Environmental Research Communications

PAPER

Real-time feedback improves multi-stakeholder design for complex environmental systems

Niles Guo1, Alex Davis, Meagan Mauter and Jay Whitacre

1 Department of Engineering & Public Policy, Carnegie Mellon University, Pittsburgh, PA, United States of America
2 Department of Civil & Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA, United States of America
3 Department of Material Science & Engineering, Carnegie Mellon University, Pittsburgh, PA, United States of America

Author to whom any correspondence should be addressed.

© 2021 The Author

Keywords: group decision making, real time feedback, group design

Abstract

We test whether providing quantitative real-time feedback relating design decisions to system objectives improves group solutions in an interdependent energy-water design task. While prior research suggests an important role of real-time feedback on task performance, few studies have examined the role of real-time feedback in the design of complex environmental systems. We tested a real-time feedback approach using a mixed within- and between-subject experiment (n = 88 Carnegie Mellon University students, divided into 22 groups of four). When compared to individual designs and informal collaborations, real-time performance feedback yielded solutions closer to the Pareto frontier and reduced both financial (by 26% and 21%) and environmental cost (by 34% and 12%). In addition, informal collaboration did not improve group decision-making when compared to individual designs. The results suggest that optimal solutions to meeting energy and water demand while minimizing cost and environmental impact can be obscured in informal collaborations, but that real-time feedback to system designers can help avoid waste of public resources.

The design of complex systems involves coordination between multiple experts who have their own perspectives, motivations, objectives, and expertise [1, 2]. These heterogeneous objectives increase the difficulty of design tasks, risking bounded rationality, where individual experts ignore global complexities and instead rationally solve a focal subproblem [3], generating a local, but not global, solution [4]. We propose an approach called Concurrent Assessment and Design of Systems (CADS) that combines a concurrent design process with real-time feedback to facilitate design decisions, with the goal of helping decision-makers converge on a globally optimal design through structured interaction and collaboration [5–7]. In the present research, we experimentally examine the effect of one component of CADS, real-time performance feedback during group collaboration, to evaluate that component’s importance in an environmental engineering design tool.

Compared to teams that work in isolation on their subproblem and then try to come to agreement on a global solution, real-time feedback provides each team member with the team’s current total performance on a task [8]. This real-time feedback does not require another expert to evaluate and provide information to teams, as is typically done in after-the-fact performance evaluations [8]. Prior research finds that when decision-makers work in teams, biases such as groupthink, egocentrism, and social loafing can lead to suboptimal solutions [9–12]. While some findings have suggested that real-time feedback can reduce those biases, there is an open question about whether it is necessary for improving group design performance [13–15]. While one study found little benefit of real-time system-level feedback for students designing a complex interconnected system (a space mission), the small sample limited the study’s statistical power [16]. Other researchers have focused on either perceptual tasks (identifying visual cues), or simple decision-making tasks (selecting the right candidate for a hiring decision), as opposed to the design of complex interconnected systems [9, 13, 17–25]. Studies that evaluate the performance of design teams in large samples, as in the case of Gonzalez-Mule et al (2016), evaluated primarily the effect of ex-post feedback (self-reported performance reviews) rather than real-time feedback, and

© 2021 The Author(s). Published by IOP Publishing Ltd
also have lacked objective criteria for evaluating the team’s solution [18]. Although prior research suggests a strong role of real-time feedback on performance, the results may not generalize to the design of complex systems, where correctness is evaluated as efficient tradeoffs between different objectives [26].

We hypothesize that groups receiving real-time feedback will generate better solutions in a complex design task than participants working independently or through informal collaboration. Real-time feedback can improve group decision-making through several mechanisms. First, researchers have found that feedback clarifies the goals and interactions between subsystems for individuals within a team [17, 18]. In a survey of 110 defense industry manufacturing firms in South Korea, respondents indicated that feedback provided clarity on group goals and allowed individuals to understand their interactions with the overall system [18]. Teams that received feedback also gained a better understanding of the roles of other team members [20]. Second, feedback (when provided objectively and in a timely manner), allows team members to evaluate their assumptions about the system [19–23]. In a study of student groups making hiring decisions, teams that received feedback performed better than teams with no feedback, as feedback allowed team members to evaluate their preconceived ideas and assumptions [14]. Finally, real-time feedback can improve the group’s collaboration process [22–24, 27, 28]. In two experiments that tasked students to generate ideas for improving the campus parking system, researchers found that feedback provided additional motivation for all participants to increase their effort [22]. Researchers also found that feedback can reduce social loafing by identifying individual contributions in a group [10, 28].

Building on this prior research, we created a group design task for complex interdependent energy-water systems to test two hypotheses about group performance with versus without real-time feedback:

**H1:** Group members in a complex system design task receiving real-time feedback will generate system-level solutions that are close to the Pareto optimal solution.

**H2:** Group members receiving real-time feedback will generate system-level solutions that are better than solutions generated independently or through informal collaborations.

To test our hypotheses, our working example is a wastewater treatment system for unconventional gas exploration operation in the Marcellus region (a region spanning New York, Pennsylvania, Maryland, Ohio, Virginia and West Virginia) [29]. The problem is formally characterized using a multi-objective, mixed integer linear programming (MILP) model [29]. In this case study, potential designs had two competing system objectives, either to reduce project lifetime financial cost, or reduce the human health impacts from air emissions (see figure 1). To design a system that meets those two objectives, analysts need to gather information from a range of experts on parameters such as wastewater flowback rate, wastewater composition in frack fluid, and wastewater treatment efficiency. Decision-makers also need to provide information on the set of system constraints, including fracking schedule, mass-balance, capacity, and finances. Finally, decision-makers need to make design and policy decisions about freshwater source use, wastewater reuse, and water transportation options.

We used this case study to examine the effect of providing real-time feedback on system performance in a multi-participant, multi-objective case study. We compare real-time feedback against two alternatives: an independent design approach where each participant had to make decisions on their own, and an informal collaboration approach where participants worked together without real-time feedback.

**Method**

Participants were asked to design a wastewater management system for shale gas exploration in the Marcellus region [29]. Initially formulated as a MILP, changes were made to allow research participants to complete the tasks in a timely manner. These include:

- Pipelines can no longer be leased on a weekly basis.
- Water must be transported through trucking or constructing a new pipeline.
- Transportation option decisions are not made on a weekly, individual connection basis, but are applied across the entire system for the duration. This reduces the number of binary decision variables from 1750 to 13, making it tractable as a design exercise.
- Increased the number of freshwater sources from one to three. Participants needed to select the freshwater source for the system.
Increased the number of pipeline options from one to three. Participants needed to choose a specific pipeline capacity if they chose to use it for water transportation.

Due to these changes, the problem transformed into these two objective functions:

\[
\text{Cost}_{\text{Financial}} = (\text{Cost}_{\text{WaterWithdrawal}} + \text{Cost}_{\text{FreshwaterTransportation}} + \text{Cost}_{\text{WastewaterTransportation}} + \text{Cost}_{\text{Storage}} + \text{Cost}_{\text{WastewaterTreatment}})
\]

Where:

- \(\text{Cost}_{\text{WaterWithdrawal}}\) is the cost of freshwater withdrawal,
- \(\text{Cost}_{\text{FreshwaterTransportation}}\) is the cost of freshwater transportation,
- \(\text{Cost}_{\text{WastewaterTransportation}}\) is the cost of wastewater transportation,
- \(\text{Cost}_{\text{Storage}}\) is the cost of water storage,
- \(\text{Cost}_{\text{WastewaterTreatment}}\) is the cost of water treatment.

And:

\[
\text{Cost}_{\text{HHE}} = \sum_{e} CE_e (M_{e}^{\text{Storage}} + M_{e}^{\text{Pipeline}} + M_{e}^{\text{Tracking}} + M_{e}^{\text{CentralTreatment}} + M_{e}^{\text{Disposal}})
\]

Where:

- \(CE_e\) is the cost of emissions for each air pollutant \(e\),
- \(M_{e}^{\text{Storage}}, M_{e}^{\text{Pipeline}}, M_{e}^{\text{Tracking}}, M_{e}^{\text{CentralTreatment}}, M_{e}^{\text{Disposal}}\) are the amount of air pollutant \(e\) generated for each activity.

Further details for each objective function can be found in appendix A.

The model also consists of a series of constraints, ranging from mass balance constraints, to flow capacity constraints, to scheduling constraints. It is important to note that even with these constraints, there are \(~2,200\) different possible designs for the system, making it infeasible for participants to iterate through the entire solution space under a time constraint.

The study employed a mixed within- and between-subjects design, with 88 participants drawn from the Carnegie Mellon University student population. Our experimental design was reviewed and approved by Carnegie Mellon University’s Institutional Review Board prior to subject recruitment. There were no exclusion criteria. Students were recruited through posters around campus, and participants were compensated with a lottery of five $50 Amazon gift cards. Each participant in the group was randomly assigned an expert role and provided with briefing material on their expertise. There were four expert roles: Well-Pad Operator, Freshwater Expert, Wastewater Expert, and Environmental Regulator. The briefing material was unique to each expert role.
and included information on the parameters of their areas of expertise, their individual goals, their motivations, and the expected interactions with other members of the group. During the experiment, the experimenter was present in the room observing participant behavior and ensured that participants did not share their briefing material with other members of the group. Carnegie Mellon University students do not accurately represent the population of experts because they did not have years of experience working in the oil and gas industry, nor the technical expertise in designing a wastewater treatment system for shale gas exploration. To address their lack of expertise on the subject matter, the briefing material included a general overview of unconventional gas exploration, specific domain knowledge such as pipeline capacity and transportation distance, and historical system performance that provides a useful marker for the participants. We also added contextual information, including the motivation of their role as an expert. Additional details on the briefing material (including the specific briefing material for each role) can be found in appendix B. Participants conducted the research tasks through a web-app (built through R Shiny) specific to their expert role. From the participant’s perspective, these apps represented the submodules that provided all the information they needed to make decisions, as well as the performance of their submodules as a function of the group’s decisions. Participants didn’t have to enter their design decisions in a specific order (e.g., regulator first).

**Research tasks**

*Pre-study task*

Each participant was briefed with the setup and the relevant information related to their role. To verify that participants sufficiently understood the setup prior to the tasks, three True or False validation questions were asked. If the participant did not answer the validation question correctly, the correct answer was provided, and the participant was given an opportunity to ask clarifying questions.

*Task 1 (Independent task)*

Each participant was independently tasked with developing one design for their submodule that they believed would best achieve the goals outlined in their briefing. The results for both their individual design and the overall system were recorded and used as the benchmark to compare against in later tasks. Participants were given ten minutes to complete their design and they could not communicate with other members of the group during this task.

*Task 2 (Collaboration task)*

In this task, participants were asked to collaborate informally and generate a design solution. The results for both their individual design and the overall system were recorded. In addition to in-person dialogue, the participants could use markers and dry-erase whiteboards. When compared to the results from Task 1, this task measured the effects of informal discussion on both individual and group performance. The participants had 20 min to complete this task.

*Task 3 (CADS task)*

In this task, participants were asked to collaborate and generate a design solution with additional access to a dashboard with real-time performance feedback. The feedback included overall financial and environmental costs of the design they generated, along with a component-by-component breakdown of the cost. Finally, the feedback included the decisions each participant made for each design, allowing participants to understand how their collective decisions affected the overall system. The feedback did not provide suggestions for future designs, nor did it indicate whether the solution was on the Pareto Frontier, meaning participants needed to actively link the feedback they received with the decisions they made to make improvements to their design. Their results for both their individual design and the overall system were recorded. When compared with both Task 1 and 2, this task measured the effect of real-time performance feedback on both individual and group performance. The participants had 20 min to complete this task.

Each group consisted of four students who completed all three research tasks, and we compared the performance between each group and within each group, across all three tasks. To explore potential order effects, the order of Tasks 2 and 3 was randomized between-subjects. Task 1 always came first.

**Results**

H1: We hypothesized that participants receiving real-time feedback would generate system solutions that are close to optimal.

Because the task was a multi-objective optimization problem (both environmental and financial cost), there are multiple potential optimal solutions that generate a Pareto frontier. We classified distance to the optimal
solution as the smallest Euclidean distance to the Pareto frontier across the multi-objective outcome space. Our hypothesis test constructed a 20% margin of non-inferiority surrounding each solution on the Pareto frontier. We assessed whether solutions generated by participants across the different tasks fell within that margin of non-inferiority [30].

At the group level, participants receiving real-time feedback (CADS task) generated solutions that were closer to the Pareto frontier than the other non-CADS tasks. As shown in figure 2, system outcomes created by groups with real-time feedback (figure 2 panel B) overlapped much more closely with the Pareto non-inferiority margin than solutions generated through independent design (figure 2 panel A) and through informal collaboration (figure 2 panels C and D). The mean distance to the Pareto Frontier was $2.4M for the CADS task (SD = $2.2M, N = 22), compared to informal collaboration task of $5.9M (SD = $9.7M, N = 22), or independent task of $14.5M (SD = $14.7M, N = 22). The one-tailed test of whether the mean distances fell within the non-inferiority region is:

\[ H_0: \mu_T \geq \mu_R + M_{NI} \]
\[ H_a: \mu_T < \mu_R + M_{NI} \]

Where:

\( \mu_T \) is the sample mean distance to the Pareto Frontier

---

**Figure 2.** Black dots represent Pareto outcomes, and gray area represents the 20% non-inferiority margin. (A) Design solutions created by participants from the individual design task. (B) Design solutions created by participants from the CADS design task. (C) Design solutions created by participants from informal collaboration design task before the CADS task. (D) Design solutions created by participants from informal collaboration design task after the CADS task. Red dotted lines denote 95% CI for the participant generated design solutions in the feasible region.
Our second hypothesis was that participants receiving real-time feedback would generate system solutions that are better than solutions generated independently or through informal collaboration.

We tested H2 by constructing a linear regression model that evaluated the effect of providing participants with real-time feedback on their group performance, as measured by the system’s financial and environmental impact costs. We found that participants receiving real-time feedback generated better solutions for both objectives than the independent design and informal collaboration design. Further, there was an asymmetric transfer effect, where participants who completed the real-time feedback task before the informal collaboration task performed better on the informal collaboration task than when the order of tasks was reversed. Figure 3 shows that the distribution of the financial cost (figure 3 panel A) and environmental cost (figure 3 panel B) are skewed with a long right tail. To account for the skew in the underlying distributions, we log transformed the dependent variables. The resulting distributions appear to be less skewed, with a shorter tail and fewer outliers. To address the concerns raised by Lo and Andrew (2015) about the log-normal assumption, we repeated the
analysis with the model specification using a log link function, and there was very little difference between the log transformed and the log link result [31]. Additional details can be found in appendix C.

The model specification is detailed in (3):

$$
\log(\text{ObjCost}) = B_0 + \delta_1 \times \text{CollabTask} + \delta_2 \\
\times \text{CADSTask} + \gamma \times 1(\text{CADSFirst}) \\
+ \tau_1 \times \text{CollabTask} \times 1(\text{CADSFirst}) \\
+ \tau_2 \times \text{CADSTask} \times 1(\text{CADSFirst}) + X_i \beta + \epsilon_i
$$

(3)

Participants receiving real-time feedback generated solutions that are better than the independent design and informal collaboration design for both objectives (shown in figure 3). As shown in table 1, participants who received the real-time feedback generated solutions that were on average $-17.2M lower in financial cost (26% (percentages are calculated using the coefficients from the log-transformed models with methods from Halvorsen, R., & Palmquist, R. (1980) [32]) df = 58, t = −3.29, p < 0.01) and $-8.4M lower in environmental cost (34%, df = 58, t = −2.39, p < 0.05) when compared to the same group’s performance in the independent design task, and $-11.6M (21%, df = 58, t = −3.21, p < 0.05) lower in financial cost and $-3.2M (12%, df = 58, t = −1.03, p < 0.30) lower in environmental cost when compared to the same group’s performance in the informal collaboration task. The difference between the real-time feedback group and the informal collaboration task is not statistically significant for environmental cost, potentially because many groups chose to trade off environmental cost to increase their gains in the financial cost category between the informal collaboration stage and the feedback stage.

The cost reductions relative to independent design were statistically significant for the CADS task both when the CADS task was before informal collaboration and after. However, there was no improvement on the informal collaboration task relative to the independent design task when participants completed the informal collaboration task before the CADS task. This suggests there is significant learning and improvement from real-time feedback, and that improvement was transferred to the informal collaboration group when CADS came before informal collaboration, but not after. Without real-time feedback, there was little learning, suggesting the presence of asymmetric transfer effect. This can be directly observed from the regression results. We conducted a Z-test on the collaboration order effect ($\tau_1 = -0.29$) against the CADS order effect ($\tau_2 = -0.01$) for financial cost and found that the two coefficients are statistically different ($Z = 2.24, p < 0.05$). This showed the order effect for collaboration is significantly larger than the order effect for CADS, resulting in an additional decrease of approximately $11.6M (~25\%)$ on top of the effect of collaboration. However, this asymmetric transfer effect was not significant for environmental cost ($\tau_1 - \tau_2 = 0.06 Z = 0.36, p = 0.36$), showing that the order effect for collaboration was approximately the same for CADS. This could be the result of the design setup, where only one participant’s primary objective in each group was environmental cost, and therefore most of the group emphasized financial cost.

We conducted an analysis of the residuals and found the residuals for the models are approximately normally distributed with conditional mean around 0. However, there is some heteroskedastic behaviour in the residuals that warranted the use of heteroskedastic and clustered robust standard errors that increased the p-values of the results. It is important to note that despite the use of heteroskedastic and clustered robust standard errors, the CADS treatment coefficient remained statistically significant for both objectives at $p < 0.05$ level. Additional details of this analysis without the clustered standard errors are reported in appendix C.

Due to the multi-objective nature of the problem, each group can have different weights for each objective [32]. Certain groups might weigh financial cost more than environmental cost, and their design decisions would reflect that choice. To verify the robustness of the results from table 1 with different objective weights, we applied relative weightings to the dependent variables, starting at 0% environmental cost, and varied it at a 5% interval until a weighting of 100% environmental cost. The regression model, adapted from equation (3), is as follows:

$$
\text{WeightedObjective}_t = B_0 + \delta_1 \times \text{CollabTask} + \delta_2 \\
\times \text{CADSTask} + \gamma \times 1(\text{CADSFirst}) + \tau_1 \times \text{CollabTask} \\
\times 1(\text{CADSFirst}) + \tau_2 \times \text{CADSTask} \times 1(\text{CADSFirst}) + X_i \beta + \epsilon_i
$$

(4)

$$
\text{WeightedObjective} = \omega_1 \times \log(\text{FinancialCost}) + \omega_2 \times \log(\text{EnvironmentalCost})
$$

(5a)

$$
\omega_1 + \omega_2 = 1
$$

(6a)

The CADS task treatment coefficients showed a consistent progression from the financial objective model to the environmental objective model. In addition, the t-statistics show that, regardless of the objective weights, the CADS Task variable remained statistically significant (figure 4). This shows that the model results are robust regardless of the relative weighting assigned to each objective.
Table 1. Treatment effects of each task for both objectives, comparing against independent task with \( \text{CADSFirst} = 0 \). \textit{Basic} refers to models with only the treatment variable, \textit{Covar} refers to models with treatment variable and group-level covariates, \textit{Order} refers to models with treatment variable and task order dummy variables, and \textit{Order + Covar} refers to models with the specification in (3). Standard errors are heteroskedastic and auto-correlated robust. *** \( p < 0.001 \), ** \( p < 0.01 \), * \( p < 0.05 \).

| Financial cost models | Coefficient | Basic  | Covar  | Order  | Order + Covar |
|-----------------------|-------------|--------|--------|--------|--------------|
| Independent (\( \text{CADSFirst} = 0 \)) | \( B_0 \) | 4.02*** | 3.94*** | 4.04*** | 3.98*** |
| Independent (\( \text{CADSFirst} = 1 \)) | \( B_0 + \gamma \) | -0.06 | -0.09 |
| Collaboration (\( \text{CADSFirst} = 0 \)) | \( B_0 + \delta_1 \) | -0.18* | -0.18*** | -0.07 | -0.07 |
| Collaboration (\( \text{CADSFirst} = 1 \)) | \( B_0 + \delta_1 + \gamma + \tau_1 \) | -0.42** | -0.45* |
| \( \text{CADS} \) (\( \text{CADSFirst} = 0 \)) | \( B_0 + \delta_2 \) | -0.31*** | -0.31*** | -0.31*** | -0.31*** |
| \( \text{CADS} \) (\( \text{CADSFirst} = 1 \)) | \( B_0 + \delta_2 + \gamma + \tau_2 \) | -0.38** | -0.41*** |

| Environmental cost models | Coefficient | Basic  | Covar  | Order  | Order + Covar |
|---------------------------|-------------|--------|--------|--------|--------------|
| Independent (\( \text{CADSFirst} = 0 \)) | \( B_0 \) | 2.92*** | 3.12*** | 2.94*** | 3.15*** |
| Independent (\( \text{CADSFirst} = 1 \)) | \( B_0 + \gamma \) | -0.06 | -0.05 |
| Collaboration (\( \text{CADSFirst} = 0 \)) | \( B_0 + \delta_1 \) | -0.30* | -0.30*** | -0.28 | -0.28* |
| Collaboration (\( \text{CADSFirst} = 1 \)) | \( B_0 + \delta_1 + \gamma + \tau_1 \) | -0.42 | -0.41* |
| \( \text{CADS} \) (\( \text{CADSFirst} = 0 \)) | \( B_0 + \delta_2 \) | -0.41*** | -0.41*** | -0.41*** | -0.41*** |
| \( \text{CADS} \) (\( \text{CADSFirst} = 1 \)) | \( B_0 + \delta_2 + \gamma + \tau_2 \) | -0.49** | -0.48* |
Discussion

In this study, we tested the effect of providing quantitative, real-time feedback on the relationship between design decisions and task objectives, then compared it against independent designs and informal collaborations in a multi-objective wastewater treatment system design task. Participants with real-time feedback generated solutions that are both closer to the Pareto frontier and with lower cost for both environmental and financial objectives than participants generating solutions independently or through informal collaboration. When participants attempted to generate solutions independently, they lacked insight about how their actions affected other members of the team and the overall system, leading to suboptimal outcomes. Even though the informal collaboration task allowed participants to communicate with each other and share information, the complexity of the system made it difficult for participants to understand the consequences of their actions. Some groups tried to understand the relationships between their roles and the overall system using the tools available to them in the room (e.g., whiteboards) during the informal collaboration task. However, we observed that no group was able to accurately map out the interdependent system relationship during the informal collaboration task.

Finally, some groups decided to focus their attention on one individual module during the informal collaboration task (akin to a depth-first search algorithm) and attempted to find the optimal solution with the lowest cost for that individual’s role. However, that strategy did not always yield the best system solution, as sacrifices were made in other modules that made it globally inefficient.

When provided with real-time feedback, participants were able to better understand the impact of their decisions on the overall system through the display of objective metrics. We also observed that participants became more motivated as they grasped the connection between their decisions and the overall solution, which manifested in participants wanting to find better solutions. In addition, participants were able to use the real-time feedback to validate their assumptions for their decisions and recognize when those assumptions were false. Finally, real-time feedback improved the results of groups who had uninterested members (observed through their lack of interaction with other members in their group), where the real-time feedback allowed groups to use that feedback as cues to point out where the unengaged members can improve the overall system.

There are several limitations to this study that are worth noting. First, the use of convenience sampling methods meant that participants are not representative of the intended population of industry experts [33]. We attempted to mitigate this limitation by providing briefing material to research participants that mimicked expert knowledge and behaviour (more detail in appendix B), however there is still likely differences between how our participants responded to the research tasks compared to industry experts. Secondly, experimental design constraints such as time and cost meant that the design task is a facsimile of the real-world design challenge. There will be differences between the research task in this study and real design sessions, such as stronger personal motivations and familiarity with each participant due to prior experience.

Performance improvements could also have been due to activation of visual cues made available through the feedback mechanism, where participants performed better not because of the system-level feedback they received, but because of additional visual cues that activated their attention [34, 35]. While this is a potential
confounding factor, because we know that attention activation is also a function of time, and we observed no difference when the CADS task order was switched, some of this concern can be alleviated [36, 37].

This study finds a different effect of providing feedback during complex decision-making processes compared to simple perceptual tasks [19]. In our study, the effect of informal collaboration is not a statistically significant indicator of improved group performance. In contrast, the quantitative, real-time feedback mechanism provided to the groups had a large and statistically significant effect. This may be attributable to inherent differences between complex design tasks and perceptual tasks. In addition to advantages that real-time feedback provided to participants, shared information bias had a greater effect on team performance in complex design tasks than in perceptual tasks in prior work [38, 39]. Without the ability to confirm their assumptions during the informal collaboration design task, participants lacked the ability to understand what information is essential, instead focusing on the information that all participants possessed. When provided with feedback, however, participants were able to confirm their assumptions and understand what information is needed from other participants to generate the optimal design. For complex design decisions, it appears that feedback is an important aspect of group success.

Although the use of students as participants threatens the external validity of the study, systems that real experts must deal with also have greater complexity than the one used in this study. Combined with the informational materials designed to get students up to speed on the task, the balance between participant expertise and task complexity may be similar in our study and the real world. Secondly, institutions such as the Jet Propulsion Lab (JPL), who in 1995 established the Advanced Projects Design Team (known as Team X) to design new space mission proposals, successfully used real-time feedback to improve their design process [40]. Experts were recruited for specific system modules in a design team (avionics, battery etc.), and they collaborated to generate mission designs through a series of concurrent design sessions [41]. When compared to previous space mission designs, Team X was able to design more missions per year, with lower average time for each design, and a lower average cost of design [42]. In addition, when simulated with experts against past mission parameters, Team X results were within 5% of actual mission costs [43]. The results of our study reinforce the idea that providing real-time feedback may have been an essential component of Team X’s accuracy. Our results suggest that similar endeavors that aim to solve complex design problems, ranging from designing utility-scale grids to resilient public infrastructure, should include real-time feedback and collaboration among experts to avoid suboptimal outcomes.

Statistics were done using R 3.5.0 (R Core Team, 2018), the googledrive (1.0.0.9000; Bryan and McGowan, 2020), googlesheet (0.3.0; Bryan and Zhao, 2018), dplyr (0.8.5; Wickham et al., 2020), tidyr (1.0.2; Wickham and Henry, 2020), Rmisc (1.5; Hope, 2013), ggplot2 (3.3.2; Wickham, 2016), lm4 (1.1-21; Bates et al., 2015), car (3.0-7; Fox and Weisberg, 2019), texreg (1.36.23; Leifeld, 2013), ggrepel (0.3.1; Pedersen, 2019), sandwich (2.5-1; Berger, Graham, Zeileis, 2017), and lmtest (0.9-37; Zeileis and Hothorn, 2002) packages [44–57]. All data, code, and stimuli are available on GitHub: https://github.com/we3lab/MultiStakeholderDesign.

Acknowledgments

The authors wish to acknowledge the financial support of the Scott Institute Innovation Seed Fund, Carnegie Mellon University Graduate Student Assembly CT Fund, and EQT Industry grant.

Data availability statement

The data that support the findings of this study are openly available at the following URL: https://github.com/we3lab/MultiStakeholderDesign.

ORCID iDs

Niles Guo 🌐 https://orcid.org/0000-0002-9541-5289
Alex Davis 🌐 https://orcid.org/0000-0002-0958-305X
Meagan Mauter 🌐 https://orcid.org/0000-0002-4932-890X
Jay Whitacre 🌐 https://orcid.org/0000-0002-3439-4111

References

[1] Kling C L, Arritt R W, Calhoun G and Keiser D A 2017 Integrated assessment models of the food, energy, and water nexus: a review and an outline of research needs Annu. Rev. Resour. Econ. 9 143–63
[2] Ascough J C, Maier H R, Ravalico J X and Strudley M W 2008 Future research challenges for incorporation of uncertainty in environmental and ecological decision-making Ecol. Modell. 219 383–99
[45] Bryan J and Zhao J 2018 googlesheets: Manage Google from R. R package version 0.3.0 https://CRAN.R-project.org/package=googlesheets
[46] Wickham H, François R, Henry L and Müller K 2020 dplyr: A Grammar of Data Manipulation. R package version 0.8.5 https://CRAN.R-project.org/package=dplyr
[47] Wickham H and Henry L 2020 tidyr: Tidy Messy Data. R package version 1.0.2 https://CRAN.R-project.org/package=tidyr
[48] Hope R M 2013 Rmisc: Rmisc: Ryan Miscellaneous. R package version 1.5 https://CRAN.R-project.org/package=Rmisc
[49] Rudis B, Bolker B and Schulz J 2017 ggalt: Extra Coordinate Systems, 'Geoms', Statistical Transformations, Scales and Fonts for 'ggplot2'. R package version 0.4.0 https://CRAN.R-project.org/package=ggalt
[50] Wickham H 2016 ggplot2: Elegant Graphics for Data Analysis (New York: Springer)
[51] Bates D, Maechler M, Bolker B and Walker S 2015 Fitting Linear Mixed-Effects Models Using lme4 Journal of Statistical Software 67 1–48
[52] Fox J and Weisberg S 2019 An {R} companion to applied regression Third Edition (Thousand Oaks CA: Sage) https://socialsciences.mcmaster.ca/jfox/Books/Companion/
[53] Leifeld P 2013 texreg: conversion of statistical model output in R to LaTeX and HTML tables Journal of Statistical Software 55 1–24 http://jstatsoft.org/v55/i08/
[54] Pedersen T 2019 L. ggforce: Accelerating ‘ggplot2’. R package version 0.3.1. https://CRAN.R-project.org/package=ggforce
[55] Zeileis A 2004 Econometric computing with HC and HAC covariance matrix estimators Journal of Statistical Software 11 1–17
[56] Zeileis A, Koll S and Graham N 2020 Various versatile variances: an object-oriented implementation of clustered covariances in r J Stat Softw. 95 1–36
[57] Zeileis A and Hothorn T 2002 Diagnostic checking in regression relationships R News 2 7–10 https://CRAN.R-project.org/doc/Rnews/