial and professional share	Capital- building share of headcount
-3	2.1*** (3.04)	-1.3*** (-4.18)	2.0*** (2.99)
-2	2.5*** (4.01)	-1.3*** (-4.55)	2.5*** (4.09)
-1	2.8*** (4.88)	-1.3*** (-5.04)	2.6*** (4.79)
Listing year	2.4*** (4.65)	-0.8*** (-3.46)	2.0*** (4.18)
+1	2.0*** (4.35)	-0.4* (-1.96)	1.7*** (3.83)
+2	1.5*** (3.62)	-0.1 (-0.67)	1.3*** (3.27)
+3	1.2*** (3.44)	-0.2 (-1.13)	1.0*** (2.97)
+4	0.9*** (3.08)	-0.2 (-1.55)	0.7*** (2.74)
+5	0.4* (1.68)	-0.1 (-0.63)	0.4* (1.77)
Observations	28,807	28,807	28,807
Capital-building minus other managerial and professional share			
-3	3.4*** (4.13)		
-2	3.8*** (5.10)		
-1	4.1*** (6.05)		

Table A8. The Organizational Build-Out at Listing: Fiscal-Year Panel

This table reports workforce changes around listing on the fiscal-year panel for the 2,262 firms that list during the sample period. A listing firm is one whose first Compustat fiscal year is no more than one year after the calendar year of its Compustat IPO date, and that calendar year is the listing year. Event time is the fiscal year minus the listing year. The sample contains these firms' fiscal-year observations from 2005 through 2024. Outcomes are the capital-building share of SG&A labor, capital-building hiring, and net capital-building hiring. Hiring is the payroll of positions beginning in capital-building occupations during the year scaled by lagged total payroll. Net hiring subtracts the payroll of positions ending in those occupations. Event-time indicators mark the listing year and each of the five following years, and the same firm's years six or more after listing form the omitted reference period. Regressions include firm and fiscal-year fixed effects, and standard errors are clustered by firm. Outcomes are winsorized at the 1st and 99th percentiles within fiscal year on each estimation sample. Coefficients are in percentage points, with *t*-statistics in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively, using two-tailed tests.

	Firm- years	Listing year	<i>t</i>+1	<i>t</i>+2	<i>t</i>+3	<i>t</i>+4	<i>t</i>+5
Capital-building share of SG&A labor	19,297	2.3*** (3.79)	1.5*** (2.97)	1.0** (2.21)	0.7* (1.73)	0.5* (1.73)	0.4* (1.70)
Hiring into capital-building occupations	18,932	16.4*** (14.44)	7.9*** (8.32)	3.5*** (4.24)	1.5** (2.18)	0.5 (0.84)	1.3** (2.08)
Net capital-building hiring	18,932	19.1*** (16.98)	8.6*** (9.24)	3.6*** (4.38)	1.4** (2.14)	0.7 (1.08)	1.4** (2.16)

Table A9. Composition of the Organizational Build-Out at Listing

This table decomposes the listing-year build-out for the 2,262 listing firms on the fiscal-year panel from 2005 through 2024. A listing firm is one whose first Compustat fiscal year is no more than one year after the calendar year of its Compustat IPO date, and that calendar year is the listing year. Each estimate is the listing-year coefficient from a regression of the outcome on indicators for the listing year and each of the five following years, with firm and fiscal-year fixed effects and standard errors clustered by firm. Years six and later form the omitted reference period. Outcomes are winsorized at the 1st and 99th percentiles within fiscal year. Panel A uses the capital-building share of total payroll, built from occupation cells aligned by calendar year. Total payroll is used in Panel A so the outcome can be decomposed by occupation subset. The finance and compliance set includes financial management, financial specialist, and compliance occupations, and the broader exclusion additionally removes general-administration occupations. Panel B decomposes capital-building payroll across the three O*NET activity classes. Panel C reports capital-building hiring scaled by lagged total payroll. Coefficients are in percentage points, with *t*-statistics in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively, using two-tailed tests.

Panel A: Listing-year change in the capital-building share of payroll

Measure	Listing-year coefficient
All capital-building occupations	2.16*** (4.75)
Excluding finance and compliance	1.95*** (4.24)
Excluding finance, compliance, and general administration	2.00*** (4.33)

Panel B: Listing-year change by activity

Activity	Listing-year coefficient	Firm-years
Training and development	0.58*** (4.29)	19,297
Management and organization	0.78*** (4.80)	19,297
Customer and external relations	0.78*** (4.63)	19,297

Panel C: Listing-year change in capital-building hiring

Hiring measure	Listing-year coefficient	Firm-years
All capital-building occupations	16.4*** (14.44)	18,932
Excluding finance and compliance	12.1*** (13.59)	18,932

Table A10. Workforce Structure and Subsequent Capital-Building Hiring

This table examines whether capital-building workforce structure predicts subsequent net hiring into capital-building occupations. The outcome is net capital-building hiring in the single fiscal year $t+1$, $t+3$, or $t+5$, not cumulative hiring through that year. Net capital-building hiring is the payroll of positions beginning in capital-building occupations minus the payroll of positions ending in those occupations, scaled by prior-year total payroll. Workforce structure is standardized so that coefficients represent a one-standard-deviation difference, and each specification also includes its annual change. All specifications include SG&A intensity, leverage, asset tangibility, employment, return on assets, workforce coverage, and industry-year fixed effects. Panel A uses the baseline occupation classification. Panel B uses, at each horizon, the subset of Panel A's sample with workforce structure under both classifications and hiring built from occupation cells under both, so its firm-years are a subset of Panel A's. The first row repeats the Panel A specification on this subsample. The second and third rows use hiring built from occupation cells, with structure and hiring under the baseline classification and under the independently re-estimated classification. Measures and controls are standardized and winsorized on each panel's sample. The sample covers fiscal 2005 through 2024, and continuous variables are winsorized at the 1st and 99th percentiles within fiscal year. Standard errors are clustered by firm and fiscal year, and t -statistics are in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively, using two-tailed tests.

Panel A: Baseline classification

Measure	$t+1$	$t+3$	$t+5$
Capital-building workforce structure	0.0031*** (2.83)	0.0035*** (4.02)	0.0042*** (3.93)
Firm-years	39,200	30,277	23,144

Panel B: Baseline and re-estimated classifications on a common subsample

Measure	$t+1$	$t+3$	$t+5$
Baseline classification, Panel A hiring measure	0.0030*** (2.69)	0.0034*** (3.96)	0.0041*** (3.84)
Baseline classification, cell-built hiring	0.0026** (2.32)	0.0038*** (4.13)	0.0042*** (3.61)
Re-estimated classification, cell-built hiring	-0.0010** (-2.44)	0.0002 (0.52)	0.0003 (0.85)
Firm-years	39,173	30,254	23,122

Table A11. Workforce Measures, Published Capitalization Measures, and Industry Parameters

This table relates the workforce measures to published capitalization measures and industry parameters. Panel A compares the workforce measures with five published measures of intangible investment or capital, those of Peters and Taylor (2017), Eisfeldt and Papanikolaou (2013), Iqbal et al. (2025), Enache and Srivastava (2018), and Ewens, Peters, and Wang (2025), each expressed as a firm-level amount scaled by lagged total assets. SG&A intensity is SG&A net of R&D divided by lagged total assets. Workforce structure is the capital-building share of SG&A payroll, and the labor-core fraction is estimated own-account labor investment in organization capital divided by SG&A net of R&D. Rank correlations are pooled Spearman correlations. Partial correlations are Pearson correlations between the labor-core fraction and each published measure after both are residualized on SG&A intensity and fiscal-year effects. Samples are firm-years with a labor-core fraction from 2005 through 2024, pairwise complete for each statistic, and the last column reports the number of firm-years in each row's comparisons. Continuous variables are winsorized at the 1st and 99th percentiles within fiscal year. Panel B reports Pearson and Spearman correlations between the workforce measures and the replicated Iqbal et al. (2025) industry-year capitalization parameter for 51,790 firm-years.

Panel A: Published capitalization measures and SG&A intensity

Measure	Rank correlation with SG&A intensity	Rank correlation with the labor-core fraction	Rank correlation with workforce structure	Partial correlation with the labor-core fraction, holding SG&A intensity fixed	Firm-years
Peters and Taylor (2017)	0.76	−0.26	0.09	−0.11	51,811
Eisfeldt and Papanikolaou (2013)	0.89	−0.18	0.12	−0.03	51,811
Iqbal et al. (2025)	0.71	−0.27	0.17	−0.15	50,875
Enache and Srivastava (2018)	0.18	−0.13	−0.06	−0.06	47,793
Ewens, Peters, and Wang (2025)	0.78	−0.22	0.08	−0.10	34,513
Labor-core investment fraction	−0.19				51,811
Capital-building workforce structure	0.15				51,811

Panel B: Correlations with the industry capitalization parameter (Iqbal et al. 2025)

Measure	Pearson	Spearman
Capital-building workforce structure	0.10	0.11
Labor-core investment fraction	−0.15	−0.20

Table A12. Characteristics of Firms with Discordant Activity and Expenditure Rankings

This table reports medians for firm-years whose within-industry-year rankings by capital-building workforce structure and SG&A intensity disagree. Workforce structure is organization-capital payroll allocated to SG&A divided by SG&A payroll. SG&A intensity is SG&A net of R&D divided by lagged total assets. The labor-core fraction is estimated own-account organization-capital labor investment divided by SG&A net of R&D. Quintiles are formed within Fama-French 48 industry-year cells with at least ten firm-years. The first group is in the top quintile of workforce structure and the bottom two quintiles of SG&A intensity. The second is in the top quintile of SG&A intensity and the bottom two quintiles of workforce structure. Return on assets is income before extraordinary items over total assets. The three most common Fama-French 12 industries are other (22 percent), wholesale and retail (19 percent), and healthcare (17 percent) in the first group, and business equipment (21 percent), healthcare (20 percent), and wholesale and retail (17 percent) in the second. The sample is firm-years with a labor-core fraction from 2005 through 2024.

Statistic	High structure, low SG&A intensity	Low structure, high SG&A intensity	All firm-years
Firm-years	2,907	2,814	50,810
Share of firm-years (percent)	6	6	100
Log total assets	5.78	4.62	5.89
Firm age (years)	13	12	14
Labor share of SG&A	0.61	0.46	0.57
SG&A intensity	0.13	0.83	0.26
R&D intensity	0.000	0.013	0.014
Sales growth	0.052	0.089	0.067
Return on assets	0.006	−0.046	0.007
Capital-building workforce structure	0.63	0.37	0.46
Labor-core investment fraction	0.061	0.029	0.043

APPENDIX B. DETAILED CONSTRUCTION OF THE WORKFORCE MEASURES

This appendix documents the occupation-level construction of the workforce measures. It covers the organization-capital and research classifications, the task-frequency intensity weights, the activity-class decomposition, and the mapping from workforce payroll to the labor-core investment fraction. It does not restate the motivation or empirical results. Complete occupation lists, code lineages, and the implementation code are in the replication package.

B.1. Organization-Capital Occupation Selection

The baseline uses the published task-based classification of Squicciarini and Le Mou  l (2012). Their classification starts from the O*NET generalized work activities, a common set of 41 activities rated for each occupation. Eleven activities are designated as organization building because they describe strategy, process design, coordination, team formation, training, and relationship maintenance. Table B1 lists them. The baseline holds this externally published occupation boundary fixed rather than selecting occupations from the firm-level accounting outcomes studied here.

Each occupation receives an Importance rating and a Level rating for every activity. Importance ranges from 1, not important, to 5, extremely important. Level ranges from 0 to 7 and measures the level at which the activity is performed. The two ratings are rescaled to the unit interval and multiplied to form the occupation-activity index below, where o indexes occupations and a indexes activities.

$$I_{oa} = \frac{IM_{oa} - 1}{4} \times \frac{LV_{oa}}{7}. \quad (B1)$$

The index is high only when an activity is both important to the occupation and performed at a high level. Squicciarini and Le Mou  l (2012) then apply two complementary selection rules. Under the distributional rule, an occupation qualifies when both its summed Importance across the

eleven activities and its minimum Level across them fall in the top quartile. Under the clustering rule, occupations are grouped by their work-activity profiles, and the clusters with the highest organization-building scores are retained. Their Table A.2 lists 94 occupations under the first rule and 112 under the second. The 84 occupations that satisfy both rules form the published set and the starting point for the baseline.

The re-estimated classification in Section 5.4 reapplies both rules to O*NET 25.1. After ratings that O*NET flags as suppressed or not relevant are dropped, 395 occupations in that release have complete Importance and Level ratings for all 41 activities. The distributional rule selects 70 of them, a complete-linkage clustering on all 41 activities selects 80, and 51 satisfy both (Appendix Table A3).

B.2. Taxonomy Lineage and the 84-to-250 Mapping

The 84 occupations are stated in the 2000 Standard Occupational Classification (SOC), whereas the workforce records and the O*NET 25.1 task data use O*NET-SOC 2019, which is based on the 2018 SOC. I carry each published code through the Bureau of Labor Statistics 2000-to-2010 and 2010-to-2018 crosswalks and then expand each resulting 2018 SOC code to every detailed O*NET-SOC 2019 code that shares it. The mapping follows codes through splits, merges, and changes of occupational group, and it does not match occupations by title. The distinction matters because 57 of the 250 final codes are reached through a code whose broad occupational group changed between the 2000 and 2018 classifications, eight of them through a change of major group.

B.3. Research Occupations and the 48 Overlapping Codes

I identify research occupations separately, following the Frascati Manual definition of R&D personnel (OECD 2015). The occupational implementation includes every code in three SOC

major groups, computer and mathematical, architecture and engineering, and life, physical, and social science occupations, together with three management occupations that direct technical or scientific work, namely computer and information systems managers, architectural and engineering managers, and natural sciences managers. It yields 169 O*NET-SOC 2019 codes, 163 from the three major groups and six from the management occupations. Occupations identify researchers, technicians and equivalent staff, and research managers.

Forty-eight codes belong to both the 250-code organization-capital classification and the 169-code research classification. Most are engineering, life-science, and social-science occupations, together with computer and information systems managers and architectural and engineering managers. The baseline rule is firm-year specific. When a firm reports R&D expense, the overlapping codes are assigned to research, and when it reports none, they are assigned to organization capital. The rule recognizes that the same occupation may perform formal research in one firm and internal technology or organization-building work in another.

The baseline assignment is not used in the R&D-initiation validation (Section 5.2) because it would make reported R&D status part of the validating measure. For that test, the 48 overlapping codes are removed from both the research and organization-capital groups, so no occupation enters both sides and neither measure depends on the firm's reported R&D status. Appendix Table A4 reports the test with the overlapping occupations included in both groups. Across the two classifications, 202 codes are organization capital only, 121 are research only, and 48 belong to both.

B.4. Occupation-Specific Task-Frequency Intensity

Occupation selection determines which occupations contain organization-building work. It does not determine how much of their work is organization building. I therefore construct an

occupation-specific capital-building task share from a separate O*NET construct, task frequency. For each task, O*NET reports the distribution of respondents across seven frequency categories, from yearly or less to hourly or more. I convert each category to a representative number of occurrences per year (Table B3) and compute each task's expected annual frequency as the response-weighted average.

Each O*NET task is linked to one or more detailed work activities, each of which belongs to one of the 41 generalized work activities. For task j in occupation o , let f_{oj} denote its expected annual frequency and q_{oj} the proportion of its linked detailed work activities that belong to the eleven organization-building activities. A task linked only to organization-building activities has q_{oj} equal to one, and a task with mixed links has a fractional value. A task without a work-activity link has q_{oj} equal to zero, so unlinked tasks are conservatively counted as non-building. Ratings that O*NET recommends suppressing are dropped. The task-frequency share is the frequency-weighted average

$$\tau_o = \frac{\sum_j f_{oj} q_{oj}}{\sum_j f_{oj}}. \quad (B2)$$

Of the 250 organization-capital codes, 217 have their own task-frequency ratings. The other 33 are mainly residual (“all other”) codes and recently split codes. I impute their task shares from the nearest rated occupations in a fixed hierarchy, first other O*NET-SOC codes within the same six-digit SOC occupation, then the SOC broad group, and then codes sharing the first four digits of the SOC code. The first level containing a rated occupation determines the mean. Nine codes are filled at the six-digit level, 21 at the broad-group level, and three at the four-digit level. These 33 codes account for 17 percent of organization-capital payroll in the sample. The imputation uses only O*NET taxonomy and rating information.

Across all 250 organization-capital codes, the task-frequency share has a mean of 0.17, a median of 0.15, and an interquartile range of 0.08 to 0.25. Among the 217 directly rated codes, the mean is 0.18 and the median is 0.16. At the median firm-year, the payroll-weighted capital-building task share is 17.6 percent. Because O*NET measures frequency rather than duration, I interpret the task share as an occupation-level proxy for the share of working time devoted to capital building, not as observed hours. Importance and Level ratings select occupations, and task frequency supplies the intensity weight within the selected set.

B.5. Activity Classes

The R&D-initiation validation (Section 5.2) and the composition of the listing build-out (Appendix Table A9) allocate organization-capital payroll across three activity classes, namely training and development, management and organization, and customer and external relations. The allocation does not change the 250-code baseline classification. It is a researcher-defined allocation based on O*NET Importance and Level ratings. The classes draw on ten activities (Table B4). Seven are among the eleven selection activities. Three others, guiding, directing, and motivating subordinates, communicating with persons outside the organization, and selling or influencing others, sharpen the management and customer content. The remaining four selection activities, judging qualities, communicating with supervisors and peers, resolving conflicts, and staffing, are not assigned because they do not distinguish among the three purposes.

For each occupation, I average the rescaled Importance-by-Level index over the activities in each class and divide each class score by the sum of the three scores. The resulting weights sum to one, so an occupation's organization-capital payroll is divided across all three classes rather than assigned to one. An occupation without usable ratings for the class activities receives equal weights. In the R&D-initiation validation, the 48 overlapping codes are removed, the remaining

organization-capital payroll is allocated with these weights, each class is expressed as a share of total payroll, and the shares are standardized within industry-year. The test asks whether the research workforce share is more strongly associated with the subsequent start of reported R&D than each of the three organization-capital classes.

B.6. Payroll Allocation and the Labor-Core Investment Fraction

Payroll is constructed at the position level. In each calendar year, a position contributes the fraction of the year in which it is active, multiplied by the Revelio occupation-country sampling weight and the model-predicted annual base salary. Calendar-year values are aligned to fiscal years by month weighting. Observed payroll is then adjusted for workforce coverage so that modeled payroll corresponds to the reported workforce. The adjustment applies symmetrically when modeled employment is below or above Compustat employment and is subject to fixed bounds. It assumes that observed and unobserved employees have similar average compensation after sampling weights are applied. I do not construct the labor-core fraction when workforce coverage is below 5 percent.

Each position is classified on two separate dimensions, the economic activity implied by its occupation and the financial-statement expense line implied by its functional role. Table B5 states the expense-line rules. The two dimensions are combined only after each has been assigned. Organization-capital payroll enters the numerator only when the position is assigned to SG&A. Organization-capital labor assigned to R&D or cost of sales lies outside the labor-core measure by construction.

Let s_{oit} be occupation o 's share of firm i 's SG&A payroll in year t , b_o indicate membership in the organization-capital set after the overlap rule, and τ_o be the occupation-specific task-frequency

share. Capital-building workforce structure and task-weighted capital-building activity, restated from equations (1) and (2) of Section 3.1, are

$$Structure_{it} = \sum_o s_{oit} b_o. \quad (B3)$$

$$Activity_{it} = \sum_o s_{oit} b_o \tau_o. \quad (B4)$$

Workforce structure records where organization-capital labor resides within SG&A labor. Task-weighted activity converts that structure into an estimate of how intensively the SG&A workforce performs organization-building work. Own-account labor investment equals SG&A payroll multiplied by task-weighted activity, as in equation (3).

The labor share of SG&A is coverage-adjusted SG&A payroll divided by SG&A net of R&D, capped at one. When Compustat SG&A is unavailable, it is reconstructed from operating expense less cost of goods sold. Multiplying task-weighted activity by the labor share expresses the measured activity as a fraction of the reported expense base, as in equations (4) and (5) of Section 3.1:

$$LaborShare_{it} = \frac{Payroll_{it}^{SG\&A}}{SG\&A_{it} - R\&D_{it}}. \quad (B5)$$

$$Fraction_{it} = LaborShare_{it} \times \sum_o s_{oit} b_o \tau_o. \quad (B6)$$

The labor-core investment fraction is therefore estimated own-account labor investment in organization capital divided by SG&A net of R&D. It excludes purchased services and other non-labor inputs, organization-building labor assigned to R&D or cost of sales, and externally produced intangible investment. The uniform benchmark replaces every occupation-specific task share with 0.20. It changes the scale of the measure but leaves the allocation logic unchanged.

Table B1. The Eleven Organization-Capital Work Activities

This table lists the O*NET generalized work activities that Squicciarini and Le Mouél (2012) designate as organization building. Element identifiers and activity names follow the O*NET 25.1 content model.

Element ID	O*NET work activity
4.A.2.a.1	Judging the Qualities of Things, Services, or People
4.A.2.b.4	Developing Objectives and Strategies
4.A.2.b.6	Organizing, Planning, and Prioritizing Work
4.A.4.a.2	Communicating with Supervisors, Peers, or Subordinates
4.A.4.a.4	Establishing and Maintaining Interpersonal Relationships
4.A.4.a.7	Resolving Conflicts and Negotiating with Others
4.A.4.b.1	Coordinating the Work and Activities of Others
4.A.4.b.2	Developing and Building Teams
4.A.4.b.3	Training and Teaching Others
4.A.4.b.5	Coaching and Developing Others
4.A.4.c.2	Staffing Organizational Units

Table B2. Occupational Taxonomy Lineage

This table summarizes the path from the published occupations to the codes used in the measurement. The replication package records the complete path from each published occupation to every final code. Codes outside the mapped set remain in payroll denominators but receive zero organization-capital weight.

Stage	Taxonomy	Construction role
Published selection	2000 SOC	84 occupations satisfying both selection rules
Crosswalk 1	2010 SOC	Bureau of Labor Statistics 2000-to-2010 crosswalk
Crosswalk 2	2018 SOC	Bureau of Labor Statistics 2010-to-2018 crosswalk
Detailed expansion	O*NET-SOC 2019	250 detailed codes

Table B3. Conversion of O*NET Task-Frequency Categories

This table reports the representative annual frequency assigned to each O*NET task-frequency category based on a 250-day working year. The conversion is applied uniformly across occupations. It places the ordered categories on a common annual scale and does not measure task duration. Replacing them with ordinal category scores (1 to 7) leaves the occupation-level ranking of task shares largely unchanged (Spearman correlation 0.92 across the 250 organization-capital codes).

Scale point	Frequency category	Occurrences per year
1	Yearly or less	1
2	More than yearly	6
3	More than monthly	26
4	More than weekly	104
5	Daily	250
6	Several times daily	750
7	Hourly or more	2,000

Table B4. Activity Classes Used to Allocate Organization-Capital Payroll

This table lists the O*NET generalized work activities that define each activity class. The three classes use ten activities. Seven are among the eleven organization-capital selection activities, and three additional activities, guiding, directing, and motivating subordinates, communicating with persons outside the organization, and selling or influencing others, sharpen the management and customer classifications. The remaining four selection activities, judging qualities, communicating with supervisors and peers, resolving conflicts, and staffing, are not allocated to a class. Each occupation's organization-capital payroll is divided across the three classes in proportion to its normalized Importance-by-Level scores averaged over each class's activities. The classes are researcher-defined and are not defined by O*NET.

Activity class	O*NET work activities
Training and development	Training and Teaching Others; Coaching and Developing Others
Management and organization	Developing Objectives and Strategies; Organizing, Planning, and Prioritizing Work; Coordinating the Work and Activities of Others; Developing and Building Teams; Guiding, Directing, and Motivating Subordinates
Customer and external relations	Communicating with Persons Outside Organization; Selling or Influencing Others; Establishing and Maintaining Interpersonal Relationships

Table B5. Functional Allocation of Position Payroll

This table reports the rules that assign each position's payroll to an expense line, based on the vendor's functional role for the position and, for engineering roles, whether the role develops products or performs research. Goods-producing firms are those with SIC codes below 4000. Operations payroll in goods-producing firms is charged to cost of sales because production management is absorbed into inventory overhead. Each position has one functional role, so the first four rules do not overlap. The internal information-technology rule is applied last and overrides them, so an internal information-technology position is charged to SG&A whatever its functional role. In aggregate, 13.2 percent of organization-capital payroll is assigned outside SG&A, 10.2 percent to R&D and 2.9 percent to cost of sales. The components do not sum exactly because of rounding.

Functional role	Expense line
Selling, marketing, administrative, and finance	SG&A
Research (scientist roles)	R&D
Engineering	R&D when the role indicates product development or research; SG&A otherwise
Operations	Cost of sales in goods-producing firms; SG&A otherwise
Internal information technology, in any function	SG&A