Decoding Work-Life Balance: Dimensionality Reduction and Predictive Classification

Lance Heinrich S. Lim[1], Lance Y. Natividad [2] and Shirlee R. Ocampo[3,*]
[123] De La Salle University
**shirlee.ocampo @dlsu.edu.ph*

Abstract: Work-life balance is important for a better quality of life. While its contributing factors are continuously being studied, influential lifestyle and well-being factors affecting it have not been concretely studied or identified. Using the Lifestyle and Wellbeing Data from Dalat (2021), which include work-life balance scores and its contributing well-being variables, this study explored the complexity of work-life balance by applying exploratory factor analysis to condense the numerous variables into a smaller, representative set of underlying latent factors. Then, the study utilized discriminant analysis to classify individuals by work-life balance scores in quintiles using the previously extracted factors, providing a structured approach to understanding and predicting work-life balance. This study identified five key latent factors that significantly influence work-life balance: *Purposeful Engagement*, *Social Support and Contribution*, *Achievement and Experience*, *Emotional Strain*, and *Physical Health*. Among these, Achievement and Experience emerged as the most influential factor in distinguishing between work-life balance groups, followed by Social Support and Contribution and Physical Health. Results further confirmed that the five extracted factors effectively differentiate individuals across the quintiles, providing a framework for understanding the relationship between lifestyle and well-being dimensions and work-life balance. Overall, the study suggests that improving work-life balance can be achieved through policies promoting physical health, emotional well-being, and professional growth. Employers can design wellness programs, flexible work arrangements, and employee recognition initiatives. Personalized interventions can enhance satisfaction and productivity. Policymakers can also consider these findings when developing labor regulations and workplace standards.

Key Words: work-life balance; lifestyle; well-being; exploratory factor analysis; discriminant analysis

1. INTRODUCTION

In the late 1970s, the United Kingdom coined the term *work-family balance* to describe the concept of striking a balance between a professional's work and family life. This balance is affected by factors including fatigue/ tiredness, frequent overtime work, proximity to the spouse, involvement in household affairs, impact of work demand, social participation, obscurity about rules, overload of work commitment, low compensation, work role overload, work/person conflict, and family/work conflict. Since then, numerous factors have been discovered to contribute to this balance. Consequently, the term *work-life balance* emerged to acknowledge other roles beyond family responsibilities, revolving around three key elements: *work* refers to any paid job or professional responsibility; *life* refers to personal pursuits, hobbies, and activities outside of work; while *balance* refers to distributing time and energy between the two in a way that minimizes conflict (Dewan & Mehendale, 2023). Thus, work-life balance is the concept of effectively managing the time and energy between professional and personal life to maintain satisfaction and well-being.

Intuitively, work-life balance is important for a better quality of life. As stated by Stoilova et al. (2020), work-life balance is perceived as an essential component and determinant of a person's job quality and broader well-being, respectively. Likewise, Vyas and Shrivastava (2017) stated that work-life balance or work-family conflict affects job stress and satisfaction in different aspects (i.e., family, life, and career). Notably, Dewan and Mehendale (2023) stated that professional and personal lives are also co-dependent, where an employee's job impacts their personal lives and vice versa. For example, a flexible work environment can boost morale, increase job satisfaction, and promote a healthier lifestyle while reducing stress. On the other hand, tension between professional and personal life caused by the rise of technology (e.g., project management platforms, communication tools, adaptive Wi-Fi) and the prominence of remote work due to COVID-19 can lead to physical and mental strain, leading to burnout and decreased productivity. Similarly, Håkansson et al. (2020) found that a combination of poor lifestyle and well-being generally reduced work ability, with self-rated health being the most significant effector. Thus, professional and personal factors are equally important in achieving work-life balance, leading to better well-being. Work-life balance and its contributing factors are continuously being studied. According to the study of Vyas and Shrivastava (2017), these factors could be related to an individual, family-related, work-related, or all. The same study provided an overview of these factors by reviewing existing literature, namely social support, organizational issues, stress issues, information technology, work issues, family issues, social issues, supportive factor, work overload, individual issues, and lack of knowledge. Building on that, the study of Stoilova et al. (2020) explored work-life balance and its aspects in the context of gender with more institutional or familial variables. Since then, Warren (2021) has argued for a broader and more inclusive approach to the gender-central work-life balance model, considering low wages and unstable hours as factors. Lastly and more recently, a study by Dewan and Mehendale (2023) found that four main factors (i.e., *Response to COVID-19*, *Support at Work Place*, *Satisfaction of Work-life balance*, *Interferences in work-life*) significantly contributed to work-life balance during the recent COVID-19 pandemic through the factor analysis of nineteen variables. Nonetheless, the investigation of work-life balance factors is traditionally focused on workplace and familial factors, with lifestyle and well-being factors remaining minimal in studies.

Despite the significance of achieving a work-life balance for a better quality of life, influential lifestyle and well-being factors affecting it have not been concretely studied or identified. This study aims to explore the complexity of work-life balance by applying Exploratory Factor Analysis (EFA) to condense numerous lifestyle and well-being variables into a smaller, representative set of underlying latent factors. Then, the study utilizes Discriminant Analysis (DA) to classify individuals by work-life balance levels using the extracted factors, providing a structured approach to understanding and predicting work-life balance. It seeks to provide insights into understanding and predicting work-life balance using the lifestyle and well-being variables.

This study is unique in its scope on individual lifestyle and well-being factors rather than traditional workplace-related variables in assessing work-life balance. By using EFA, the study identifies key well-being and lifestyle variables affecting work-life balance then performs DA to predict work-life balance levels. The findings provide personalized insights that can help individuals improve their well-being beyond workplace interventions.

Additionally, the results of this study support the United Nations Sustainable Development Goals 3 (Good Health and Well-being), 8 (Decent Work and Economic Growth), and 12 (Responsible Consumption and Production), which are focused on improving resource efficiency and promoting sustainable lifestyles.

2. METHODOLOGY

2.1 Data

Lifestyle and Wellbeing Data recorded from January 2015 to March 2021 has 15,972 observations, with 3,675 of them being recorded from 2020 onwards, was sourced from Kaggle, a website containing data sets, notebooks, and courses. The dataset includes survey results conducted by Authentic Happiness—an online platform run by the Positive Psychology Center at the University of Pennsylvania—to compute the "work-life balance score" of respondents by an undisclosed algorithm from their responses. Based on the responses to the 43 questions, the dataset has a total of 24 variables. TIMESTAMP and WORK_LIFE_BALANCE_SCORE were identified as interval, AGE as ordinal, and GENDER as nominal. The rest of the 20 variables are assumed to be ordinal based on the 11-point Likert scale of the original survey.

To ensure relevance and that the responses are apt for statistical analysis, data cleaning was first performed. Specifically, entries outside January 2020 to March 2021 were excluded to ensure timeliness. A total of 21 entries were identified as outliers and subsequently removed, leaving 3,654 valid observations for further analysis. Then, the WORK_LIFE_BALANCE_SCORE variable responses were reclassified from 0 to 4 to indicate the quintile where an individual belongs (with 0 being the first quintile and 4 being the last quintile) to facilitate subsequent DA after performing EFA. Finally, a total of 21 ordinal variables were included in the analysis, excluding TIMESTAMP, AGE, and GENDER as these are not included in the scope of lifestyle and well-being variables. Table 1 compiles the dataset's variables and their descriptions.

Table 1. Table of variables and descriptions

Variables	Descriptions
TIMESTAMP	Date when the survey was completed
BMI_RANGE	What is your body mass index (BMI) range?
DAILY_STEPS	How many steps (in thousands) do you typically walk every day?
FRUITS_VEGGIES	How many fruits or vegetables do you eat every day?
SLEEP_HOURS	About how long do you typically sleep?
DAILY_SHOUTING	How often do you shout or sulk at somebody?
DAILY_STRESS	How much stress do you typically experience every day?
SUFFICIENT_INCOME	How sufficient is your income to cover basic life expenses?
WEEKLY_MEDITATION	In a typical week, how many times do you have the opportunity to think about...
TODO_COMPLETED	How well do you complete your weekly to-do lists?

Table 1 continuation

Variables	Descriptions
FLOW	In a typical day, how many hours do you experience "flow"?
ACHIEVEMENT	How many remarkable achievements are you proud of?
PERSONAL_AWARDS	How many recognitions have you received in your life?
CORE_CIRCLE	How many people are very close to you?
LOST_VACATION	How many days of vacation do you typically lose every year?
PLACES_VISITED	How many new places do you visit?
SOCIAL_NETWORK	With how many people do you interact during a typical day?
TIME_FOR_PASSION	How many hours do you spend every day doing what you are passionate about?
SUPPORTING_OTHERS	How many people do you help achieve a better life?
DONATION	How many times do you donate your time or money to good causes?
LIVE_VISION	For how many years ahead is your life vision very clear for?
AGE	Age groups
GENDER	Male or female
WORK_LIFE_BALANCE_SCORE	Score calculated by Authentic Happiness

2.2 Statistical Analyses

All subsequent analyses were performed in SAS 9.4. EFA was first performed to reduce the number of ordinal variables that may be used to predict the WORK_LIFE_BALANCE_SCORE variable. Then, DA was conducted to identify which key life dimension contributes the most to a better work-life balance and to be able to predict a new individual's work-life balance. The level of significance used for all statistical tests is 5%.

3. RESULTS AND DISCUSSION

To explore the underlying structure of the 20 qualitative variables affecting WORK_LIFE_BALANCE_SCORE, an Exploratory Factor Analysis (EFA) using iterative principal factor extraction method with oblique equamax rotation was conducted.

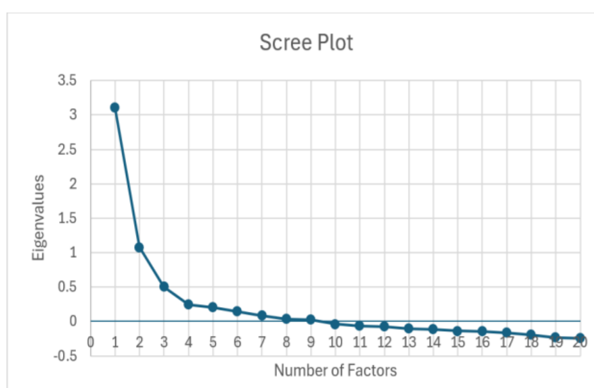

Fig. 1. Scree plot of the 20 qualitative predictor variables

A scree plot was generated to determine the number of factors to be retained. As seen in Figure 1, the scree plot indicated that retaining around five factors would be parsimonious and appropriate for meaningful interpretation. Thus, five factors were chosen as also supported by Seligman (2011), hypothesizing that these well-being variables could be grouped into five latent factors: *Physical Health* (e.g., BMI_RANGE, DAILY_STEPS), *Emotional Well-Being* (e.g., DAILY_STRESS), *Social Connections* (e.g., SOCIAL_NETWORK, CORE_CIRCLE), *Expertise* (e.g., ACHIEVEMENT, PERSONAL_AWARDS), and *Purpose* (e.g., SUPPORTING_OTHERS, TIME_FOR_PASSION).

Since the survey questionnaire consistently used an 11-rating scale, the raw Cronbach Coefficient Alpha was utilized to measure internal consistency. The measured raw Cronbach Coefficient Alpha was 0.7031, implying that the scale used by the survey questionnaire is reliable as the coefficient was greater than 0.7. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was also assessed, revealing that all individual variables and overall MSA values were above 0.5. These values confirmed that the dataset was suitable for factor analysis.

The initial factor analysis, conducted without rotation, resulted in significant cross-loadings, particularly for WEEKLY_MEDITATION, and disproportionately high loadings on the first factor. To improve factor interpretability, an oblique equamax rotation was applied. Following this adjustment, the factor pattern matrix revealed that certain variables—SLEEP_HOURS, TODO_COMPLETED, SUFFICIENT_INCOME, WEEKLY_MEDITATION, and LOST_VACATION—exhibited very low factor loadings (i.e., less than 0.3). Additionally, SLEEP_HOURS, SUFFICIENT_INCOME, and LOST_VACATION had extremely low raw item-total correlations and communalities. After excluding these three variables, a second factor analysis was performed. However, WEEKLY_MEDITATION continued to demonstrate low factor loadings and raw item-total correlations. Consequently, it was removed in a subsequent iteration. Finally, TODO_COMPLETED was also eliminated due to persistently low factor loadings across all factors.

Following these refinements, EFA yielded variance explained by each factor higher than 0.85 in Table 2. Satisfactory loadings across all retained variables are shown in Table 3.

The five extracted factors were then identified as follows:

- **Factor 1: Purposeful Engagement**
- **Factor 2: Social Support and Contribution**
- **Factor 3: Achievement and Experience**
- **Factor 4: Emotional Strain**
- **Factor 5: Physical Health**

These factors provide a meaningful representation of the latent constructs underlying the dataset. The results somewhat support the hypothesized groupings of variables of Seligman (2011) and suggest that well-being can be understood through these five distinct but interrelated dimensions. Furthermore, Factors 2 and 3 resemble named factors found by Dewan and Mehendale (2023) (i.e., *Support at Work Place* and *Satisfaction of Work-life balance*). Lastly, these factors also align with Vyas and Shrivastava (2017)'s statement that work-life balance factors could be related to an individual (i.e., Factors 1, 3, 4, 5), family-related (i.e., Factor 2), or work-related (i.e., Factor 2). Notably, Factor 2 relates strongly with a named factor of said study (i.e., *Social Support*).

The final factor pattern matrix and associated statistical outputs confirm the robustness of these factors in capturing the underlying structure of the data. Factor scores were then generated for subsequent analysis.

Table 2. Variance explained by each factor ignoring other factors

Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
1.87557	1.85976	1.75079	0.73046	0.84958

Table 3. Final Rotated Factor Pattern Matrix

Factor	1	2	3	4	5
BMI_RAN	0.064	0.146	-0.012	0.074	-0.315
DAILY_ST	0.082	0.112	-0.023	0.020	0.496
FRUITS_V	-0.047	0.129	0.175	-0.090	0.302
DAILY_SH	0.019	-0.042	0.060	0.474	0.003
DAILY_ST	-0.127	0.057	-0.024	0.559	-0.030
FLOW	0.637	0.092	0.014	-0.006	0.016
ACHIEVE	0.367	-0.031	0.475	0.125	0.043
PERSONAL	0.053	0.147	0.515	0.043	-0.044
CORE_CIR	0.105	0.322	0.148	-0.077	0.132
PLACES_VI	-0.071	0.029	0.383	-0.128	0.275
SOCIAL_N	0.198	0.430	-0.098	0.171	0.226
TIME_FOR	0.587	0.094	0.050	-0.130	-0.029
SUPPORTI	0.121	0.548	0.174	0.024	-0.034
DONATIO	-0.049	0.425	0.229	-0.059	-0.073
LIVE_VISIO	0.393	0.034	0.175	-0.060	-0.011

3.2 Discriminant Analysis

Following EFA, DA was conducted to determine the ability of the extracted factors to classify individuals into different quintiles of WORK_LIFE_BALANCE_SCORE. Before proceeding, assumptions for DA were first tested. Normality and homogeneity of covariance matrices were not satisfied. No evidence of multicollinearity was also detected among the factors, allowing for meaningful discriminant analysis. Hence, nonparametric discriminant analysis was employed using the *proc discrim* procedure with normal kernel using Silverman's Rule of Thumb and radius = 0.05. Silverman's Rule of Thumb is a widely used method for selecting the bandwidth in kernel density estimation (KDE), particularly for a Gaussian (Normal) kernel. The bandwidth controls the smoothness of the estimated density function—smaller values lead to more detail (potentially overfitting), while larger values produce a smoother estimate (potentially over-smoothing). Silverman's formula provides an optimal bandwidth under the assumption that the data is normally distributed, balancing bias and variance. It is computed as $1.06 * \sigma * n^{-1/5}$, where σ is the sample standard deviation and n is the sample size (Oser, n.d.). The results yield bandwidth estimates between 0.15 to 0.17, hence bandwidths around these estimates were tested. Lastly, cross-validation was implemented to assess model performance.

The multivariate test statistics indicated that the Wilks' lambda had a p-value of less than 0.0001, suggesting a statistically significant discriminant function for classifying WORK_LIFE_BALANCE_SCORE quintiles based on the five extracted factors. Furthermore, the univariate F-tests for each factor in Table 4 showed p-values of less than 0.0001, confirming that all five factors contributed significantly to discriminating between groups. Notably, Factor 3, interpreted as "Achievement and Experience," exhibited strong classification power, as evidenced by an Rsq/(1-Rsq) value exceeding 1. Notably, this factor is not emphasized by other studies exploring the predictors of work-life balance. Factors 2 and 5, interpreted as "Social Support and Contribution" and "Physical Health," respectively, were the next most significant contributors to group differentiation.

Table 4. Univariate test statistics

Variable	R-Square	R-Square/ (1-RSq)	F Value	Pr > F
Factor1	0.3546	0.5495	501.25	<.0001
Factor2	0.4531	0.8286	755.89	<.0001
Factor3	0.5784	1.3716	1251.28	<.0001
Factor4	0.1793	0.2185	199.31	<.0001
Factor5	0.3905	0.6406	584.43	<.0001

Cross-validation results further assessed the classification accuracy of DA. As seen in Tables 6 and 7, the cross-validation procedure yielded an error rate of 10.62%. Most misclassifications occurred within the 2nd to 4th quintiles, indicating that individuals with moderate WORK_LIFE_BALANCE_SCORE values were more challenging to classify accurately. This outcome aligns with our expectations, as the likelihood of underestimation or overestimation is higher for those in the middle quintiles compared to those in the extreme ends.

Overall, the discriminant analysis confirmed that the five extracted factors effectively differentiate individuals across the WORK_LIFE_BALANCE_SCORE quintiles, providing a robust framework for understanding the relationship between well-being dimensions and work-life balance.

Table 5. Cross-validation error count estimates

	0	1	2	3	4	Total
Rate (%)	6.72	14.7	14.0	11.9	5.73	10.6

Table 6. Cross-validation classification results

From quintile	0	1	2	3	4	Total
0	680	44	5	0	0	729
1	35	626	57	16	0	734
2	4	40	627	49	9	729
3	0	13	40	642	34	729
4	0	0	6	36	691	733
Total	719	723	735	743	734	3654
Rate (%)	19.7	19.8	20.1	20.3	20.1	100

4. CONCLUSION

This study identified five key latent factors that significantly influence work-life balance: *Purposeful Engagement*, *Social Support and Contribution*, *Achievement and Experience*, *Emotional Strain*, and *Physical Health*. Among these, Achievement and Experience emerged as the most influential factor in distinguishing between work-life balance groups. These factors provide valuable insights into how individuals' well-being and lifestyle choices affect their perceived work-life balance. The results underscore the importance of addressing multiple dimensions of well-being to foster a healthier and more productive workforce.

The findings hold practical implications for various stakeholders, particularly working professionals seeking to improve their work-life balance. Employers can use these insights to design policies that promote physical health, emotional well-being, and professional growth, such as wellness programs, flexible work arrangements, and employee recognition initiatives. Furthermore, personalized interventions, such as mentoring programs, mental health support, and physical activity incentives, could enhance overall satisfaction and productivity. Individuals can leverage these factors to assess and adjust their daily habits, focusing on aspects like stress management, social connections, and purposeful engagement. Lastly, policymakers may also consider these findings when developing labor regulations and workplace standards aimed at improving the well-being of the workforce.

For future research, further investigation is recommended to explore additional variables, along with lifestyle and well-being variables discussed in this paper, that may influence work-life balance, such as industry-specific factors, cultural differences, or the impact of remote work. Additionally, alternative analytical techniques could be employed to validate the findings and enhance predictive accuracy. Lastly, it may be worthwhile to investigate quantifying work-life balance scores through other means.

6. REFERENCES

Dalat, Y. (2021, March 14). Lifestyle_and_Wellbeing_Data. Kaggle. https://www.kaggle.com/datasets/ydalat/lifestyle-and-wellbeing-data/data

Dewan, A., & Mehendale, S. (2023). The Factors Impacting Work-Life Balance amidst COVID-19. Cardiometry, 25, 594–602. https://www.researchgate.net/publication/370398632_The_Factors_Impacting_Work-Life_Balance_Amidst_Covid-19

Håkansson, C., Gard, G., & Lindegård, A. (2020). Perceived work stress, overcommitment, balance in everyday life, individual factors, self-rated health and work ability among women and men in the public sector in Sweden – a longitudinal study. Archives of Public Health, 78(1). https://doi.org/10.1186/s13690-020-00512-0

Oser, S. (n.d.). Kernel density estimation & bandwidth selection [Lecture notes]. University of British Columbia. https://phas.ubc.ca/~oser/p509/Lec_23.pdf

Schreiber, J. B. (2020). Issues and recommendations for exploratory factor analysis and principal component analysis. Research in Social and Administrative Pharmacy, 17(5), 1004–1011. https://doi.org/10.1016/j.sapharm.2020.07.027

Seligman, M. E. P. (2011). Flourish: A Visionary New Understanding of Happiness and Well-being. Simon and Schuster.

Stoilova, R., Ilieva-Trichkova, P., & Bieri, F. (2020). Work–life balance in Europe: institutional contexts and individual factors. International Journal of Sociology and Social Policy, 40(3/4), 366–381. https://doi.org/10.1108/ijssp-08-2019-0152

Vyas, A., & Shrivastava, D. (2017). Factors affecting work life balance-a review. Pacific Business Review International, 9(7).

Warren, T. (2021). Work–life balance and gig work: 'Where are we now' and 'where to next' with the work–life balance agenda? Journal of Industrial Relations, 63(4), 522–545. https://doi.org/10.1177/0022185621100716

Woods, C. M., & Edwards, M. C. (2007). 12 factor analysis and related methods. In Handbook of statistics (pp. 367–394). https://doi.org/10.1016/s0169-7161(07)27012-9