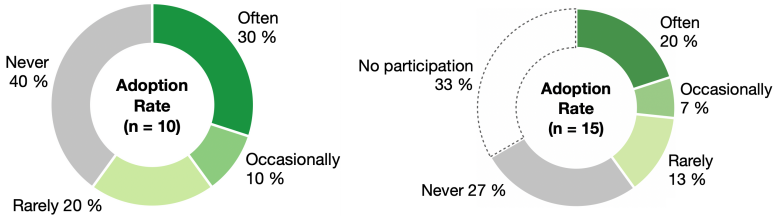
Revisiting PAVED: Studying Tool Adoption After Four Years

L. Cibulski[1,2] and T. May[2]

[1]Technical University of Darmstadt, Germany
[2]Fraunhofer Institute for Computer Graphics Research IGD, Germany



(a) *Study participants* **(b)** *Entire target group*

Figure 1: RQ_1 *Adoption Rate: (a) 60% of participants reported self-initiated use. (b) Considering all potential target users, this rate is 40%.*

Abstract

Design studies create visualizations that provide lasting solutions to real-world problems. Yet, they rarely validate this goal. Validation of domain usefulness typically stops shortly after the end of a project. Following up on the long-term acceptance, however, can provide important indications of how well a tool addresses the true needs of target users. For an existing decision support tool, we close this gap by revisiting its adoption in the target domain after four years. Our survey reveals a small number of power-users and helps carve out factors that influence whether and how a tool is adopted in the intended work environment.

CCS Concepts

• ***Human-centered computing*** → ***Empirical studies in visualization;*** • ***Applied computing*** → *Engineering;*

1. Introduction

In a previous paper, we presented a design study on PAVED, an interactive parallel coordinates visualization to support cost-benefit trade-offs in decision-making [CMMK20]. We confirmed its short-term domain usefulness in a field study with eight domain experts. After the design study, our primary domain experts decided to integrate PAVED into their daily work flows. However, the decision was made during a period of intense collaboration and tool use likely decreases after such a period [BISM14]. Evidence of short-term usefulness might thus not generalize to long-term routine use. After four years, we therefore ask: is PAVED still being used in the experts' daily work? An adoption in the sense of repeated, self-initiated use for the engineers' daily activities (as opposed to requested use in the summative evaluation) would provide strong evidence that PAVED in fact addresses the true needs of the target users. In our target domain, this might be particularly meaningful, because motor designers are free to choose any tool they consider useful. A review of related works revealed that, despite the relevance of adoption, few visualization (design study) papers follow up on the long-term usage of their proposed tools.

We close this gap by revisiting the usefulness of our decision support tool in the targeted domain after four years. The results help develop an understanding of factors that influence whether and how a tool is adopted in the intended work environment, which can be a starting point for organizing visual designs and collaborations accordingly. The primary contributions of this work are:

- The first long-term study of decision support that investigates the self-initiated use of PAVED in the target domain after four years.
- A reflection on the results showing that the target users in fact have the characterized problem and that PAVED solves it.

2. Aim of the Study

Our objective is to assess the day-to-day usage of PAVED and derive insights about its long-term benefits for the engineers' tasks. Of the adopters, we want to know for which tasks they use PAVED, what functionality is (not) useful, what challenges they face, how they rate its usability, and how well the tool blends with existing work flows. Of the non-adopters, we want to learn about the reasons. Our research questions can be summarized as follows:

DOI: 10.2312/evs.20241067

RQ$_1$ Adoption Rate – What portion of target users did adopt PAVED for regular use in their daily work? Who are they?
RQ$_2$ Usage Context – What are the circumstances, under which PAVED is still used in the application domain?
RQ$_3$ Usability – How does the perceived usability compare to that four years ago?
RQ$_4$ Reasons for Refusal – Why did some users not adopt it?

3. Related Work

Munzner recommends to observe the adoption and long-term use of a visualization design as a downstream validation to the domain and abstraction threats [Mun09]. It helps to assess factors that influence how a tool is adopted in the intended work environment [LBI*11], e.g., whether the barrier to adoption is an integration issue or an indication that the tool failed to address the true needs of the target users [BISM14]. Most design studies do not follow up on that.

One exception is a study to clarify the true adoption of a tool after empirical studies had shown promising results [GK03]. It revealed that domain experts gradually lost interest in using the tool as part of their current work routines, which the authors attributed to a potential misunderstanding of the experts' work flow. Similarly, Kang and Stasko complemented an earlier lab study by interviewing analysts who they knew had used their tool on their own initiative for two to 14 months [KS12]. Studying the adoption of *Overview* over a period of two years helped Brehmer et al. refine their understanding of why and how domain experts used their tool, ultimately leading to revised task abstractions and design rationales [BISM14]. Kincaid et al. collected user feedback from research labs that had employed their visualization tool for scientific studies, one of them for over a year [KBDY05]. McKeon observed the activities on a public deployment of their wiki-like visualization dashboard system over half a year [McK09]. While we can gain methodological insights from these works, they do not study decision support tools.

Longitudinal field studies have also investigated insight generation [SNLD06], task abstractions [MMKN08], the integration of statistics with visualization [PS08], and early stages of geovisualization design [LD11]. Guidelines help evaluate the prolonged use of visualizations over weeks or months [SP06]. Except for the *Overview* tool [BISM14], studies involve little reflection on how feedback evolves in response to changes made to the tool across different post-deployment stages. Most post-deployment studies also suffer from survivorship bias, i.e., they solely focus on participants who have used the tool [BISM14,KBDY05,McK09]. We also cover unsuccessful cases in which experts lost interest in using a tool.

To summarize, visualization tools are typically evaluated with respect to their short-term usage only. Works on long-term post-deployment evaluation are remarkably rare [Mun09, BISM14, SP06], given that lasting solutions to domain problems are a core goal of design studies. Finally, we identified a lack of studies that investigate the adoption of decision support tools.

4. Study Methodology

The target users of PAVED are mechatronic engineers who are involved in the design of electric drives using their in-house optimization tool *SyMSpace* [SKW*18]. Our primary domain expert named 15 engineers to whom this applies. Some of them had already participated in the evaluation four years ago [CMMK20]. We collected the engineers' experience with PAVED based on their voluntary use since the deployment four years ago. Using an anonymous online survey, we collected data for a qualitative analysis, which we augmented with a quantitative usability scale where applicable. Contrary to the previous evaluation, the researcher did not engage in an observation. Without any further announcement, we invited the target users via e-mail. By inviting all target users regardless of their usage behavior, we avoided survivorship bias.

The aim of our study was purely academic. To minimize the refusal rate, it was thus important to develop a time-effective yet meaningful questionnaire. We drew inspiration from sample questions provided by other visualization researchers to assess the adoption of a visualization. These included 1) the intended practices and (long-term) experience with a visualization [LBI*11], 2) the usage purpose, comparison to traditional methods, (not) useful and missing features, and barriers [KS12], and 3) who adopted the tool, how it had been used, whether it is still in use, and what problems users reported [BISM14]. On these sample questions, we performed a thematic analysis to collect aspects to cover in our questionnaire:

- **Demography** – What responsibilities the target users have and for how long they have been employed in their job role (**RQ$_1$**)
- **Usage Behavior** – For what daily activities and data PAVED has been used, whether it is still in use, and how important and recommendable it is for the daily work (**RQ$_2$**)
- **Features** – What functionalities did (not) work well or were missing, what challenges were encountered, and how PAVED compares to existing methods (**RQ$_2$**)
- **Integration** – How well PAVED blended into existing work flows in terms of access, data handover, and functionality coordination (**RQ$_2$**)
- **Barriers** – What limitations prevented an adoption (**RQ$_4$**)

To assess how the perceived usability has evolved in the past four years (**RQ$_3$**), we additionally included **System Usability** as a theme. We replicated the use of the System Usability Scale (SUS) [Sau11] from our design study [CMMK20].

We started the questionnaire by asking the participants whether they were aware of the possibility to use PAVED before taking part in the questionnaire and, if yes, how regularly they used PAVED. Only those who reported to have used PAVED at least rarely were asked in detail about their experience by working through the above themes. Figure 2 depicts how participants were directed along different branches in the questionnaire depending on their answers. In total, we ended up with 31 questions, of which 21 were closed-ended (yes/no and rating scales). Assuming an average of 25 seconds completion time per question [LKK16], this matches our ambition to require a maximum of 10 to 15 minutes time commitment.

We validated the questionnaire in a preliminary study with five research fellows. The comments addressed the clarification of intentions, the scope of questions, and the layout (e.g., placement of consent or labeling of rating scales). Subject to a detailed discussion has been whether participants should be informed that a certain answer will end the questionnaire and whether questions should contain examples of what kind of answer is expected. The final questionnaire can be found in the supplemental material.

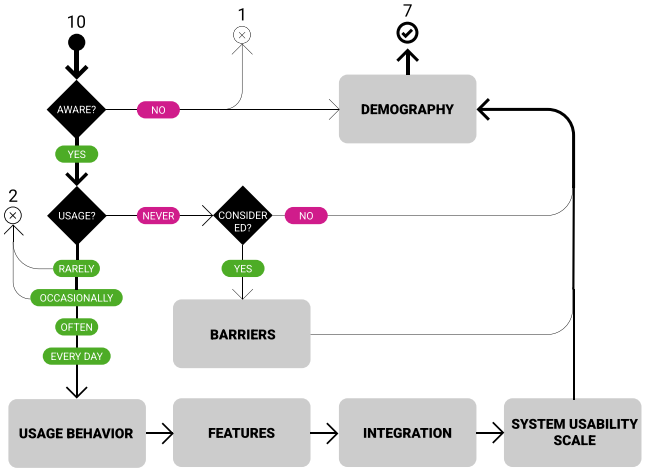


Figure 2: *Flowchart of how participants are directed along different routes. Four adopters traversed the outermost loop on details of their usage. Three participants took a shorter route because they had not used PAVED. Another three dropped out in the middle ⊗.*

5. Results

Of the 15 target users, 10 participated in our study. This is above the average response rate of 44.1% [WZFA22]. The completion times ranged from 00:30 to 5:00 min. for the short routes and from 10:30 over 22:30 to 51:30 min. for answering the detailed theme blocks.

RQ_1 Adoption Rate We define adoption as *"Did a [user] freely choose the tool for their own investigation, rather than trying out the tool in response to direct solicitation by the researchers?"* [BISM14]. Adoption occurred in six of ten target users, who tried out PAVED on their own initiative (Figure 1a). The other participants' reasons for not using PAVED will be presented later.

Of the six adopters, three are still using PAVED for their work today. Two of the six left the questionnaire immediately after they stated how often they used PAVED such that we do not know about their on-going use. Finally, one reported to not currently use the tool for the daily work. This might be attributed to the participant's responsibilities as a manager. In this position, the role of PAVED is to demonstrate the high-level potential of optimization software in customer discussions rather than being used for actual projects.

Five, or one third, of the invited target users did not participate in the study, such that we cannot make reliable statements about their adoption of PAVED (Figure 1b). Still, we know that at least one of the five has used PAVED one and a half years after the deployment based on questions and feedback we received via e-mail.

RQ_2 Usage Context Those four adopters who did not drop out of the questionnaire answered the detailed questions regarding usage behavior, features, and integration. Three reported to have used PAVED often in their daily work, while the other one used it rarely. We characterize the usage context of PAVED based on the participants' ratings of the closed-ended questions (see supplemental material) and a coding of their responses to the open-ended questions.

PAVED is perceived as a *"very useful tool, especially for presentations"* (A3). In line with the task it was designed for [CMMK20], PAVED was mainly used for exploring motor designs and choosing an optimal one (A1, A2, A3). An integral part of this was to analyze the relations between the different properties (A2) and to communicate what performance is achievable under which conditions to the customer (A3). Revisiting decisions with customers was already identified as a hot topic in the first evaluation [CMMK20]. Rather than for actual motor design projects, one participant used PAVED with a demo data set in customer meetings to showcase the general potential of visualization for the design of electric motors (A4). The other three participants, who still use PAVED today, expressed that the tool is very important for their daily work. All four participants would recommend the tool to a new colleague in a similar job role. This aligns with the high usability we observed, where tools with a score above 80.3 are more likely to be recommended [Sau11].

PAVED's primary view, the parallel coordinates visualization, was mentioned as most useful (A4), especially for exploring the dependencies and conflicts between attributes (A2). In line with identified requirements for value retrieval and decision transparency [CMMK20], the adopters agreed that the parallel coordinates provide a *"very fast overview of an optimization"* (A3). One participant appreciated the tabular view of the options' raw attribute values (A2), confirming the relevance of tabular visualizations for decision-making [DBD17]. The participants proposed to extend the parallel coordinates to view additional result sets (A3), e.g., to compare multiple Pareto fronts (A1). Being able to show and hide axes was perceived as particularly helpful, followed by the suggestion of a semantic grouping of attributes to facilitate multi-selection (A2). In line with our view that interaction is essential for an effective use of parallel coordinates [CMMK20], the participants also highlighted the dynamic filtering as particularly useful (A2, A4). Contrary to the first evaluation [CMMK20], our preference brushes for optimization criteria did not provoke comprehensibility issues this time. Rather, one participant found the restriction of design parameters *"not very intuitive"* (A1). This conflicts with our learning from the first evaluation that the target users were familiar with standard range brushes. Overall, the participants rated PAVED as *"somewhat more"* helpful for their work than traditional approaches.

Integration issues between PAVED and the in-house software are not likely a reason for failed adoption. The responses indicate that PAVED is used in a well-functioning symbiosis with their domain software. The participants found the PAVED interface *"easy"* to open from within SyMSpace and were *"very satisfied"* with the handover of data and selections between both tools. They also agreed that PAVED and SyMSpace are *"complementing"* each other in terms of functionalities. The strongest disagreement occurred on how well PAVED blends into the SyMSpace work flows, where one participant said *"poorly"* (A1) and two said *"very well"* (A2, A3).

RQ_3 Usability The same four adopters filled out the quantitative System Usability Scale. While we hoped to confirm the previously high usability, we hypothesized that the self-initiated use might reveal issues that did not occur in the observed walk-through four years ago. With a score of 86.9 out of 100 (Table 1, n = 4), the self-initiated and undirected use of PAVED during the adoption phase scored similarly to the requested and prescribed use directly after the tool's deployment, which received a score of 89.4 (n = 8). This

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Total
A1	10	7.5	7.5	7.5	7.5	10	7.5	10	7.5	10	85
A2	7.5	7.5	7.5	7.5	7.5	10	7.5	10	7.5	7.5	80
A3	10	10	10	10	10	10	10	10	10	10	100
A4	7.5	10	5	7.5	7.5	10	10	10	7.5	7.5	82.5
Avg 2023	8.8	8.8	**7.5**	8.1	8.1	**10**	8.8	**10**	8.1	8.8	**86.9**
Avg 2019	**7.8**	9.4	9.7	8.8	8.1	8.1	9.4	**10**	8.4	9.7	**89.4**

Table 1: RQ$_3$ *Usability: results of the long-term System Usability Scale [Sau11]. The tool achieved a total score of 86.9 out of 100 (n = 4) compared to 89.4 (n = 8) four years ago [CMMK20].*

indicates that the perceived usability of PAVED does not depend on the setting, in which it was obtained [GBKP13]. Given the generally high level of scores received, the differences to the previous SUS assessment remain comparatively small. The previous interest in frequent use (Q1, 7.8 out of 10) was confirmed by the four adopters who have indeed used PAVED sometimes or often. The interest increased to a strong agreement of 8.8 out of 10 among those who adopted PAVED. The ratings for tool complexity (Q2) and learnability for oneself (Q4 and Q10) and others (Q7) [LS09] have slightly decreased. In contrast, the consistency rating (Q6) raised from 8.1 out of 10 in the first evaluation to the highest possible score of 10 in the present case. This might be attributed to changes that we made in response to the initial feedback.

RQ$_4$ Reasons for Refusal Two of the four non-adopters reported that they had not known of the existence of PAVED. One of them did not use the in-house tool where PAVED is integrated. Technically, this participant does not belong to our group of target users. The other unaware participant dropped out of the questionnaire immediately after saying so, such that we cannot learn about the reasons. The remaining two participants reported that they were aware of PAVED but did not use it. One of them stated to not have considered its use at all. The other participant, a senior researcher with 18 years of experience, stated to have decided against using PAVED. The barrier in this case was an integration issue. The participant said that it was unclear how to use PAVED within the in-house optimization tool. To some extent, this contradicts our finding above where integration issues were not reported among adopters.

6. Discussion

The adopters applied the tool in the way we envisioned it to be used, suggesting that our domain characterization successfully informed the visual design. Allowing target users to adjust to a new visualization support in their work environment [HN15] before assessing its usefulness is advisable. In fact, we argue that studying the self-initiated use of a proposed visualization should be considered as a source of information in design studies whenever applicable.

The low number of study participants is due to a limited set of 15 target users that was given in the context of our project collaboration. While this helped mitigate a selection bias that is inherent to many case studies [BISM14], we could only analyze the responses of ten participants due to a refusal rate of 33%. Consequently, the absolute number of known adopters at this point remains fairly low. In principle, PAVED is generalizable to decision tasks outside of motor design [CMMK20]. Further adoption studies could indicate to what extent the data and task abstractions also apply to other application scenarios. This requires promotion of PAVED as a tool for making cost-benefit trade-offs within the respective communities.

A limitation of our study is the choice of an indirect survey over observational methods as used in previous works [BISM14, KS12]. With a moderate number of target users, direct methods are generally feasible and might have resulted in richer feedback. However, the domain experts' resources available for this follow-up evaluation are likely less compared to the resources available for the short-term evaluation conducted within the original project. Consequently, we needed to keep the time effort low. Independent of the research method, the comparability between our short-term and long-term evaluations is limited. While both are conducted on the users' daily decision tasks, the initial evaluation requested tool use whereas the adoption study involved voluntary, self-initiated use.

For a reliable adoption rate, it is important to only consider responses of target users as a reference. Despite an appropriate initial selection, we recommend to query all information needed to discriminate target users from other participants. As we invited supposed target users, we only discriminated participants based on their awareness and consideration of the possibility to use PAVED. We did not query whether they were target users at all. However, we noticed that this was not the case for all participants from their self-reports. Still, we computed the adoption rate based on all responses, potentially resulting in an underestimation. Similarly, the regularity of tool use needs to be considered in relation to the regularity of task occurrence. While rare tool usage might be a result of misled task abstraction, it might also be a result of the participant not facing a relevant task more often. We only verified this for reported non-usage and assumed daily confrontation with a relevant task elsewhere. Again, this might have resulted in underestimation.

7. Conclusion

Design studies rarely include an observation of their work's long-term usefulness in the target domain. In this work, we extend our previous design study on a decision support tool [CMMK20] by an adoption study after four years. It revealed a small number of power-users, who routinely use PAVED for choices about motor designs in their daily work and *"would not want to miss it"* (A2). These users also reconfirmed its high usability. They appreciate PAVED for what it was designed for: a fast overview of all options and attributes with reduced interaction mechanisms that help learn what performance is achievable under which conditions. This supports the design study idea of carefully designing or combining simple visual encodings rather than striving for novelty. One target user abandoned PAVED due to an integration issue. The responses of 10 participants suggest an adoption rate of 60% after four years.

We hypothesize that there is more to designing for permanent adoption than addressing the true needs of target users. Understanding how domain experts earn praise for their work and supporting them in being successful might be one strategy. In our study, one benefit was that PAVED intensified the communication with customers. Other factors might include involving gatekeepers in the collaboration, discussing integration possibilities from early on, or raising awareness for the tool through promotion activities.

References

[BISM14] BREHMER M., INGRAM S., STRAY J., MUNZNER T.: Overview: The design, adoption, and analysis of a visual document mining tool for investigative journalists. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (2014), 2271–2280. 1, 2, 3, 4

[CMMK20] CIBULSKI L., MITTERHOFER H., MAY T., KOHLHAMMER J.: Paved: Pareto front visualization for engineering design. In *Computer Graphics Forum* (2020), vol. 39, Wiley Online Library, pp. 405–416. 1, 2, 3, 4

[DBD17] DIMARA E., BEZERIANOS A., DRAGICEVIC P.: Conceptual and methodological issues in evaluating multidimensional visualizations for decision support. *IEEE Transactions on Visualization and Computer Graphics 24*, 1 (2017), 749–759. 3

[GBKP13] GRIER R. A., BANGOR A., KORTUM P., PERES S. C.: The system usability scale: Beyond standard usability testing. In *Proceedings of the human factors and ergonomics society annual meeting* (2013), vol. 57, SAGE Publications, pp. 187–191. 4

[GK03] GONZÁLEZ V., KOBSA A.: A workplace study of the adoption of information visualization systems. In *Proc. I-KNOW* (2003), vol. 3, pp. 92–102. 2

[HN15] HATRY H. P., NEWCOMER K. E.: Pitfalls in evaluations. *Handbook of Practical Program Evaluation* (2015), 701–724. 4

[KBDY05] KINCAID R., BEN-DOR A., YAKHINI Z.: Exploratory visualization of array-based comparative genomic hybridization. *Information Visualization 4*, 3 (2005), 176–190. 2

[KS12] KANG Y.-A., STASKO J.: Examining the use of a visual analytics system for sensemaking tasks: Case studies with domain experts. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2869–2878. 2, 4

[LBI*11] LAM H., BERTINI E., ISENBERG P., PLAISANT C., CARPENDALE S.: Empirical studies in information visualization: Seven scenarios. *IEEE Transactions on Visualization and Computer Graphics 18*, 9 (2011), 1520–1536. 2

[LD11] LLOYD D., DYKES J.: Human-centered approaches in geovisualization design: Investigating multiple methods through a long-term case study. *IEEE transactions on visualization and computer graphics 17*, 12 (2011), 2498–2507. 2

[LKK16] LEE S., KIM S.-H., KWON B. C.: Vlat: Development of a visualization literacy assessment test. *IEEE Transactions on Visualization and Computer Graphics 23*, 1 (2016), 551–560. 2

[LS09] LEWIS J. R., SAURO J.: The factor structure of the system usability scale. In *Proceedings of the Human–Computer Interaction International Conference* (2009), pp. 94–103. 4

[McK09] MCKEON M.: Harnessing the information ecosystem with wiki-based visualization dashboards. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 1081–1088. 2

[MMKN08] MCLACHLAN P., MUNZNER T., KOUTSOFIOS E., NORTH S.: Liverac: Interactive visual exploration of system management time-series data. In *Proceedings of the SIGCHI conference on human factors in computing systems* (2008), pp. 1483–1492. 2

[Mun09] MUNZNER T.: A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 921–928. 2

[PS08] PERER A., SHNEIDERMAN B.: Integrating statistics and visualization: Case studies of gaining clarity during exploratory data analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2008), pp. 265–274. 2

[Sau11] SAURO J.: Measuring usability with the system usability scale (sus). https://measuringu.com/sus, 2011. [Accessed 23 July 2023]. 2, 3, 4

[SKW*18] SILBER S., KOPPELSTÄTTER W., WEIDENHOLZER G., SEGON G., BRAMERDORFER G.: Reducing development time of electric machines with symspace. In *8th International Electric Drives Production Conference (EDPC)* (2018), IEEE, pp. 1–5. 2

[SNLD06] SARAIYA P., NORTH C., LAM V., DUCA K. A.: An insight-based longitudinal study of visual analytics. *IEEE Transactions on Visualization and Computer Graphics 12*, 6 (2006), 1511–1522. 2

[SP06] SHNEIDERMAN B., PLAISANT C.: Strategies for evaluating information visualization tools: Multi-dimensional in-depth long-term case studies. In *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization* (2006), pp. 1–7. 2

[WZFA22] WU M.-J., ZHAO K., FILS-AIME F.: Response rates of online surveys in published research: A meta-analysis. *Computers in Human Behavior Reports 7* (2022), 100206. 3