A Statistical Evaluation of Combining Human Productivity Metrics in the Indoor Environment

Keene, Kevin[1]
Pacific Northwest National Laboratory
PO Box 999 Richland, Washington 99352
kevin.keene@pnnl.gov

Jung, Wooyoung
Pacific Northwest National Laboratory
PO Box 999 Richland, Washington 99352
wooyoung.jung@pnnl.gov

ABSTRACT

The potential of improving human productivity by providing healthy indoor environments has been a consistent interest in the building field for decades. This research field's long-standing challenge is to measure human productivity given the complex nature of office work. Previous studies have diversified productivity metrics, allowing greater flexibility in collecting human data; however, this diversity complicates the ability to combine productivity metrics from disparate studies within a meta-analysis. This study aims to categorize existing productivity metrics and statistically assess which categories show similar behavior when used to measure the impacts of indoor environmental quality. The 106 productivity metrics compiled were grouped into six categories: neurobehavioral speed, accuracy, neurobehavioral response time, call handling time, self-reported productivity, and performance score. Then, this study set neurobehavioral speed as the baseline category given its fitness to the efficiency-based

[1] Corresponding author

definition of productivity (i.e., output versus input) and conducted three statistical analyses with the other categories to evaluate their similarity. The results showed the categories of neurobehavioral response time, self-reported productivity, and call handling time had statistical similarity with neurobehavioral speed. This study contributes to creating a constructive research environment for future meta-analyses to understand which human productivity metrics can be combined with each other.

Keywords: Indoor Environmental Quality, Productivity, Office Buildings

INTRODUCTION

Healthy building certification systems have quickly grown their project portfolios [1]. This movement demonstrates the growing interest in creating healthy indoor environments in buildings, considering the value of occupants' health and productivity. From the employer's perspective, if their employees could become more productive, it would help justify the upfront refurbishment costs or additional operational costs. The cost of employees to an employer accounts for 92% of building costs, while utility bills account for only 6% [2]. This scale gap suggests a large opportunity for financial benefits by providing indoor environmental conditions that contribute to occupant productivity. Specifically, potential office worker productivity gains from improved indoor environments have been estimated at up to $230 billion (in 2016 U.S. dollars) at the national level [3].

The human productivity impacts due to variations in indoor environments in office buildings has been a consistent research interest for several decades (e.g., [4]). Building systems are the main vehicle for facility managers to control the indoor environmental quality (IEQ) of their building and therefore physical comfort and well-being of the occupants.

Therefore, gaining the knowledge of how IEQ contributes to human productivity is important to determine which IEQ parameters to target and to what extent. For example, Gupta et al. [5] demonstrated that task performance (numerical calculation and proofreading) was impacted by air temperature and carbon dioxide (CO_2) concentration, and the occupants in naturally ventilated offices had higher tolerance to IEQ variations. Kawamura et al. [6] asked human subjects to perform three-digit multiplication tasks under different IEQ conditions and showed that subjects prioritized the thermal and acoustic environments and perceived that they performed better when satisfied with IEQ. Niemelä et al. [7] used a computerized monitoring system to compute call center workers' productivity (the number of telephone communications divided by the active work time) under different IEQ conditions and found that the average productivity decreased when the air temperature exceeded 25°C.

Although the previous studies demonstrated the trends of human productivity under different environmental conditions, they used a variety of human productivity metrics, including office workers' task speed, reaction time, scores of cognitive performance tests, and self-reported productivity. The use of diverse human productivity categories in this research field allows flexibility in conducting individual studies, but it poses several challenges. First, each productivity metric accounts for different aspects of performance (e.g., speed, accuracy). Second, meta-analyses are often conducted in this research field (e.g., [8-11]) due to the small participant sizes in individual studies and difficulty of studying human subjects. Hence, when productivity data using different metrics are compiled, it is challenging to conduct a meaningful meta-analysis. Specifically, Seppänen et al. [12] used productivity data from literature with a wide range of metrics, from objectively reported work performance to cognitive performance test scores, to create a human productivity prediction model with respect to air temperature. Even though all these metrics are important components of human productivity, they have different scales, units, and relationships to overall performance. Therefore, the results from such a model could become impracticable without careful consideration. In the end, this research aims to address the question of whether different types of human productivity metrics can be combined directly.

In order to tackle such challenges, this study categorized productivity metrics diversely employed in previous studies based on common attributes and units of measurement (hereinafter, these are called productivity metric categories). Then, the productivity metric categories were compared statistically using summary statistics, graphical representation, and a pairwise t-test. This approach used productivity metrics that measure the *efficiency* or *speed* (i.e., a ratio of output to input) of office work, categorized as neurobehavioral speed, as a baseline for comparing other productivity metric categories. An example of an efficiency-based productivity metric is the time it takes to complete a cognitive task, such as a visual learning module, measured in units completed per hour. Examples of other productivity metric categories are accuracy and performance score (more details in the Methodology section). If statistical similarity is found between two productivity datasets, there is some evidence supporting the ability to directly combine the productivity metrics in these categories within a human productivity meta-analysis. If there are incongruencies in the statistical behavior of two categories, it is suggested that careful consideration is needed to combine these categories in research studies. In summary, this study aims to evaluate the statistical similarity of the data from various productivity metrics. In the end, this research contributes to creating a constructive research environment for future meta-analyses in this field.

This paper is structured as follows. The Literature Review section provides a summary of human productivity metrics used in previous studies and meta-analysis efforts on office performance influenced by IEQ. The Methodology section details the steps taken to evaluate productivity metrics in this study, including how the data from existing productivity and IEQ studies were collected, how these data were processed in preparation for a meta-analysis, and an explanation of the statistical methods used to analyze the data. The Results section presents a detailed list of literature studies compiled to support this study, the results of the statistical tests conducted, and a discussion of the findings. In the Conclusion section, the limitations and the future directions of this research are discussed.

LITERATURE REVIEW

Office work fits in diverse contexts ranging from manufacturing-based to knowledge-based [13]. This diverse nature of office work adds complexity because many office workers conduct different types of tasks and skills on a daily basis. Ultimately, analyzing an individual office worker's productivity becomes challenging.

Studies tried to solve this by focusing on easily quantifiable office work or diversifying productivity metrics. For example, many field studies observed the call handling time of call center representatives [7, 14] or nurses [12], given the ease of collecting these data. In laboratory-based studies, cognitive performance tests have been widely used, given the association of cognitive function with office work. Even though this approach does not directly measure office work in a field setting, it offers a simulated situation where participants' productivity can be quantified and external conditions can be controlled. The tests to measure cognitive performance include numerical calculation, typewriting, memory (e.g., word, number, image), and reasoning tests (e.g., numerical, alphabetical, conditional, spatial, etc.) [15-17]. Studies have measured speed, response time, and/or accuracy as the human productivity metrics in such tests. Other studies utilized self-reported productivity, a subjective metric. The National Aeronautics and Space Administration (NASA) Task Load Index (TLX) [18] has been employed in a number of studies [19, 20]. In this index, participants report their productivity on a 20-level Likert scale (from perfect to failure). Similar questions were posed in other studies [21, 22].

To take a holistic perspective of office work it is important to look at a variety of productivity metrics; however, as discussed above, combining metrics may confound the results in a meta-analysis. For example, response time, speed, score, and accuracy are widely used metrics from a cognitive performance test. Each metric shows an important aspect of productivity with regard to IEQ [15], but their scales and units might not align in a way that makes them directly compatible with each other for an overarching productivity metric. As an attempt to mitigate this issue, some studies came up with new metrics to combine accuracy and response time [11, 14], normalized the values measured in each metric as a percent improvement [22], or applied weights to the different types of tasks based on their application to general office work [12, 23]. However, such attempts did not quantitatively analyze the applicability of joining the various metrics to each other.

This challenge has not been addressed in any previous meta-analyses. Seppänen et al. [23] collected studies that correlated ventilation rate to office productivity and created a regression model. This study weighted various productivity metrics based on their relative relevance to real work using the authors' judgment. A similar approach was taken in [12]. Some studies directly combined existing data regardless of productivity metrics without discussing their compatibility [24, 25]. Another research study compiled existing regression models relating thermal comfort and air temperature to productivity decrement to recreate a model that can be adapted to various work tasks [26]. This study also did not discuss the suitability of the productivity metrics themselves to be combined directly within a single model.

Hence, this research aims to support future studies where human productivity data are measured with diverse metrics that are simultaneously combined to gauge the impacts of IEQ in office environments. This research will help inform which types of data should be combined and which types of data require caution before combining with other types.

METHODOLOGY

Literature Collection

In order to collect the articles that investigated the impact of IEQ on human productivity in office buildings, we leveraged literature databases like Science Direct, Wiley, and Google Scholar to identify peer-reviewed publications. In such databases, the following keywords were individually and collectively used: office, productivity, performance, cognitive score, self-reported productivity, IEQ, thermal comfort, temperature, ventilation, CO_2, lighting, and horizontal illuminance. Then the title and the abstract of each study were read to see whether they aligned with the research interests. In addition, existing literature review papers and meta-analyses (e.g., [12]) were used to track down the studies in those publications to build an extensive collection of studies.

The criteria for including a study in this literature collection were the following: (1) it measured a specific change to the IEQ parameter (such as a general retrofit to the building (e.g., [27]) was excluded); (2) it defined the productivity metric(s); and (3) it recorded the impacts. Regression results from meta-analysis studies were not included; rather, the original studies were identified and reported directly in this collection. The IEQ metrics that were most available in literature studies that looked at productivity and performance variation were CO_2 [28], ventilation [29], thermal comfort [30], and horizontal illuminance [31]. These IEQ metrics are easily quantifiable in buildings and cover a wide range of IEQ categories (indoor air quality, thermal comfort, and lighting). In the end, the literature collection included 32 studies.

Data Preprocessing

After compiling the literature collection, the number of data points, number of subjects, study duration, and number of unique productivity metrics were extracted. The number of data points refers to the number of unique changes in IEQ conditions multiplied by the number of productivity metrics recorded in the study. For example, if a study changed CO_2 levels from 600 ppm to 1,000 ppm and recorded three productivity metrics at each condition, that would be three testing data points that could be analyzed individually. If the study tested three CO_2 conditions (e.g., 600 ppm, 1,000 ppm, and 1,500 ppm) and recorded three productivity metrics at each condition, that would be six data points (one of the CO_2 conditions is the baseline and the other two can be the experimental conditions). Note that the number of subjects and duration recorded refer to each data point, not the entire study. For example, if the subject population spent one week at 1,500 ppm CO_2 and one week at 700 ppm CO_2, the duration recorded would be one week (seven days) even though the entire study lasted two weeks. Studies that took place in less than one day (i.e., several hours) were recorded as one day.

There were a number of confounding factors (e.g., outdoor air temperature, season, acoustic conditions) that could be controlled for within studies but not between studies. It was found that studies did not consistently report all IEQ conditions when they were not the independent variables (e.g., a study varied lighting levels, but did not collect CO_2 or air temperature data). Therefore, the effects of different baseline IEQ and other conditions on the experimental parameters could not be included in this research.

Similar to the analyses in [12] and [23], it was determined to apply a weighting factor that would not completely overshadow studies with small sample sizes. Based on the judgment of the authors, weight on a scale of 1 to 10 was applied to each data point based on the number of subjects and the study duration using a simple logarithmic weighting function (Equation (1)).

$$W_n = 9 * \frac{LN(P_n * D_n)}{LN(P_{MAX} * D_{MAX})} + 1 \tag{1}$$

W	weighting factor
n	n^{th} data point in the set
D	duration of measurement (days)
P	number of participants
Max	maximum value in the set

In the literature review, some important caveats were identified in IEQ parameters. In the case of CO_2, some studies found a difference in artificially introduced CO_2 into a space and human-produced CO_2. The studies that measure both have found that artificially introduced CO_2 may not cause a significant decrement in productivity compared to human-produced CO_2 [32]. It is suggested that CO_2 itself is not a harmful pollutant at the levels typically found in buildings, but rather a proxy for human bioeffulents and other comorbid indoor air pollutants that are not being circulated out due to low ventilation [33]. For that reason, artificially and human-produced CO_2 were considered as separate IEQ parameters.

In the case of thermal comfort, two prominent variables were used in studies: predicted mean vote (PMV) and ambient temperature. PMV is a more holistic variable, taking into account ambient and radiant temperature, relative humidity, airflow, clothing level, and metabolic rate into its calculation. Eight of the 19 thermal comfort studies in our literature collection provide both PMV and temperature data and are included under both of these IEQ parameters. Some of these studies only gave temperature but included enough information to infer PMV. If the study included temperature and humidity values and at least a qualitative description of metabolic rate and clothing level, we calculated the PMV values using Center for the Built Environment's Thermal Comfort Tool [34].

For the last step of preprocessing, the productivity metrics were grouped into categories based on contextual similarity in order to strengthen the statistical analyses. For example, time to complete an addition task and time to complete a subtraction task were grouped into one category. More details on the attributes used to group metrics are included in the Results section. As noted, the approach in this research prioritized the productivity metrics interpreted as *efficiency* (i.e., a ratio of input to output) for a basis of reference for comparing to other categories.

Analyses of Productivity Metrics

Before comparing the impacts of each productivity metric category, the relative productivity was calculated using Equation (2). To maintain consistency in the positive and negative orientation of the results between IEQ parameters, the criteria for "worse" and "better" IEQ conditions were determined as the following:

- The greater the CO_2, the worse the IEQ.
- The greater the ventilation rate, the better the IEQ.
- The temperature farther from thermal neutral, the worse the IEQ (23℃ was used for thermal neutral based on the findings of optimal temperatures for performance in other studies [11, 25, 35, 36]).
- The greater the horizontal illuminance, the better the IEQ.

It is set up this way to show what the productivity improvement potential is for a building that is looking to improve IEQ.

$$RP_w = {(P_b - P_w)}/{P_w} \text{ (when higher P values are considered as 'better' performance)}$$

$$RP_w = {-(P_b - P_w)}/{P_w} \text{ (when lower P values are considered as 'better' performance)} \quad (2)$$

RP relative productivity improvement (%)

P absolute value of productivity metric (e.g., units/minute)

w worse IEQ condition

b better IEQ condition

If the units of P are considered better if they are less (e.g., number of errors, reaction time), then the value of RP is multiplied by -1 so that a positive value of RP corresponds to an improvement in productivity. If the units from the study are already in a percentage (e.g., percent errors, percent self-reported productivity), then the difference between the percentages is used instead of the percent change.

Three statistical methods were used to compare the datasets: summary statistics (mean, quartiles, variance); the density of values of each dataset represented graphically; and a pairwise t-test. The summary statistics convey the characteristics of the datasets in an easily interpretable fashion, the density plots graphically represent the data and their distribution, and the t-tests provided a more definitive conclusion concerning the similarities of the datasets.

The R programming language was employed to process these steps. These three methods were applied to (1) each productivity metric category and (2) each category separated by IEQ parameter. Throughout the remainder of this publication, they are referred to as the first analysis and second analysis.

To curtail possible statistical biases, several procedures were used. The outliers from each productivity grouping are defined as less than or greater than 1.5 multiplied by the interquartile range of the data from the 25th or 75th quartile respectively and removed. In the end, 14% of the total data points were removed. In the case of the first analysis, the results could be skewed by the range of IEQ values in the studies; for example, call handling time might include more studies with worse IEQ conditions whereas performance score might look at more studies with better IEQ conditions, making the productivity impacts relatively smaller. With a large sample of studies, the odds of an entire category being skewed is reduced. In response, for the second analysis, the data were separated by IEQ category to normalize based on change in the IEQ parameter (e.g., the change in relative productivity per-unit decrease in CO_2). Equation (3) shows the normalization by IEQ variable.

$$NRP_w = {}^{RP_w}/_{ABS(E_b - E_w)} \tag{3}$$

NRP	normalized relative productivity improvement (e.g., %/ppm CO_2)
RP	relative productivity improvement (%)
E	IEQ value (e.g., ppm CO_2)
w	worse IEQ condition
b	better IEQ condition

At the end of the Results section is a summary of the three methods of comparison for each productivity metric category. In this comparison, the summary statistics and density plots are assigned a rating on a six-point scale from very similar to very different based on qualitative observations of the data relative to the performance of the other categories. The t-test is assigned a rating on the same scale based on the p-value.

RESULTS

Data Preprocessing

Table 1 shows the data collected, attributes of the studies that were extracted and the weights calculated using Equation (1).

From the literature collection, a total of 106 productivity metrics were identified and grouped based on common attributes and units of measurement into the following six categories (Table 2). This categorization allowed for a greater sample size for the comparison of productivity metrics. Neurobehavioral speed most closely aligned with the definition of efficiency-based productivity described in the Introduction section, because the units were in output (tasks completed) per input (employees' time). Call handling time was also aligned with efficiency-based productivity in terms of units (e.g., calls completed per hour), but it was kept as a separate category because it was unique to the other metrics in that it required interpersonal communication and was typically collected in field studies as opposed to laboratory studies. Neurobehavioral response time looked at small timescale (seconds or milliseconds) compared to the neurobehavioral speed metrics (minutes, hours) and does not measure any type of work output. When human productivity was measured in a testing program with a score (e.g., on a scale of 10 or 100, as determined by the test creator), regardless of the focus area, it was categorized in the performance score category. These tests included the Strategic Management Simulation (SMS) test (e.g., basic/applied/focused activity level, task orientation, crisis response information seeking/usage [59, 60]) and cognitive performance tests like digit-span memory, picture recognition, symbol-digit modalities, text typing, Tsai-Partington, creative thinking, executive function, and cognitive flexibility. Five out of the 16 studies that have performance score metrics used SMS [28, 40, 43, 45, 46].

Statistical Evaluation of Productivity Metric Categories

Table 3 shows the summary statistics of the datasets for each productivity metric category. The metrics included are the mean, variance, and quartile values after applying the weighting factors from Table 1.

The results indicated that neurobehavioral speed was most similar to neurobehavioral response time and somewhat similar to self-reported productivity and call handling time, considering primarily the value of the means and also looking at the percentiles and variance. Performance score was the least similar to speed, with a mean almost three times larger and a large variance. Figure 1 shows the same results graphically. The curve for performance score showed a wide distribution, skewed to the right and extending past 20% on the x-axis where the graph is cut off. Accuracy shows a very narrow distribution compared to the other categories and neurobehavioral response time, neurobehavioral speed, self-reported productivity, and call handling time were fairly similar in distribution.

Table 4 shows the unweighted p-value results of the pairwise t-test. The six productivity grouping datasets did not meet the heteroskedasticity assumption of an ANOVA test (Levene's test p-value < $2.2e^{-16}$) because the sample size of each group did not have equal variance. Therefore, a non-parametric pairwise t-test adjusted by the Benjamini-Hochberg with no assumption of equal variance [61] was conducted. Neurobehavioral speed did not show a significant difference (i.e., showed similarity) to call handling time, neurobehavioral response time, and self-reported productivity.

Statistical Evaluation of Productivity Metric Category Behavior by IEQ Parameter

The second analysis divided the data by IEQ metrics. Although this reduced the sample size, it gave the ability to look at per-unit changes to productivity for comparison. Horizontal illuminance did not have enough data points to analyze individually and was not included in this section. Some of the productivity metric categories no longer had a robust sample size within certain IEQ metrics and those categories were omitted from the analyses and discussion. It should be noted that there are no conclusions being made in this research about the behavior of productivity metric categories within IEQ metrics; rather these observations will be compiled to draw insights about the productivity categories as a whole, outside of any specific IEQ metric.

Within human-produced CO_2, ventilation, and PMV, the productivity category datasets did not meet the heteroskedasticity assumptions of the standard ANOVA test (Levene's test p-value $< 4.6e^{-06}$, $< 8.8e^{-05}$, and < 0.032 respectively). Therefore, a non-parametric pairwise t-test adjusted by the Benjamini-Hochberg method with no assumption of equal variance was conducted, just as was performed in the first analysis. Although within artificially introduced CO_2 and ambient temperature the productivity categories met the heteroskedasticity assumptions of the standard ANOVA test, we used the same non-parametric pairwise t-test for consistency.

Carbon Dioxide

In Table 5, for human-produced CO_2, call handling time and neurobehavioral speed had similar means but dissimilar quartiles. Call handling time had a relatively small sample size for human-produced CO_2 and neurobehavioral response time had a small sample size in both tables. These categories were therefore excluded from the pairwise t-test and graphical representations in this section. No self-reported productivity data were identified for CO_2. Accuracy and neurobehavioral response time had means very close to zero, suggesting there was not a noticeable relationship between CO_2 and these productivity metrics. Performance score had a very large mean, suggesting it is not directly compatible with the other metrics. For artificially introduced CO_2, there was also a large variance for performance scores. This means that the various metrics within the performance score category may not be compatible even with each other.

Figure 2 shows the density distribution plot for productivity improvement results of studies by productivity metric category and by human-introduced CO_2 (left) and artificially introduced CO_2 (right). The horizontal extents of the two plots were limited to -20 to +40%/1,000 ppm to better show the behavior of the plots around the means, even though the data of performance score extended farther. The narrow curves with a high spike signify cohesiveness in the performance impacts, whereas the flat wide curves indicate there is high variability in the results.

Call handling time and neurobehavioral response time were removed from Tables 7 and 8 due to small sample size. Table 7 shows that performance score had low similarity to all other datasets for human-produced CO_2. All other datasets for human-produced CO_2 and all datasets within artificially introduced CO_2 were similar to each other, respectively.

Ventilation Rate

Within ventilation rate, call handling time and neurobehavioral speed had similar means and similar quartiles, as shown in Table 9. Performance score had a much larger mean and percentiles than the other groupings. Call handling time and neurobehavioral response time had small sample sizes and were therefore excluded from the pairwise t-test and graph in this section. Figure 3 shows that performance score had a very wide distribution compared to neurobehavioral speed, and accuracy had a slightly narrower distribution.

In Table 10, all the categories included showed significant differences from each other.

Thermal Comfort

In Table 11, for temperature, neurobehavioral speed aligned somewhat well with call handling time. In Table 12, for PMV, neurobehavioral speed did not align with any of the other groupings. Performance score had a negative mean, suggesting an inverse relationship to PMV.

Figure 4 shows the density distribution plot for productivity improvement results of studies by productivity category and by ambient temperature (left) and PMV (right).

Tables 13 and 14 show the results of the t-test for temperature and PMV respectively. Call handling time did not show significant similarity to accuracy or neurobehavioral response time for temperature. Accuracy did not show significant similarity to neurobehavioral response time for PMV. The remaining categories presented did not show any significant differences from each other.

Summary of Results and Discussion

Table 15 shows the summary of the results using neurobehavioral speed as a basis of reference. This table includes the statistical summary statistics (focusing on the mean as a value for comparison), a comparison through graphical observation (focusing on the distribution of data), and the pairwise t-test results. Each comparison in the table is given a rating: very similar, similar, somewhat similar, somewhat different, different, or very different. For the summary statistics and graphs, these ratings are assigned qualitatively based on the results presented. For the pairwise t-test, the rating is assigned based on the p-value.

Artificially introduced CO_2, with the except of performance score, had minimal impact on productivity results. This could be, as other studies have suggested, because CO_2 is an indicator of comorbid indoor air pollutants and ventilation rate and is not a harmful pollutant itself in the concentrations typically found in buildings. The data we collected for this IEQ variable are therefore not used in the comparison of productivity metrics categories in Table 15. Temperature and PMV were combined into thermal comfort for the summary because there is some overlap in the studies these data come from and because the results were generally similar.

Based on the summaries in Table 15, all of the p-values for call handling time, neurobehavioral response time, and self-reported productivity showed no significant differences when compared to neurobehavioral speed. In addition, the majority of qualitative ratings based on the statistical summaries and graphical comparisons for these categories were somewhat similar, similar, or very similar. The majority of the p-values for accuracy and performance score (with the exception of thermal comfort analysis) showed significant difference compared to neurobehavioral speed. In addition, the majority of the qualitative ratings are somewhat different, different, or very different.

CONCLUSIONS

This study has analyzed human productivity metrics used to measure the impact of IEQ in office buildings. The approach taken defines productivity as an efficiency (ratio of input to output), which has paved the way for investigating each of the productivity metric categories. In previous studies and meta-analyses, the absence of this analysis confused the feasibility of combining productivity metrics in meta-analyses.

The studies that aligned with our research interest were compiled, and each study's data points, IEQ conditions, productivity metrics, and productivity impact were extracted. The 106 productivity metrics were grouped into six categories and neurobehavioral speed category was set as the baseline given its fitness to our definition of productivity. The other five categories (neurobehavioral response time, accuracy, call handling time, performance score, and self-reported productivity) were analyzed statistically. The neurobehavioral response time, self-reported productivity, and call handling time metrics were found to have statistical similarity to the baseline. Based on the results of this study, regression models from meta-analyses that incorporate performance score or accuracy metrics with other productivity categories should be scrutinized. In future studies, more care should be given to defining productivity and evaluating how studies fit into this definition, so the results of those studies are meaningful.

Data scarcity is a significant limitation of this research. All the categories would benefit from more data points due to the complex nature of this field and from studies with larger participant sizes and study durations. The authors note that this research is not conclusive and only suggestive of the true nature of productivity metrics given the limited data that were used for the findings. Furthermore, this research would benefit from the inclusion of more IEQ parameters, such as volatile organic compounds, particulate matter, daylight, and acoustics, given their impacts on human productivity in buildings. This would provide a larger picture of the building-occupant nexus.

There are a number of confounding factors that could be controlled for within studies but not between studies. For example, one study could look at the effects of CO_2 at 30℃ and another could run the experiment at 21℃. The studies were conducted at various times of the year and in various climate zones, which can impact the indoor environment and human behavior. It is expected that, in general, the results will be similar in nature; for example, decreasing CO_2 at a constant 30°C will likely improve productivity if decreasing CO_2 by the same amount improves productivity at 21℃; however, the magnitude of this productivity improvement may be dampened or exacerbated due to the interaction of environmental parameters. Because these confounding factors are not reported consistently across studies, it is not possible to take them into consideration and still maintain a large sample size for the meta-analysis.

This research is applicable to productivity in general for office workers. In future research, we plan to explore how the diversity of work functions and types could be addressed to make this research more applicable to individual buildings. Different types of cognitive tasks may be more or less applicable to different types of office work. For example, a digit-span memory test may be a suitable metric for a worker who regularly exercises executive function skills but may not be applicable to a worker who relies mostly on innovation and creativity skills. In a future where there is uncertainty around return-to-work status and a potential norm of increased remote working, there is the possibility of applying research like this to remote workers to understand how the IEQ conditions of employees' homes could be creating an opportunity for employer-sponsored home improvements.

ACKNOWLEDGMENTS

The research is funded by the U.S. Department of Energy (DOE) Federal Energy Management Program. The authors would like to thank the DOE project managers Jefferey Murrell and Allison Ackerman for sponsoring this research and their continued support throughout the project. Special thanks go to our General Services Administration collaborators Judith Heerwagen, PhD, and Brian Gilligan for sharing their technical expertise and domain knowledge. We would like to thank Erik Mets at Pacific Northwest National Laboratory for contributing to this research and providing feedback.

References

1. WELL. *WELL projects*. 2021 [cited 2021 06/28]; Available from: https://account.wellcertified.com/directories/projects.

2. Sieglinde Fuller. *Life-Cycle Cost Analysis (LCCA)*. 2016 [cited 2019 April 30]; Available from: https://www.wbdg.org/resources/life-cycle-cost-analysis-lcca.

3. William J. Fisk, *HEALTH AND PRODUCTIVITY GAINS FROM BETTER INDOOR ENVIRONMENTS AND THEIR RELATIONSHIP WITH BUILDING ENERGY EFFICIENCY.* Annual Review of Energy and the Environment, 2000. **25**(1): p. 537-566.

4. C.-E. A. Winslow, James Alexander Miller, D. D. Kimball, Earle B. Phelps, Frederic S. Lee, Edward Lee Thorndike, George T. Palmer, and Joseph Herzstein, *Ventilation: Report of the New York State Commission on Ventilation*. 1923: E.P. Dutton & Company.

5. Rajat Gupta, Alastair Howard, and Sahar Zahiri, *Investigating the relationship between indoor environment and workplace productivity in naturally and mechanically ventilated office environments.* Building Services Engineering Research and Technology, 2020. **41**(3): p. 280-304.

6. Akihiro Kawamura, Shin-ichi Tanabe, Naoe Nishihara, Masaoki Haneda, and Masanori Ueki. *Evaluation method for effects of improvement of indoor environmental quality on productivity*. in *Clima 2007 WellBeing Indoors*. 2007. Citeseer.

7. Raimo Niemelä, Mika Hannula, Sari Rautio, Kari Reijula, and Jorma Railio, *The effect of air temperature on labour productivity in call centres—a case study.* Energy and Buildings, 2002. **34**(8): p. 759-764.

8. John F. Wing, *A review of the effects of high ambient temperature on mental performance*. 1965, AIR FORCE AEROSPACE MEDICAL RESEARCH LAB WRIGHT-PATTERSON AFB OH.

9. C. Roelofsen, *The impact of office environments on employee performance: The design of the workplace as a strategy for productivity enhancement.* Journal of Facilities Management, 2002. **1**: p. 247-264.

10. Olli Seppänen, William J. Fisk, and David Faulkner, *Cost benefit analysis of the night-time ventilative cooling in office building*. 2003.

11. Li Lan, Pawel Wargocki, and Zhiwei Lian, *Quantitative measurement of productivity loss due to thermal discomfort.* Energy and Buildings, 2011. **43**(5): p. 1057-1062.

12. Olli Seppänen, William J Fisk, and QH Lei, *Effect of temperature on task performance in office environment*. 2006, Lawrence Berkeley National Laboratory.

13. Barry P. Haynes, *Office productivity: a theoretical framework.* Journal of Corporate Real Estate, 2007.

14. Lisa Heschong, *Windows and Offices: A Study of Office Worker Performance and the Indoor Environment*. 2003, California Energy Commission: Fair Oaks, CA.

15. Li Lan and Zhiwei Lian, *Use of neurobehavioral tests to evaluate the effects of indoor environment quality on productivity.* Building and Environment, 2009. **44**(11): p. 2208-2217.

16. Li Lan, Zhiwei Lian, Li Pan, and Qian Ye, *Neurobehavioral approach for evaluation of office workers' productivity: The effects of room temperature.* Building and Environment, 2009. **44**(8): p. 1578-1588.

17. Cheng Sun, Yunsong Han, Lin Luo, and Huixuan Sun, *Effects of air temperature on cognitive work performance of acclimatized people in severely cold region in China.* Indoor and Built Environment, 2020: p. 1420326X20913617.

18. Sandra G. Hart and Lowell E. Staveland, *Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research*, in *Advances in Psychology*, P.A. Hancock and N. Meshkati, Editors. 1988, North-Holland. p. 139-183.

19. Naoe Nishihara, Shin-ichi Tanabe, Masaoki Haneda, Masanori Ueki, Akihiro Kawamura, and Kouei Obata, *Effect of overcooling on productivity evaluated by the long term field study.* 2007.

20. M. Haneda, Pawel Wargocki, M. Dalewski, and Shin-ichi Tanabe, *The effects of thermal discomfort on task performance, fatigue and mental work load examined in a subjective experiment.* 9th International Conference and Exhibition - Healthy Buildings 2009, HB 2009, 2009.

21. Weilin Cui, Guoguang Cao, Jung Ho Park, Qin Ouyang, and Yingxin Zhu, *Influence of indoor air temperature on human thermal comfort, motivation and performance.* Building and Environment, 2013. **68**: p. 114-122.

22. S. Lamb and K. C. S. Kwok, *A longitudinal investigation of work environment stressors on the performance and wellbeing of office workers.* Applied Ergonomics, 2016. **52**: p. 104-111.

23. O. Seppänen, W. J. Fisk, and Q. H. Lei, *Ventilation and performance in office work.* Indoor Air, 2006. **16**(1): p. 28-36.

24. Pawel Wargocki, Jose Ali Porras-Salazar, Sergio Contreras-Espinoza, and William Bahnfleth, *The relationships between classroom air quality and children's performance in school.* Building and Environment, 2020. **173**: p. 106749.

25. Yang Geng, Wenjie Ji, Borong Lin, and Yingxin Zhu, *The impact of thermal environment on occupant IEQ perception and productivity.* Building and Environment, 2017. **121**: p. 158-167.

26. Paul Roelofsen, *A computer model for the assessment of employee performance loss as a function of thermal discomfort or degree of heat stress.* Intelligent Buildings International, 2016. **8**(4): p. 195-214.

27. Joseph J. Romm and William D. Browning, *Greening the building and the bottom line.* Rocky Mountain Institute. Snowmass, colorado, 1994.

28. Joseph G. Allen, Piers MacNaughton, Usha Satish, Suresh Santanam, Jose Vallarino, and John D. Spengler, *Associations of Cognitive Function Scores with Carbon Dioxide, Ventilation, and Volatile Organic Compound Exposures in Office Workers: A Controlled Exposure Study of Green and Conventional Office Environments.* Environmental health perspectives, 2016. **124**(6): p. 805-812.

29. M. Haneda, Shinichi Tanabe, N. Nishihara, and S. Nakamura. *The combined effects of thermal environment and ventilation rate on productivity*. 2008.

30. Fan Zhang, Shamila Haddad, Bahareh Nakisa, Mohammad Naim Rastgoo, Christhina Candido, Dian Tjondronegoro, and Richard de Dear, *The effects of higher temperature setpoints during summer on office workers' cognitive load and thermal comfort.* Building and Environment, 2017. **123**: p. 176-188.

31. Iason Konstantzos, Seyed Amir Sadeghi, Michael Kim, Jie Xiong, and Athanasios Tzempelikos, *The effect of lighting environment on task performance in buildings – A review.* Energy and Buildings, 2020. **226**: p. 110394.

32. X. Zhang, P. Wargocki, Z. Lian, and C. Thyregod, *Effects of exposure to carbon dioxide and bioeffluents on perceived air quality, self-assessed acute health symptoms, and cognitive performance.* Indoor Air, 2017. **27**(1): p. 47-64.

33. Kenichi Azuma, Naoki Kagi, U. Yanagi, and Haruki Osawa, *Effects of low-level inhalation exposure to carbon dioxide in indoor environments: A short review on human health and psychomotor performance.* Environment International, 2018. **121**: p. 51-56.

34. Federico Tartarini, Stefano Schiavon, Toby Cheung, and Tyler Hoyt, *CBE Thermal Comfort Tool: Online tool for thermal comfort calculations and visualizations.* SoftwareX, 2020. **12**: p. 100563.

35. Raimo Niemela, Jorma Railio, Mika Hannula, Sari Rautio, and Kari Reijula. *Assessing the effect of indoor environment on productivity*. in *Clima 2000*. 2001. Napoli.

36. C. Federspiel, G. Liu, M. Lahiff, D. Faulkner, D. DiBartolomeo, W. Fisk, P. Price, and D. Sullivan, *WORKER PERFORMANCE AND VENTILATION: ANALYSES OF INDIVIDUAL DATA FOR CALL-CENTER WORKERS.* Lawrence Berkeley National Laboratory, 2002.

37. H. Maula, V. Hongisto, V. Naatula, A. Haapakangas, and H. Koskela, *The effect of low ventilation rate with elevated bioeffluent concentration on work performance, perceived indoor air quality, and health symptoms.* Indoor Air, 2017. **27**(6): p. 1141-1153.

38. Pawel Wargocki, David P. Wyon, and P. Ole Fanger, *The performance and subjective responses of call-center operators with new and used supply air filters at two outdoor air supply rates.* Indoor Air, 2004. **14**(s8): p. 7-16.

39. Junseok Park and C. Yoon, *The effects of outdoor air supply rate on work performance during 8-h work period.* Indoor air, 2010. **21**: p. 284-90.

40. R. Maddalena, M. J. Mendell, K. Eliseeva, W. R. Chan, D. P. Sullivan, M. Russell, U. Satish, and W. J. Fisk, *Effects of ventilation rate per person and per floor area on perceived air quality, sick building syndrome symptoms, and decision-making.* Indoor Air, 2015. **25**(4): p. 362-370.

41. László Kajtár, Levente Herczeg, E. Láng, T. Hrustinszky, and L. Bánhidi, *Influence of carbon-dioxide pollutant on human well-being and work intensity.* HB 2006 - Healthy Buildings: Creating a Healthy Indoor Environment for People, Proceedings, 2006. **1**: p. 85-90.

42. Stephen Snow, Amy S. Boyson, Karlien H. W. Paas, Hannah Gough, Marco-Felipe King, Janet Barlow, Catherine J. Noakes, and m c schraefel, *Exploring the physiological, neurophysiological and cognitive performance effects of elevated carbon dioxide concentrations indoors.* Building and Environment, 2019. **156**: p. 243-252.

43. Usha Satish, Mark J. Mendell, Krishnamurthy Shekhar, Toshifumi Hotchi, Douglas Sullivan, Siegfried Streufert, and William J. Fisk, *Is CO2 an indoor pollutant? Direct effects of low-to-moderate CO2 concentrations on human decision-making performance.* Environmental health perspectives, 2012. **120**(12): p. 1671-1677.

44. Weiwei Liu, Weidi Zhong, and Pawel Wargocki, *Performance, acute health symptoms and physiological responses during exposure to high air temperature and carbon dioxide concentration.* Building and Environment, 2017. **114**: p. 96-105.

45. Christopher D. Rodeheffer, Sarah Chabal, John M. Clarke, and David M. Fothergill, *Acute Exposure to Low-to-Moderate Carbon Dioxide Levels and Submariner Decision Making.* Aerospace Medicine and Human Performance, 2018. **89**(6): p. 520-525.

46. Robert R. Scully, Mathias Basner, Jad Nasrini, Chiu-wing Lam, Emanuel Hermosillo, Ruben C. Gur, Tyler Moore, David J. Alexander, Usha Satish, and Valerie E. Ryder, *Effects of acute exposures to carbon dioxide on decision making and cognition in astronaut-like subjects.* npj Microgravity, 2019. **5**(1): p. 17.
47. Pawel Wargocki, David Wyon, and P. Fanger, *Productivity is affected by the air quality in offices.* Proceedings of Healthy Buildings 2000, 2000. **1**.

48. K. W. Tham, *Effects of temperature and outdoor air supply rate on the performance of call center operators in the tropics.* Indoor Air, 2004. **14**(s7): p. 119-125.

49. Michael Humphreys and Fergus Nicol, *Self-assessed productivity and the office environment: Monthly surveys in five European countries.* ASHRAE Transactions, 2007. **113**: p. 606-616.

50. Shin-ichi Tanabe, Masaoki Haneda, and Naoe Nishihara, *Workplace productivity and individual thermal satisfaction.* Building and Environment, 2015. **91**: p. 42-50.

51. Thomas Witterseh, *Environmental perception, SBS symptoms and the performance of office work under combined exposures to temperature, noise and air pollution*, in *Mechanical Engineering*. 2001, Technical University of Denmark: Kgs. Lyngby.

52. L. G. Berglund, R. R. Gonzales, and A. P. Gagge. *Predicting human performance decrement from thermal discomfort and ET*. in *the 5th International Conference on Indoor Air Quality and Climate*. 1990. Toronto, Canada.

53. S. Tanabe and N. Nishihara, *Productivity and fatigue.* Indoor Air, 2004. **14**(s7): p. 126-133.
54. Alan Hedge, Wafa Sakr, and Anshu Agarwal. *Thermal effects on office productivity*. 2005. SAGE Publications Sage CA: Los Angeles, CA.

55. Hui Zhang, Edward Arens, DongEun Kim, Elena Buchberger, Fred Bauman, and Charlie Huizenga, *Comfort, perceived air quality, and work performance in a low-power task–ambient conditioning system.* Building and Environment, 2010. **45**(1): p. 29-39.

56. D. P. Wyon, *The Effects of Moderate Heat Stress on Typewriting Performance.* Ergonomics, 1974. **17**(3): p. 309-317.

57. D. P. Wyon, *Individual microclimate control: required range, probable benefits, and current feasibility.* Proc. IndoorAir 96, Tokyo, 1996. **1**: p. 1067-1072.

58. J. F. Barnaby, *Lighting for productivity gains.* Lighting Design & Application, 1980. **10**(2): p. 20-28.

59. Siegfried Streufert, Rosanne Pogash, and Mary Piasecki, *SIMULATION-BASED ASSESSMENT OF MANAGERIAL COMPETENCE: RELIABILITY AND VALIDITY.* Personnel Psychology, 1988. **41**(3): p. 537-557.

60. U. Satish, S. Streufert, M. Dewan, and S. Vande Voort, *Improvements in simulated real-world relevant performance for patients with seasonal allergic rhinitis: impact of desloratadine.* Allergy, 2004. **59**(4): p. 415-420.

61. Wouter van Loon, *The Power of the Benjamini-Hochberg Procedure*. 2017, Leiden University: Leiden, The Netherlands. p. 83.

Figure Captions List

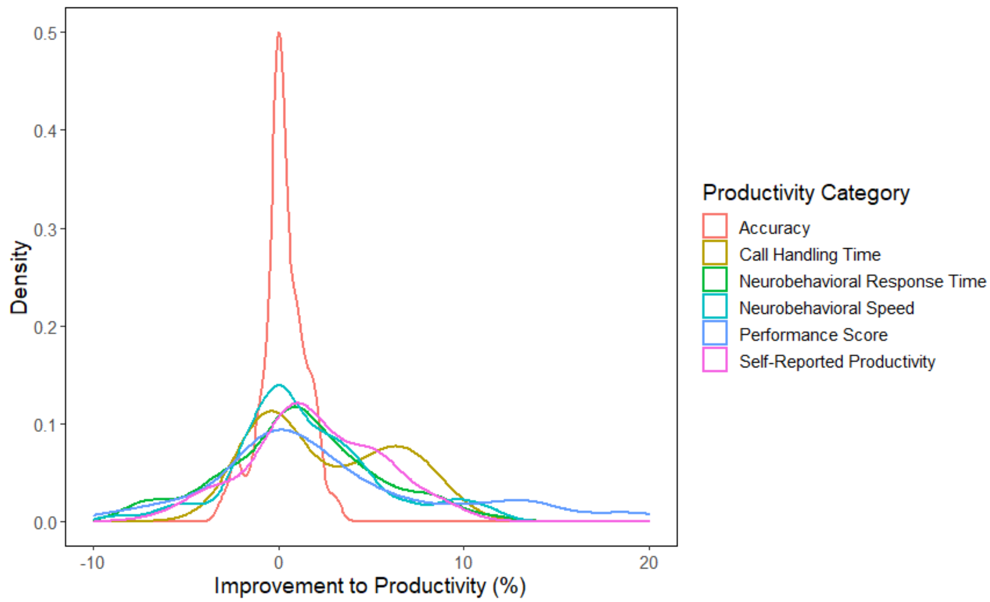

Fig. 1 Density distribution plot for improvement by productivity category

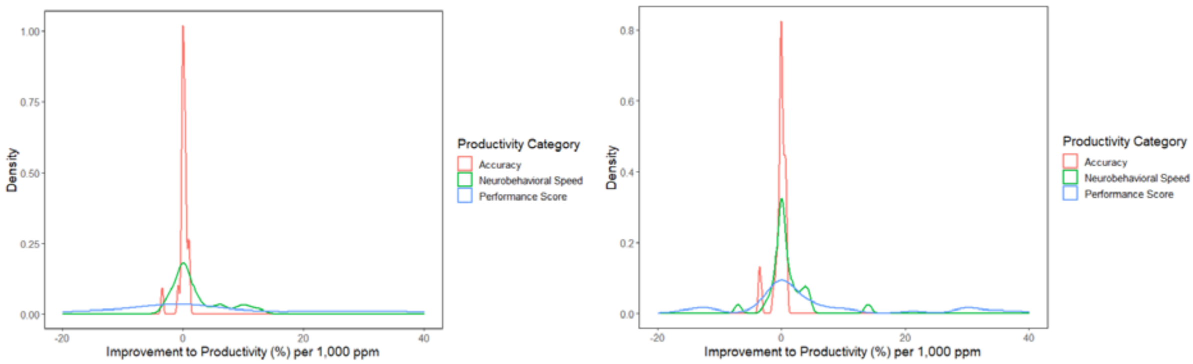

Fig. 2 Density distribution plot of improvement by productivity metric category for human-introduced CO_2 (left) and artificially introduced CO_2 (right)

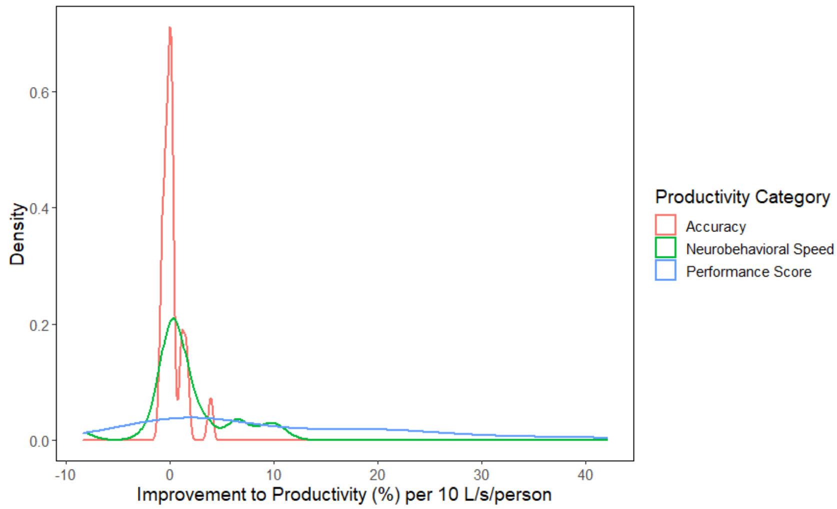

Fig. 3 Density distribution plot of improvement by productivity metric category for ventilation rate

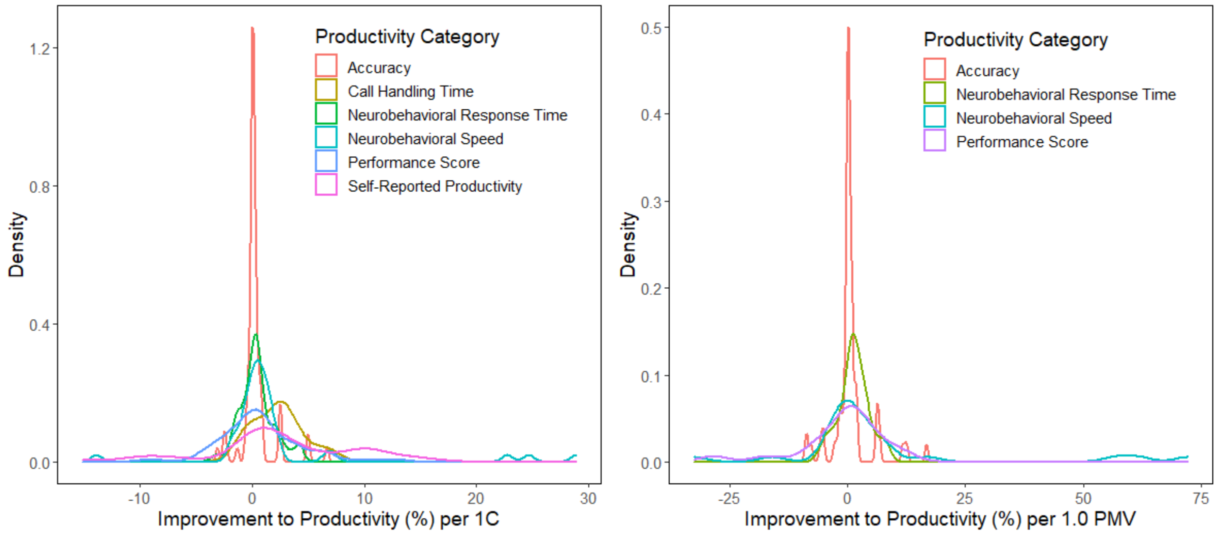

Fig. 4 Density distribution plot of improvement by productivity metric category for ambient temperature (left) and PMV (right)

Table Captions List

Table 1 List of studies comprising the literature collection and relevant details

Reference	# of subjects	Duration (days)	# of data points	# of productivity metrics	Weight					
					Human CO_2	Artificial CO_2	Vent. rate	Amb. Temp.	PMV	Hor. Illum.
Allen et al. [28]	24	1	27	9	3.93	3.93	3.93			
Maula et al. [37]	36	1	14	14	4.30		4.30			
Federspiel et al. [36]	290	60	3	1	10.0		10.0	10.0		
Wargocki et al. [38]	26	14	1	1	6.44		6.44			
Park and Yoon [39]	24	1	14	7	3.93		3.93			
Zhang et al. [32]	25	1	48	24	3.97	3.97	3.97			
Maddalena et al. [40]	16	1	8	8	3.56		3.56			
Kajtár et al. [41]	10	1	8	2		3.12				
Snow et al. [42]	31	1	7	7		4.17				
Satish et al. [43]	22	1	18	9		3.85				
Liu et al. [44]	12	1	21	21		3.29		3.29	3.29	
Rodeheffer et al. [45]	36	1	9	9		4.30				
Scully et al. [46]	22	2	18	9		4.49				
Wargocki et al. [47]	30	1	6	2			4.13			
Tham [48]	26		1	1			4.00	4.00		
Heschong [14]*	50	40	2	1			8.01			
Heschong [14]*	33	40	2	1				7.62		
Heschong [14]*	17	20	6	1						6.37
Humphreys and Nicol [49]	931^{+}	1	45	1				7.16		
Cui et al. [21]	36	1	8	2				4.30		
Geng et al. [25]	21	1	30	5				3.81		
Tanabe et al. [50]*	11	6	18	6				4.86	4.86	
Tanabe et al. [50]*	56^{+}	3	11	1				5.32^{+}		
Witterseh [51]	30		3	3				4.13		
Niemelä et al. [7]	17	30	10	1				6.75		
Berglund et al. [52]	11		15	1				3.21	3.21	
Tanabe and Nishihara [53]*	40	1	4	1				4.40	4.40	
Tanabe and Nishihara [53]*	16	1	1	1						3.56

Reference	# of subjects	Duration (days)	# of data points	# of productivity metrics	Weight					
					Human CO_2	Artificial CO_2	Vent. rate	Amb. Temp.	PMV	Hor. Illum.
Hedge et al. [54]	9	2	22	2				3.66	3.66	
Zhang et al. [55]	18		12	3				3.66	3.66	
Lan and Lian [15]	21	1	46	23				3.81	3.81	
Lan et al. [16]	24	1	51	17				3.93	3.93	
Lan et al. [11]	12	14	12	12				5.72		
Wyon [56]	15	10	2	1				5.62		
Wyon [57]	22	1	36	4				3.85		
Barnaby [58]	10	2	2	1						3.76

*Studies that had different subjects/durations/metrics for different data points.
+Average value presented due to varying subject numbers for different data points.
Amb. Temp.: Ambient Temperature, Hor. Illum.: Horizontal Illuminance.

Table 2 Details of productivity metric categories used for this study

Productivity Metric Category	Common Attribute	Units	Sample Metrics	Unique Metrics
Accuracy	Any type of accuracy (correct response, incorrect responses) of a test or task	Number of errors, percent correct, percent incorrect, etc.	Accuracy on reading comprehension test, accuracy on typing task, accuracy on proof-reading task, accuracy on grammatical reasoning task, etc.	30
Call Handling Time	Productivity of employees in call centers	Minutes per call, calls completed per hour	Average call handling time, average call wrap-up time, number of calls completed	3
Neurobehavioral Response Time	Time to complete fast-paced cognitive exercise	Seconds, milliseconds	Spatial image reaction time, letter search response time, visual choice response time, etc. measured in seconds typically	20
Neurobehavioral Speed	Speed or output on complex cognitive tasks	Units/minute, units/hour, characters/minute	Text typing char/hour, text typing total output, addition test units/min, attention test units/10 mins, judgment test units/min, etc.	21
Performance Score	Score on performance or cognitive test	Score out of 10, score out of 100, etc.	Score on strategy test, simulated management software score (e.g., results from the SMS [59]), concentration score, other cognitive performance tests	31
Self-Reported Productivity	Survey-based	Percent	Self-reported productivity	1
			Total number of metrics	106

Table 3 Summary statistics by productivity metric category for the first analysis

Productivity Metric Category	# of Data Points	25th Percentile	Mean	75th Percentile	Variance
Accuracy	253	-0.30%	0.22%	1.00%	0.01%
Call Handling Time	33	-0.50%	2.31%	6.00%	0.12%
Neurobehavioral Response Time	140	-1.82%	1.12%	3.55%	0.16%
Neurobehavioral Speed	184	-0.56%	1.62%	3.55%	0.15%
Performance Score	242	-1.45%	4.27%	9.71%	1.77%
Self-reported productivity	61	0.00%	2.02%	5.00%	0.12%

Table 4 p-value results of a pairwise t-test by productivity metric category for the first analysis

Productivity Metric Category	Accuracy	Call Handling Time	Neurobehavioral Response Time	Neurobehavioral Speed	Performance Score
Call Handling Time	0.002**				
Neurobehavioral Response Time	0.028*	0.071			
Neurobehavioral Speed	0.001**	0.153	0.467		
Performance Score	0.000***	0.153	0.003**	0.007**	
Self-Reported Productivity	0.002**	0.421	0.208	0.421	0.033*

* p-value of less than 0.05; ** p-value of less than 0.01; *** p-value of less than 0.001

Table 5 Summary statistics by productivity metric category for human-produced CO_2 in units of productivity impact (%) per 1,000 ppm

Productivity Metric Category	# of Data Points	25^{th} Percentile	Mean	75^{th} Percentile	Variance
Accuracy	22	-0.15%	0.00%	0.39%	0.00%
Call Handling Time	4	-8.31%	2.21%	11.22%	0.00%
Neurobehavioral Response Time	10	0.00%	0.38%	0.73%	0.00%
Neurobehavioral Speed	29	-0.04%	2.04%	2.93%	0.00%
Performance Score	34	-1.42%	27.97%	44.67%	0.02%

Table 6 Summary statistics by productivity metric category for artificially introduced CO_2 in units of productivity impact (%) per 1,000 ppm

Productivity Metric Category	# of Data Points	25th Percentile	Mean	75th Percentile	Variance
Accuracy	27	-0.19%	-0.23%	0.46%	0.00%
Neurobehavioral Response Time	8	-2.92%	0.05%	0.63%	0.00%
Neurobehavioral Speed	29	0.00%	0.90%	1.85%	0.00%
Performance Score	58	-1.78%	113.94%	30.15%	1.88%

Table 7 Results of pairwise t-test by productivity metric category in human-produced CO_2

Productivity Metric Category	**Accuracy**	**Neurobehavioral Speed**
Neurobehavioral Speed	0.014*	
Performance Score	0.002**	0.002**

* p-value of less than 0.05; ** p-value of less than 0.01

Table 8 Results of pairwise t-test by productivity metric category in artificially introduced CO_2

Productivity Metric Category	Accuracy	Neurobehavioral Speed
Neurobehavioral Speed	0.108	
Performance Score	0.073	0.073

Table 9 Summary statistics by productivity metric category for ventilation rate in units of productivity impact per 10 Liters/second/person

Productivity Metric Category	# of Data Points	25^{th} Percentile	Mean	75^{th} Percentile	Variance
Accuracy	23	-0.39%	0.24%	0.27%	0.00%
Call Handling Time	7	0.00%	2.21%	5.06%	0.01%
Neurobehavioral Response Time	9	-1.33%	-0.74%	0.74%	0.01%
Neurobehavioral Speed	32	-0.11%	1.72%	2.91%	0.01%
Performance Score	29	0.00%	9.54%	18.50%	0.14%

Table 10 Results of pairwise t-test by productivity metric category in ventilation rate

Productivity Metric Category	Accuracy	Neurobehavioral Speed
Neurobehavioral Speed	0.042*	
Performance Score	0.001***	0.003**

* p-value of less than 0.05; ** p-value of less than 0.01;
*** p-value of less than 0.001

Table 11 Summary statistics by productivity metric category for ambient temperature in units of productivity impact per 1°C

Productivity Metric Category	# of Data Points	25th Percentile	Mean	75th Percentile	Variance
Accuracy	69	0.00%	0.34%	0.45%	0.02%
Call Handling Time	16	0.00%	2.05%	3.04%	0.05%
Neurobehavioral Response Time	61	-0.51%	0.42%	0.90%	0.02%
Neurobehavioral Speed	43	-0.17%	1.73%	1.34%	0.41%
Performance Score	59	-1.22%	0.68%	2.12%	0.13%
Self-Reported Productivity	55	0.00%	2.59%	6.25%	0.41%

Table 12 Summary statistics by productivity metric category for PMV in units of productivity impact per 1.0 change in PMV

Productivity Metric Category	# of Data Points	25th Percentile	Mean	75th Percentile	Variance
Accuracy	61	-0.27%	0.49%	0.84%	0.17%
Neurobehavioral Response Time	44	-0.03%	1.34%	3.46%	0.10%
Neurobehavioral Speed	29	-2.40%	6.42%	5.56%	4.27%
Performance Score	21	-4.63%	-1.69%	3.87%	0.86%

Table 13 Results of pairwise t-test by productivity metric category for ambient temperature

Productivity Metric Category	Accuracy	Call Handling Time	Neurobehavioral Response Time	Neurobehavioral Speed	Performance Score
Call Handling Time	0.039*				
Neurobehavioral Response Time	0.871	0.039*			
Neurobehavioral Speed	0.242	0.871	0.255		
Performance Score	0.618	0.115	0.741	0.421	
Self-Reported Productivity	0.081	0.871	0.083	0.871	0.187

* p-value of less than 0.05

Table 14 Results of pairwise t-test by productivity metric category for PMV

Productivity Metric Category	Accuracy	Neurobehavioral Response Time	Neurobehavioral Speed
Neurobehavioral Response Time	0.006**		
Neurobehavioral Speed	0.218	0.457	
Performance Score	0.386	0.341	0.341

** p-value of less than 0.01;

Table 15 Summary of results for the three methods of comparison by productivity metric category

Productivity Metric Category	Analysis	Statistical Summary	Graphical Comparison	Pairwise t-Test
Accuracy	1 (all IEQ)	Different	Different	Different (p=0.001**)
	2 (CO_2)	Different	Somewhat different	Somewhat different (p=0.014*)
	2 (Vent.)	Different	Somewhat different	Somewhat different (p=0.042*)
	2 (Thermal)	Different	Somewhat different	Similar (p=0.242/0.218)
Call Handling Time	1 (all IEQ)	Somewhat different	Somewhat similar	Similar (p=0.153)
	2 (CO_2)	Low sample size		
	2 (Vent.)	Low sample size		
	2 (Thermal)	Somewhat similar	Somewhat different	Similar (p=0.871)
Neurobehavioral Response Time	1 (all IEQ)	Very similar	Very similar	Similar (p=0.467)
	2 (CO_2)	Low sample size		
	2 (Vent.)	Low sample size		
	2 (Thermal)	Somewhat different	Similar	Similar (p=0.255/0.457)
Neurobehavioral Speed		Baseline for comparison		
Performance Score	1 (all IEQ)	Very different	Somewhat different	Different (p=0.007**)
	2 (CO_2)	Very different	Different	Different (p=0.002**)
	2 (Vent.)	Very different	Very different	Different (p=0.003**)
	2 (Thermal)	Different	Similar	Similar (p=0.421/0.341)
Self-Reported Productivity	1 (all IEQ)	Somewhat similar	Somewhat similar	Similar (p=0.421)
	2 (CO_2)	No data		
	2 (Vent.)	No data		
	2 (Thermal)	Somewhat different	Somewhat similar	Similar (p=0.871)

Significant differences: * p-value of less than 0.05 (somewhat different); ** p-value of less than 0.01 (different); *** p-value of less than 0.001 (very different)