



A Journey Through SPACE

Unpacking the Perceived Productivity of GitHub Copilot

Viggo Tellefsen Wivestad¹(✉) and Rasmus Ulf snes^{1,2}

¹ SINTEF Digital, 7034 Trondheim, Norway
viggo.wivestad@sintef.no

² Norwegian University of Science and Technology, 7491 Trondheim, Norway

Abstract. This study examines the influence of perceived changes in productivity in the context of introducing AI Coding Assistants, specifically GitHub Copilot, within two large-scale agile organizations. Using a cross-sectional survey, we measured self-reported changes in productivity using the SPACE framework. Our comparative analysis suggests several perceived benefits of AI Code Assistant adoption, though with a more conservative impact than previously reported. Further, a correlational analysis employing Kendall's tau and PLSR with 10-fold cross-validation suggests that perceived changes to productivity are moderately associated only with four of the nine features tested, namely job satisfaction, flow, task completion speed, and ability to focus on satisfying work. However, the SPACE framework's ability to fully capture perceived productivity was further challenged, indicating discrepancies in its dimensions of “Performance” and “Communication and collaboration”.

Keywords: Developer Productivity · AI Code Assistants · SPACE · GitHub Copilot

1 Introduction

In the evolving landscape of software development, Artificial Intelligence (AI) tools are becoming increasingly prevalent. GitHub Copilot, an AI Coding Assistant, has emerged as a notable example of this trend [9], aiding developers with coding tasks in real-time [14].

Recent studies have looked at the effects of using generative AI in software development practices [18]. One notable study was released by GitHub Research, where participants were asked to self-report various perceived benefits using a 5-point Likert agreement scale. This large-scale survey indicated a solid increase in the perceived productivity by users of Copilot [7]. However, an agreement scale is susceptible to overestimating the agreement ratio [10], which motivated this study to replicate their findings instead of measuring *perceived change* rather than agreement toward statements.

While productivity itself and its driving factors are studied and disputed concepts in software development [2, 5, 13, 19], the SPACE framework represents

one attempt to unpack this using five so-called dimensions. This framework is already being used in research, with one notable example being Ziegler et al., who explored whether measuring developers' interactions with GitHub Copilot could predict their self-reported productivity according to SPACE [21]. However, SPACE itself lacks empirical support, and the assumption that it provides a good proxy for productivity is still an open question.

We therefore set out to answer the following two research questions:

1. *Are the perceived benefits of adopting GitHub Copilot altered by transforming the Likert scales from an agreement scale to a change scale?*
2. *Which factors of SPACE influence perceived changes in productivity when adopting an AI Coding Assistant?*

To answer these and address the lack of empirical data on potential productivity gains from adopting AI Coding Assistants, we employed a cross-sectional survey involving two large-scale agile organizations. The survey extends the survey framework developed by GitHub Research [7].

2 Background

Developers and agile organizations have long been interested in how different modes of organizations, teams, and technology can boost developer productivity. In the new age of AI Coding Assistants, some studies proclaim multiple benefits for adopting such technology, [7], while others find that the main difference between users and non-users is mainly related to work satisfaction dependence on colleagues, not productivity [20].

Multiple attempts have been made to develop frameworks for measuring productivity [19], with an ongoing debate related to how to best measure it, using either objective measurements (e.g. lines of code or time spent writing code [2]) or subjective measurements (e.g. self-ratings or peer evaluations) [13]. To consolidate these perspectives, Forsgren et al. proposed the SPACE framework [5], which attempts to describe productivity holistically using five distinct dimensions of the developer's work life.

GitHub Research used this framework to guide their study when assessing potential productivity gains for their own product, GitHub Copilot [7]. SPACE itself is an acronym for its five dimensions: *Satisfaction and well-being* refers to the developer's satisfaction with work, work-life balance, and general happiness. Studies on hackathons have e.g. shown that having fun at work can influence productivity [12]. Emotional states in general are also shown to affect the perceived productivity of the developers [6]. *Performance* relates to outcomes of a system or process, mainly focusing on quality and impact. *Activity* typically consists of more objective measures, like lines-of-code, number of completed actions or outputs during work, *Communication and collaboration* covers communication, coordination, and collaboration within and between teams. Such characteristics are recognized as particularly important for large-scale inter-team coordination. [3, 17], especially in the post-pandemic world, with hybrid teams becoming being

normalized [16]. The last dimension, *Efficiency and Flow*, focuses on the ability to work focused and uninterrupted. This resonates with previous findings where unplanned meetings and interruptions were found to be the main detriments to unproductive work [11].

3 Method and Study Design

Our research design consisted of a cross-sectional survey, with a target population consisting of software developers with access to GitHub Copilot. Participants were selected from two companies, NAV IT and MarComp (pseudonym). While both are large-scale agile software organizations, NAV IT is a national organization belonging to the public sector in Norway and employs around 500 developers, while MarComp is an international company with developers in Norway, Poland, and India, with a developer headcount of around 80.

We collected a total of 120 responses (70 from NAV IT, 50 from MarComp), consisting of 84% males, 13% females, and 4% who did not wish to reply. 73% of the developers were in-house, and 27% were external consultants. The complete survey instrument is available online at: <https://doi.org/10.5281/zenodo.10987170>.

The survey instrument extends a subset of GitHub Research's survey [7, 22], which utilizes the SPACE framework [5] to assess perceived productivity. However, the original questionnaire utilized a 5-point Likert agreement scale, which is susceptible to overestimating the agreement ratio [10]. Several questions also posed analytical challenges, with statements like "I am more productive when using Copilot" being somewhat ambiguous to interpret (does "strongly disagree" indicate an explicit reduction in productivity, or no change, i.e. disagreeing with the statement that productivity has increased?). To combat this, we transformed the survey items into a 5-point bidirectional Likert evaluation scale which we call a "change scale", ranging from "Major increase" to "Major decrease", with "No change" being a neutral center point. This transformation required slight modifications to survey item formulations, e.g. "Since getting access to Copilot, have you noticed a change in the following: Your own productivity?".

The final questionnaire consisted of 57 questions, with 10 questions utilizing the transformed change scale. For this study, we are focusing on the questions intersecting GitHub Research's survey. The survey items and their mapping to the SPACE framework can be found in Table 1.

3.1 Comparing Likert Scales Results

In our study, we employed a Top 2-Box (T2B) analysis to compare the outcomes from GitHub Research's study [7] with our survey results. This method concentrates on the proportion of a subset of the Likert scale, e.g. agreement ("strongly agree" + "agree") from GitHub Research's study and increase ("major increase" + "minor increase") from our survey. While this approach omits parts of the response distribution and thus provides an incomplete view of the data, it

is recognized for its utility in highlighting areas of interest within data sets [15]. Finally, since this was the only data provided by [7], it was our only choice for comparison.

To examine potential relationships between perceived productivity and the SPACE-based variables, and compensate for the incomplete view given by the Top 2-Box analysis, we employed a bivariate histogram to inspect the distribution between pairs of independent variables and the dependent variable. To test our hypotheses that there exist relationships between the SPACE-based variables and perceived productivity, Kendall's τ coefficient was utilized to examine both strength and relational direction. This non-parametric method is suitable for ordinal data, which is inherent in Likert scale responses. To correct for multiple hypothesis testing, our threshold for statistical significance was reduced using Bonferroni correction.

Finally, to explain the independent variables' impact on perceived productivity, we treated the data as interval scales and employed Partial Least Squares Regression (PLSR) with 10-fold cross-validation. The model's robustness and predictive capabilities were validated using the coefficient of determination (R^2) and Root Mean Squared Error (RMSE) metrics. The aggregated Variable Importance in Projection (VIP) scores statistics quantifies the contribution of each independent variable, with scores above 1 generally indicating a significant influence. Uncertainty is reported using standard deviation.

4 Results

Our study surveyed 120 GitHub Copilot users about their perceived change in productivity, using a survey instrument where one item directly measured perceived productivity and nine items assessed changes according to SPACE.

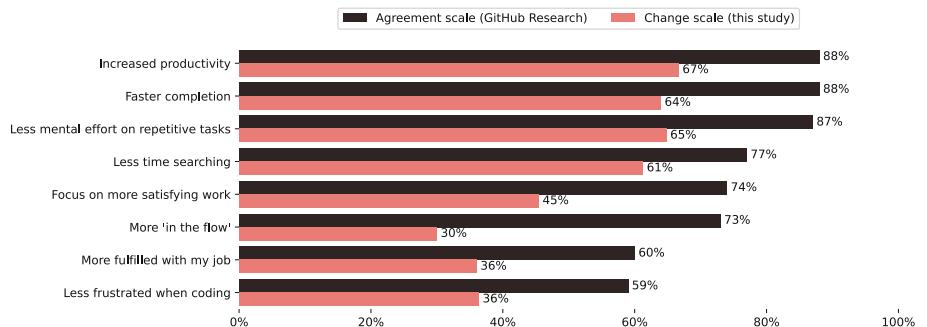


Fig. 1. Top 2-Box plot comparing the proportion of positive agreement vs. increased change related to various attitudes among Copilot-users (the word “less” indicates that the bars focus on disagreement/decrease rather than agreement/increase)

Comparative Analysis (Agreement vs. Change Scale): Figure 1 compares our change scale results with GitHub Research's agreement scale findings.

Table 1. Results from testing associations between SPACE-based variables and ‘perceived productivity’. Independent correlations are tested using Kendall’s τ rank correlation with associated P-values for significance with a Bonferroni threshold of $\alpha = 0.0056$.

SPACE Dimensions	Survey Items (“Change in...”)	τ	P-value	Accept H_A
Satisfaction and well-being	Job satisfaction	0.468	<0.001	✓
	Focus on satisfying work	0.453	<0.001	✓
	Frustration boring tasks	-0.221	0.014	
Performance	Code quality	0.256	0.005	
Efficiency and flow	Task completion speed	0.598	<0.001	✓
	Flow	0.452	<0.001	✓
	Searching for information	-0.256	0.003	
	Mental work	-0.377	<0.001	
Communication and collaboration	Dependence on colleagues	-0.279	0.002	

Although both scales recorded positive impacts, the change scale showed significantly more conservative outcomes, most notably for the ability to stay in the “flow”, where less than a third (30%) of our respondents reported an increase, compared to nearly three-quarters (73%) in GitHub research’s study. Our results further showed that about two-thirds of the users reported increased productivity (67%) and completion speed (64%), and decreased mental effort spent on repetitive tasks (65%) and time spent searching for information (61%). However, only about one-third report being more in the flow (30%), feeling more fulfilled with their job (36%), and feeling less frustrated when coding (36%). About half of the users experienced an increase in their ability to focus on satisfying work (45%).

Distribution and Correlation Analysis: Figure 2 presents the distribution of responses for all survey items, visualizing the relationships between perceived changes in productivity and the SPACE variables. We only accept the alternative hypothesis (H_A : there is a correlation) for cases that show *both* a significant correlation towards perceived productivity after Bonferroni correction, and show at least moderate correlation levels ($|\tau| \geq 0.4$) using Dancey and Reidy’s thresholds [1, 4]). As detailed in Table 1, we were able to accept the alternative hypothesis, and thereby confirm a moderate correlation towards perceived productivity, in four instances: (1) **task completion speed**, (2) **ability to stay in the flow**, (3) **job satisfaction** and (4) **the ability to focus on satisfying work**.

PLSR Model Performance and Feature Importance: The PLSR model, evaluated using 10-fold cross-validation, demonstrated modest predictive capability, with an R^2 of 0.214 ± 0.244 , indicating that the model explains 21.4% of the variance on average. An RMSE of 0.616 ± 0.070 suggests moderate predictive errors, with low RMSE scores normally indicating better predictive accuracy. However, the high standard deviation suggests substantial variability in R^2 , hinting at potential model instability. Nevertheless, the VIP scores in Fig. 3 align with the Kendall τ results, identifying the same four dimensions as the

most influential predictors of perceived changes in productivity, supporting their relevance in the SPACE framework.

5 Discussion and Implications

Our findings show that developers who adopt an AI Coding Assistant report notable increases in productivity, both when asked directly, and indirectly using SPACE variables. This resonates with the findings of GitHub Research [7], and other studies on Generative AI where developers report spending less time searching for information and being blocked by repetitive and boring tasks [18].

Using our proposed change scale, the proportion of users reporting increased benefits declined significantly across all aspects when compared to GitHub Research's findings [7]. This result was expected, as the change scale was designed to counteract potential false inflation of agreement. Most notable was the big drop in users who reported increased “flow”, which resonates with other work that found little difference in perceived “flow” when comparing users to non-users of GitHub Copilot [20]. It should be noted that these two Likert scales

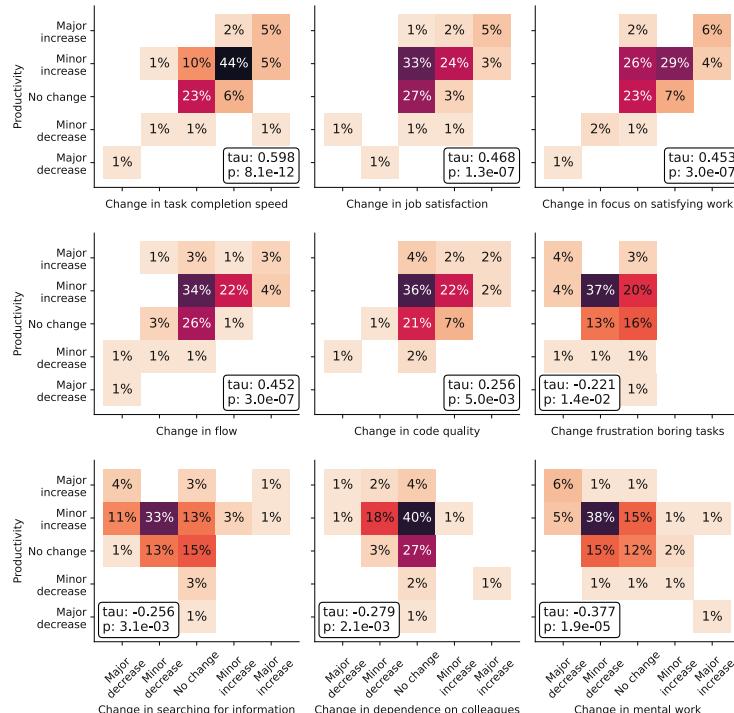


Fig. 2. Bivariate histograms showing the joint distribution of perceived change in productivity and the SPACE-based variables. Kenball's τ (**tau**) measures correlation, while P-value (**p**) measures statistical significance.

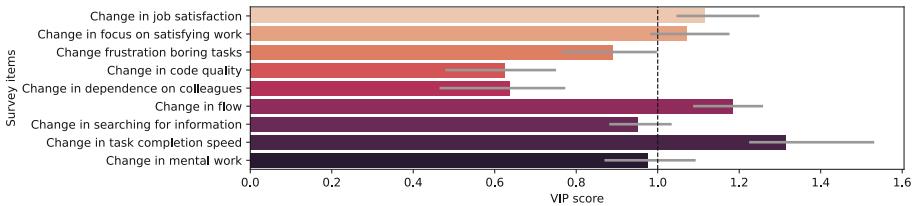


Fig. 3. VIP scores showing the contributions from each variable on the perceived change in productivity. The dotted vertical line marks the significance threshold at 1.0, and error bars show the min/max values from the 10-fold cross-validation

measure slightly different concepts and are not directly comparable, providing only an imprecise comparison. Furthermore, our target population is limited, and our findings might therefore not generalize to a global population.

When looking closer at which factors influence perceived changes in productivity, our study reveals both correlations and discrepancies between developers' perceived productivity and the SPACE dimensions. Both "Satisfaction and well-being" and "Efficiency and flow" contained facets that showed moderate, positive correlation and impact towards perceived productivity. These results confirm previous findings on how different states of mind and contexts might affect perceived productivity [2, 8, 11]. The dimensions "Performance" and "Communication and collaboration" were not found to be important, though it should be noted that "Performance" was measured with only a single variable.

Overall, our findings suggest that the dimensions of SPACE either do not fully describe productivity, or that productivity, as defined by SPACE, does not align with developers' definition of productivity. This echoes similar discrepancies found by Beller et al., which indicates that perceived productivity does not always align with the type of productivity managers are interested in [2].

Our study provides the following implications for practice

- Utilizing AI Coding Assistants is associated with perceived benefits, most notably increased levels of productivity, efficiency and flow, and satisfaction and well-being, though more conservative than those proclaimed by GitHub Research.
- Some dimensions of SPACE seem more influential in changing perceived productivity. Facilitating "Efficiency and flow" and "Satisfaction and well-being" therefore seems more impactful than "Performance" and "Communication and collaboration", if the goal is to increase perceived productivity.

As for research implication, our results indicate that the SPACE framework might be incomplete, misaligned, or inappropriate in terms of capturing perceived productivity, with two of the SPACE dimensions showing little correlation with how developers define productivity. To extend this study, future work should aim for greater coverage of each SPACE dimension, with a special focus on including the "Activity" dimension, when assessing their influence on productivity.

References

1. Akoglu, H.: User's guide to correlation coefficients. *Turk. J. Emerg. Med.* **18**(3), 91 (2018). <https://doi.org/10.1016/j.tjem.2018.08.001>, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6107969/>
2. Beller, M., Orgovan, V., Buja, S., Zimmermann, T.: Mind the gap: on the relationship between automatically measured and self-reported productivity. *IEEE Software* **38**(5), 24–31 (Sep 2021). <https://doi.org/10.1109/MS.2020.3048200>, <https://ieeexplore.ieee.org/abstract/document/9311217/references#references>
3. Berntzen, M., Hoda, R., Moe, N.B., Stray, V.: A taxonomy of inter-team coordination mechanisms in large-scale agile. *IEEE Trans. Softw. Eng.* **49**(2), 699–718 (2022). <https://ieeexplore.ieee.org/abstract/document/9739868/>
4. Dancey, C.P., Reidy, J.: Statistics Without Maths for Psychology: Using SPSS for Windows, 3rd edn. Pearson Prentice Hall, Harlow, Munich (2006). [repr.] edn
5. Forsgren, N., Storey, M.A., Maddila, C., Zimmermann, T., Houck, B., Butler, J.: The SPACE of developer productivity. *Queue* **19**(1), 20–48 (2021). <https://doi.org/10.1145/3454122.3454124>
6. Girardi, D., Lanubile, F., Novielli, N., Serebrenik, A.: Emotions and perceived productivity of software developers at the workplace. *IEEE Trans. Softw. Eng.* **48**(9), 3326–3341 (2022). <https://doi.org/10.1109/TSE.2021.3087906>, <https://ieeexplore.ieee.org/abstract/document/9449979>
7. Kalliamvakou, E.: Research: quantifying GitHub Copilot's impact on developer productivity and happiness (2022). <https://github.blog/2022-09-07-research-quantifying-github-copilots-impact-on-developer-productivity-and-happiness/>
8. Kim, Y.H., Choe, E.K., Lee, B., Seo, J.: Understanding personal productivity: how knowledge workers define, evaluate, and reflect on their productivity. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–12. ACM, Glasgow (2019). <https://doi.org/10.1145/3290605.3300845>, <https://dl.acm.org/doi/10.1145/3290605.3300845>
9. Liang, J.T., Yang, C., Myers, B.A.: A large-scale survey on the usability of AI programming assistants: successes and challenges. In: Proceedings of the IEEE/ACM 46th International Conference on Software Engineering, pp. 1–13. ICSE 2024. Association for Computing Machinery, New York (2024). <https://doi.org/10.1145/3597503.3608128>, <https://dl.acm.org/doi/10.1145/3597503.3608128>
10. Lohr, S.L.: Sampling: Design and Analysis. Chapman & Hall/CRC Texts in Statistical Science, 3rd edn. CRC Press, Taylor & Francis Group, Boca Raton London New York (2022)
11. Meyer, A.N., Barr, E.T., Bird, C., Zimmermann, T.: Today was a good day: the daily life of software developers. *IEEE Trans. Softw. Eng.* **47**(5), 863–880 (2019)
12. Moe, N.B., Ulfsnes, R., Stray, V., Smite, D.: Improving productivity through corporate hackathons: a multiple case study of two large-scale agile organizations. In: Proceedings of the 55th Annual Hawaii International Conference on System Sciences (HICSS) (2021)
13. Murphy-Hill, E., et al.: What predicts software developers' productivity? *IEEE Trans. Softw. Eng.* **47**(3), 582–594 (2021). <https://doi.org/10.1109/TSE.2019.2900308>, <https://ieeexplore.ieee.org/abstract/document/8643844>
14. Ross, S.I., Martinez, F., Houde, S., Muller, M., Weisz, J.D.: The programmer's assistant: conversational interaction with a large language model for software development. In: Proceedings of the 28th International Conference on Intelligent User Interfaces, pp. 491–514. ACM, Sydney (2023). <https://doi.org/10.1145/3581641.3584037>

15. Shull, F., Singer, J., Sjøberg, D.I.K. (eds.): Guide to advanced empirical software engineering. Springer, London (2008). <https://doi.org/10.1007/978-1-84800-044-5>
16. Smite, D., Moe, N.B., Hildrum, J., Huerta, J.G., Mendez, D.: Work-from-home is here to stay: call for flexibility in post-pandemic work policies. *J. Syst. Softw.* **195**, 111552 (2023). <https://doi.org/10.1016/j.jss.2022.111552>
17. Ulfnes, R., Berntzen, M., Moe, N.B., Sporsem, T., Stray, V.: Exploring the organizational models for data science in agile software development: challenges and strategies from a multi-case study. In: Proceeding of the 57th Hawaii International Conference on System Sciences (HICSS 2024) (2024)
18. Ulfnes, R., Moe, N.B., Stray, V., Skarpen, M.: Transforming software development with generative AI: empirical insights on collaboration and workflow, pp. 219–234. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-55642-5_10
19. Wagner, S., Deissenboeck, F.: Defining productivity in software engineering. In: Sadowski, C., Zimmermann, T. (eds.) Rethinking Productivity in Software Engineering, pp. 29–38. Apress, Berkeley (2019). https://doi.org/10.1007/978-1-4842-4221-6_4
20. Wivestad, V., Barbala, A., Stray, V.: Copilot's island of joy: balancing individual satisfaction with team interaction in agile development. In: Agile Processes in Software Engineering and Extreme Programming - Workshops. 25th International Conference on Agile Software Development, XP 2024, Bolzano, Italy, 4–7 June 2024, p. 7. Springer, Cham (2024)
21. Ziegler, A., et al.: Productivity assessment of neural code completion. In: Proceedings of the 6th ACM SIGPLAN International Symposium on Machine Programming, MAPS 2022, pp. 21–29. Association for Computing Machinery, New York (2022). <https://doi.org/10.1145/3520312.3534864>, <https://dl.acm.org/doi/10.1145/3520312.3534864>
22. Ziegler, A., et al.: Measuring GitHub copilot's impact on productivity. *Commun. ACM* **67**(3), 54–63 (2024). <https://doi.org/10.1145/3633453>, <https://dl.acm.org/doi/10.1145/3633453>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

