Scholars Journal of Engineering and Technology
Abbreviated Key Title: Sch J Eng Tech
ISSN 2347-9523 (Print) | ISSN 2321-435X (Online)
Journal homepage: https://saspublishers.com

OPEN ACCESS

Productivity and Wellness Can be scored on a Scale Based on Biases and Relevance

Nitish Shrivastava[1*], Neel Tiwari[2], Elaiyaraja Thangavel[3]

[1]Senior Vice President at Persistent Systems, Cupertino, CA - USA
[2]Senior Product Manager at Persistent Systems, Pune, India
[3]Associate Vice President at Persistent Systems, Pune, India

DOI: 10.36347/sjet.2022.v10i10.001 | **Received:** 29.08.2022 | **Accepted:** 06.10.2022 | **Published:** 11.10.2022

***Corresponding author:** Nitish Shrivastava
Senior Vice President at Persistent Systems, Cupertino, CA - USA

Abstract **Original Research Article**

The research conducted in this article was used to prove that it's possible to score productivity and wellness on a scale based on biases and relevance to individual subjects. Previous research conducted to prove that workplace stress can affect productivity and wellness was used to establish profiles under which working professionals could be assessed for success/failure criteria based on profile-specific activities and their outcomes. In this research article, we show that by assessing specific activities based on formulae derived from success/failure criteria for such activities, it is possible to identify an associated score ranging from 1 to 10, 1 being the worst outcome possible, and 10 being the most successful. It is further shown that scores can be tailored to individuals, groups, and organizations under profiles by using a numeric system that captures *bias* and *relevance* of a particular activity to the assessed subject, thereby increasing its utility and applicability as a measure of productivity and wellness.

Keywords: Work-life, productivity, health, wellness, flow, organizational resources, personal resources, positive, bias, relevance, score, formulae.

INTRODUCTION

With the renewed focus on productivity and wellness, organizations the world over have recognized threats to productivity and wellness as fundamental threats to public health and safety of employees (Goubin Dai, 2021), and that it has long term consequences for individuals and organizations alike.

One area of focus in this regard has been attempts to come up with a system for measuring and assessing productivity; tied to this idea has been attempts to understand if there exists a state in which individuals achieve peak productivity and wellness. One of the most famous examples is the concept of *flow* during work that was introduced in *Optimal Experience in Work and Leisure* by *Csikszentmihalyi* and *LeFevre;* it utilized methods like the Experience Sampling Method (ESM) to assess productivity of individuals in their day-to-day tasks. Previous research by the authors of this study also delved into this area; most notably turning up evidence that workplace stress is a key factor in degrading productivity and wellness. During the course of that research, the authors ended up gathering data on activities that were eventually organized by "profile" of working professionals; it included information about what outcomes (tied to specific profiles) could be considered *productive* or *unproductive* for each of the identified activities.

It became important (as the next step) to determine if the data being gathered (along with the indications of productivity success/failure) could be used in a reliable manner to evaluate performance as an individual. It should be noted that such a methodology could open up opportunities to advance an individual in terms of social comparison, which was shown to potentially lead to hostile attention from peers and sub-ordinates (Chuang Zhuang, 2020), and in certain instances, superiors (Yu, Duffy, 2018); it was therefore important to come up with a *universal* system that could be used to fairly assess individuals on a criteria built up by consensus on profile, expected tasks, responsibilities, priorities, etc. to engender maximum co-operation and acceptance.

The question we seek to answer here is, is it possible to come up with a consistent scoring system that can *relate* productive and unproductive activity outcomes based on the profile of the individual? An additional question we seek to answer here is, can scores generated from such a system be tailored to individuals based on their relevance to individual experiences, and the importance that the individual itself places on specific activities that go towards generating a score?

There are obvious advantages from an organizational perspective as well; employees who can utilize such a system in order to achieve a work-related flow can result in a feedback loop with efficient utilization of organizational resources facilitating work-related flow in return (Salanova, Bakker, LLorens, 2005). This is in addition to personal benefits from the self-efficacy that such a system would foster in individuals; this self-efficacy has been shown to be a great driver for increased personal success, providing a good example for individuals following in their footsteps, and positive feedback (Ma, Tschirhart, 2021). Even from a wellness perspective, it has been shown that self-efficacy is a significant factor in reducing the likelihood and intensity of burnout across individuals (Van Deusen, 2002).

Ultimately, a scoring system must look to maximize use of limited resources available to individuals (time, health, attention, cognitive ability, etc.). It has been found that inefficient usage of such resources (in other words, a "loss" of resources) is strongly co-related with the production of stress; such a hypothesis has already been widely discussed in the Conservation of Resources (CORS) theory (Hobfoll, 1989, 2021). A scoring system that results in minimization of resource loss while maximizing success outcomes would (in the opinion of the authors) make a significant dent in the effects and magnititude of workplace stress, and would thus be a worthy endeavour.

MATERIALS AND METHODS

The authors of this research study relied on data gathered from a previous research study; the preceding study was built out of extensive research data that was collected from working professionals. We will first re-introduce materials and methods utilized for data from the preceding study, and then supplement additional information that explains our methodology for answering the relevant questions for *this* study.

STATEMENT OF HYPOTHESIS

Null Hypothesis (Ho)

Productivity scoring systems would be unable to generate scores that (at least 95% of the time) is within 0.5 to 1 points of a score that the subjects would intuitively assess to encapsulate their day.

Alternate Hypothesis (Ha)

Productivity scoring systems are able to generate scores that (at least 95% of the time) is within 0.5 to 1 points of a score that the subjects would intuitively assess to encapsulate their day.

CONTROL VARIABLES EMPLOYEED

In order to ensure widest spread of data, data was collected in the following proportions from subjects:

- Across multiple job disciplines (criteria explained in section *organizing subjects based on profile* below).
- Equal numbers of men and women.
- Equal proportions of shift-based employees (morning shift and night shift).
- Split in equal proportions across managerial/individual contributor roles.
- Spanning multiple countries (United States, India).

EVOLUTION OF METHODOLOGY

The next set of sections describes how the authors evolved the methodology to be employed for testing the hypothesis.

Organizing Subjects based on Profile

(Note: At the outset, we obtained explicit permission from our subjects to collect each data attribute that we utilized in our research).

Some of the information that we collected at the beginning was still useful. Since it was important that we have a wide spread of working professionals represented in our subject list, we identified some key "profiles" of working professionals that we wished to study and aligned them to specific jobs/professions to aid in classification and segregation. These included:

- Sales professionals.
- Software developers.
- Support engineers.
- Product Managers.
- QA engineers.

It should be noted that most of these subjects were engaged over a period of *5 years* in terms of collection of data for our study.

Evolving Quantifiable Categories of Information to Collect

We spent some time interviewing subjects from each of these profiles. Instead of trying to rely on manual recorded observations, we first started with a set of specific questions:

- *What measures do you usually employ in your job profile to determine that your goals have been achieved?*
- *What conditions/outcomes during the course of your specific job profile would be considered as a failure to achieve your goals?*
- *Based on previous questions, how would you categorize your job performance?*

Asking these questions and arriving at a consensus resulted in a set of “criteria” for productivity/wellness for each job profile being evaluated.

General Wellness Criteria among Profiles

For individuals belonging to these profiles, we intended to make a general case for analyzing wellness; accordingly, we captured certain common data across all profiles (like Heart rate, Sleep, Steps and Fitness activity).

Sales Professionals

This profile corresponds to working professionals who work in sales to win deals that result in additional revenue for their employer.

For such individuals, success would be categorized under the following categories:

- Bringing in new sales leads that result in opportunities for increased revenue for the company.
- Successfully closing deals with customers to realize additional revenue (and doing it as quickly as possible).

Conversely, there are a few scenarios that could be judged as a “failure” in productivity:

- Failure to bring in new leads over the course of a financial year.
- Failure to close out deals, resulting in dropping these opportunities and preventing revenue from being realized.
- Failure to *adequately* pursue open opportunities through available methods like customer in person meets calls, emails/meetings, etc.

Armed with objectives for “gauging” productivity, we captured data from specific data sources:

- Sales related data (CRM).
- Emails/Meetings.
- Travel information.

Software Developers

This profile corresponds to software engineers who are directly or indirectly responsible for maintaining the code base of products/services/projects in an organization, whether it be by contributing to new features or fixing existing issues.

For such individuals, success would be categorized under the following categories:

- Timely contributions to the source code management system (which would imply quick closure of assigned defects, fast closure of requests for enhancements or new features).
- Good quality contributions (which would imply minimizing of defects arising from changes/fixes to the product, infrequent changes happening on touched source code files, etc.).

Failure in such cases in terms of productivity would include:

- Leaving open assigned defects opened for a long period of time without closure.
- Bad quality of code contributions resulting in increased issues, and requiring more code rewrites, slowing down development, etc.

An attempt was made to capture data in these categories:

- Code check-ins in source code management systems utilized by subjects.
- Bugs/feature requests in project management software.
- Emails/Meetings.
- Software App Usage.

Support Engineers

This profile corresponds to working professionals who are responsible for directly interfacing with customers utilizing products/services from their organization (with the purpose of customer assistance/support, preliminary analysis, communication with backend teams, and closure of reported issues).

For such professionals, success criteria would include:

- Number of customer issues (aka “tickets”) resolved.
- Reduced time to resolve filed tickets.

On the flip side, failure would include scenarios like:

- Taking too long to resolve tickets.
- Having a higher number of critical/high priority tickets open without resolution.

To account for such scenarios, we captured data in the following categories:

- Incident management system tickets.
- Emails/Meetings.

Laying out Patterns for a Working Professional’s day

To properly assess the data required to answer this question, it must be framed in a manner that can be tied to typical patterns of work and leisure that working professionals undergo. We started by taking a time interval of 1 day (24 hours) out of the life of a person; we can roughly categorize periods of the day in the manner below. Note that in this instance we are assuming that the subject works in the morning shift; in the case of working professionals who work in different shifts, the time periods and associated tasks/behaviours would change accordingly.

- ***Sleep period:*** This is the number of hours of the day during which a subject would typically be in a sleep state. The actual quality of sleep during this period would have to be judged by multiple factors (i.e., REM periods, number of times that subject woke up, amount of sleep, how closely it fit circadian rhythms, etc.).

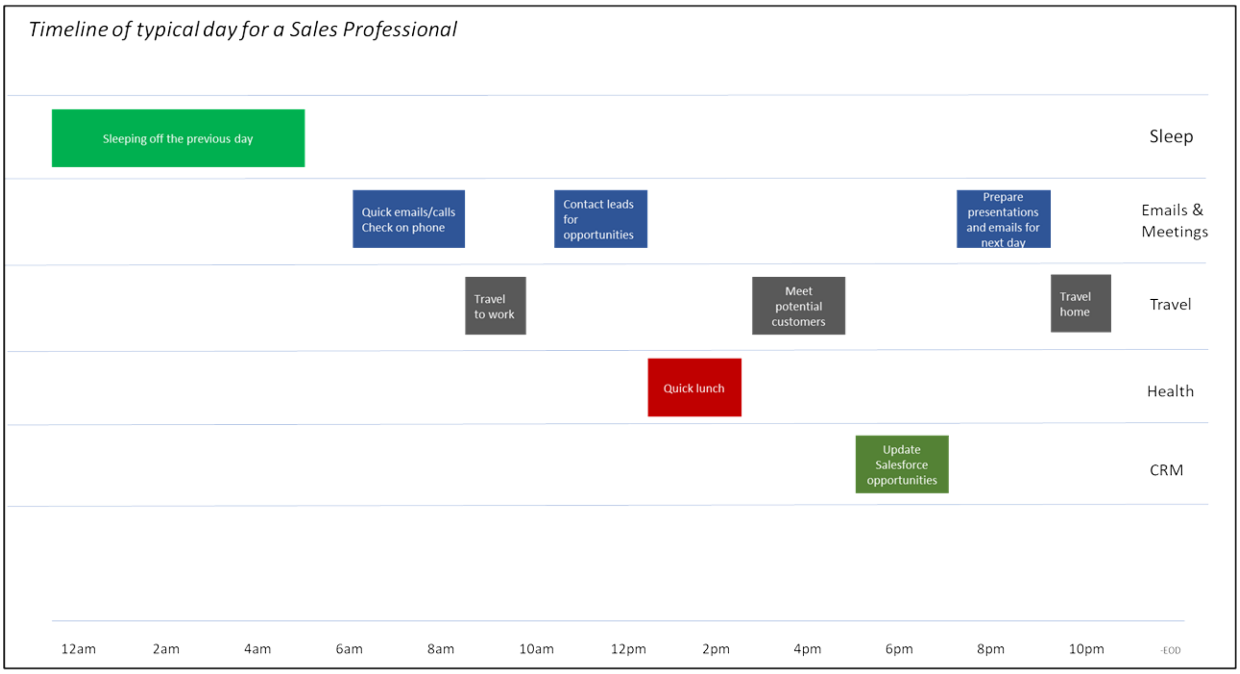
- ***Post sleep morning period:*** Typically, this is the time just after the subject has woken up, where the subject would indulge in activities that would eventually transition into typical activities during the day; this could include time to brush, take a shower, morning constitutional, breakfast, etc. Note that it's possible that activities during this period may also include preparatory work for the rest of the day or may include physical activities from the point of view of exercise.
- ***Office commute period:*** This can vary from person to person (and may not even exist for a subject who works remotely 100% of the time). The time period can vary depending on the commute to work distance, condition of traffic based on time of day, etc.
- ***Working hours:*** This would be the period where a subject is *expected* to engage in most productive activities from the context of the job/profession.
- ***Return commute period:*** At the conclusion of the day, if the subject is working from an office, this period would coincide with the return journey back to subject's home.
- ***Pre sleep period:*** Usually associated with a "winding down" of the day (and can also include physical activities), including dinner and relaxation activities followed eventually by commencement of the sleep period.

At this point of the methodology, we had essentially constructed a "picture" of a person's day based on time periods of *presumed activity*. This point is crucial; even aside from the fact that these patterns of time periods can vary depending on "shifts" in which a working professional can operate, it also doesn't fully consider the *quality* of activities that are undertaken during these periods. We instinctively (which is to say, without the need for explicit measurement) can ascertain that it is very rare for a working professional to have periods of activity that occur with such consistency, and even in the event of said periods actually coinciding with the "expected" activities, it is rare that there not be some sort of interruption or negative effect on quality of the activity, be it physical, mental, psychological or otherwise.

The decision was then made to more finetune our picture of working professionals by relying on *sources of data* that are available throughout our subject's day.

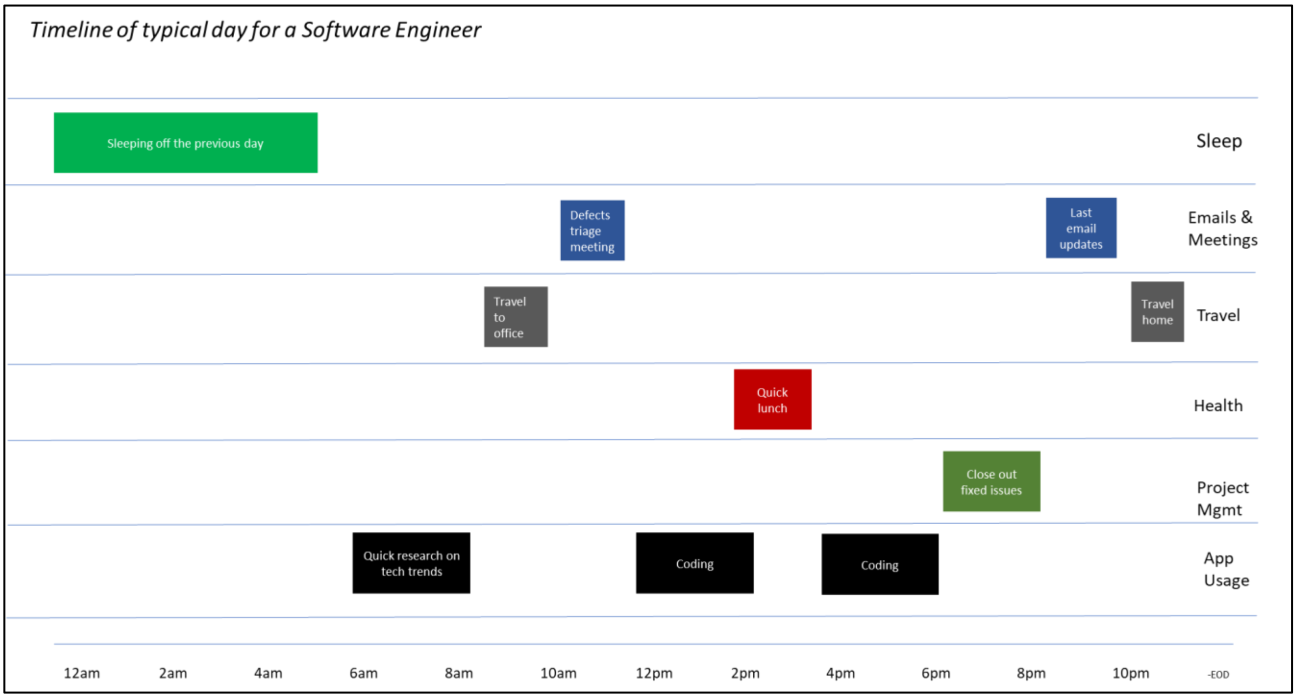
Timeline for a Sales Professional

To accomplish this, we interviewed sales professionals to build a "picture" of the sales professional's day. To do this, we create a timeline that models all 24 hours of a person's day, and then placing (based on their feedback) typical periods of activity. Accordingly, we come up with the following diagram for particular sales professional:



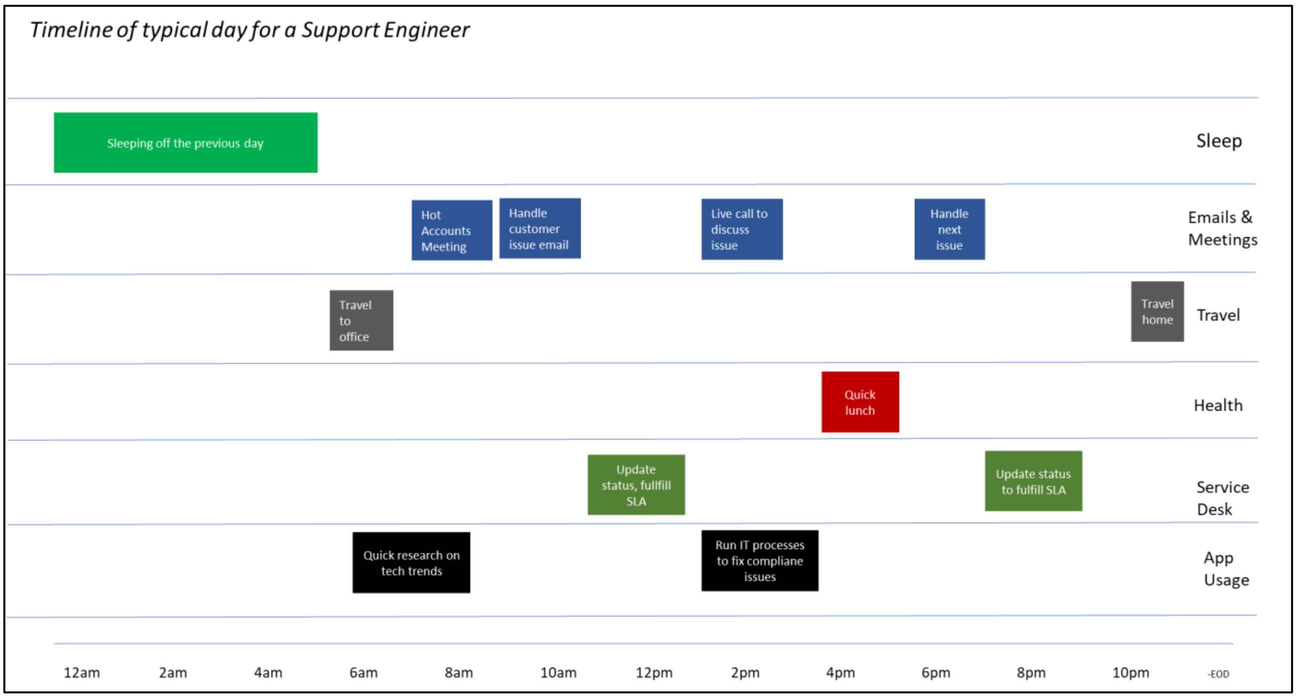
Timeline for a Software Developer

In a similar fashion, we came up with a representative timeline for software developers:



Timeline for a Support Engineer

A similar timeline was created for support engineers as follows:



Organizing the Set of Subjects and Data Capture

As we mentioned before, our original plan was to have our subjects maintain work diaries that they would write into overtime. There were several problems with this approach; for one thing, it wasn't very reliable as a *comprehensive* record of activities since it depended on frequency and accuracy of written entries; for another, it could potentially *detract* from the efficiency with which work activities were undertaken, thus potentially undermining the study.

Once we had this realization, we realized that our only recourse would be to *automate* the collection of data from the subjects. Accordingly, we went back to the subjects, and determined what productivity tools were utilized for our subjects to do their day-to-day jobs? We gathered the answers to this question across all profiles and came up with a list of services/information to be gathered. Over time, we researched methods for gathering this information in an automated fashion (web services, APIs, software/apps, etc.), and began the process of monitoring and collection of information from subjects.

Accordingly, with appropriate disclosures of our intentions and with explicit permission obtained, we arranged for the capture of data from multiple subjects, using appropriate data sources to feed into our research. To briefly summarize the extent of data capture from subjects:

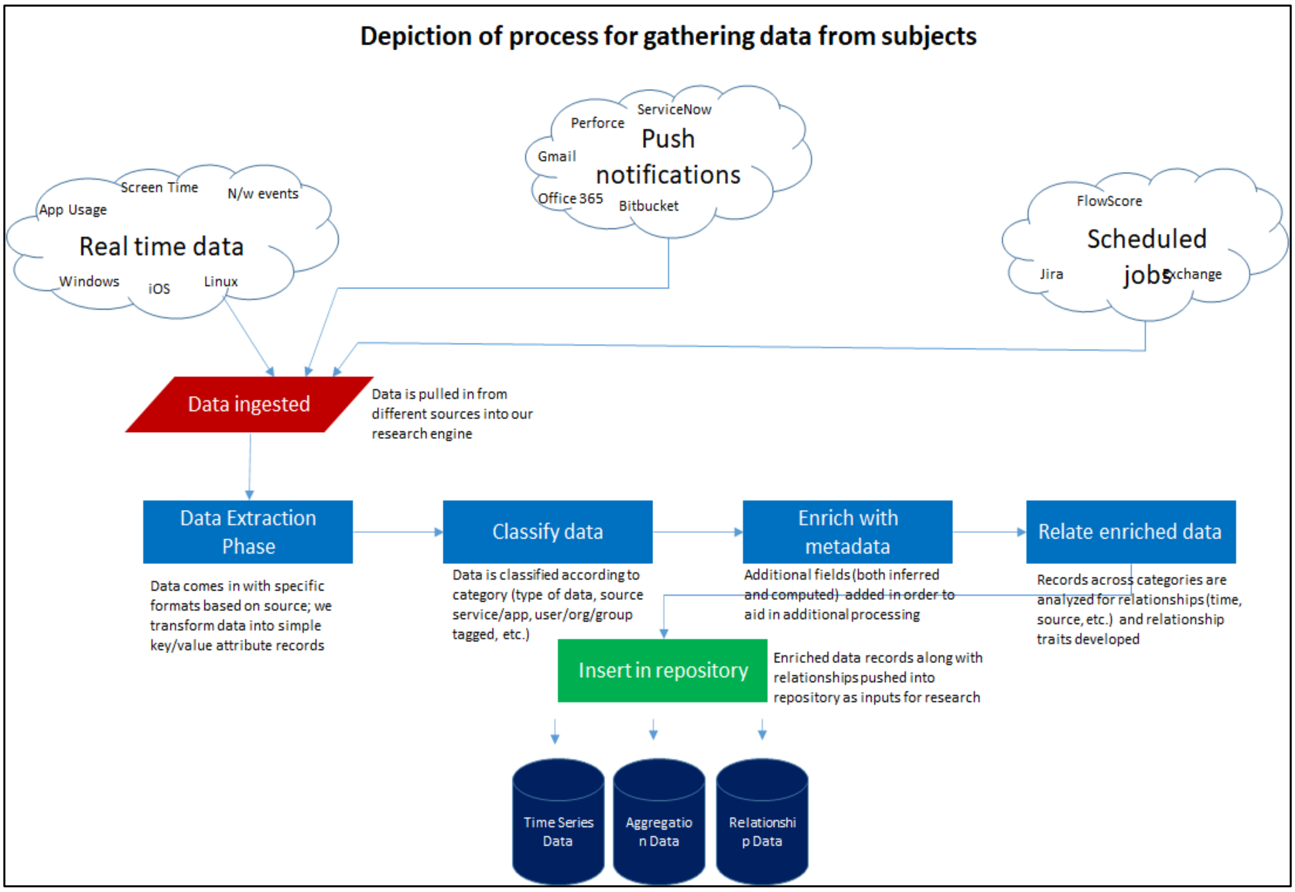
- **Health information:** We collected information about sleep, heartrate, fitness activities, etc. In addition, we also sync a health *score* that fitness tracker tools that subjects make use of.
- **Emails/meetings:** We collected emails/meetings information from popular office suites like Microsoft Office 365, Microsoft Exchange, Gmail, etc. with subject's consent.
- **Business desktop apps:** We collected information about screen time (defined as time spent actively working on a screen of a desktop PC, laptop, mobile, etc.), actual software/processes that the subject was working on, and factors that can influence degradation of work undertaken using business apps (i.e., network interruptions, machine restarts, etc.).
- **Business critical services:** We collected information from services that are utilized in some form or the other by subjects for specific purposes, i.e., CRM data from Salesforce, service desk tickets, tasks/issues filed in project management systems.
- **Mobile phone apps:** Given that cellular phones are now a critical medium of communication and work for professionals, we collected information from phones related to screen time, apps used and duration of usage, etc.

Relating Collected Data based on Specific Organizing Criteria

Finally, we evolved a system for *relating* the data was collected on multiple criteria to put them into proper context. These organizing criteria fell into the following:

- **Time based**: Data is related based on time of day, hour, day, week, month, year, etc.
- **Organized groups:** Data is aggregated and related based on groups of subjects, by role (i.e., engineers versus managers), by geographic proximity, by organization, etc.

The diagram below summarizes the data collection process:



Representing Aspects of a Scoring System: Attributes, Sub-Attributes

In order to properly introduce our scoring system, it is necessary to explain how factors that go into scoring were identified and computed.

To keep things simple, our scoring system envisioned having *top-level* scoring parameters called *attributes;* these attributes would roughly correspond to *categories* of work/apps that our subjects used in their day-to-day work activities (i.e. Emails, Service Desk, Business Apps Usage, etc.). The ultimate goal of the scoring system is to generate a number between 1 to 10 for each attribute; 10 indicating highest productivity being reached for that attribute, and 1 indicating a total failure to achieve productivity in that attribute.

For each attribute identified, specific *measurable criteria* (which may in fact apply to one or more attributes) are identified collectively as "fulfilling" the evaluations necessary for each attribute.

Taking the example of Business Apps Usage. It can be stated that scoring of "Business Apps Usage" as an attribute is a combination of:

- How much time was spent in a day working with apps.
- What percentage of that time was spent on *actual business apps* (as opposed to "non-business" apps like games. Note here that "non-business" as an adjective is subjective, i.e. the professional may be a professional video games QA tester, in which case this adjective would not apply).

In the above example, two *sub-attributes* have been identified:

- Time spent on apps in a day.
- Percentage of time spent in business apps versus non-business apps.

Introducing Formulae and Score Tables for Scoring

The next step is to *map* possible ranges of values for each sub-attribute to entries in a *score table*. Take the example of the "percentage of time spent in business apps versus non-business apps" sub-attribute. A possible score table could look like this…

Percentage of time spent in business apps	Sub-attribute score
100%	10
90%	9
80%	9
75%	7
60%	6
55%	6
50%	5

This can be interpreted in the following ways:

1. A subject who managed to spend 100% of his time exclusively working on business apps is a stellar achiever and gets the highest score (10).
2. A subject who spends between 80-90% of his time exclusively on business apps is pretty close to accomplishing the desired goal; he gets a score of 9.
3. When encountering subjects who have spent less than 80% of their time, the exact percentage determines how much their score is reduced (i.e. someone who doesn't spend more than half his working time using business apps gets a score of 5).

In addition, there may be a possibility that (for an attribute), one contributing sub-attribute may be *more important* than another contributing sub-attribute for a subject, or profile, or any other organizing criteria. For instance, it may be that the percentage of time spent on business apps (sub-attribute 1) may be more important than the total time spent on apps (sub-attribute 2), even if total time spent is part of the computation.

In this case, a scoring *weight* (a value between 0 to 1) is assigned to sub-attributes, such that the total *weight* of all sub-attribute weights is 1.

Carrying forward our last example, it would then be possible, say, to assign a weight of 0.2 to "total time spent on apps", but a weight of 0.8 to "percentage of time spent on business apps" (notice that the sum of combined weights evaluates to 1, this is important for our scoring system).

In such a scenario, the *attribute score* would be:

Attribute score = (*weight of sub-attribute 1 * score of sub-attributes 1)* + (*weight of sub-attribute 2 * score of sub-attribute 2)* + …. + (*weight of sub-attribute n * score of sub-attribute n)*

Where *n* = total number of sub-attributes that are used to compute the score for the attribute.

Introducing Bias and Relevance for Overall Scores

Now that each category of productivity was representable via attribute scores, further work was done to account for these factors:

- It's possible that attributes don't have the same *relevance* across job profiles. For example, it may be the case that a sales professional doesn't really rely on a Code Quality category to understand his productivity (when it would be very relevant to a software engineer), but may find more relevance in a good Emails score or Meetings score (for instance, measuring number of emails resulting in confirmed sales).
- Even in relevant attributes, a particular individual may have a *bias* for specific attributes over others. Take, for example, a support engineer who considers having a good Service Desk score to be more important that Business Apps Usage.

For this reason, when looking at the *overall* score based on all attributes, we provide a formula that incorporates individual attributes' bias and relevance as follows:

Overall score = sum (attribute score * bias * relevance) / sum (relevance of each attribute)

This gives us an overall score (between 1 to 10) by which a subject can get a calculated assessment of overall productivity for a particular day.

Scoring Accuracy Evaluation Via "Declared" Scores Collection from Subjects

In order to verify that calculated scores (as per our devised scoring system) were accurate in measuring productivity, subjects in the associated group were asked to maintain a day-to-day diary of their activities, primarily "declaring" a score between 1 to 10 for each day. The purpose of this exercise was to provide a "reference" point against which to assess computed scores to determine degree of accuracy.

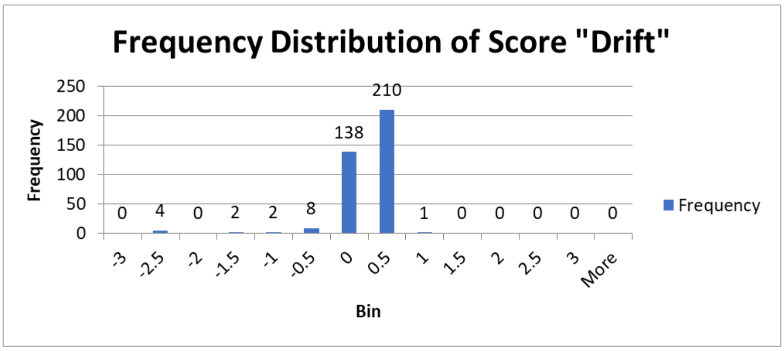
The following steps were then followed:

- The "difference" between declared and calculated scores was extracted.
- These were then plotted in a frequency distribution.

The frequency distribution was generated using a "bin size" of 0.5 points per bin. The idea was to extract the percentage of points that fell within 0.5-1 points of 0 (as an indicator of "calculated score is close *enough* to the declared score to be considered accurate") using the formula:

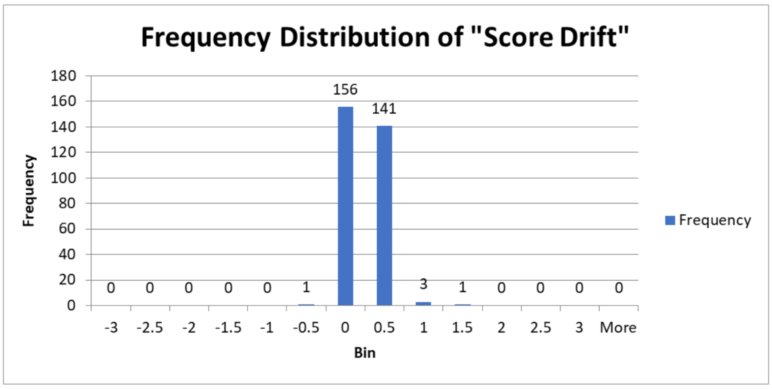
- ((frequency distribution in bins from +1 to –1) / (frequency distribution in all bins) * 100
- The value derived would be used to verify the hypothesis.

TABLES AND GRAPHS



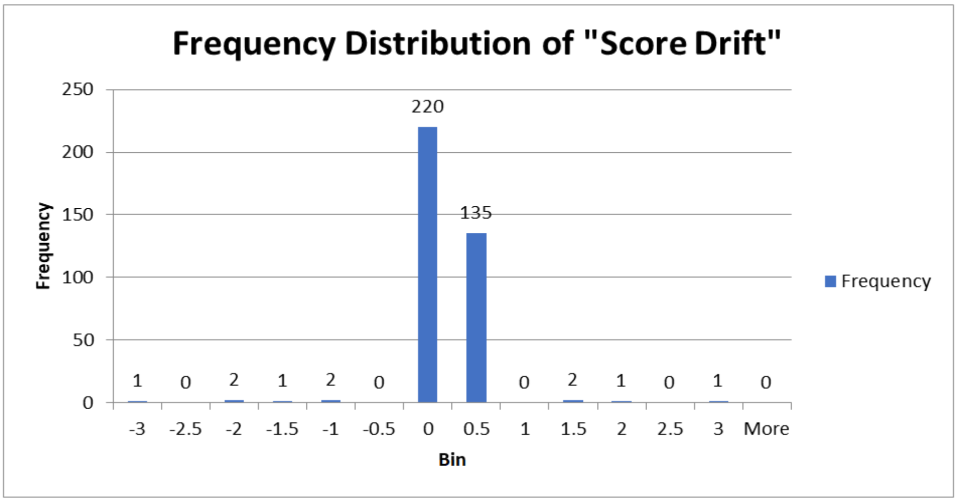
Graph 1.1: Frequency Distribution of "Drift" In Scores for Developer Scoring

Based on data gathered, the percentage of drifts within *1 point* of "declared" scores was found to be 98.352%.



Graph 1.2: Frequency Distribution of "Drift" In Scores for QA Scoring

Based on data gathered, the percentage of drifts within *0.5 points* of "declared" scores was found to be 99.003%.



Graph 1.3: Frequency Distribution of "Drift" In Scores for Support Scoring

Based on data gathered, the percentage of drifts within *0.5 points* of "declared" scores was found to be 97.26027%

RESULTS

Computed scores from formulae co-relate to "declared" score from surveyed subjects:

Based on the research, the authors found that "calculated" scores that processed raw data captured from subjects across 3 profiles were *within 0.5 to 1 points of what subjects "declared" their score to be*.

Based on this result, the authors believe that the null hypothesis has been disproven, and therefore the alternate hypothesis has been proven to be true.

DETAILED ANALYSIS

While the authors did observe outliers in the generated histograms, these could be adjusted for by factors like the subjective nature of the declared score mechanism (being base on the *informed* opinion of the subject supplying the score), minor measurement errors in collected data that goes towards the calculated scores, etc.

In general, it was observed that the primary goal of the research experiment was reached: the authors *were* able to evolve a scoring system, and *were* able to get comparable results to arbitrarily declared scores (it could be argued that their data-driven nature would make them even more accurate indicators of a possible score as compared to the subjective declared scores by participants in the study).

CONCLUSIONS

It is clear that productivity and wellness can be scored; these scores can be derived from data collected by individuals, which has been properly categorized, enriched, and organized in terms of relevance and importance to individuals.

It should be noted that utilizing a scoring model that relies on efficient utilization of time combined with focus on high reward patterns can result in increase in leisure periods; it is even possible to utilize quick break periods at specific times of the day after intense periods of activity (like *cyber loafing*) to serve as periods of leisure; it has been show that such patterns of "quick" bursts of inactivity can have a positive effect on employees' ability to pursue innovative strategies at their workplace (Zhong, 2022). This bodes well for organizations who understand the importance of elaborate employee health and wellness programs (Ramraj, Amolo, 2021); a scoring system (in the opinion of the authors) should be a critical part of any strategy utilized by corporate programs targeting employee health and wellness.

REFERENCES (LITERATURE CITED)

- Dai, G., Zeng, H., Luo, W., & Jiang, Y. (2021, January). The Relationship between Employees' Body Dissatisfaction and Job Burnout in the Workplace: A Management Research Proposal. In *International Conference on Decision Science & Management* (pp. 167-173). Springer, Singapore.
- Hobfoll, S. E. (1989). Conservation of resources: a new attempt at conceptualizing stress. *American psychologist*, *44*(3), 513.
- Hobfoll, S. E. (2001). The influence of culture, community, and the nested-self in the stress process: Advancing conservation of resources theory. *Applied psychology*, *50*(3), 337-421.
- Ma, Y., & Tschirhart, M. (2021). Enhancing self-efficacy during community service: Factors influencing AmeriCorps' members' change in self-

efficacy. *Nonprofit and Voluntary Sector Quarterly*, *50*(5), 1009-1028.

- Perrewé, P. L., Hochwarter, W. A., Rossi, A. M., Wallace, A., Maignan, I., Castro, S. L., ... & Van Deusen, C. A. (2002). Are work stress relationships universal? A nine-region examination of role stressors, general self-efficacy, and burnout. *Journal of International management*, *8*(2), 163-187.
- Ramraj, A. B., & Amolo, J. (2021). Appraising the future of employee health and wellness programmes in the Fourth Industrial Revolution: Wellness programmes. In *Future of work, work-family satisfaction, and employee well-being in the Fourth Industrial Revolution* (pp. 133-149). IGI Global.
- Salanova, M., Bakker, A. B., & Llorens, S. (2006). Flow at work: Evidence for an upward spiral of personal and organizational resources. *Journal of Happiness studies*, *7*(1), 1-22.
- Yu, L., Duffy, M. K., & Tepper, B. J. (2018). Consequences of downward envy: A model of self-esteem threat, abusive supervision, and supervisory leader self-improvement. *Academy of Management Journal*, *61*(6), 2296-2318.
- Zhang, C. (2020). The effects of upward social comparison on workplace envy: Moderate of gender identification and self-esteem. *American Journal of Industrial and Business Management*, *10*(3), 531-543.
- Zhong, J., Chen, Y., Yan, J., & Luo, J. (2022). The mixed blessing of cyberloafing on innovation performance during the COVID-19 pandemic. *Computers in Human Behavior*, *126*, 106982.