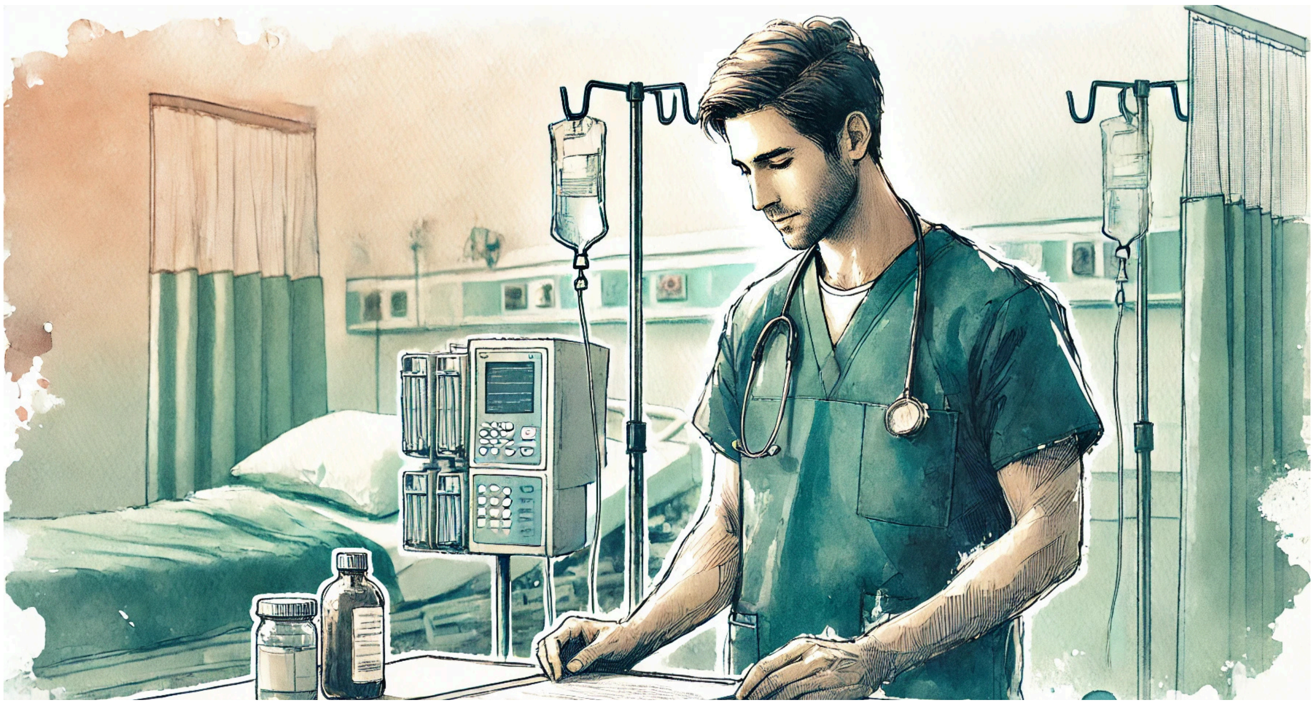
HEAL ECONOMY: TECHNICAL APPENDIX

This appendix provides additional details on our data sources, alternative classification methods, and methodological and analytical choices.



CONTENTS

Data sources

This research brief introduces the HEAL (health, education, and literacy) classification, drawing on methodologies from prior efforts to classify STEM occupations. HEAL occupations are identified using three complementary frameworks: task-based, skill-based, and knowledge-based classifications. Each framework builds on publicly available occupational data from the following sources:

Occupational Information Network (O*NET)

Developed by the U.S. Department of Labor, O*NET surveys incumbent workers to describe occupations across multiple dimensions including worker characteristics (abilities, interests, values), requirements (skills and knowledge), occupational contexts (work conditions and activities), and job-specific tasks and tools. O*NET offers a comprehensive view of the competencies and conditions associated with hundreds of occupations, making it a valuable tool for workforce analysis.

While O*NET has extensive occupational data, not all occupations have data on all measurements. We merged O*NET occupation data to produce a dataset of work activities, work context, work styles, work values, interests, skills, and knowledge for 734 occupations from an initial list of 774 occupations. All classification methods used this dataset of 734 occupations.

U.S. Bureau of Labor Statistics (BLS)

We also merged O*NET data with the BLS's Employment Projections (EP) to capture economic characteristics, including employment levels, projected growth, and wages. However, not all O*NET occupations align with the SOC-based classifications used in BLS datasets, resulting in dropped occupations during the merge process. Overall, our merged O*NET-BLS dataset captures 91% of total employment.

The O*NET-SOC taxonomy is more granular than the BLS Standard Occupational Classification (SOC) system, providing detailed classifications for specialized roles within broader occupational categories. For example, while the BLS SOC groups all “registered nurses” under the single code 29-1141, the O*NET-SOC further differentiates specific nursing roles like “acute care nurses” (29-1141.01), “advanced practice psychiatric nurses” (29-1141.02), “critical care nurses” (29-1141.03), and “clinical nurse specialists” (29-1141.04).

The BLS Employment Projections (EP) data includes employment projections for 832 occupations. When merged with O*NET-SOC data, we excluded O*NET-SOC occupations that lack corresponding entries in the BLS dataset, resulting in a final dataset with employment projections for 730 occupations. For the purposes of our analysis, where we rely on BLS data, we used the broader O*NET attributes associated with occupations like “Registered Nurses” (29-1141.00) and excluded data for more granular O*NET-SOC classifications, such as “Acute Care Nurses” (29-1141.01) and “Critical Care Nurses” (29-1141.03), to align with the BLS framework.

American Community Survey (ACS)

The gender composition of occupations was derived from an analysis of ACS five-year data, a robust dataset produced by the U.S. Census Bureau. This dataset aggregates responses collected over a five-year period, providing detailed demographic, social, economic, and housing information about the U.S. population. The 2023 ACS five-year estimates combine data from 2019 to 2023, offering a larger sample size and more precise estimates than annual data. By using ACS five-year data, we can analyze detailed occupation-level characteristics, such as gender representation or educational attainment, with the understanding that these estimates reflect averages across the entire five-year period rather than a single year.

However, gender composition data was not available for all O*NET-SOC and BLS occupations. When linking ONET-SOC and BLS employment projections with the 2023 ACS five-year data—specifically Tables B2115 and B2116 on gender composition—a total of 390 occupations remained for analysis.

For our analysis of the share of men and women in HEAL and STEM occupations over time (1970-2020), we used the 2018-2022 ACS five-year data, as this was the most recent microdata available.

We also use the ACS to analyze racial compositions for each occupation and classification. We use a published 2022 ACS table to calculate the racial composition of HEAL and STEM occupations. Like the process for calculating gender compositions, occupations are dropped in the merging process, again resulting in 390 occupations for analysis, or 73% of total employment.

Occupational classifications

Unlike STEM, where different classification methods can yield widely varying workforce estimates, these HEAL classifications produced relatively similar results in terms of size, wages, and projected growth. This consistency suggests that HEAL occupations share fundamental characteristics that make them identifiable across multiple classification methods, reinforcing the idea that HEAL represents a coherent and meaningful category of work rather than being dependent on any single approach.

Using the original 734 occupations available in O*NET, we then constructed three potential classification indices.

Overall, depending on the classification, the HEAL economy encompasses between 24.3 and 33 million jobs and is expected to produce 1.6 million to 2.7 million new jobs by 2033.

Comparing HEAL classification methods

Characteristics of HEAL occupations by classification

	Skill-Based	Knowledge-Based	Task-Based	STEM
# of Occupations	130	134	128	105
Current # Employed (2023)	33M	24.3M	24.9M	10.7M
Projected Job Growth (2023-2033)	2.7M	1.6M	2.2M	1.1M
Median Wages	$64K	$64K	$61K	$102K
Median Wages (BA+)	$86K	$80K	$74K	$105K
Median Wages (no-BA)	$40K	$41K	$39K	$59K

Wage estimates are in 2023 dollars. Wages are calculated based on median wages of individual occupations. Occupations were classified as STEM according to the BLS OEWS May 2023 definition.

Source: Authors' Analyses of O*NET & BLS EP

aibm AMERICAN INSTITUTE FOR BOYS AND MEN

Figure 1

To complement our knowledge-based classification, we explored two additional approaches for classifying HEAL occupations: based on tasks and based on skills. While these methods are not the focus of our primary analysis, they provide valuable context and help highlight the strengths and limitations of different approaches. The task-based classification relies on the Standard Occupational Classification (SOC) system, which groups occupations based on the primary work performed and offers a straightforward and widely used framework for workforce analysis.

The skill-based classification, on the other hand, uses detailed data from O*NET to identify occupations that emphasize interpersonal and non-cognitive skills critical to care-related work. Both approaches contribute to our understanding of HEAL roles, and a closer examination of their methodologies and results is included below.

Task-based

The Standard Occupational Classification (SOC) was established by the Office of Management and Budget (OMB) to aid comparisons across federal agencies who publish occupational data. One of the principles of SOC is that occupations are primarily classified based on "work performed." For example, under the current 2018 SOC, physician assistants "conduct complete physicals, provide treatment, and counsel patients" and "may, in some cases, prescribe medication."

Our task-based HEAL classification include all occupations in the following "major groups":

- 21-0000: Community and social service occupations
- 25-0000: Educational instruction and library occupations
- 29-0000: Healthcare practitioners and technical occupations
- 31-0000: Healthcare support occupations

A task-based HEAL classification has the advantage of being straightforward, easy to understand, and a common way of classifying occupations used by a range of federal agencies and researchers.

On the other hand, focusing solely on tasks may not capture the full complexity of an occupation and its underlying knowledge or skill requirements. Other scholars have also noted a movement away from task-based classifications toward those based on skills or aptitudes that may be more transferable across occupations and that may better support workforce development in today's labor market where job switching is more common.

Skill or aptitude-based

Recent research has highlighted the increasing labor market demand for non-cognitive skills like social perceptiveness and communication, with more jobs requiring social interaction across industries. Social skills can be especially important in the "care economy" where outputs tend to be individually tailored and involve high levels of personal connection, intrinsic motivation, and responsibility for the wellbeing of others.

Influenced by literature on the care economy, pink-collar workforce, and the increasing importance of non-cognitive skills in the labor market, we created a "human-centered" index of O*NET attributes to capture essential care-related job attributes like human interaction and a responsibility for others as well as worker-specific traits and dispositions like social perceptiveness and a concern for others.

The process of developing this index began with a data-driven, agnostic approach to uncover the key axes of variation within a broad spectrum of occupational characteristics. We compiled a dataset of 734 occupations, each described by 331 O*NET variables spanning "work styles," "work activities," "work context," "skills," and "abilities." We then performed a principal component analysis (PCA) on these occupational attributes to extract the primary latent dimensions, with four components collectively explaining nearly 63% of the overall variance. Although each component encompasses multiple variables, we interpret them as follows:

1. **Communicative and cognitive (38% of variance):** Emphasizes speaking, active listening, judgment, and complex decision-making.
2. **Technical (16% of variance):** Involves process monitoring, equipment inspection, perceptual speed, and quality control
3. **Social (6% of variance):** Aligns with our human-centered index, reflecting social orientation, concern for others, and caregiving.
4. **Precision (3% of variance):** Focuses on task repetition, accuracy, and selective attention.

This four-component solution served as the foundation for further refinement and index construction. Below we highlight the top 10 variables that load most heavily on each of the above dimensions.

O*NET attribute loading on top principal components

Top 10 highest loaded O*NET attributes

PC1 - Cognitive and Communicative	Loading	PC2 - Technical	Loading	PC3 - Social	Loading
Speaking_LV	0.108	Operations Monitoring_LV	0.130	Social Orientation	0.166
Active Listening_LV	0.106	Inspecting Equipment, Structures, or Materials_LV	0.130	Assisting and Caring for Others_IM	0.165
Critical Thinking_LV	0.105	Repairing and Maintaining Mechanical Equipment_LV	0.129	Performing for or Working Directly with the Public_IM	0.165
Active Learning_LV	0.103	Inspecting Equipment, Structures, or Materials_IM	0.129	Concern for Others	0.163
Reading Comprehension_LV	0.103	Troubleshooting_LV	0.127	Self-Control	0.160
Writing_LV	0.103	Operations Monitoring_IM	0.127	Assisting and Caring for Others_LV	0.158
Judgment and Decision Making_LV	0.100	Troubleshooting_IM	0.126	Programming_LV	−0.154
Active Listening_IM	0.100	Repairing and Maintaining Mechanical Equipment_IM	0.125	Physical Proximity	0.152
Critical Thinking_IM	0.100	Exposed to Hazardous Conditions	0.124	Performing for or Working Directly with the Public_LV	0.151
Active Learning_IM	0.100	Equipment Selection_LV	0.124	Deal With Unpleasant or Angry People	0.150

aibm
AMERICAN INSTITUTE FOR BOYS AND MEN

Source: Authors' Analyses of O*NET

Figure 2

What we identified as a “social orientation dimension” (PC3) seemed to capture many of the same attributes described as unique or important in the literature on care work, the care economy, and the role of noncognitive factors in labor market outcomes.
Based on these results, we created an index of human-centered job attributes:

- **Contact with others (work context):** How much does this job require the worker to be in contact with others (face-to-face, by telephone, or otherwise) in order to perform it?
- **Assisting and caring for others (general work activity):** Providing personal assistance, medical attention, emotional support, or other personal care to others such as coworkers, customers, or patients.
- **Service orientation (skills):** Actively looking for ways to help people.
- **Social perceptiveness (skills):** Being aware of others’ reactions and understanding why they react as they do.
- **Concern for others (work style):** Job requires being sensitive to others’ needs and feelings and being understanding and helpful on the job.
- **Establishing and maintaining interpersonal relationships (general work activity):** Developing constructive and cooperative working relationships with others, and maintaining them over time.
- **Interpreting the meaning of information for others (general work activity):** Translating or explaining what information means and how it can be used.
- **Relationships (interests and values):** Occupations that satisfy this work value allow employees to provide service to others and work with co-workers in a friendly non-competitive environment.
- **Social (interest and values):** Work involves helping, teaching, advising, assisting, or providing service to others.

We calculated the average score for each attribute across 734 O*NET occupations and then standardized them. This involved subtracting the mean score for each attribute from the occupation-specific score and dividing by the standard deviation. Occupations with a standardized score exceeding 1.0 (i.e., one standard deviation above the mean) on the combined index were classified as HEAL.

Knowledge-based classification

O*NET measures knowledge requirements for occupations through structured surveys completed by subject matter experts, job incumbents, and analysts. Knowledge is defined as the organized set of principles and facts in specific domains necessary for job performance. The survey asks respondents to evaluate the importance of thirty-three standardized knowledge domains (e.g., "education and training", "medicine and dentistry", "psychology") and the level of knowledge required for a given occupation. Ratings are provided on a scale from 1 (not important) to 5 (extremely important) when rating the importance of such knowledge, and from 1 to 7 when rating the required level of knowledge with specific anchor questions to calibrate the scale. For example, when asking about the level of "economics and accounting" knowledge required in an occupation, a 2 is anchored as "answers billing questions from credit card customers." These ratings are averaged across respondents to produce occupation-specific knowledge profiles.

The resulting data provide a nuanced understanding of the knowledge essential to perform various jobs, which is particularly useful for workforce analysis and policy development. By standardizing knowledge domains across occupations, O*NET enables comparisons and aggregations that inform curriculum design, training programs, and occupational classification frameworks.

To standardize these ratings, we calculated the mean and standard deviation for each knowledge domain across all 734 occupations in our dataset. Each occupation's score was then standardized with the same method used for our skill-based classification. Occupations scoring at least 1.5 standard deviations above the mean in any one of these domains were classified as HEAL:

- Education and training
- Medicine and dentistry
- Therapy and counseling

This threshold ensures that only those occupations with significantly higher-than-average knowledge requirements in these areas are included in the knowledge-based HEAL classification. Standardization allowed us to compare knowledge requirements across occupations on a consistent scale and create a robust framework for identifying HEAL roles.

While other knowledge domains are potentially relevant—psychology, sociology and anthropology, personnel and human resources, and customer and personal service—these were excluded due to overlap or strong correlation with other fields, though we acknowledge that this may overlook their contribution to HEAL-related fields.

For more detail and analysis of O*NET knowledge scores see Todd Gabe's paper, "Knowledge and Earnings," where he examines the marginal effect of specific domain knowledge on earnings. Of note, he finds that medicine and dentistry knowledge has large and significant marginal effects on wages, the therapy and counseling domain has smaller but still positive effects, and the education and training knowledge domain has small negative effects on earnings.

Overlap across classifications

Overall, approximately 51% of occupations are common to all three classifications and an additional 28% are present in at least two, underscoring the methodological trade-offs.

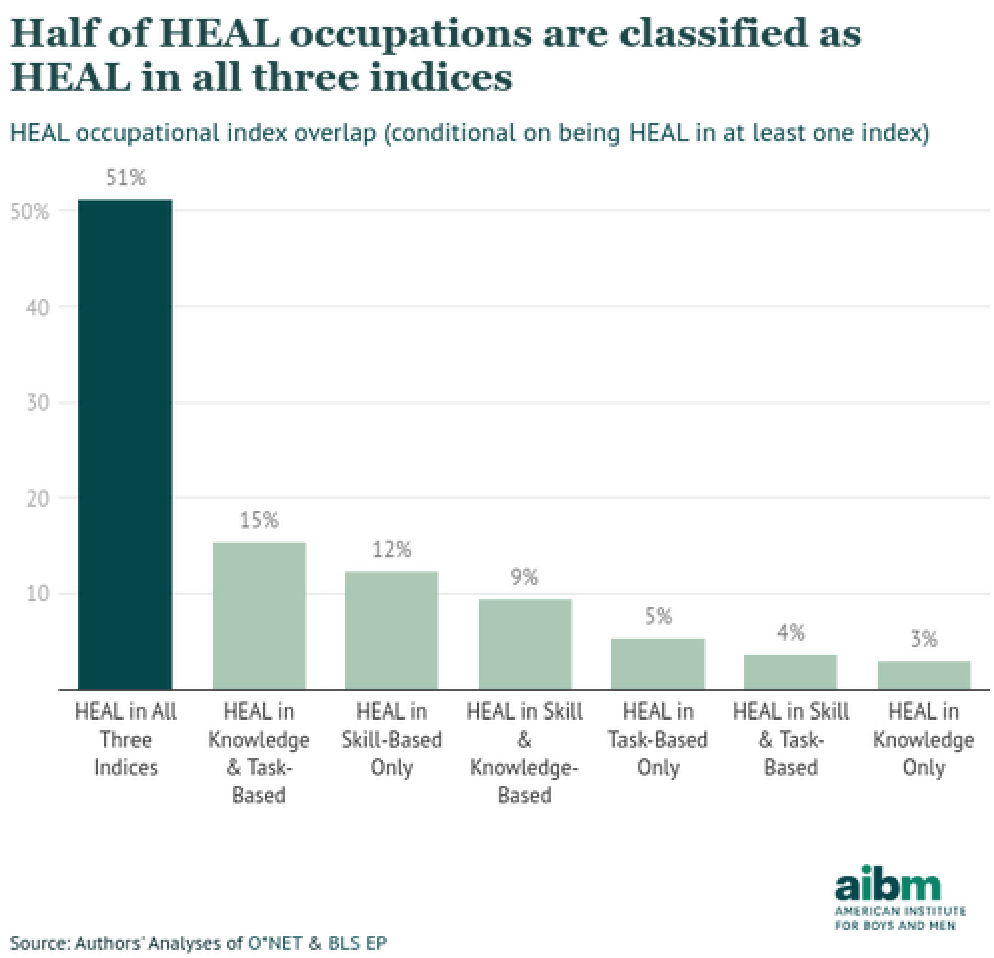


Figure 3

STEM estimates

We use the U.S. Bureau of Labor Statistics' Occupational Employment and Wage Statistics (OEWS) definition of STEM, which includes the following guidance:

"This data set uses only one of many possible definitions of STEM. The Standard Occupational Classification Policy Committee (SOCPC) has provided options for defining STEM under the 2018 SOC system at www.bls.gov/soc/2018/#crosswalks. The definition used in this data set consists of all occupations in subdomains 1 and 3, as shown in Appendix B of the SOCPC guidance."

This definition encompasses 105 occupations, which we use to estimate the size of the STEM workforce, projected job growth, median wages, the percentage of jobs requiring a bachelor's degree, and the male share of employment (though male share is not available for all 105 occupations).

Additionally, not all STEM occupations have corresponding O*NET data. After merging with O*NET, our working set consists of 91 occupations. As a point of comparison, the total size of the STEM workforce based on the full BLS classification is estimated at 10.7 million workers, with projected growth of 1.1 million. Using the smaller universe of STEM occupations with available O*NET data, these estimates decrease to 7.1 million workers and 650,000 in projected growth.

Since knowledge and human-centered averages are derived from O*NET, they are calculated using only these 91 occupations. Additionally, because some STEM occupations are missing from our dataset, they are not included in our index constructions. This means that standard deviation (SD) calculations may not fully account for the entire STEM occupation pool. Given that the omitted occupations are likely to be on the lower end of our knowledge index, their inclusion—if it were possible—could shift classification thresholds, potentially causing some occupations that are currently excluded from HEAL to be reclassified as HEAL. However, due to O*NET data limitations, these occupations cannot be incorporated.

Wage calculations

To analyze wages across HEAL, STEM, and other occupations, we used the following methodology:

1. Extract occupation-level wages: We obtained the median annual wage for each detailed occupation from the BLS EP dataset.
2. Assign knowledge index classifications: Each occupation was assigned a classification based on the knowledge-based HEAL framework to categorize it as HEAL, STEM, or Other.
3. Calculate group-level wages: Within each classification group (HEAL, STEM, other), we computed a weighted median wage, where individual occupation medians were weighted by their 2023 employment counts.

This method provides a central estimate of earnings within each group but does not capture the full wage distribution across all occupations, because we use the median wage within each occupation and then take a weighted median across categories.

An alternative approach would be to calculate the actual median wage across all occupations using the 2023 monthly CPS. However, this would exclude some occupations due to data availability. In order to capture all 700+ occupations consistently across analyses, we rely on the BLS EP dataset to ensure comparability between occupation-level and classification-level wage estimates.

Male share in HEAL and STEM occupations over time

Harmonized occupational coding: The OCC2010 scheme

To determine changes in the share of men and women in HEAL and STEM occupations over time, we make extensive use of a harmonized occupation classification variable in IPUMs, OCC2010, based on the Census Bureau's 2010 occupation classification system. In the interest of harmonization, the 2010 scheme has been modified so that the occupational categories are as comparable as possible over time. That is, detailed categories available in the 2010 data are sometimes grouped together because earlier censuses coded multiple occupations under a single combined category.

Data processing and merging

The analysis begins by extracting a detailed breakdown by occupation and sex from samples from each decade between 1950 and 2010 using the online IPUMs data analysis system. This provides the total number and proportion of men and women in each OCC2010 occupation.

Next, the 1950-2010 data are merged with the same occupation by sex summary obtained from the ACS 5-year microdata (2018–2022 because the 2019-2023 microdata was not available at the time of analysis). This ACS data includes not only the OCC2010 code but also the SOC code (using the OCCSOC variable), allowing us to create a crosswalk between the OCC2010 occupation codes and their corresponding 2018 SOC codes. This is necessary because STEM and HEAL roles are classified according to 6-digit SOC codes.

One complication is that 60 OCC2010 codes map to multiple OCCSOC codes. For example SOC occupations 11-2011 (advertising and promotions managers), 11-2021 (marketing managers), and 11-2022 (sales managers) all map to a single OCC2010 code for "managers in marketing, advertising, and public relations." A second, similar complication, is that our classification of STEM and HEAL roles only looked at the detailed (6-digit) occupations but some 6-digit SOC occupations which we might have classified as STEM or HEAL are not captured in that level of detail in the ACS.

For example, 15-1231 (computer network support specialists) and 15-1232 (computer user support specialists) are both classified as STEM but the broader 15-1230 (computer support specialists) is not.

To correct these issues, we reviewed the crosswalks between OCC2010, OCCSOC, and the 2018 SOC, then adjusted classifications by assigning a broader occupation the same label as its detailed occupations when consistent, basing the classification on the dominant workforce size when the detailed occupations were mixed (for example, “medical secretaries and administrative assistants” are HEAL, but only a small proportion of the broader “secretary and administrative assistant” category and so this broader category was classified as “other”), and using judgment when workforce size data was unavailable (such as classifying all post-secondary teachers as HEAL despite some chemistry instructors being STEM).