Do Common Performance Metrics in Indoor Environmental Quality Research Represent Actual Office Work? A Survey-Based Assessment

Wooyoung Jung[1*], *Prosper* Babon-Ayeng[1], *Kevin* Keene[2], *Sophia* Witt[1] and *Jessica* Baweja[3]

[1]Civil and Architectural Engineering and Mechanics, The University of Arizona, 85719 Tucson, Arizona, USA
[2]Building and Connected Systems, Pacific Northwest National Laboratory, 99352 Richland, Washington, USA
[3]AI and Data Analytics, Pacific Northwest National Laboratory, 99352 Richland, Washington, USA

Abstract. Existing studies in the field of indoor environmental quality (IEQ) have relied on a limited set of objective performance metrics, including call handling time, keystrokes, and cognitive performance tests, to represent office worker performance. However, these metrics may capture narrow and task-specific aspects of performance and may not represent the broader, multi-faceted nature of office work. This study assessed the applicability of such objective metrics by surveying office workers on whether and to what extend these metrics are relevant to their work and how much of their office work can be quantitatively measured. A total of 387 survey responses from a variety of occupations were analyzed and we found significant variability in the relevance of some performance metrics across occupations (e.g., call handling time, attention and concentration, executive functioning, language skills, and processing speed) through analysis of variance (ANOVA). The results indicate the heterogeneity of office work and the difficulty in developing a universal office worker performance prediction model applicable to all types of office workers. This study sheds light on the needs of adaptability in office worker performance data analyses and interpretations and paves the way for customized performance prediction models with respect to IEQ, specific to occupation types or individuals.

Keywords. Office Worker Performance, Indoor Environmental Quality, Performance Evaluation Metrics, Applicability Assessment

1 Introduction

The idea that indoor environmental quality (IEQ) influences office workers' performance has been around for more than a century, with numerous studies subsequently investigating its diverse effects on different dimensions of performance. To date, studies have utilized various metrics to assess performance, including call handling time among center representatives [1], [2], [3], [4], [5], keystroke activity [6], [7] and results from cognitive performance tests [8], [9], [10], which are considered objective indicators. These metrics are typically quantitative and relatively viable to capture, making them practical proxies for assessing the effects of IEQ on work-related outcomes. However, these metrics are limited in their scopes, as they only capture narrow aspects of performance – often task-specific measures or isolated components of cognitive function – and may not reflect the broader, multifaceted nature of human performance, including factors such as creativity, complex problem-solving, communication, and emotional resilience.

This limited spectrum of objective office worker performance metrics presents a challenge of the IEQ field, as the nature of office work spans from repetitive, task-based activities to complex, knowledge-oriented tasks [11], and organizations and teams often prioritize different competencies among office workers [12]. In recent years, self-assessed performance metrics have been increasingly adopted, as demonstrated in [13], owing to the belief that workers may have a more accurate understanding of their own performance when assessed subjectively. This shift highlights the inherent difficulty in objectively quantifying office workers' performance, especially in knowledge-intensive roles where output is less tangible and harder to standardize than in procedural, methodical, and routine-based tasks. However, subjective measures come with notable limitations. They are susceptible to individual biases and social desirability effects [14] – where individuals tend to respond to self-assessment in a manner that they believe will be viewed favorably by others – which can compromise the accuracy and consistency of the data. Moreover, self-assessments may lack standardization across respondents, making it difficult to compare results objectively or replicate findings across different studies and contexts.

The goal of this study is to evaluate existing objective performance metrics for office workers that are commonly

* Corresponding author: wooyoung@arizona.edu

used in the field of IEQ research, with a focus on assessing their suitability and relevance for quantifying office workers' performance in varying IEQ conditions. We specifically seek to answer two research questions (RQs).

RQ 1. To what extent do actual office workers believe that their work performance can be objectively quantified?

RQ 2. To what extent can the objective performance metrics used in IEQ field actually represent actual office workers' performance?

We developed a survey to address these RQs in a systematic manner. Then, this survey was distributed after institutional review board (IRB) approval from the University of Arizona (STUDY00003041). This research shares the results of this survey and discusses their implications for the field of IEQ-based office work performance assessment. We firmly believe that this effort offers a novel approach to understanding office worker performance in relation to IEQ conditions.

This study did not consider any subjective metrics such as self-reported work performance. Also, for the purpose of this research, the term 'objective' for metrics was used in contrast to 'subjective' to describe metrics that were externally observable and quantifiable.

This paper is structured as follows. In Section 2, we elaborate on the survey development and distribution. In Section 3, the survey results and their meaning are presented. The fourth section discusses the key takeaways and limitations of this study and shares the remaining work in this IEQ field. The last section is the conclusion of this study, including future research direction.

2 Methodology

2.1 Survey Development and Distribution

2.1.1 Synthesis of objective office work performance metrics

To identify existing office worker performance metrics in the literature, an extensive search was conducted. Articles were retrieved using the following keywords in Google Scholar and Web of Science: office, office environment, office worker, office work, productivity, performance, indoor environmental quality, indoor air quality, air temperature, humidity, CO2, cognitive performance, buildings, and built environments. Each article's abstract was then reviewed to determine whether it reported any office worker performance data. Also, the authors' Google Scholar profiles were thoroughly examined to ensure a comprehensive inclusion of relevant publications. The following criteria were established to guide this process.

- The term 'productivity' was often used in field studies, and 'performance' was frequently used for cognitive performance tests. Therefore, both terms were our interest.
- When there was no clear IEQ modification, such studies were excluded (e.g., [15]).

Table 1. Cognitive performance tests used in existing studies of interest

Name	Description
Typewriting	Participants are asked to retype the texts presented in a display
Numerical calculation	Participants are asked to calculate the numbers from one to three digits (e.g., addition)
Memory test	After presenting a set of words, participants should sort them out from a large group of words [16].
Cue-utilization test or event sequencing	Subjects are asked to link circles containing letters or numbers, placed at random, in order [17].
Reaction/ response	A small dot at the center of the oscilloscope moves up or down. Participants are asked to detect the occurrence and discriminate between long and short signals.
Search	Detect the presence of letters, symbols, etc., as quickly as possible.
Overlapping	Geographic shapes (e.g., triangular) are randomly fallen onto one on top of the other. Participants number the figures according to their position in a pile.
Sentence comprehension	Participants are asked to mark a set of sentences as true of false through grammatical reasoning
Proofreading	Subjects are asked to identify misspelled words in sentences.
Reasoning	Participants are asked to rearrange numbers, alphabets, etc., presented at a random order, in order.
Sudoku	A puzzle that requires numerical reasoning to fill out the numbers in a 9×9 grid
Hand-eye coordination	A tracking test that participants should follow and copy the motion of a dot or wave in a display.
Stroop test	Participants are asked to report the color of a target word [18].
Finger tapping	Tapping a key as quickly as possible for a certain time.

We identified two types of objective office worker performance metrics: direct (objective metrics that measure work output directly) and indirect (those that relied on proxy indicators like cognitive performance tests to infer performance levels). Most direct performance measurements were from studies focusing on call handling times by call center representatives [1], [2], [3], [4], [5] with a few additional examples based on keystrokes from programmers [6], [7]. Indirect performance metrics – namely, cognitive performance tests – were employed in various studies to assess aspects of mental functioning, as synthesized in Table 1. These tests were designed to capture a range of cognitive abilities through outputs such as response time (time to complete fast-paced cognitive exercise), speed, and accuracy (or error rates) [15].

2.1.2 Survey development and distribution

With the metrics synthesized through literature review, a survey was developed. First, we asked respondents to identify their occupation using categories defined by the U.S. Bureau of Labor Statistics [19]. Participants were provided a link to review definitions for each occupation. Next, we inquired about the quantitative nature of their work and its perceived relevance to the synthesized performance metrics. Asking participants to evaluate the relevance of every individual test' result metric was deemed impractical, as it would have overwhelmed them with excessive information. Instead, broader domains of cognitive function (Table 2) – representative of the synthesized cognitive performance metrics – were used as proxies and they reflect a continuum of cognitive abilities ranging from basic sensory and perceptual processes to more complex executive functions that regulate, coordinate, and integrate these foundational processes.

Table 2. Domains of cognitive ability [20]

Cognitive ability	Description
Attention and concentration	Humans are capable of processing information based on relevance and importance (attention) and have the ability to sustain attention over time (concentration)
Memory	Humans are capable of maintaining and manipulating information in their consciousness. This cognitive function contains encoding (processing information for long-term storage), storing (retention of information), and retrieving.
Language skills	Language skills include receptive and productive abilities related to language, such as assessing semantic memory and identifying the names of objects
Sensation and perception	Sensation refers to the ability to detect a stimulus through five sensory modalities, and perception is the function of identifying a meaningful stimulus
Executive functioning	This function manifests control over other cognitive abilities to perform complex tasks (e.g., problem-solving)
Processing speed	This domain refers to the speed required to perform a task
Motor skills	Motor skills refer to basic elements of motor activity such as manual dexterity and motor speed

We implemented the questionnaires with Qualtrics' platform and distributed the survey in Oct, 2023 utilizing the LinkedIn post and emails to the authors' professional network. This participant recruitment process is intended for convenience sampling to maximize the sampling size in our data analysis. To confirm the applicability of the potential survey participants, we clearly stated that you are welcome to take the provided survey if you work in an office environment.

2.1.3 Survey Response

The online survey was open for seven days, and a total of 434 responses were received initially. After the exclusion process (e.g., dropping incomplete responses), a total of 387 responses remained to be analyzed in this study. Figure 1 presents the occupations of the respondents surveyed in this study. A variety of occupations were included in this survey, reflecting the diversity of contemporary office environments. These ranged from administrative and clerical roles to managerial, technical, and creative positions. This diversity allowed for a more holistic understanding of how different types of work may influence or interact with performance metrics in the context of varying indoor environmental conditions.

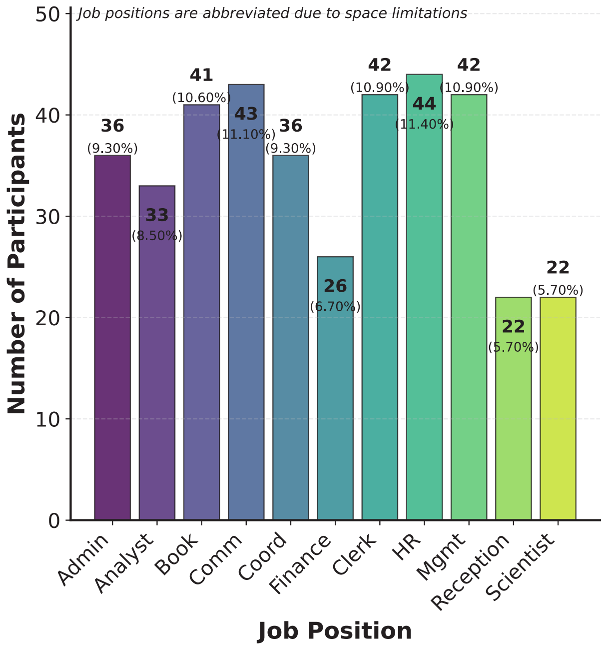

Figure 1. Distribution of Occupations among Survey Participants

2.2 Survey Analysis Method

This subsection elucidates the methodology employed in analyzing the survey data, highlighting the inclusion criteria for responses and the statistical techniques applied. Firstly, responses that were incomplete were excluded from the analysis. Also, to ensure the statistical validity of the results, a minimum response count of 20 for each occupation was set as a threshold for inclusion in the results. We decided this specific threshold given (1) the descriptive and inferential statistics and (2) the inclusion of diverse occupations in our analyses. This approach aligns with recommendations by [21], who emphasizes the importance of adequate sample size in providing reliable statistical results and reducing the likelihood of errors.

The foundation of the analysis was laid with descriptive statistics, which provided a preliminary overview of the survey results. Key statistical measures including means and standard deviations were computed, offering an initial understanding of the data distribution and central tendencies. To further explore the representativeness and similarity of office work across various occupations, a one-way Analysis of Variance (ANOVA) [22] was conducted with occupation as the independent variable, at a significance level of $\alpha = 0.05$. ANOVA is generally robust to moderate violations of normality and homogeneity of variance when group sample sizes are sufficiently large, which is the case in this study with a total of 387 responses. The primary purpose of the ANOVA was to decide whether occupation-level differences exist in the perceived relevance of each performance metric, rather than to identify specific pairwise differences between occupations. As such, the interpretation focuses on the overall F-statistics and their associated p-values.

3 Results

3.1 Point of View Toward Performance Quantification of Office Work

Overall, respondents said that approximately 39.77% of their office work could be quantitatively measured, with a relatively large standard deviation of 24.10%, highlighting the diversity of perspectives on this topic (Figure 2). Reception and Finance professionals reported higher mean scores of 48.46%. In contrast, Administration and Human Resources positions reported the lowest mean values of 33.33% and 33.86%, respectively. This variation could be attributed to the differing nature of tasks across roles, where some involve more structured, outcome-driven activities, while others rely on interpersonal or process-oriented responsibilities that are harder to quantify. Another key takeaway is that the standard deviation within each occupation ranged from 18.21% to 29.98%, demonstrating a significant variation in perceptions of quantifiability even among individuals in the same field.

We probed into the reasons why certain tasks of office work are perceived as difficult to quantify (Figure 3). The challenge most frequently mentioned by respondents was the subjective nature of many office tasks. The second most cited reason was the non-linear progression of work. The complexity of tasks and their resistance to numerical measurement was also highlighted as a concern. The survey included an open-ended question seeking to know the reasons why office work cannot be quantitatively measured. Only one response to the open-ended question was received, which stated that, “Something that takes 3 hours for me one day can take 1 hour for me. For some the task might take five hours and for some it could take 10 minutes. The only metric I can think of is time devoted, and this is not a great metric.” The respondent’s observation that the time it takes to complete the same task can vary significantly, even for the same individual, is a crucial insight.

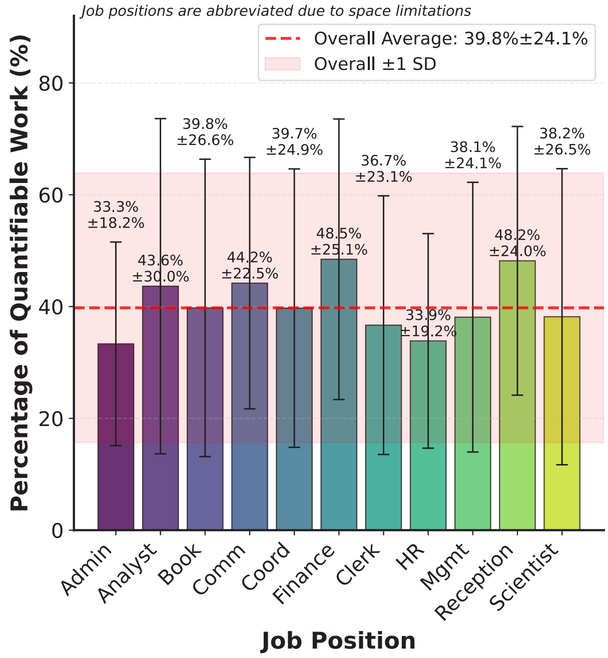

Figure 2. Quantifiable work portions for each office occupation

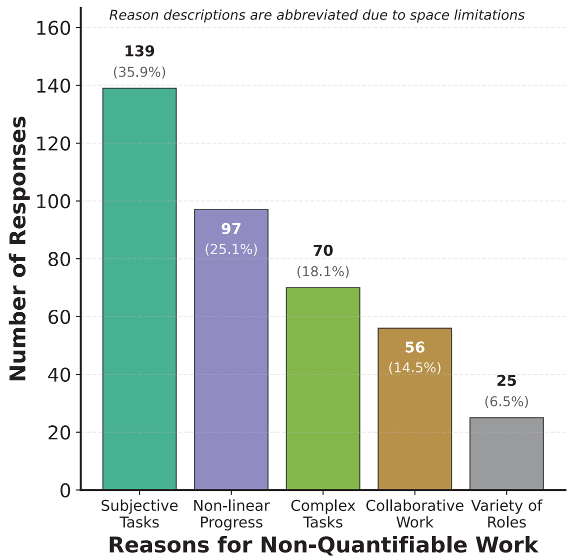

Figure 3. Responses to the question why your office work cannot be quantitatively measurable.

3.2 Representativeness of Existing Office Worker Performance Metrics

With regard to direct office worker performance metrics, the average rating for similarity to keystrokes on a scale of 0 to 10 was 6.90, with a standard deviation of 1.96. This suggests that participants perceived their office tasks as somewhat similar to the keystroke metric, reflecting the importance of computer and typing tasks in their roles. Call handling times received a slightly lower mean rating of 6.50, with a standard deviation of 2.20. The higher standard deviation in this case indicates greater variability in participants' perceptions, lightly due to differences in task types and communication demands depending on individual job functions.

As in Table 3, a one-way ANOVA with occupation as the independent variable showed an F = 0.97, p = 0.47, indicating that there were no statistically significant differences between the occupations in their perceived similarity to programmers' keystrokes. For the metric of call handling time, the F-value was 3.81 with a p-value < 0.001. This suggests that the extent to which respondents' work relates to call handling time significantly varies among different roles.

Table 3. ANOVA results for office worker performance metrics by occupation

#	Index	Degree of Freedom	F score	P value
1	Call handling times	10	3.81	$0.70\times e^{-4}$
2	Keystrokes	10	0.97	0.47

Light yellow background: Statistically significantly different

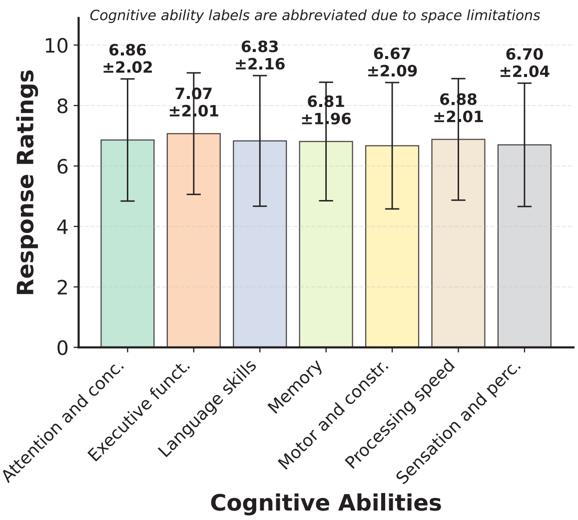

Figure 4. Importance ratings of cognitive abilities to office workers

Regarding the perceived importance of various cognitive abilities in office work, the results are presented in Figure 4. The descriptive statistics show that Executive Functioning was considered the most important cognitive ability. The other abilities (Processing speed, Attention and concentration, and Language skills, Memory, Sensation and perception, and Motor skills and construction) all scored between 6.67 and 6.88, suggesting a somewhat significant perceived relevance in office settings.

Figure 5 provides a breakdown of the mean values of cognitive abilities by occupation, revealing notable variations across different roles. There were positions with very big circles across the board (management and finance) and small circles (bookkeeping, coordinator, and communication). In contrast, there were the ones that have uneven circles: attention stands out for analyst, executive function for coordinator, general clerk, human resources, language is low for reception, and language, attention, executive function, and memory are all high for scientists, but the rest are low.

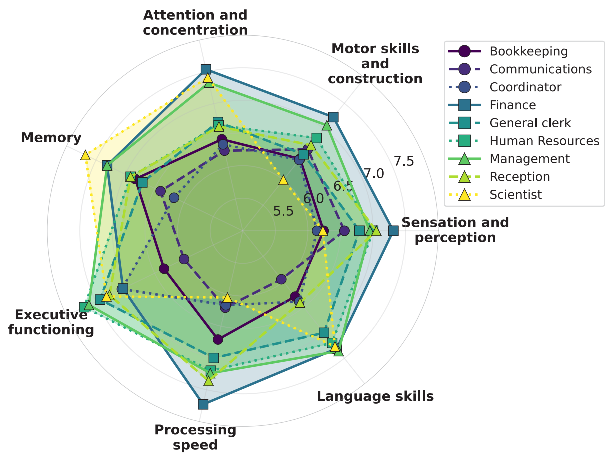

Figure 5. Mean values of importance of cognitive abilities by occupation

Table 4. ANOVA of cognitive abilities by occupation

#	Cognitive Abilities	Degree of Freedom	F score	P value
1	Sensation and perception	10	1.14	0.33
2	Motor skills and construction	10	0.78	0.64
3	Attention and concentration	10	2.03	$2.92\times e^{-2}$
4	Memory	10	1.64	0.09
5	Executive functioning	10	2.85	$1.95\times e^{-3}$
6	Processing speed	10	2.34	$1.10\times e^{-2}$
7	Language skills	10	1.95	$3.82\times e^{-2}$

Light yellow background: Statistically significantly different

As in Table 4, an ANOVA, evaluating the significance of various cognitive abilities across different office occupations, demonstrated no statistically significant differences among occupations for cognitive abilities such as sensation and perception, motor skills and construction, and memory. Conversely, abilities such as attention and concentration, executive functioning, processing speed, and language skills had significant variations across distinct occupations.

4 Discussion and Limitation

In this section, we would like to bring up several discussion points in the IEQ field and limitations of this study. The first point is the question of how objective performance metrics should be used in the prediction models. On average, respondents felt that only 40% of their work could be quantitatively measured, and so this study raises the question of the applicability of existing prediction models that were developed using objective performance metrics. Second, we considered 'occupation' as the decisive factor in interpreting the applicability and fitness of existing performance metrics. However, as presented in Figure 2, individuals in the same occupation shared a variety of ideas of how such metrics play a role in their work. In other words, the ultimate decisive factor(s) possibly lies in individuals such as their work types, expertise, mind, rather than occupations. Third, the use of self-assessed performance metrics remains as a potential direction but there is little research investigating the alignment between actual office worker performance and self-assessed office worker performance. The last point is that IEQ is one of the factors in office environments that impacts office worker performance. Explaining the changes of office worker performance with only IEQ changes might demonstrate a less definitive relationship (e.g., a low R-squared value in the regression model) given its absence of other critical factors when the performance data were collected. Hence, a high performance office worker performance prediction model would require high quality data that cover many aspects of office environments.

Several limitations of this study should be acknowledged. First, the convenience sampling approach may introduce sampling bias. Second, the survey responses are themselves subject to the same self-report biases discussed in the Introduction, including social desirability effects. Respondents may have over- or under-estimated the quantifiable nature of their work based on how they wished to present their roles. Lastly, the domains of cognitive ability used in this survey (Table 2) were synthesized from existing IEQ literature and may not encompass all cognitive dimensions relevant to office work. Similarly, non-IEQ factors that influence office worker performance (e.g., organizational culture, management practices, and individual motivation) were beyond the scope of this study.

5 Conclusion

This study contributes largely to the IEQ field of measuring the variation of office worker performance by revisiting existing metrics and investigating their applicability. This research was initiated by asking the level of quantifiable nature of office worker performance through existing objective metrics, and a survey was conducted to address them. The critical takeaways of this study are following.

- Survey respondents reported that a somewhat small portion of their work can be quantified objectively, about 40% on average with a large standard deviation of 24%. Many participants responded that their work is subjective and non-linear in nature.
- The metrics (call handling time, attention and concentration, executive functioning, processing speed, and language skills) were found to have significant variation between occupations, suggesting the importance of including occupation as a factor in a work performance regression model.

As a follow-up study, we will investigate the possibility of devising office worker performance prediction models that reflect the user perspectives on, for example, which performance metrics should be weighted more. In addition, as discussed in Section 4, the variability within the same occupation suggests that individual-level factors – such as specific work types, expertise, and task complexity – may be more decisive than occupation alone. Future research should explore these individual-level variables as predictors and take specific IEQ parameters, such as lighting, noise, ventilation, and thermal comfort, into consideration in performance models to better capture the diversity of office work. Furthermore, longitudinal field studies that track performance metrics alongside IEQ changes over time in real office settings would help establish stronger causal relationships than cross-sectional survey data alone. Lastly, given that many existing metrics were found to be limited in representing the breadth of office work, future efforts should explore the development of new performance metrics – such as task-completion quality, collaborative output, or project milestones – that may better capture the multifaceted nature of knowledge work in relation to IEQ conditions.

The authors would like to appreciate the survey participants and the librarians at Virginia Polytechnic Institute and State University who helped find articles of our interest.

References

1. R. Niemelä, M. Hannula, S. Rautio, K. Reijula, and J. Railio. (2002). The effect of air temperature on labour productivity in call centres—a case study. Energy and Buildings. **34**(8): p. 759–764. https://doi.org/10.1016/S0378-7788(02)00094-4
2. R. Niemela, J. Railio, M. Hannula, S. Rautio, and K. Reijula. Assessing the effect of indoor environment on productivity. in Clima 2000. (2001). Napoli.
3. L. Heschong, Windows and Offices: A Study of Office Worker Performance and the Indoor Environment. (2003), California Energy Commission: Fair Oaks, CA.
4. K.W. Tham. (2004). Effects of temperature and outdoor air supply rate on the performance of call center operators in the tropics. Indoor Air. **14**(s7): p. 119–125. https://doi.org/10.1111/j.1600-0668.2004.00280.x
5. C.C. Federspiel, W.J. Fisk, P.N. Price, G. Liu, D. Faulkner, D.L. Dibartolomeo, D.P. Sullivan, and M. Lahiff. (2004). Worker performance and ventilation in a call center: analyses of work performance data for registered nurses. Indoor Air. **14**(s8): p. 41–50. https://doi.org/10.1111/j.1600-0668.2004.00299.x
6. A. Hedge, W. Sakr, and A. Agarwal. Thermal effects on office productivity. (2005). SAGE Publications Sage CA: Los Angeles, CA.
7. N. Nishihara, S.-i. Tanabe, M. Haneda, M. Ueki, A. Kawamura, and K. Obata. (2007). Effect of overcooling on productivity evaluated by the long term field study.
8. L. Lan and Z. Lian. (2009). Use of neurobehavioral tests to evaluate the effects of indoor environment quality on productivity. Building and Environment. **44**(11): p. 2208–2217. https://doi.org/10.1016/j.buildenv.2009.02.001
9. L. Lan, Z. Lian, L. Pan, and Q. Ye. (2009). Neurobehavioral approach for evaluation of office workers' productivity: The effects of room temperature. Building and Environment. **44**(8): p. 1578–1588.
10. S. Snow, A.S. Boyson, K.H.W. Paas, H. Gough, M.-F. King, J. Barlow, C.J. Noakes, and m.c. schraefel. (2019). Exploring the physiological, neurophysiological and cognitive performance effects of elevated carbon dioxide concentrations indoors. Building and Environment. **156**: p. 243–252. https://doi.org/10.1016/j.buildenv.2019.04.010
11. B.P. Haynes. (2007). Office productivity: a theoretical framework. Journal of Corporate Real Estate.
12. M. Haneda, S.-i. Tanabe, N. Nishihara, M. Ueki, and A. Kawamura. (2007). Development of Survey Tools for Indoor Environmental Quality and Productivity. Proceedings of Clima 2007. **1**.
13. D.R. Ilgen and J. Schneider. (1991). Performance measurement: a multi-discipline view. International Review of Industrial and Organizational Psychology. **6**: p. 71–108.
14. D.L. Paulhus. (1984). Two-component models of socially desirable responding. Journal of personality and social psychology. **46**(3): p. 598.
15. D.P. Wyon, Individual control at each workplace: the means and the potential benefits, in *Creating the productive workplace*. (2000): London. doi:
16. Y. Geng, W. Ji, B. Lin, and Y. Zhu. (2017). The impact of thermal environment on occupant IEQ perception and productivity. Building and Environment. **121**: p. 158–167. https://doi.org/10.1016/j.buildenv.2017.05.022
17. D. Wyon, I. Andersen, and G. Lundqvist. (1980). The effects of moderate heat stress on mental performance. Scandinavian journal of work, environment & health. **5**: p. 352–361. https://doi.org/10.5271/sjweh.2646
18. J.R. Stroop. (1935). Studies of interference in serial verbal reactions. Journal of Experimental Psychology. **18**(6): p. 643–662. https://doi.org/10.1037/h0054651
19. U.S.B.o.L. Statistics. Occupational Employment and Wage Statistics. 2022 [cited 2024 03/12]; Available: https://www.bls.gov/oes/current/oes_stru.htm
20. P.D. Harvey. (2019). Domains of cognition and their assessment Dialogues in clinical neuroscience. **21**(3): p. 227–237. https://doi.org/10.31887/DCNS.2019.21.3/pharvey
21. A. Field, Discovering statistics using IBM SPSS statistics. (2013): sage.
22. S. Abbasi. (2022). The degradation rate study of methyl orange using MWCNTs@ TiO2 as photocatalyst, application of statistical analysis based on Fisher's F distribution. Journal of cluster science. **33**(2): p. 593–602.