More and Better: Utility-Aware Adaptive Survey Timing for Human-Centric Data Collection

by

Shashank Chauhan

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science
(Artificial Intelligence)
in the University of Michigan-Dearborn
2026

Master's Thesis Committee:

Professor Zheng Song (Co-Chair)

Professor Qiang Zhu (Co-Chair)

Professor Natalia V. Czap

Professor Hans J. Czap

Shashank Chauhan

shashanc@umich.edu

ORCID iD: 0009-0007-5450-7631

© Shashank Chauhan 2026

Acknowledgements

I would like to express my sincere gratitude to my advisors, Professor Qiang Zhu and Professor Zheng Song, for their guidance, support, and encouragement throughout this research. Their insights and feedback were invaluable in shaping both the technical contributions and the overall direction of this work.

I would also like to thank my committee members, Professor Natalia V. Czap and Professor Hans J. Czap, for their thoughtful feedback and constructive suggestions that significantly improved this thesis.

I am grateful to my co-author Siddhi Baravkar for their contributions to the early stages of this project.

I would also like to thank the 121 participants who took part in the three-week study. Their time and commitment made this research possible.

This research is supported by a Bold Challenges Initiative grant from the University of Michigan.

Table of Contents

Acknowledgements 3

List of Figures 7

Abstract 8

Chapter 1 Introduction 9

Chapter 2 Related Work 14

2.1 Human-Centered Data Collection Applications 14

2.2 Sensor-Agnostic Survey Triggering Strategies 15

2.3 Increasing Data Collection Rate 15

2.4 Sensor-Aware Survey Triggering 16

Chapter 3 Challenges and Solution Overview 17

3.1 Case Study: Air Quality and Human Wellbeing 17

3.2 Challenges and Key Novelties 18

3.3 Solution Overview 22

Chapter 4 Adaptive Survey Timing Method (ASTM) 23

4.1 Utility Estimation via Confidence Index 23

4.2 Participant Willingness Modeling 24

4.3 Scoring Function 24

4.4 Time Slot Evaluation 25

4.5 Learning Layer: Exploration-Exploitation Policy 25

Chapter 5 HCDC System Design and Implementation 28

5.1 Backend Design 29

5.2 SMS Delivery 29

5.3 Data Collection Pipeline 29

5.4 Gamification Component 30

5.4.1 User Categorization 31

5.4.2 Personalized Feedback Design 34

Chapter 6 Experiments and Results 35

6.1 Experimental Setup 35

6.1.1 Study Design 35

6.1.2 Participants and Recruitment 36

6.1.3 Data Collection Protocol 36

6.1.4 Evaluation Metrics 37

6.2 Results 37

6.2.1 Model Quality and Learning Efficiency 37

6.2.2 Participation and Engagement 40

6.3 Alpha Sensitivity Analysis 43

6.3.1 User Behavior Model 44

6.3.2 Simulation Setup 45

6.3.3 Results 46

6.4 Key Interpretations 49

Chapter 7 Conclusion 51

7.1 Limitations and Future Directions 52

References 54

List of Figures

Figure 1 HCDC: combining sensor data with subjective survey responses 10

Figure 2 Hypothetical example of high utility data collection in HCDC 11

Figure 3 Overview of the proposed survey timing method 21

Figure 4 Workflow of the proposed adaptive survey timing method (ASTM) 27

Figure 5 System architecture of the HCDC system 28

Figure 6 Gamification leaderboard showing participant rankings, response rate, and achievement badges for Category A participants 32

Figure 7 Personalized progress page for Category B participants, showing survey completion milestones and improvement over time 33

Figure 8 Learned PM2.5 and wellbeing relationships under EE and random timing 38

Figure 9 Model MAE over time for EE and random timing conditions 39

Figure 10 Missed survey rates across the four experimental conditions 41

Figure 11 Response rates at three-day intervals for gamified and non-gamified conditions over the three-week study 42

Figure 12 User behavior model used in the simulation 45

Figure 13 Simulated response rates for different alpha configurations 47

Figure 14 AQ coverage under different alpha configurations 48

Abstract

Human-centric data collection integrates objective sensor measurements with subjective human input (e.g., perceptions, experiences, or feelings) to understand how objective conditions relate to human experiences and behaviors. A key challenge in such systems is improving data quality without increasing participant burden. Existing approaches, such as fixed or random sampling, incentives, reminders, and gamification, primarily aim to increase response rates. They often collect redundant data that contributes limited new information for modeling. As a result, higher participation does not necessarily lead to better understanding. To address this, we propose an adaptive survey timing framework that determines when to request user input by jointly considering data utility and participant willingness. The framework prioritizes moments when responses are expected to be most informative while accounting for when participants are most likely to respond. To handle uncertainty under limited data, we adopt an exploration-exploitation (EE) strategy to guide survey timing decisions. We evaluate the proposed approach through a 3-week randomized controlled experiment with 121 participants residing in Southeast Michigan, using a between-subjects factorial design with four conditions: baseline (random timing), EE-based timing, gamification, and EE + gamification. Results show that adaptive timing improves model learning efficiency and reduces missed surveys compared to random scheduling. Furthermore, combining adaptive timing with gamification achieves the highest participation. These findings demonstrate that optimizing when data is collected can simultaneously improve participation and data utility without increasing participant burden.

Keywords: human-centered data collection, air quality, exploration-exploitation, adaptive survey timing, gamification

Chapter 1 Introduction

Understanding how objective conditions relate to human experiences is essential for advancing public health, workplace productivity, and overall quality of life. Objective signals such as environmental factors (e.g., air quality, noise, temperature), physiological states (e.g., heart rate, activity level), and behavioral patterns (e.g., daily exercise or screen time) can influence mood, stress, engagement, and cognitive performance. For example, workplace studies examine how noise exposure affects concentration [1], physical activity research explores how exercise relates to mood and stress [2], and environmental health studies investigate how air pollution impacts subjective wellbeing [3, 4, 5, 6]. These examples highlight a common goal: understanding the relationship between measurable conditions and human experiences.

Human-**C**entric **D**ata **C**ollection (HCDC) addresses this need by integrating sensor-based objective measurements with subjective human input. In this work, we define HCDC as a data collection paradigm performed close to individuals that combines objective measurements obtained via portable or wearable sensors with subjective self-reports, enabling joint analysis of objective conditions and subjective human experiences. Related paradigms appear in the literature under terms such as participatory sensing [7], experience sampling [8], and human-in-the-loop sensing [9]. Compared to these approaches, HCDC emphasizes both proximity to the individual and the explicit integration of subjective and objective data streams. Advances in IoT (Internet of Things) and mobile sensing have made HCDC increasingly practical and widely adopted. These technologies enable researchers to combine continuous sensing with real-time self-reports to study behavior and wellbeing in everyday settings [10, 11, 12, 13]. Figure-1 illustrates this process,

showing how objective sensor measurements and subjective survey responses are jointly collected to analyze the relationship between objective conditions and subjective human experience.

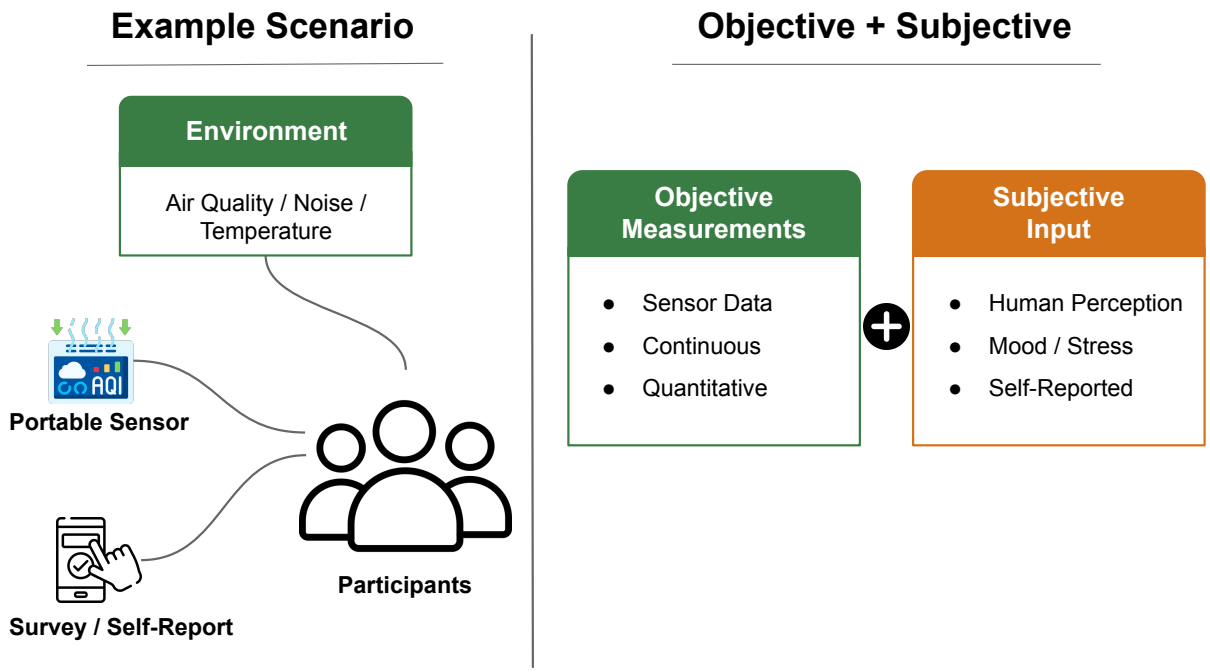


Fig. 1: HCDC: combining sensor data with subjective survey responses.

Despite its promise, HCDC faces a fundamental challenge: the effectiveness of the collected data depends on participants' willingness to provide subjective responses. From a modeling perspective, learning accurate relationships requires sufficient paired data points that combine objective measurements with corresponding subjective reports. However, unlike objective sensor data, subjective responses require active human involvement. Participants must interrupt their activities, reflect on their current state, and respond to question prompts. As a result, human attention becomes a limited and costly resource. A key challenge in HCDC is therefore how to collect informative subjective responses while respecting limited human effort and minimizing participant burden. Simply increasing the number of survey prompts risks fatigue,

disengagement, and reduced response reliability [14], ultimately weakening the learned relationship between objective conditions and subjective human experiences.

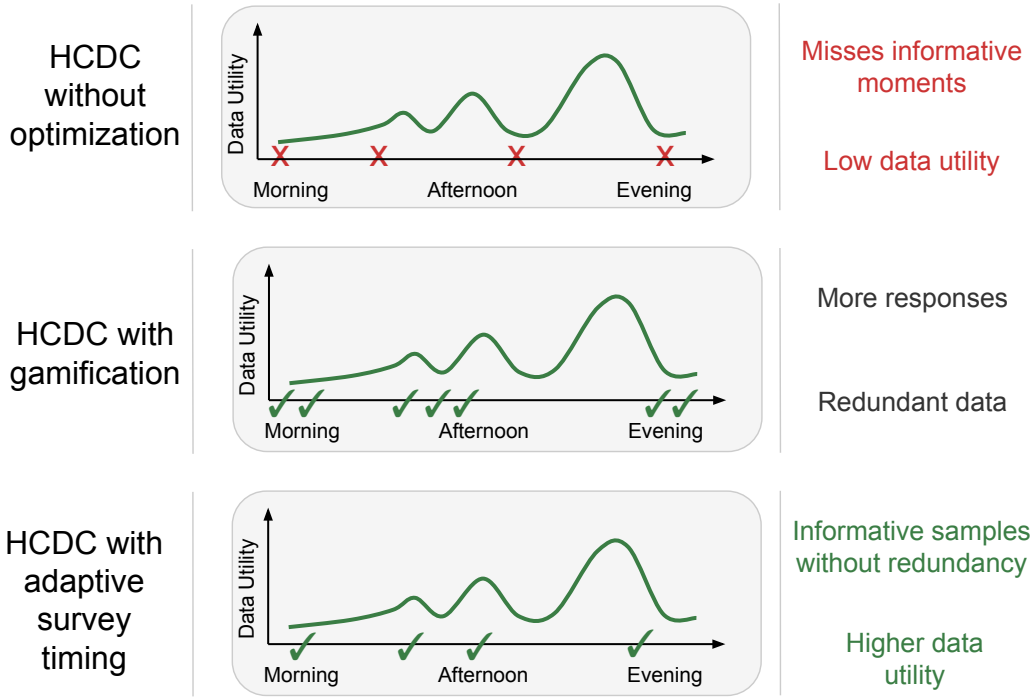


Fig. 2: Hypothetical example of high utility data collection in HCDC. Adaptive timing focuses on informative moments without overburdening participants.

As illustrated in Figure-2, the usefulness of collected responses can vary significantly across different situations. For example, responses collected under stable conditions (e.g., during quiet periods of the day) may repeatedly describe similar states, providing limited new information. In contrast, responses collected during periods of change or less frequently observed conditions can reveal important variations in both objective measurements and subjective human experiences. As a result, two datasets with similar numbers of responses may differ substantially in how much they help explain the relationship between objective conditions and subjective experiences.

Most existing approaches [15, 16, 17, 18, 19] to improving participation in HCDC focus on increasing the number of responses rather than improving their informativeness. Techniques

[10, 15, 16, 20, 21, 22] such as fixed or random sampling, incentives, and gamification are widely used in longitudinal studies to boost response rates. While these methods can increase participation, they often generate redundant data that contributes little new information for modeling. As a result, higher response rates do not necessarily lead to better understanding of the relationship between objective measurements and subjective responses. This limitation highlights the need to promote informative responses rather than simply maximizing response volume.

The timing of subjective data collection plays a critical role in determining both participation and informativeness. On one hand, question prompts must be delivered when participants are available and willing to respond; otherwise, they may result in delayed responses or non-participation, reducing the number of usable data points. On the other hand, the informativeness of a response depends on the underlying objective conditions at the time it is collected. Responses obtained under stable or well-represented conditions may contribute limited new information, whereas responses collected during periods of change or higher uncertainty can provide substantially greater value for model learning. This dual role of timing is particularly important in domains where objective conditions evolve over time. For example, in environmental health studies, changes in air quality, noise, or temperature can affect human experience within short time scales [23]. Despite its importance, how to systematically leverage timing to balance participant responsiveness and data utility remains largely unexplored in existing HCDC systems.

This work proposes a utility-aware adaptive survey timing framework for human-centric data collection. The framework determines when to issue surveys by balancing two key factors: data utility and participant willingness. Data utility is estimated from model uncertainty, allowing the system to prioritize moments when a new response is expected to provide the greatest informational gain. Participant willingness is learned from historical response behavior, enabling the system to avoid sending surveys at inconvenient times and improve response likelihood. To operationalize this design, we develop an exploration–exploitation (EE) strategy that adaptively selects survey times by jointly considering these two signals. In addition, we incorporate a

lightweight gamification component that provides periodic feedback to sustain participant engagement over time. We implement the proposed framework as an end-to-end data collection system that integrates portable air quality sensing with real-time subjective reporting. We evaluate the system through a multi-week randomized experiment. The results show that adaptive survey timing improves model learning efficiency and reduces missed responses compared to standard scheduling approaches, while the addition of gamification further enhances participation without overburdening participants.

Preliminary work on this problem was reported in [24], where we introduced the basic air quality and wellbeing data collection system and an initial exploration-exploitation (EE) method for deciding when to send survey messages. That work mainly focused on system design and dataset-based evaluation. This thesis substantially extends our early study. First, we improve the timing method by combining data utility with participant willingness through a scoring function, instead of relying only on the earlier EE formulation. Second, we add a gamification component to support sustained participation. Third, rather than only using existing data, we validate the full system through a real-world study with 121 participants and four experimental conditions (including control).

The remainder of the thesis is organized as follows. Chapter 2 reviews related work. Chapter 3 presents the main challenges and an overview of the solution. Chapter 4 describes the proposed adaptive survey timing method. Chapter 5 introduces the system design and implementation. Chapter 6 presents the experimental setup, results, sensitivity analysis, and key interpretations. Chapter 7 concludes the thesis and discusses limitations and future directions.

Chapter 2 Related Work

This Chapter reviews prior work related to human-centered data collection (HCDC) and survey-based sensing systems. We organize the discussion into four parts. First, we review existing HCDC applications that combine sensor measurements with self-reported data. Second, we examine commonly used survey triggering strategies that operate independently of sensor data and primarily focus on temporal coverage. Third, we discuss techniques aimed at increasing data collection rates, such as incentives and gamification, and their limitations in improving data quality. Finally, we review recent efforts that leverage sensor data to trigger data collection and highlight how they differ from our approach. This progression reveals a key gap: existing approaches either ignore data utility or fail to jointly optimize data utility and participant behavior.

2.1 Human-Centered Data Collection Applications

There is a growing demand for human-centered data collection systems that integrate objective sensor measurements with subjective self-reports to study real-world behavior and experience. These systems have been widely adopted in applications such as health monitoring, stress detection, sleep tracking, and environmental exposure analysis. In such settings, sensor data provides continuous and objective observations of environmental and behavioral conditions, while self-reported data captures perceptions, emotions, and experiences that cannot be directly measured by sensors [10, 11, 12].

Recent work further demonstrates the value of combining wearable sensing with subjective reporting in domains such as mental health, anxiety, and wellbeing assessment [25, 26, 27]. These studies establish HCDC as a powerful paradigm for understanding the relationship between objective conditions and human experience. However, while these systems enable data collection at scale, they largely rely on predefined or heuristic strategies to determine when subjective data should be collected.

2.2 Sensor-Agnostic Survey Triggering Strategies

A large body of prior work focuses on survey triggering strategies that are independent of sensor data. Common approaches include fixed-interval sampling, random sampling, and participant-initiated self-reports. The Experience Sampling Method (ESM) and Ecological Momentary Assessment (EMA) are widely used frameworks for collecting in-situ self-reports in natural environments [8, 10, 28, 29].

In fixed-interval sampling, surveys are delivered at predefined times to ensure regular data collection. Random sampling aims to improve temporal coverage by distributing prompts across different time periods. Participant-initiated reporting allows users to submit responses voluntarily when they perceive meaningful events. These approaches are popular due to their simplicity and ease of deployment [20].

However, these strategies do not consider the underlying sensor data or the informativeness of the collected responses. Survey timing is typically designed to maximize coverage or convenience rather than to capture informative or under-represented conditions. As a result, many responses may be collected during stable or redundant states, while critical moments of change remain unobserved.

2.3 Increasing Data Collection Rate

As an extension of the sensor-agnostic survey triggering strategies discussed above, a related line of work focuses on improving participation and increasing the volume of collected data. Prior studies have explored techniques such as incentives, reminders, personalization, and simplified survey design to encourage higher response rates [15, 16]. More recently, gamification mechanisms, including points, progress indicators, and feedback dashboards, have been introduced to sustain long-term engagement in mobile sensing and survey-based studies [21, 30].

While these approaches can effectively increase the quantity of collected data and reduce participant dropout, they do not explicitly improve data quality. In particular, increasing the number of responses does not guarantee that the collected data is more informative. Recent studies suggest that gamification may lead to higher response rates without necessarily improving the usefulness of the data for downstream modeling [22]. This highlights a fundamental limitation: optimizing for participation alone is insufficient for improving human-centered data collection.

2.4 Sensor-Aware Survey Triggering

With the increasing availability of mobile and IoT sensing technologies, recent work has begun to explore the use of sensor data to guide survey triggering. Context-aware and sensor-triggered EMA approaches leverage signals such as physical activity, location, or environmental conditions to determine when to prompt users [17, 18, 31, 32]. These methods represent an important step toward integrating objective data into survey timing decisions.

More recent efforts have further explored combining predicted response likelihood with model uncertainty to guide data collection [19]. However, existing approaches remain limited in several aspects. First, many studies are conducted in offline or simulation settings, without validating the approach in real-world deployments. Second, the notion of data utility is often tied to specific prediction tasks (e.g., emotion prediction) rather than to improving the learned relationship between objective and subjective data. Third, these methods typically do not jointly optimize data utility and participant behavior in an online and adaptive manner.

In contrast, our work studies human-centered data collection in a real-world air quality and wellbeing setting, explicitly defines data utility in terms of improving the relationship between objective environmental conditions and subjective responses, and develops an adaptive survey timing framework that jointly considers data utility and participant willingness. We further validate the proposed approach through a randomized controlled experiment, demonstrating its effectiveness in improving both data efficiency and model learning.

Chapter 3 Challenges and Solution Overview

This Chapter presents the main challenges that arise in developing an adaptive survey timing method, together with the key ideas behind our approach. At a high level, the problem is to determine when subjective responses should be collected so that the resulting data is both informative and likely to be obtained under limited participant attention. We first introduce a concrete example scenario to illustrate the problem setting and then describe the main challenges and the corresponding design ideas.

3.1 Case Study: Air Quality and Human Wellbeing

We consider an HCDC system for air quality and human wellbeing to illustrate the problem setting. The goal is to understand how objective air conditions relate to subjective human experiences. This problem is particularly important in regions such as Detroit and Dearborn cities, where air pollution remains a persistent public health concern [33], making it critical to collect accurate and representative data that reflects individuals' real-world exposure and experiences.

Prior work has reported qualitative and correlational relationships between air pollution, mood, stress, and wellbeing [3, 4, 5, 6]. However, effectively learning these relationships in a quantitative manner remains challenging. A key limitation is that many existing studies rely on coarse-grained air quality measurements, for example, aggregating data at the city or regional level while asking participants to assess their wellbeing. This approach overlooks the fact that individuals may experience significantly different air conditions depending on their immediate context (e.g., indoors vs. outdoors, proximity to pollution sources). As a result, subjective responses may not align with the actual environmental conditions experienced at the time of reporting. Furthermore, these studies often lack precise air quality measurements at the exact moment when survey responses are provided.

In our study, two types of data are collected. Objective data is obtained from portable sensors that continuously measure environmental signals such as PM2.5, PM10, VOC, temperature, and humidity. Subjective data is collected through online surveys, where participants receive a unique survey link via SMS and are expected to report their current feelings, such as mood and stress, shortly after receiving the SMS. While objective data can be continuously observed, subjective data must be actively provided by participants in response to these prompts, making it costly and limited. As a result, the system operates under a constrained survey budget, where only a small number of surveys can be issued to each participant every day.

This constraint introduces two key considerations. First, a response is more valuable when it is collected under conditions that are less represented in the data, as it can provide new information about the relationship between objective conditions and subjective experiences. Second, a response is only useful if the participant actually completes the survey, which depends on when the prompt is delivered and whether the participant is available to respond. These two factors correspond to data informativeness and participant willingness. Informativeness reflects how much a response contributes to improving the learned relationship, while willingness reflects the likelihood that a participant will respond at a given time.

When building such a data collection system, a key requirement is to determine survey timing under a limited budget such that the collected responses are both informative and likely to be completed. This calls for a method that leverages real-time sensor observations and past response behavior to dynamically decide when survey prompts should be issued.

3.2 Challenges and Key Novelties

When developing such a method, we identify three key challenges and present the core novelties of our approach in addressing them.

Challenge 1: Estimating Data Utility: Not all responses contribute equally to learning the relationship between objective conditions and subjective human experiences. To formalize this notion of informativeness, we utilize the concept of *data utility*, which quantifies the contribution of a data point to improving model learning. A key challenge is therefore to estimate the utility of each potential response and use this estimation to guide survey timing decisions.

This is difficult for two reasons. First, the utility of a subjective response depends jointly on the participant's real-time objective sensor measurements and the distribution of previously collected (objective, subjective) data pairs. Yet there is no established model that can effectively capture this dependency and quantify the marginal value of a new response. Second, if utility estimation is used to influence survey behavior in real time (e.g., deciding when to prompt users), it requires access to real-time sensor data and timely decision-making, which are often unavailable or difficult to integrate into existing survey systems.

Key Novelty 1: We address this challenge through a model-driven utility estimation. First, we train a predictive model that captures the relationship between objective sensor measurements and subjective human responses using historically collected data. Given a specific objective sensor reading, the model produces both a predicted subjective response and an associated uncertainty estimate. When the model's uncertainty is high, it indicates that the current data is insufficient to reliably characterize the relationship under those conditions, suggesting that additional data would be highly informative. We therefore define the utility of a potential subjective response as a function of the model's prediction uncertainty at the corresponding objective sensing state.

Second, the system periodically collects real-time sensor data and incrementally retrains the predictive model as new data becomes available. This enables continuous, real-time estimation of data utility and allows the system to make informed decisions about when issuing a survey prompt is likely to yield the greatest informational gain.

Challenge 2: Balancing Data Utility and Participation: As discussed above, collecting high-utility data is critical for effectively learning the relationship between objective sensor measurements and subjective human responses. However, collecting very few responses also makes the learned relationship unreliable. A practical human-centric data collection system must therefore achieve both goals: it should collect informative data, and it should maintain sufficient participation.

Balancing these two goals is difficult in practice for two main reasons. First, data utility and participation often pull survey timing in different directions. High-utility moments may occur at times when participants are busy or less willing to respond, which reduces participation. On the other side, surveys sent at convenient times may help increase participation but concentrate responses under stable conditions and provide low-utility data. Second, participants' routines and availability change over time. A timing strategy that works well for one person, or even for the same person on one day, may not work consistently across days. Together, these factors make it difficult to determine when surveys should be sent in order to collect more informative data while maintaining sufficient participation.

Key Novelty 2: We address this challenge by considering both data utility and participant willingness when making survey timing decisions. First, we estimate data utility using the uncertainty-driven method described in Challenge 1. Second, we estimate participant willingness, which reflects when a participant is likely to respond, based on historical response behavior such as response latency and completion patterns. Third, we combine these two estimates using a weighted formulation, so that survey timing decisions balance both how informative a response may be and how likely the participant is to respond.

Challenge 3: Learning under Limited Prior Knowledge: Even after defining how to estimate data utility and participant willingness, a key challenge remains: these estimates are initially not sufficient and improve only as more data is collected. Early in the data collection

process, the system has limited information about which times are informative and when participants are likely to respond. Both data utility and participant willingness must be refined over time based on observed survey outcomes. As a result, survey timing decisions must be made while these estimates are still uncertain, requiring the system to adaptively balance learning and usage as data collection progresses.

Key Novelty 3: We address this challenge using an exploration-exploitation algorithm [34]. Exploration allows the system to send surveys at less certain times to learn new patterns of data utility and participant willingness. Exploitation uses accumulated observations to prioritize survey timing at periods that are already known to be informative and likely to receive responses. By balancing exploration and exploitation over time, the system progressively improves survey timing decisions and achieves more effective data collection as additional information becomes available.

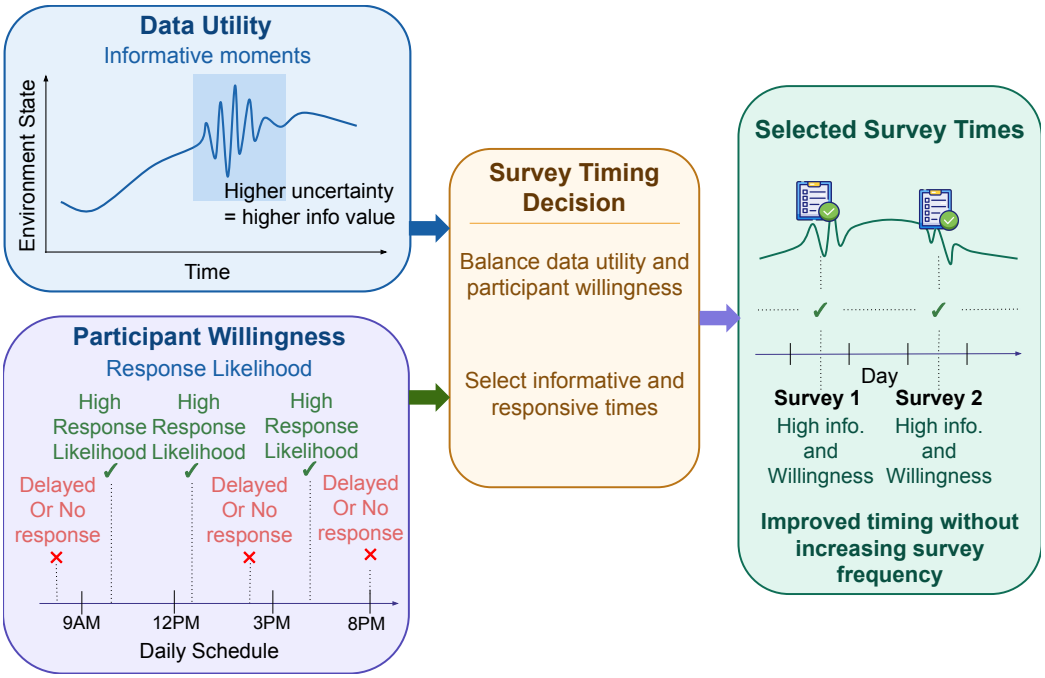


Fig. 3: Overview of the proposed survey timing method.

3.3 Solution Overview

The proposed method combines the key novelties above into a unified adaptive survey timing pipeline, as illustrated by Figure-3. At a high level, the system uses objective sensor data and historical survey responses to estimate two signals for each candidate survey time: the expected utility of a new response and the participant's willingness to respond. These signals are then combined to determine when a survey prompt should be issued.

More specifically, the method first uses a predictive model to estimate data utility under the current sensing conditions. Intuitively, candidate times associated with higher model uncertainty are treated as more valuable because an additional response is expected to provide greater informational gain. In parallel, the method estimates participant willingness from historical response behavior, capturing when a participant is more likely to complete a survey after receiving a prompt.

The method then combines these two signals into a single decision score and selects the survey time that is expected to be both informative and feasible. Because both utility and willingness estimates improve as more data is collected, the method continuously refines its decisions over time using an exploration-exploitation strategy. Our approach is also compatible with complementary engagement strategies, such as incentives or gamification, which can be used alongside adaptive timing to support sustained participation without increasing the survey budget.

Chapter 4 Adaptive Survey Timing Method (ASTM)

This Chapter presents the proposed utility-aware adaptive survey timing method (ASTM), which determines when to issue survey prompts by jointly considering data utility and participant willingness. As illustrated in Figure-4 the method consists of two components: a decision layer that evaluates candidate survey times, and a learning layer that updates these decisions over time. At each decision point, candidate times within predefined slots are scored based on (i) the expected utility under current sensing conditions and (ii) the likelihood of participant response. The final survey time is selected using an exploration-exploitation strategy that balances learning under uncertainty with prioritizing effective times.

4.1 Utility Estimation via Confidence Index

To estimate how informative a new survey response would be, we model the relationship between objective air quality measurements and subjective wellbeing responses using Gaussian Process Regression (GPR) [35]. For each participant, a separate GPR model is trained using previously collected survey responses and corresponding air quality data. Gaussian Process Regression has also been applied to environmental sensing and air quality modeling tasks where uncertainty estimation is important [36].

For each minute within a time slot, the GPR model predicts the expected subjective response and provides an associated confidence index CI(t). This confidence index reflects how certain the model is about its prediction at that moment, given the data collected so far. Lower confidence values indicate conditions that are less well represented in the existing data, where collecting an additional response is expected to provide higher informational value. This idea is conceptually related to active learning, where uncertainty sampling strategies prioritize observations expected to provide the greatest information gain [37]. The confidence index is normalized to the range [0, 1] and is used to guide survey timing decisions.

4.2 Participant Willingness Modeling

To estimate how likely a participant is to respond at a given time, we model participant willingness based on past response behavior. For each completed survey, we compute the delay between when the survey was sent and when it was completed. Shorter delays indicate higher willingness at that time of day, while longer delays or missed surveys indicate lower willingness. This assumption is supported by prior EMA research showing that response times can reflect underlying response processes and participant engagement during momentary reporting [38].

These observations are aggregated over time to produce a willingness score Willingness(t) for each minute and normalized to the range [0, 1]. This score represents the estimated likelihood that a participant will respond if a survey is sent at time t, capturing individual routines and temporal response patterns without requiring explicit input from participants. This design is also consistent with prior work showing that EMA compliance depends on prompt timing, participant context, and study design factors [39, 40].

4.3 Scoring Function

For each candidate survey time t, we compute a combined score:

$$\text{Score}(t) = \alpha_1 \cdot (1 - \text{CI}(t)) + \alpha_2 \cdot \text{Willingness}(t)$$

where CI(t) is the normalized confidence index from the GPR model at time t, and Willingness(t) is the estimated response likelihood. The term (1 - CI(t)) captures the utility signal by favoring uncertain conditions, while Willingness(t) captures the likelihood of response.

The weights α_1 and α_2 satisfy $\alpha_1 + \alpha_2 = 1$ and control the balance between informativeness and responsiveness. In this work, we use equal weighting $\alpha_1 = \alpha_2 = 0.5$ to avoid bias toward either factor and to maintain a balanced trade-off between learning informative conditions and ensuring participant response. A higher value of α_1 prioritizes more informative moments, while a higher

value of α_2 favors times when participants are more likely to respond. The framework allows these weights to be adjusted based on application-specific requirements.

4.4 Time Slot Evaluation

Survey timing decisions are limited to predefined time slots to reduce participant burden. Within each slot, candidate times are evaluated at fixed intervals and assigned scores based on the scoring function. These scores are then used by the exploration-exploitation policy to determine whether the current decision should prioritize high-scoring times or less certain times that may improve future estimates.

At each decision step, the system evaluates the score for all times within the predefined time slot and selects the time $t^* = \text{argmax Score}(t)$, which corresponds to the most suitable moment for sending the survey. If the maximum score is above a predefined threshold, the survey is sent at this selected time. Otherwise, if no candidate time satisfies the decision criteria, the survey is sent at the final minute of the time slot to ensure consistent sampling. This approach allows survey timing to adapt to both environmental conditions and participant behavior while maintaining predictable limits on survey frequency.

4.5 Learning Layer: Exploration-Exploitation Policy

Because both utility and willingness estimates depend on observed data, the method incorporates an exploration-exploitation algorithm to improve its decisions over time. As shown in Figure-4, newly collected survey responses are fed back into the system to update both the predictive model (for utility estimation) and the willingness model.

During exploration, the method prioritizes candidate times whose utility estimates are uncertain to learn about under-sampled sensing conditions and participant response patterns. The method uses a confidence threshold of 0.3 to distinguish the two modes. This threshold is chosen based on our preliminary study in [24], which showed that a 30% exploit and 70% explore balance

produced the lowest MAE in the earlier version of the system. Accordingly, times with CI(t) < 0.3 correspond to lower model confidence and higher uncertainty and are therefore treated as exploration. During exploitation, the algorithm prioritizes times that are already estimated to be both informative and likely to receive responses. Times with CI(t) >= 0.3 correspond to higher model confidence and are treated as exploitation, where learned response patterns are used to guide decisions. By balancing these two modes, the system progressively improves survey timing decisions while maintaining a fixed survey budget.

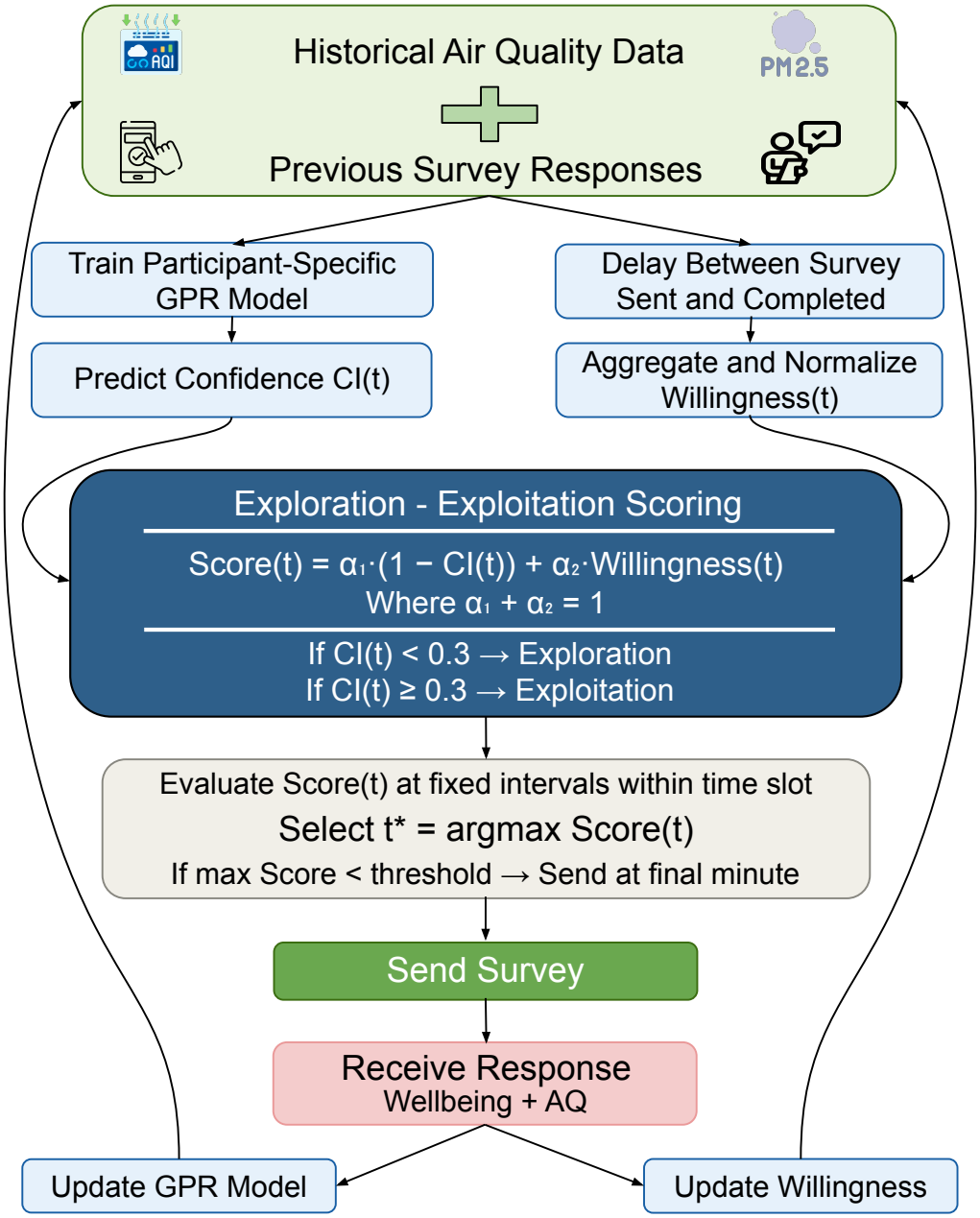


Fig. 4: Workflow of the proposed adaptive survey timing method (ASTM), showing how sensor data and response history are used to estimate utility and willingness, and how exploration-exploitation guides survey timing decisions.

Together, these components comprise the adaptive survey timing method, which dynamically balances data utility and participant willingness to improve data collection effectiveness over time.

Chapter 5 HCDC System Design and Implementation

This Chapter presents an HCDC system that demonstrates the proposed ASTM in practice. The system is built to support the air quality and human wellbeing case study introduced in Section 3.1 and serves as an end-to-end realization of the adaptive survey timing method under real deployment constraints. The implementation includes both cloud-side and user-facing components. On the cloud side, the system hosts the backend services for data ingestion, preprocessing, timestamp alignment, decision making using ASTM, model updating, and gamification feedback. On the user-facing side, participants carry portable sensors that automatically and continuously collect air quality data, receive SMS survey prompts on their mobile phones, and complete Qualtrics surveys [41] on their mood and stress levels. Figure-5 shows the overall architecture of the system and the main components involved in sensing, survey delivery, decision making, and gamification.

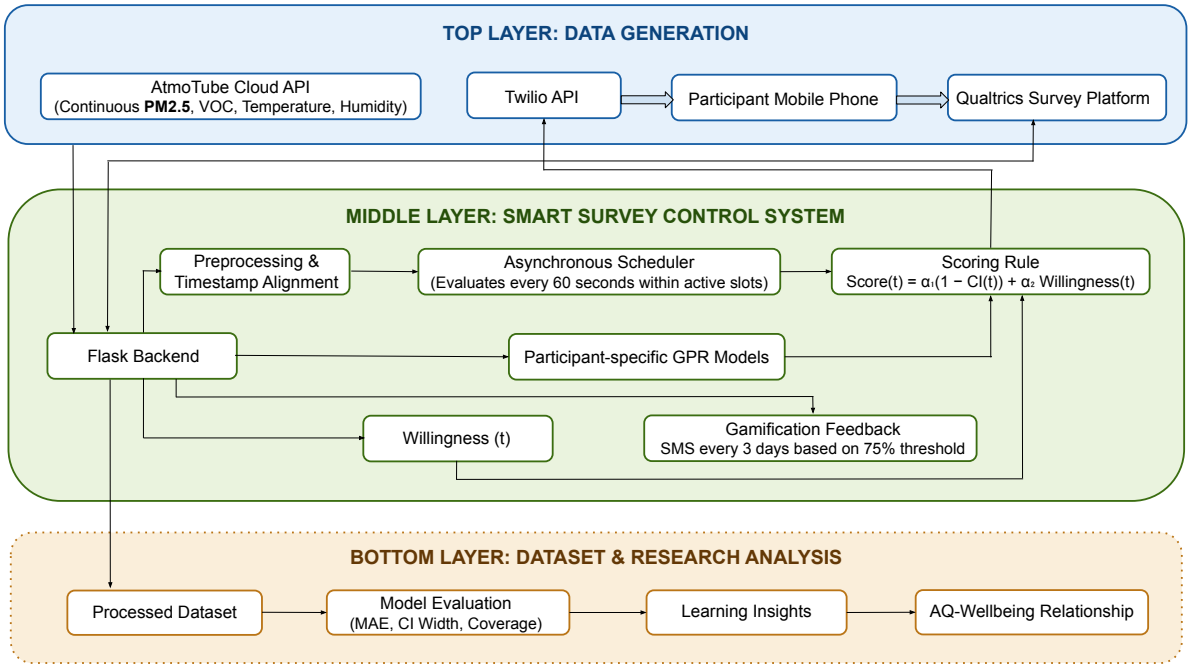


Fig. 5: System architecture of the HCDC system.

5.1 Backend Design

To support real-time adaptive decision making, we implement a Flask-based backend with an asynchronous scheduler that evaluates survey timing decisions every 60 seconds during active time slots. The backend services are deployed in the cloud so that survey scheduling, model updates, and data integration can run continuously during active study periods. For each participant, the scheduler maintains state information including recent survey outcomes, learned willingness profiles, and decision history.

At each decision point, the scheduler applies the ASTM scoring rule described in Chapter 4 to evaluate candidate survey times and select when to issue a prompt. All decisions are logged for later analysis, enabling us to examine how the ASTM method behaves in practice. Participant-specific GPR models are continuously retrained using accumulated survey responses and corresponding air quality measurements. This allows both the confidence index and willingness estimates to evolve over time, reflecting the dynamic conditions described in Chapter 3.

5.2 SMS Delivery

To deliver survey prompts in a real-world setting, surveys are sent to participants using the Twilio API [42]. Each message contains a unique Qualtrics survey link, which encodes basic metadata such as participant ID and send time. This ensures that each response can be accurately matched to the corresponding sensing context, which is critical for learning the relationship between air quality and wellbeing described in Section 3.1.

The system enforces the survey budget constraint by limiting delivery to predefined time slots. Once a participant completes a survey, the response is recorded via a Qualtrics callback and stored for subsequent model updates.

5.3 Data Collection Pipeline

The data collection pipeline implements the HCDC paradigm introduced in Chapter 1 and instantiated in the case study of Section 3.1. Objective air quality measurements are collected from users' portable sensors, transmitted to the cloud, and accessed via the AtmoTube cloud API [43]. This pipeline provides continuous readings of PM2.5.

Raw sensor data undergoes preprocessing and quality checks before being stored and aligned with survey timestamps. This process is consistent with prior work on validating lifelogging data collected from heterogeneous IoT-enabled sensing systems [44]. Specifically, the pipeline performs noise filtering and outlier removal (e.g., discarding extreme or physically implausible readings, such as sudden spikes or unrealistically high PM2.5 values that exceed typical environmental ranges for the study region) and handles missing or irregularly sampled data through interpolation or smoothing. Survey responses are validated for completeness and time-stamped at submission. These two data streams are then joined to form paired objective-subjective observations, which are used to train participant-specific GPR models for estimating CI values and guiding future survey timing decisions.

Accurate temporal alignment is critical in this system. As discussed in Chapter 3, the value of a subjective response depends on the sensing conditions under which it is collected. Any misalignment between sensor readings and survey timestamps would directly degrade the system's ability to learn meaningful relationships.

5.4 Gamification Component

While ASTM focuses on optimizing *when* to collect data, it is orthogonal to mechanisms that influence *whether* participants respond. In particular, ASTM can be naturally combined with existing incentive and engagement strategies, such as gamification, reminders, or monetary rewards. These approaches operate at different levels: ASTM improves the informativeness of collected data through timing decisions, while incentive mechanisms help sustain participant engagement over time. To demonstrate this complementary relationship, we incorporate a

lightweight gamification component alongside the timing algorithm as one form of engagement support. This design highlights that the proposed method is compatible with existing strategies such as gamification, reminders, or incentives, which can be integrated to sustain participation without altering the core timing mechanism.

5.4.1 User Categorization

Our gamification design is motivated by prior work showing that leaderboard-style feedback and game-based participation mechanisms can improve engagement in repeated survey settings [21, 22]. We categorize participants based on their survey completion rate using a 75% threshold. Participants with a completion rate above 75% belong to Category A, while those below 75% belong to Category B. This categorization enables tailored feedback based on engagement level.

Participants in Category A receive access to a leaderboard that displays their response rate, ranking, and achievement badges (e.g., Gold, Silver, Bronze), as shown in Figure-6. This design introduces competition and recognition to encourage continued participation.

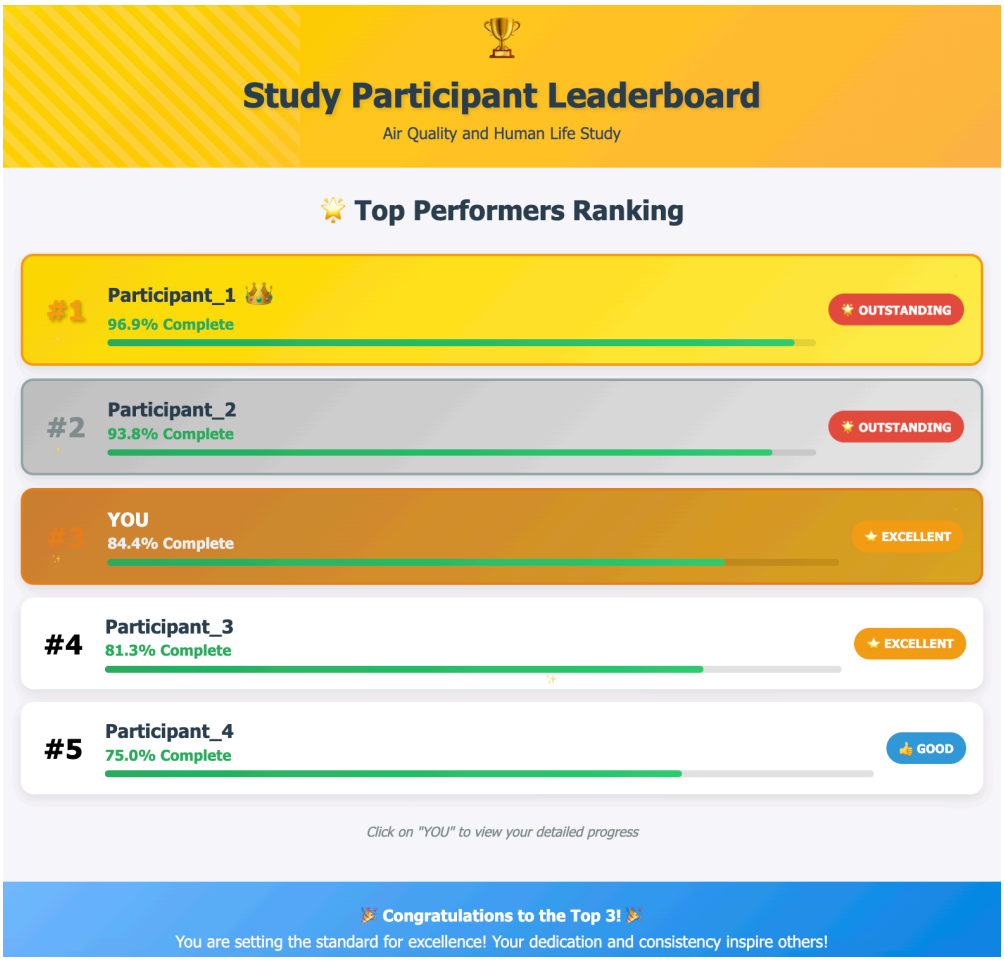


Fig. 6: Gamification leaderboard showing participant rankings, response rate, and achievement badges for Category A participants.

Participants in Category B receive a personalized progress page that shows how many surveys they have completed and how many additional responses are needed to reach the leaderboard threshold, as illustrated in Figure-7. It also highlights their improvement over time, helping participants understand their progress and encouraging increased engagement.

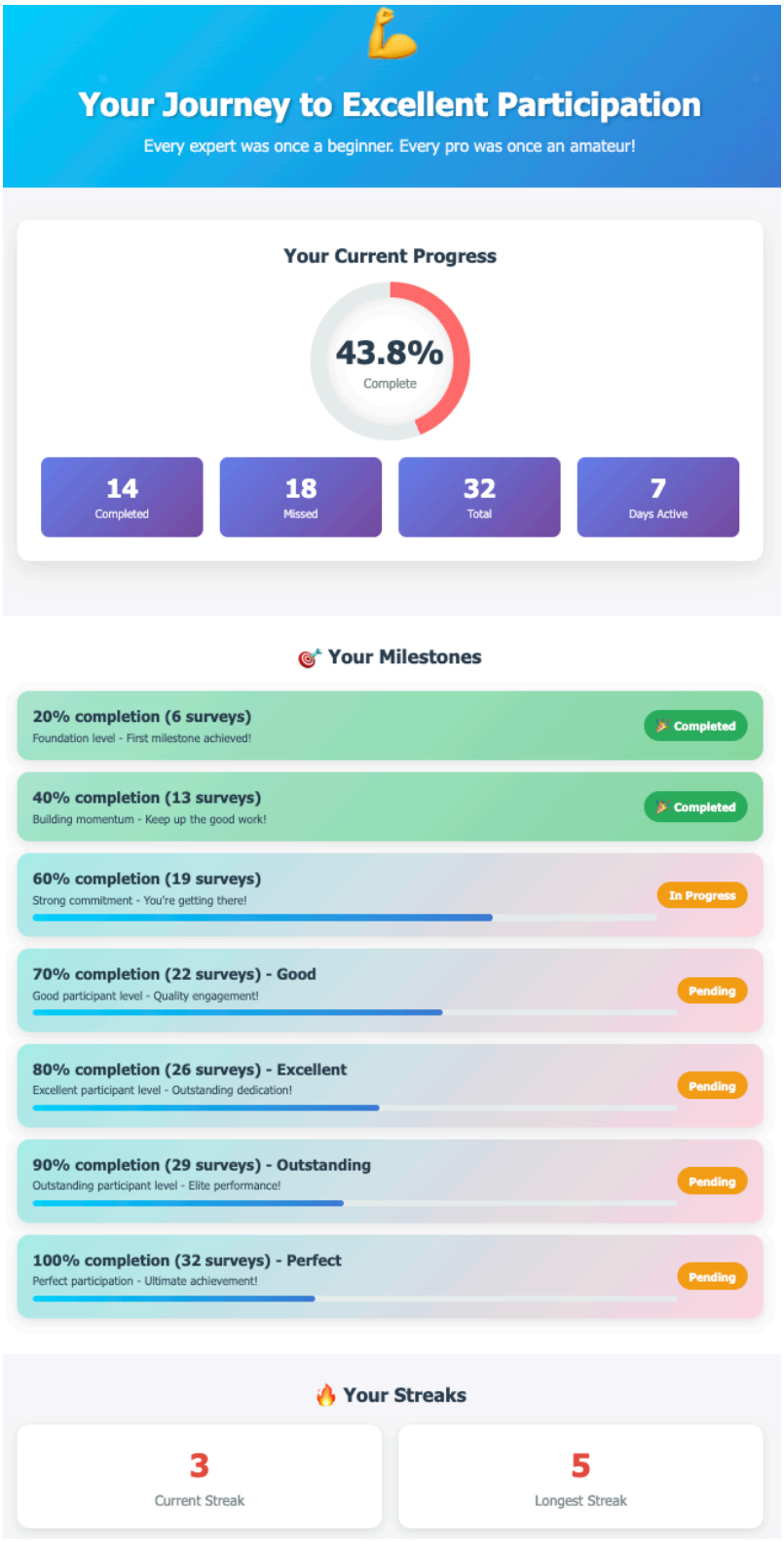


Fig. 7: Personalized progress page for Category B participants, showing survey completion milestones and improvement over time.

5.4.2 Personalized Feedback Design

To maintain engagement over time, the system sends personalized feedback messages every three days via SMS. Each message summarizes recent participation and includes a link to either the leaderboard or the progress page, depending on the participant's category.

The leaderboard reinforces participation through competition and visibility, while the progress page supports participants with lower engagement by providing clear and actionable feedback. This dual design helps maintain consistent participation throughout the study without increasing survey frequency or burden.

The bottom layer in Figure-5 represents the downstream analysis stage built on the collected dataset. After aligning sensor data with survey responses, the resulting data can be used to study the relationship between air quality and its impact on human wellbeing. The results of this analysis will be reported in a separate paper.

Together, these components demonstrate how the proposed ASTM can be deployed in a real HCDC system. The system integrates real-time sensing, adaptive survey timing, and complementary engagement mechanisms to support utility-aware data collection in practice, directly addressing the challenges identified in Chapter 3.

Chapter 6 Experiments and Results

This Chapter evaluates the proposed adaptive survey timing method through a real-world study. We first describe the experimental setup, including the study design, participant recruitment, data collection protocol, and evaluation metrics. We then present results to assess how the method affects model quality, learning efficiency, and participant engagement under different study conditions.

6.1 Experimental Setup

The experimental design evaluates the central claim of this work: that jointly considering data utility and participant willingness leads to improved participation and higher-quality data under a fixed survey budget. While the study is implemented in an air quality and wellbeing context, the experiment focuses on evaluating the effectiveness of the timing strategy rather than the underlying environmental relationships, which will be reported in a different paper.

6.1.1 Study Design

With Institutional Review Board (IRB) approval, we conduct a 3-week randomized controlled experiment with a between-subjects factorial design, with 121 participants assigned to four conditions: Random timing, EE algorithm, Gamification, and EE + Gamification. All conditions enforce identical time slot constraints. The Random condition selects survey times uniformly at random and serves as a baseline/control. The EE-only condition applies the timing optimization described in Chapter 4 without feedback. The third condition provides feedback on response (using gamification) but retains random timing. The combined condition applies both optimized timing and gamification feedback. This design isolates the effects of timing optimization and gamification on participation and data utility. By comparing these four conditions, the study directly tests whether adaptive timing, engagement support, or their combination produce the strongest improvement in data collection performance.

6.1.2 Participants and Recruitment

We recruit adult participants from Southeast Michigan who meet basic eligibility requirements, including smartphone ownership and residence in the study area for the study duration. Participants provide informed consent before enrollment. We recruit participants through outreach across the university campus. We post flyers with a QR code linking to the participation form in the academic buildings, student centers, and library. We also visit several classes and give a short 10-minute introduction explaining the purpose of the study and how students can participate. Interested students and community members scan the QR code and fill out the enrollment form to register.

After registration, we schedule participants in small batches for short on-boarding sessions of about 15 minutes. During these sessions, we hand out the air quality monitoring devices and explain how to carry and charge them. We also show participants how the SMS surveys will be sent and how they should complete them. This session helps participants clearly understand the study procedure before the data collection period begins.

6.1.3 Data Collection Protocol

Participants carry portable sensors that measure PM2.5 and other air quality variables for three weeks. The sensors record data continuously during the study period. We send SMS surveys to collect self-reports on how people feel (happy, engaged, and relaxed) and environment (whether it is sunny outside). Each participant receives four surveys per day. Participants are required to complete the survey within a short time after they receive the message. This helps us match the survey response with the sensor data at that moment.

We divide the day into four time slots: 8:00-11:00 AM, 11:00 AM-2:00 PM, 2:00-5:00 PM, and 5:00-8:00 PM. Each participant receives one survey SMS for each time slot. This structure spreads the surveys across the day and captures changes in both air quality and wellbeing. Each survey contains a small number of short questions to keep the task simple for participants. After a

participant submits a survey, we match the response time with the sensor readings. This step helps us link environmental conditions (specifically air quality) with the participant's reported wellbeing.

6.1.4 Evaluation Metrics

We evaluate the proposed method using two main outcome categories. The first category measures participation. We use response rate to measure how often participants complete the surveys that we send. The second category measures data utility. This category shows how much each survey response improves the air quality-wellbeing model. These measures help us test whether adaptive survey timing produces better data and not just more responses.

We measure model performance using standard regression metrics, such as mean absolute error (MAE). MAE calculates the average difference between the predicted wellbeing values and the survey responses reported by participants. A lower MAE means that the model predicts wellbeing more accurately. We also track how the model error changes during the study. If the error decreases faster, the system learns useful patterns earlier. This result shows that the data collection method improves the efficiency of the learning process.

6.2 Results

We evaluate the system along two key dimensions: model quality and participation. First, we analyze how exploration-exploitation (adaptive) timing affects learning efficiency and prediction performance. Next, we examine its impact on participant responsiveness and engagement across different study conditions.

6.2.1 Model Quality and Learning Efficiency

We first evaluate whether exploration-exploitation (EE) timing improves model quality compared to random survey timing.

a. Model Accuracy Comparison:

Figure-8 compares the learned PM2.5 and wellbeing relationships under EE and random timing.

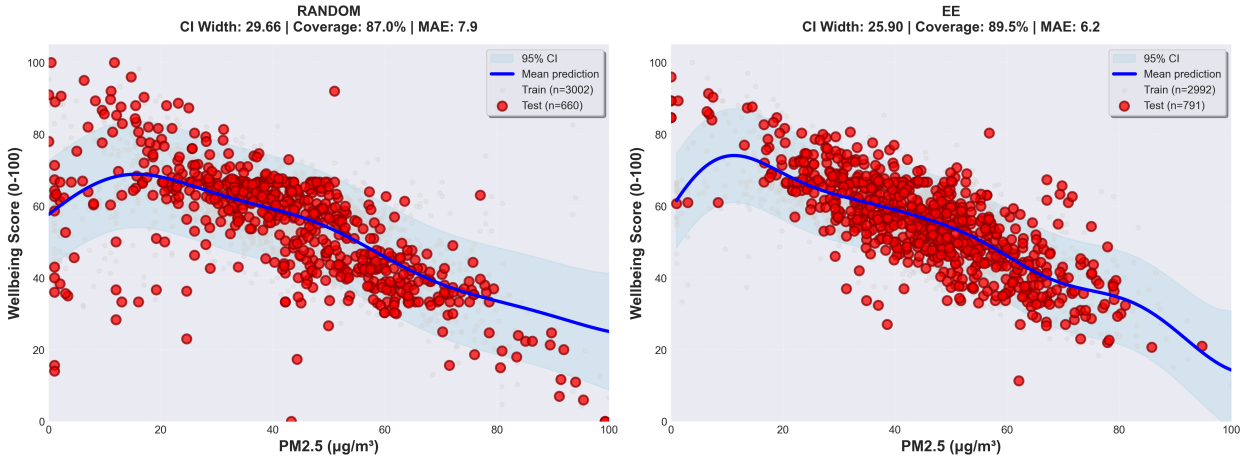


Fig. 8: Learned PM2.5 and wellbeing relationships under EE and random timing.

EE produces a narrower confidence interval (CI width = 25.90) compared to random timing (CI width = 29.66). A narrower CI reflects higher estimation precision, as the model is able to constrain the range of plausible values more tightly given the available data. In other words, the predictions are more informative and less uncertain. Importantly, both approaches achieve comparable coverage (EE: 89.5%, Random: 87.0%), demonstrating that the tighter intervals under EE do not compromise reliability. Together, these results indicate that EE improves predictive efficiency by reducing uncertainty while maintaining accurate uncertainty calibration.

In terms of prediction accuracy, EE achieves a lower mean absolute error (MAE = 6.2) compared to random timing (MAE = 7.9). The improvement in MAE (6.2 vs 7.9, a 21.5% reduction) suggests that EE-timed responses capture more informative data points. Together, these results indicate that adaptive timing improves model confidence while maintaining comparable predictive accuracy.

These results suggest that the EE strategy helps the model focus on more informative data points. By selecting survey times when the system expects higher uncertainty, the model receives

responses that improve its understanding of the air quality-wellbeing relationship. As a result, the prediction intervals become narrower while maintaining similar coverage levels. This indicates that the model becomes more confident about its predictions without sacrificing reliability.

b. MAE Learning Curves:

To evaluate how improved alignment translates into predictive performance, we analyze model error over time. Figure-9 illustrates the temporal evolution of mean absolute error (MAE) across study days for both EE and random timing. Results are reported from Day 8 onward, since the first week is used as a cold-start phase with random sampling to collect initial data for training the GPR models.

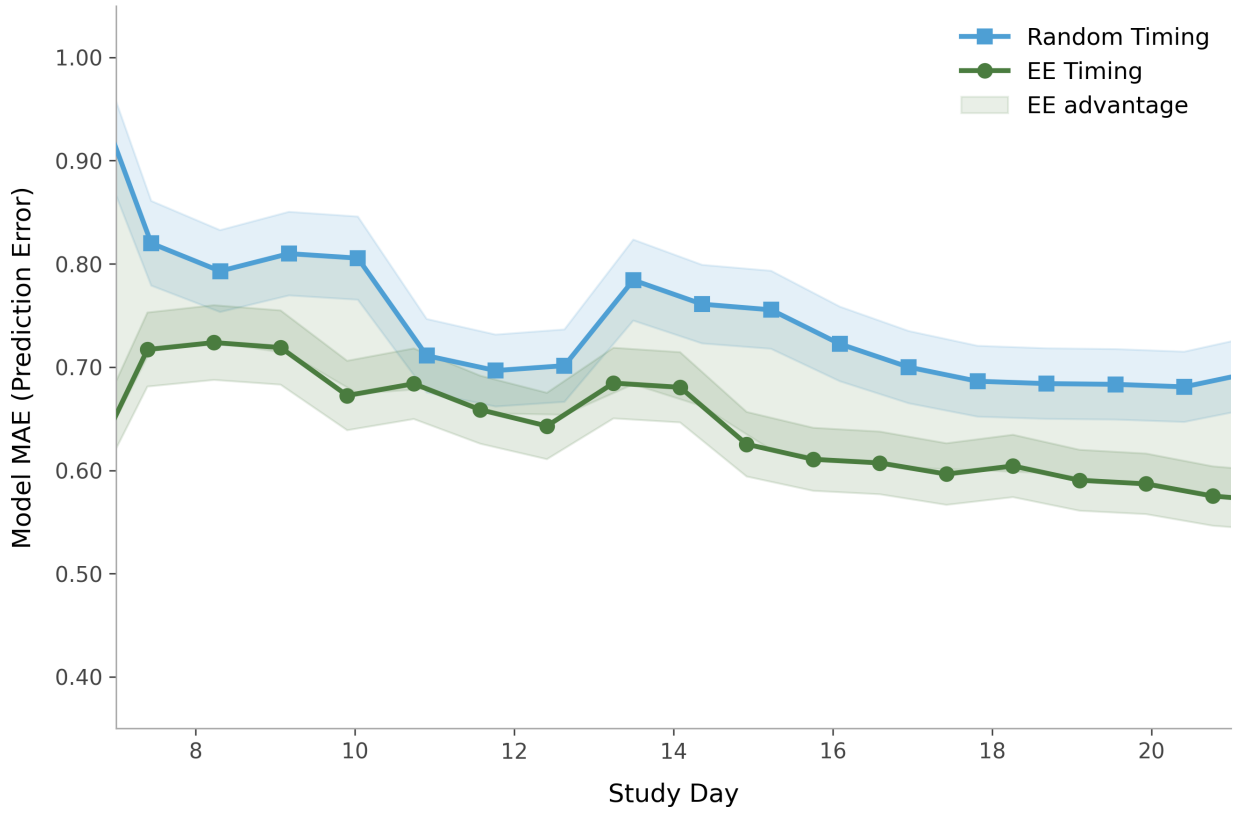


Fig. 9: Model MAE over time for EE and random timing conditions.

EE demonstrates a consistently faster reduction in prediction error compared to random sampling. In particular, EE reaches a target error level approximately twice as quickly as the random baseline. Throughout the study period, the EE curve remains below the random curve, indicating sustained improvement under equivalent survey budgets. These findings confirm that targeting high-uncertainty windows accelerates learning and improves model efficiency, supporting the claim that adaptive timing yields better data rather than simply more data.

This result further confirms that the EE strategy improves learning efficiency by prioritizing more informative observations. By selecting survey times that are expected to provide more useful information, the system gathers responses that improve the model faster. As a result, the model reaches lower prediction error with fewer observations. This improvement is important because it means the system can build accurate models without increasing the number of surveys sent to participants.

6.2.2 Participation and Engagement

This subchapter evaluates how the proposed method affects participant behavior and engagement. We analyze missed survey rates and response patterns across different study conditions to understand how adaptive timing and gamification influence participation over time.

a. Missed Survey Rate Analysis:

We next evaluate the missed survey rate across study conditions, as shown in Figure-10.

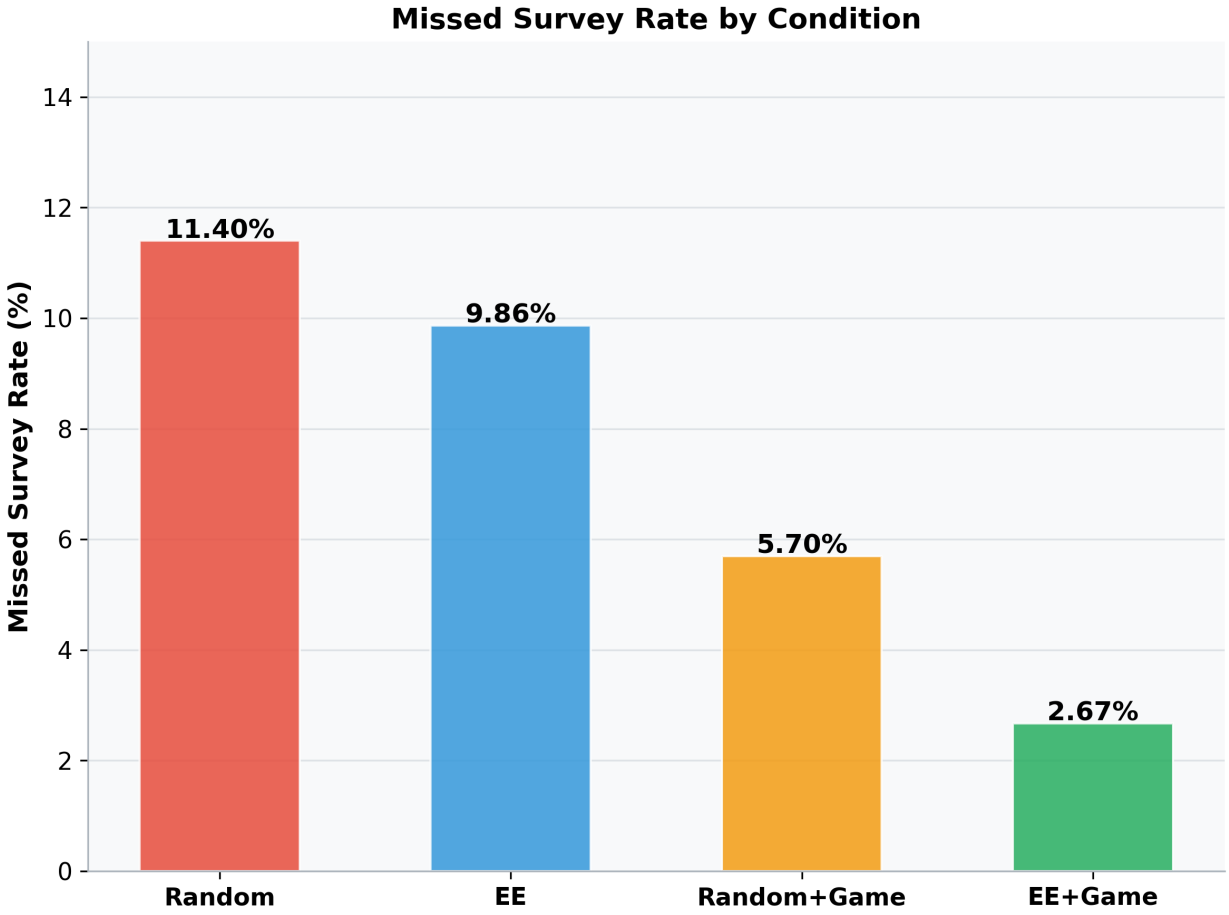


Fig. 10: Missed survey rates across the four experimental conditions.

Random timing shows the highest missed survey rate at 11.40%, while EE timing shows a lower missed survey rate of 9.86%. This result indicates that adaptive timing helps reduce the number of missed surveys. When gamification is introduced, the missed survey rate decreases further. Random + Gamification shows a missed rate of 5.70%, while EE + Gamification achieves the lowest missed survey rate at 2.67%. These results show that adaptive timing and gamification both help reduce missed surveys. The combined EE + Gamification condition achieves the best participant responsiveness.

These results show that both adaptive timing and gamification play a role in improving participation. The EE strategy reduces missed surveys by sending messages at times when

participants are more likely to respond. Gamification further improves participation by providing motivation and feedback. When both strategies are combined, the system achieves the lowest missed survey rate among all conditions.

b. Gamification Impact:

To examine engagement dynamics over time, Figure-11 presents weekly response rates for gamified and non-gamified conditions.

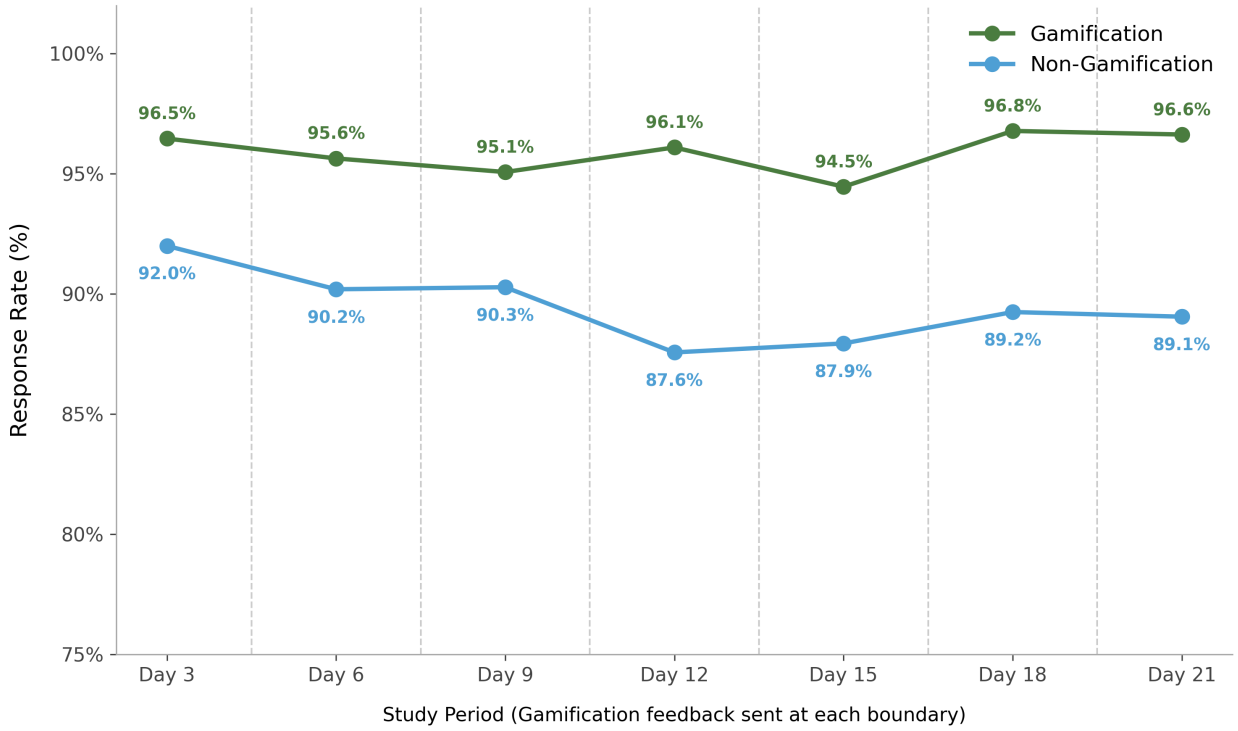


Fig. 11: Response rates at three-day intervals for gamified and non-gamified conditions over the three-week study.

At Day 3, when the first feedback was delivered, the gamified condition maintains a high response rate (96.5%), compared to 92.0% in the non-gamified condition. Over time, the non-gamified condition shows a gradual decline, reaching lower response levels around Day 12 (87.6%) and remaining below 90% thereafter. In contrast, the gamified condition remains consistently high across all intervals, with only minor fluctuations between approximately 94.5%

and 96.8%. This pattern indicates that periodic personalized feedback helps maintain consistently high participation throughout the study duration.

Another important observation is that participation in the non-gamified condition gradually decreases over time. This time-decay suggests that participants, unsurprisingly, lose motivation during longer studies. In contrast, the gamified condition maintains a consistently high response rate throughout the three weeks. This stability indicates that periodic feedback and progress visibility help participants stay engaged in the study.

Overall, the results consistently show that adaptive survey timing improves both data quality and participant responsiveness compared to random survey scheduling. The experimental results demonstrate clear differences between the four study conditions in terms of response rate, data utility, and engagement sustainability. In particular, the exploration-exploitation timing strategy improves both participation and model performance when compared to random survey timing. The addition of gamification further stabilizes participation over time without increasing survey frequency. These results not only demonstrate improvements in participation and model performance but also raise important questions about how and why timing influences data quality. The following discussion interprets these findings and examines their broader implications.

6.3 Alpha Sensitivity Analysis

The scoring function used in this study assigns equal weight to data utility and participant willingness, setting $\alpha_1 = \alpha_2 = 0.5$. To evaluate whether this choice provides a reasonable balance, we conduct an alpha sensitivity analysis using five different configurations: utility-dominant ($\alpha_1 = 0.9$, $\alpha_2 = 0.1$), utility-leaning ($\alpha_1 = 0.7$, $\alpha_2 = 0.3$), balanced ($\alpha_1 = 0.5$, $\alpha_2 = 0.5$), willingness-leaning ($\alpha_1 = 0.3$, $\alpha_2 = 0.7$), and willingness-dominant ($\alpha_1 = 0.1$, $\alpha_2 = 0.9$).

Since repeating the full user study under each alpha configuration is not feasible, we use a simulation to compare how different alpha values affect response rate and AQ coverage. The

simulation is not intended to reproduce the exact behavior of every participant in the study. Instead, it provides a controlled setting for testing how the scoring function behaves when participant response behavior is modeled from collected data.

6.3.1 User Behavior Model

To simulate participant response behavior, we use a parametric response model whose structure is defined mathematically, while its parameters are learned from the collected participant response data. This allows the simulation to generate synthetic response outcomes that reflect observed participant behavior without directly replaying the original study data.

The model captures two important factors. First, participants are more likely to respond when a survey is sent near a time when they have historically responded quickly. Second, response probability decreases as more time passes after the survey is sent. The response probability is defined as:

$$P_i(\text{respond} \mid t, d) = P_\max, i \cdot \exp\left(-(t - h_i)^2 / 2\sigma_i^2\right) \cdot \exp\left(-\lambda_i \cdot d\right)$$

where t is the survey send time, d is the elapsed delay after the survey is sent, h_i is the learned preferred response time for participant i, σ_i controls the width of the participant's comfortable response window, λ_i controls how quickly response probability decreases with delay, and P_max,i is the maximum response probability.

The parameters are estimated from the collected response data. The preferred response time h_i is estimated as the time window in which the participant historically showed the fastest response tendency. The window width σ_i is estimated from the spread of response times around h_i. The decay rate λ_i is estimated as ln(2) divided by the participant's median response delay, meaning response probability decreases to half its value at the median delay. The maximum response probability P_max,i is estimated from the condition-level survey completion rate observed in the study.

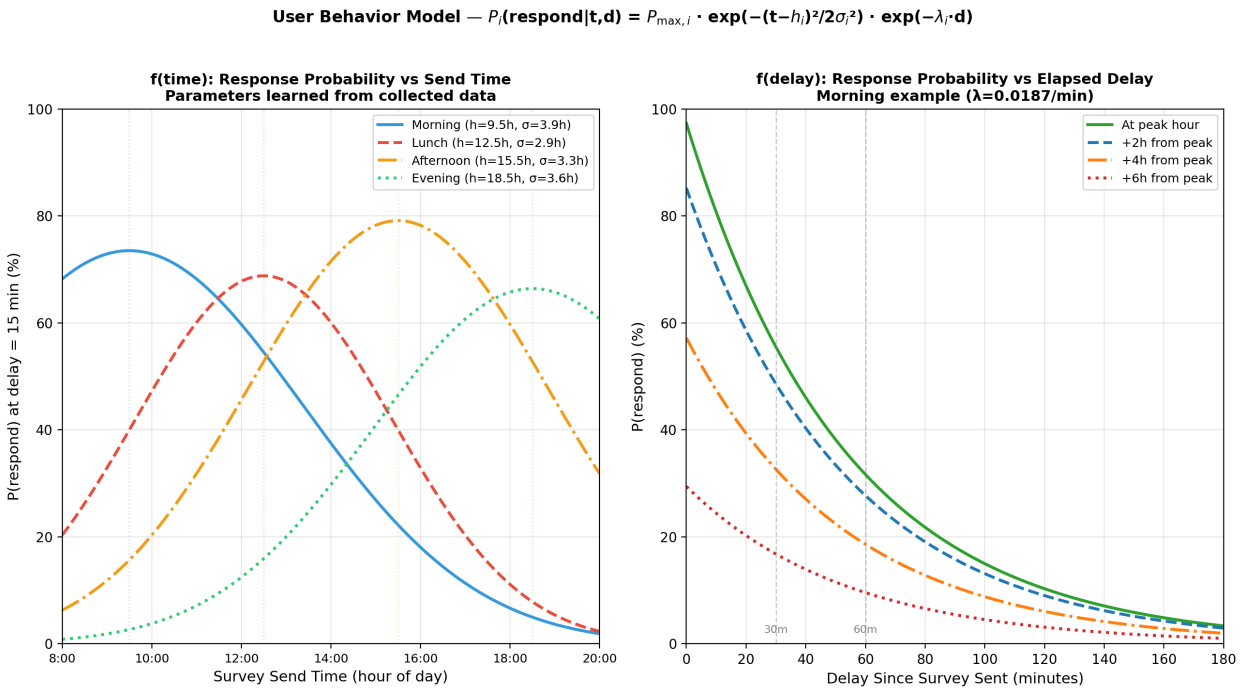


Fig. 12: User behavior model used in the simulation

Figure 12 illustrates the learned user behavior model. The left panel shows representative learned response patterns across participants, where each curve reflects a different preferred response window derived from the data. The right panel shows how response probability decreases with elapsed delay after the survey is sent. This model provides a data-learned basis for generating simulated responses in the alpha sensitivity analysis.

6.3.2 Simulation Setup

The simulation evaluates candidate survey times every 15 minutes between 8:00 AM and 8:00 PM, giving 49 candidate times per day. For each candidate time t, the system computes the same scoring function used in the proposed method:

$$\text{Score}(t) = \alpha_1 \cdot (1 - \text{CI}(t)) + \alpha_2 \cdot \text{Willingness}(t)$$

where CI(t) is the confidence index from the GPR model and Willingness(t) = P_i(respond | t, d = 0) is the response probability at send time t with zero elapsed delay, normalized to [0, 1]

across candidate times. The term $(1 - CI(t))$ represents data utility by favoring uncertain or less-represented AQ conditions, while the willingness term favors survey times when the participant is more likely to respond.

For each alpha configuration, the simulation selects the candidate time with the highest score. After a survey is sent at the selected time, the simulation checks for a response every 15 minutes. At each delay value, the response probability is computed using the learned behavior model, and a random draw is used to determine whether the participant responds. The first successful draw is treated as the response time. If no response occurs within 180 minutes, the survey is counted as missed. This setup allows us to compare how different alpha values influence two outcomes: the simulated response rate and the diversity of selected AQ conditions.

6.3.3 Results

Table 1 summarizes the simulated response rates and AQ coverage across all five configurations.

α_1	α_2	Strategy	Response Rate	AQ Coverage
0.9	0.1	Utility-Dominant	64.0%	12.32
0.7	0.3	Utility-Leaning	65.3%	12.21
0.5	**0.5**	**Balanced (Used)**	**73.9%**	**12.19**
0.3	0.7	Willingness-Leaning	81.4%	11.91
0.1	0.9	Willingness-Dominant	82.6%	11.74

The results show a clear trade-off between data utility and participant willingness.

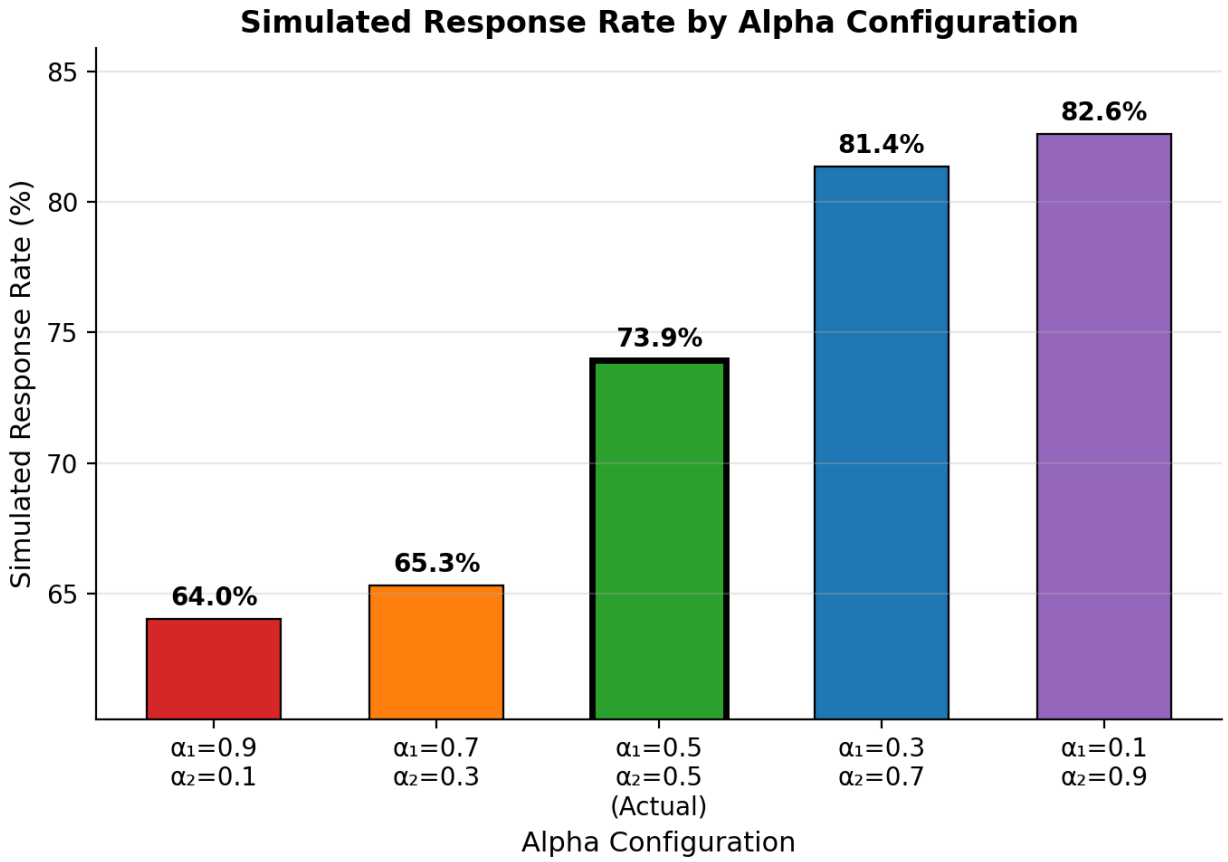


Fig. 13: Simulated response rates for different alpha configurations

As shown in Figure 13, when the willingness weight increases, the simulated response rate increases from 64.0% in the utility-dominant setting to 82.6% in the willingness-dominant setting. This happens because the system increasingly selects survey times close to the participant's learned preferred response window.

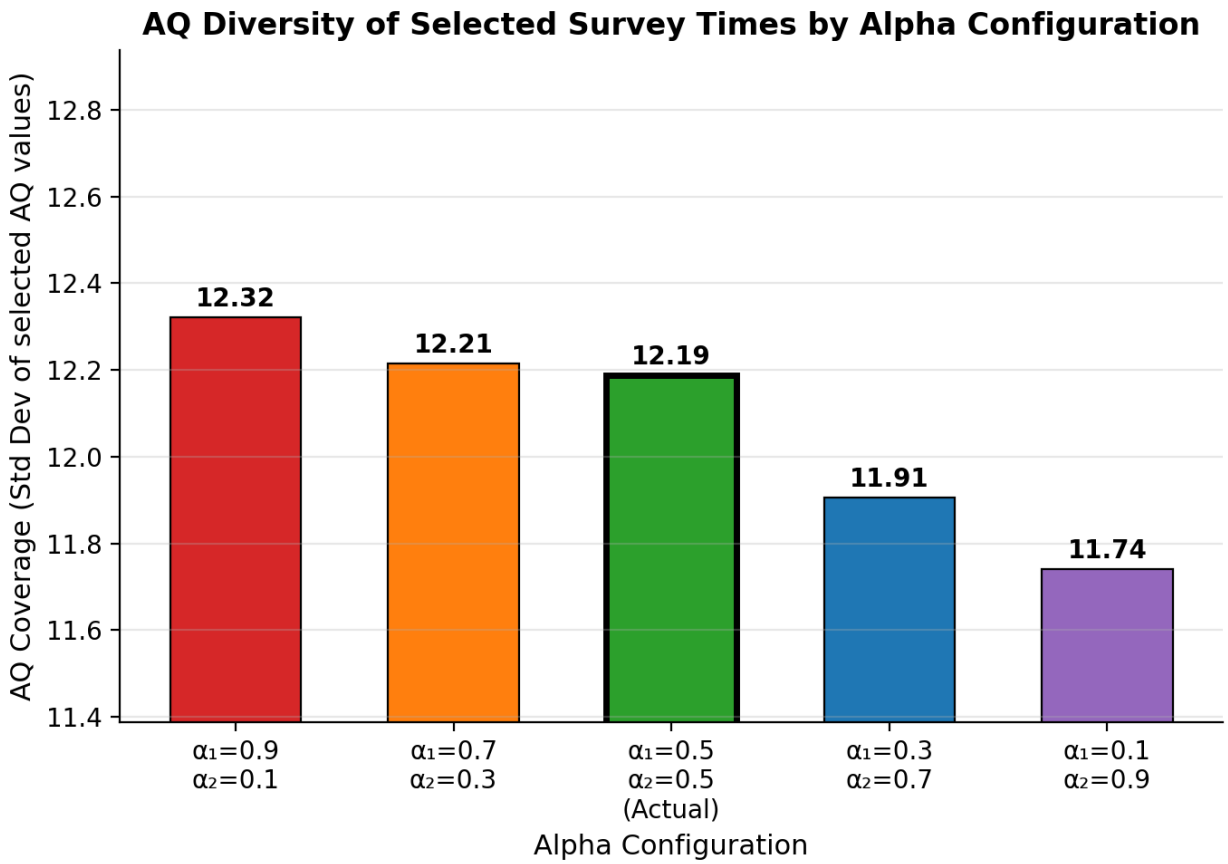


Fig. 14: AQ coverage under different alpha configurations

However, this increase in response rate comes with a reduction in AQ coverage. As shown in Figure 14, AQ coverage decreases from 12.32 in the utility-dominant setting to 11.74 in the willingness-dominant setting. This means that when the system prioritizes willingness too strongly, it tends to send surveys at convenient times, but those times cover a narrower range of AQ conditions.

The balanced configuration, where $\alpha_1 = \alpha_2 = 0.5$, achieves a response rate of 73.9% and an AQ coverage value of 12.19. This setting sits between the two extremes: it avoids the lower response rates observed in utility-dominant configurations while also avoiding the stronger reduction in AQ coverage observed in willingness-dominant configurations.

It is important to note that this simulation is designed to explore different alpha trade-off scenarios, not to derive or justify the specific alpha values used in the real study. The choice of $\alpha_1 = \alpha_2 = 0.5$ in the experiment reflects a practical starting point that gives equal importance to data utility and participant willingness. The simulation shows that this balanced setting maintains reasonable performance across both response rate and AQ coverage, while configurations that prioritize only one objective introduce a clear cost to the other. The MAE improvement reported in Section 6.2 reflects the full effect of adaptive timing across the three-week deployment and should be interpreted independently of this sensitivity analysis.

6.4 Key Interpretations

The results show that adaptive survey timing can improve both participation and data quality in human-centered data collection. They also suggest that the value of a survey depends not only on whether a participant responds, but also on when that response is collected. More broadly, these findings highlight the need to balance user burden, response likelihood, and data utility when designing human-centered data collection systems.

The exploration-exploitation timing strategy improves how efficiently the system learns the relationship between air quality and wellbeing. The EE approach produced models with lower prediction error and higher predictive accuracy compared to random timing. This suggests that the algorithm successfully targets moments when additional responses are more informative for the model. Instead of repeatedly collecting responses during stable environmental conditions, the system prioritizes moments where the model is uncertain.

Another important observation is that optimizing for informative moments does not reduce participation. The response rate under EE timing is slightly higher than random timing. This observation is consistent with prior EMA studies showing that response behavior depends on timing, context, and participant characteristics [40]. This indicates that the algorithm successfully balances two factors: collecting informative data and selecting moments when participants are

likely to respond. By incorporating participant willingness into the decision process, the system avoids sending surveys at inconvenient times while still targeting useful sampling moments.

The results also show that combining this adaptive survey system with engagement feedback provides the best overall outcome. The EE + gamification condition achieves both the highest response rate and the strongest model performance. This suggests that timing optimization and engagement support address different aspects of human-centered studies. Timing improves the usefulness of collected data, while feedback helps maintain participation throughout the study period.

Chapter 7 Conclusion

This thesis addressed the problem of improving data quality in human-centered data collection without overburdening participants. In studies that combine sensor measurements with human responses, the timing of surveys plays a critical role in determining both participation and data usefulness.

We propose an exploration-exploitation based (adaptive) timing strategy that selects survey moments by balancing two signals: data utility and participant willingness to respond. Informativeness is estimated using prediction uncertainty from the AQ-wellbeing model, while participant willingness is derived from historical response patterns. The system also incorporates a feedback-based gamification mechanism designed to maintain engagement without increasing survey frequency.

A three-week randomized controlled experiment demonstrates that adaptive timing significantly improves data collection performance. The results show higher response rates, better model accuracy, and more efficient use of collected responses compared to random survey timing. The combined EE and gamified feedback approach provides the best overall performance by improving both participation and data quality.

These findings highlight the importance of optimizing when data is collected in human-centered sensing systems. Rather than increasing the number of surveys, adaptive timing allows researchers to collect more informative data while maintaining a positive participant experience. More broadly, this work suggests that timing-aware data collection can serve as a practical design principle for future systems that combine continuous sensing with momentary self-report in everyday life. An important implication of this work is that improving data collection does not necessarily require increasing participant effort. Instead, better timing strategies can make each interaction more valuable. This perspective can guide the design of future human-centered sensing

systems that aim to balance user experience with data effectiveness. This work demonstrates that optimizing when data is collected is as important as what data is collected in human-centered sensing systems.

7.1 Limitations and Future Directions

Despite these promising results, several limitations should be considered. First, the study duration is relatively short. A three-week deployment allows evaluation of the system and participation patterns, but longer studies are needed to understand long-term engagement and possible survey fatigue.

Second, the study focuses on a single environmental context and a specific community. Environmental conditions and participation behavior may differ across other populations or geographic locations. Additional deployments would help validate the general applicability of the proposed framework.

Another limitation relates to the self-reported wellbeing responses. The study relies on participants actively responding to surveys throughout the day. Prior work has shown that higher sampling frequency and longer questionnaires can increase perceived burden and reduce response quality in experience sampling studies [45]. If participants miss surveys or respond inconsistently, the quality of the collected data may decrease. In addition, self-reported responses may be influenced by factors beyond air quality exposure, such as personal events or daily activities. As a result, variability in how participants interpret and report their feelings may introduce noise into the dataset.

These limitations also point to several directions for future work, including longer deployments, broader participant populations, and more adaptive models that respond to changing behavior over time. Future studies can also explore whether the same timing strategy remains effective over longer periods and under different environmental conditions. Overall, the results suggest that adaptive survey timing is a promising direction for improving human-centric data

collection systems. By focusing on collecting better data rather than simply more data, the proposed approach improves both model learning and participant experience.

References

[1] M. J. Jafari and others, “The Effect of Noise Exposure on Cognitive Performance and Brain Activity Patterns,” The Open Biomedical Engineering Journal, vol. 13, pp. 44–52, 2019.

[2] Y. Liao and others, “The Acute Relationships Between Affect, Physical Feeling States, and Physical Activity in Daily Life: A Review of Current Evidence,” Frontiers in Psychology, vol. 6, pp. 1975, 2015.

[3] P. Dolan and K. Laffan, “Bad air days: The effects of air quality on different measures of subjective well-being,” Journal of Benefit-Cost Analysis, vol. 7, no. 1, pp. 147–195, 2016.

[4] X. Zhang, X. Zhang and X. Chen, “Happiness in the air: How does a dirty sky affect mental health and subjective well-being?,” Journal of Environmental Economics and Management, vol. 85, pp. 81–94, 2017.

[5] Y. Liu, K. Zhu, R.-L. Li, Y. Song and Z.-J. Zhang, “Air pollution impairs subjective happiness by damaging health,” International Journal of Environmental Research and Public Health, vol. 18, no. 19, 2021.

[6] C. Herrera and P. Cabrera-Barona, “Impact of perceptions of air pollution and noise on subjective well-being and health,” Earth, vol. 3, no. 3, pp. 906–926, 2022.

[7] J. Burke and others, “Participatory sensing,” in Proceedings of the ACM SenSys Workshop on World-Sensor-Web, pp. 117–134, 2006.

[8] M. Csikszentmihalyi and R. Larson, “Validity and reliability of the experience-sampling method,” Journal of Nervous and Mental Disease, vol. 175, no. 9, pp. 526–536, 1987.

[9] M. N. K. Boulos and others, “Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management,” International Journal of Health Geographics, vol. 10, pp. 67, 2011.

[10] N. van Berkel and D. Ferreira and V. Kostakos, “The experience sampling method on mobile devices,” ACM Computing Surveys, vol. 50, no. 6, 2017.

[11] V. P. Cornet and R. J. Holden, “Systematic review of smartphone-based passive sensing for health and well-being,” Journal of Biomedical Informatics, vol. 77, pp. 120–132, 2018.

[12] R. Kraft and M. Reichert and R. Pryss, “Mobile crowdsensing in ecological momentary assessment mhealth studies: A systematic review and analysis,” Sensors, vol. 24, no. 2, pp. 472, 2024.

[13] J. Wannenburg and R. Malekian, “Physical Activity Recognition From Smartphone Accelerometer Data for User Context Awareness Sensing,” IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 47, no. 12, pp. 3142–3149, 2017.

[14] S. R. Porter and M. E. Whitcomb and W. H. Weitzer, “Multiple surveys of students and survey fatigue,” New Directions for Institutional Research, vol. 2004, no. 121, pp. 63–73, 2004.

[15] H. Nicholl and M. Tracey and J. Begley, “Strategies to improve response rates to web surveys: A literature review,” International Journal of Nursing Studies, vol. 123, 2021.

[16] M. G. Smith and M. Witte and S. Rocha and M. Basner, “Effectiveness of incentives and follow-up on increasing survey response rates and participation in field studies,” BMC Medical Research Methodology, vol. 19, no. 230, 2019.

[17] V. Mishra and K. Caine and B. Lowens and D. Kotz and S. Lord, “Investigating contextual cues as indicators for EMA delivery,” Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 2017, pp. 935–940, 2017.

[18] J. Delobelle and E. Lebuf and S. Compernolle and T. Vetrovsky and J. Van Cauwenberg and R. Cimler and J. Kuhnova and D. Van Dyck, “Sensor-triggered ecological momentary assessment in physical activity and sedentary behaviour research among Belgian community-dwelling elderly,” BMJ Open, vol. 15, pp. e096327, 2025.

[19] Z. D. King and M. Khalid and H. Yu and K. Shibuya and K. Zanna and M. Majd and R. L. Brown and Y. Shen and G. Kypriotakis and T. Vaessen and C. P. Fagundes and A. Sano, "Machine learning-based context aware EMAs: An offline feasibility study," arXiv preprint arXiv:2506.15834, 2025.

[20] S. Shiffman and A. A. Stone and M. R. Hufford, "Ecological momentary assessment," Annual Review of Clinical Psychology, vol. 4, pp. 1–32, 2008.

[21] A. Mavletova, "A gamification effect in longitudinal web surveys among children and adolescents," International Journal of Market Research, vol. 57, no. 3, pp. 413–438, 2015.

[22] E. Dejonckheere and others, "Real-time incentivizing survey completion with game-based rewards in experience sampling research may increase data quantity but reduces data quality," Computers in Human Behavior, vol. 150, 2024.

[23] M. A. Russell and J. M. Gajos, "Intensive longitudinal methods for investigating behaviors and feelings in everyday life: An introduction and overview," Journal of Social and Clinical Psychology, vol. 36, no. 8, pp. 581–597, 2017.

[24] S. Baravkar and Z. Song and Q. Zhu and N. Czap and H. Czap, "A data collection system for assessing the impact of air quality on human wellbeing," in Proc. IEEE 4th Int. Conf. Computing and Machine Intelligence (ICMI), pp. 1–6, 2025, doi: 10.1109/ICMI65310.2025.11141059.

[25] A. Velu and J. Ramesh and A. Ahmed and F. A. Aloul, "Integrating Fitbit wearables and self-reported surveys for machine learning-based state-trait anxiety prediction," Applied Sciences, vol. 15, no. 19, pp. 10519, 2025.

[26] N. Kwon and D. Lee and S. Hwang and S. Yang and I. Youn and H.-J. Moon and S. Han, "Enhancing the accuracy of mental health assessments through the integration of self-report and objective measures," Acta Psychologica, vol. 243, pp. 105432, 2025.

[27] S. Conklin and J. R. Dietch and G. Kargosha and F. Luyster and M. Atwood and M. S. Tenan and G. Zammit and N. Banerjee and J. Brooks, “Integration of sensor-based and self-reported metrics in a sleep diary: A pilot exploration,” Behavioral Sleep Medicine, pp. 1–12, 2024.

[28] L. P. de Vries and B. F. M. Baselmans and M. Bartels, “Smartphone-based ecological momentary assessment of well-being: A systematic review and recommendations for future studies,” Journal of Happiness Studies, vol. 22, pp. 2361–2408, 2021.

[29] S. S. Intille and J. Rondoni and C. Kukla and I. Ancona and L. Bao, “A context-aware experience sampling tool,” in Proc. CHI '03 Extended Abstracts on Human Factors in Computing Systems, pp. 972–973, 2003.

[30] R. Wang and F. Chen and Z. Chen and T. Li and G. Harari and S. Tignor and X. Zhou and D. Ben-Zeev and A. T. Campbell, “StudentLife: Assessing mental health, academic performance and behavioral trends of college students using smartphones,” in Proc. ACM Int. Joint Conf. Pervasive and Ubiquitous Computing (UbiComp), 2014.

[31] S. Aminikhanghahi and M. Schmitter-Edgecombe and D. J. Cook, “Context-aware delivery of ecological momentary assessment,” IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 10, pp. 2858–2866, 2020.

[32] G. F. Dunton and E. Dzubur and S. Intille, “Feasibility and performance test of a real-time sensor-informed context-sensitive ecological momentary assessment to capture physical activity,” Journal of Medical Internet Research, vol. 18, no. 6, pp. e106, 2016.

[33] Michigan Department of Health and Human Services, “Air Quality,” https://www.michigan.gov/mdhhs/safety-injury-prev/environmental-health/topics/mitracking/air, Accessed: 2026-04-01, 2026.

[34] P. Auer and N. Cesa-Bianchi and P. Fischer, “Finite-time analysis of the multiarmed bandit problem,” Machine Learning, vol. 47, no. 2-3, pp. 235–256, 2002.

[35] C. E. Rasmussen and C. K. I. Williams, Gaussian Processes for Machine Learning. Cambridge, MA, USA: MIT Press, 2006.

[36] C. McKay and others, “Gaussian process emulators for spatial air quality models: An applied research tool for communities,” Environmental Modelling and Software, vol. 52, pp. 251–264, 2014.

[37] B. Settles, “Active Learning Literature Survey,” University of Wisconsin-Madison, Computer Sciences Technical Report 1648, 2009.

[38] S. Schneider and R. Hernandez and D. U. Junghaenel and B. Orriens and P.-J. Lee and A. A. Stone, “Response times in ecological momentary assessment (EMA): Shedding light on the response process with a drift diffusion model,” Current Psychology, vol. 43, no. 7, pp. 1–19, 2023.

[39] A. L. Murray and others, “Prompt-level predictors of compliance in an ecological momentary assessment study of adolescents' socioemotional experiences,” Journal of Affective Disorders, vol. 338, pp. 494–502, 2023.

[40] D. Cook and A. Walker and B. Minor and C. Luna and S. T. Farias and L. Wiese and R. Weaver and M. Schmitter-Edgecombe, “Understanding the relationship between ecological momentary assessment methods, sensed behavior, and responsiveness: Cross-study analysis,” JMIR mHealth and uHealth, vol. 13, pp. e57018, 2025.

[41] Qualtrics, LLC, “Qualtrics Experience Management Platform,” https://www.qualtrics.com, Accessed: 2026-04-05, 2026.

[42] Twilio Inc., “Twilio API Documentation,” https://www.twilio.com/docs, Accessed: 2026-04-05, 2026.

[43] ATMO Support Center, “Cloud Dashboard API,” https://support.atmotube.com, Accessed: 2026-04-06, 2024.

[44] P. Yang and D. Stankevicius and V. Marozas and Z. Deng and E. Liu and A. Lukosevicius and F. Dong and L. Xu and G. Min, “Lifelogging Data Validation Model for Internet of Things Enabled Personalized Healthcare,” IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 48, no. 1, pp. 50–64, 2018.

[45] G. Eisele and H. Vachon and G. Lafit and P. Kuppens and M. Houben and I. Myin-Germeys and W. Viechtbauer, “The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data,” Assessment, vol. 29, no. 2, pp. 136–151, 2022.