Panopticon: The Design and Evaluation of a Game that Teaches Data Science Students Designing Privacy

Yuhe Tian*
University of California, San Diego
San Diego, California, USA
yut009@ucsd.edu

Shao-Yu Chu*
University of California, San Diego
San Diego, California, USA
shc076@ucsd.edu

Yuxuan Liu*
University of California, San Diego
San Diego, California, USA
yul208@ucsd.edu

Haojian Jin
University of California, San Diego
San Diego, California, USA
haojian@ucsd.edu

Abstract

In this paper, we describe the design and evaluation of Panopticon, an educational board game that helps data science students learn the skills of designing privacy-sensitive data practices with fun. Panopticon draws inspiration from the classic economics-themed game Monopoly, but re-imagines Monopoly's financial system as a data economy and requires players to conduct privacy design related activities as they navigate the game board. We used two learning science principles, peer learning and formative feedback, to guide the game design. We evaluated the game through a user study with 36 players (i.e., 12 game sessions) and compared their learning outcomes to a control group (n=36) who learned privacy design through paper content. To measure the learning outcomes, we developed rubrics to quantitatively assess the quality of the privacy designs, covering the level of detail, the technical feasibility, and the empathy for stakeholders. Our results suggest that Panopticon increased the learning outcomes by 354%, with significant improvements in all three dimensions. Participants also reported it as an entertaining way to learn in the post-study interview.

Keywords

privacy education, educational game, data practice design

1 Introduction

Companies are hiring privacy engineers but have trouble filling these positions with qualified candidates [16]. As a result, universities have started to offer privacy courses [9] or include privacy modules in courses on computer security [72], technology and public policy [21], and computers and society [71]. However, a key challenge with these efforts is we only have limited education methods tailored to teaching privacy [40].

Training privacy practitioners typically involves two main types of teaching activities. The first is lectures on privacy concepts, which provide students with a foundational understanding of key topics such as data protection, consent, encryption, and privacy regulations [15]. The second is privacy review essays (e.g., [2, 9, 81]), where students evaluate the privacy issues of selected technology and discuss possible solutions to address the issues in short essays.

In this paper, we introduce Panopticon, an educational board game that helps data science students learn the skills of designing privacy-sensitive data practices (i.e., designing privacy) with fun. Panopticon draws inspiration from the classic economics-themed game Monopoly, but re-imagines Monopoly's financial system as a data economy. While players in Monopoly alternate between the roles of landowners and tenants, players in Panopticon switch between digital service users and developers. As players navigate the game board, Panopticon requires players to iteratively design, critique, and revise peers' privacy designs. Panopticon will introduce **a new hands-on, interactive teaching activity** into the toolbox of privacy engineering educators.

We used two learning science principles, **peer learning** and **formative feedback**, to guide the game design. When students critique others, they engage in peer learning [36]. Evaluating others' work requires players to practice their understanding of the subject, which can enhance their learning [46]. The formative feedback [31] principle emphasizes that learning is most effective when learners receive timely, specific, and actionable feedback [68] that helps them understand their strengths, identify areas for improvement, and guide their next steps.

We used a human-centered approach to iteratively refine the game design. We held 9 pilot game sessions with 2-3 players each. In each session, we collected participant feedback to refine our game design. We then tested the modified game in the subsequent session. Our final game design includes three key artifacts: (1) a Monopoly-like paper-based game board, serving as the primary interface for player interaction (Figure 1); (2) a design worksheet, helping students create structured data practice designs (Figure 5); and (3) a task bank offering diverse data practice scenarios for players to draw from (Table 7).

We then evaluated the game through a user study with 36 players (i.e., 12 game sessions) and compared their learning outcomes to a control group (n=36) who learned privacy design through paper content. Before and after the learning activity, we assigned each participant two random data practice scenarios and asked them to design data practices using the design worksheet. We then created rubrics to assess each design based on three key dimensions: level

*These authors contributed equally to this paper.

Proceedings on Privacy Enhancing Technologies 2025(3), 398–414

https://doi.org/10.56553/popets-2025-0105

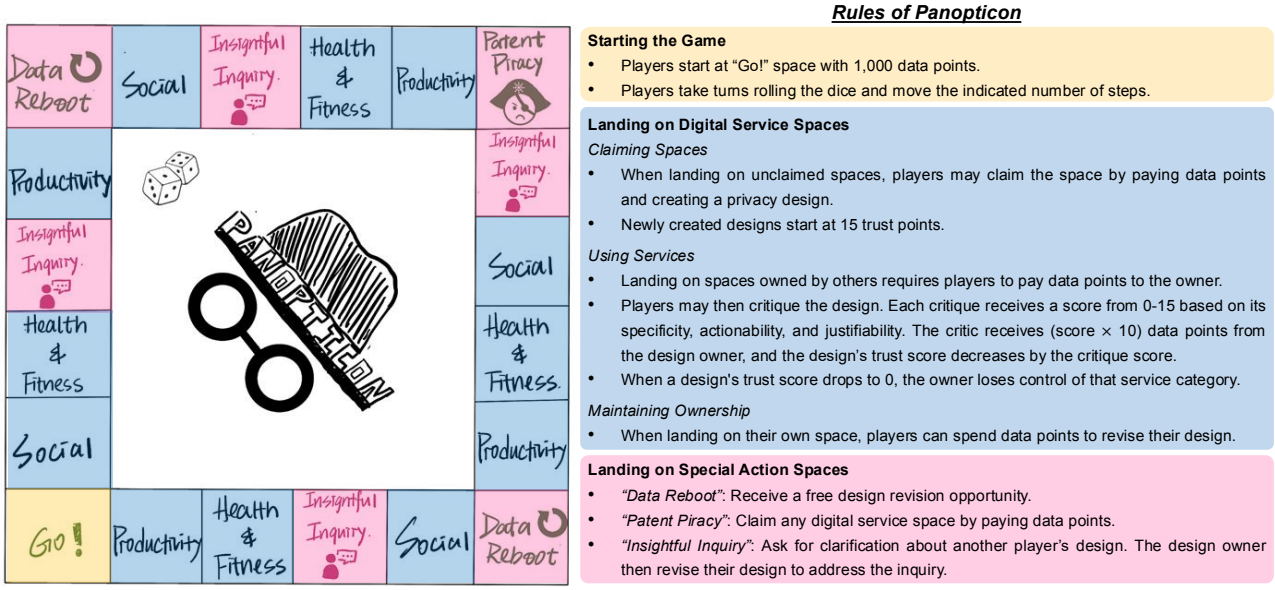


Figure 1: Panopticon game board and rules. Panopticon re-imagines Monopoly's financial system as a data economy and offers players an engaging way to learn designing privacy through iterative design, critique, and revision.

of detail, technical feasibility, and empathy towards stakeholders. We measured the learning outcome in both groups as the delta between the average score in the post-intervention group and the average score in the pre-intervention group.

We found that Panopticon can improve the learning outcome by 354%, with significant improvements in all three dimensions. We also analyzed participants' pre- and post-intervention designs. We found the designs evolved from generic technical solutions to detailed considerations of data handling processes, user acceptance, and practical implementation steps (see examples in Figure 2). Participants also reported it as an entertaining way to learn in the post-study interview, with most players expressing interest in continuing beyond the predefined time limit. The game artifacts will be made publicly available following the publication of the paper.

2 Related Work

2.1 Privacy Literacy

Researchers have been developing privacy literacy solutions to help non-technical individuals make informed decisions about sharing and protecting their personal data [49, 58, 60, 61, 80]. Multiple online libraries provide reading resources for users to learn more about online privacy issues [6, 59, 65]. Gamified approaches have also proven effective in making privacy concepts engaging and easier for laypersons to understand [14, 24, 32, 48, 67]. For instance, Sheng et al. designed an educational game to help users better identify phishing websites and not fall for phishing [67]. Digital tools, such as *The Data Detox Kit* [77] and *Terms of Service; Didn't Read* [55], guide non-technical individuals to manage their digital footprint by breaking down privacy terms into simpler summaries.

While these solutions focus on teaching non-technical audiences to make informed privacy design decisions, our work explores a less studied space: how to teach tech-savvy data science students privacy design skills.

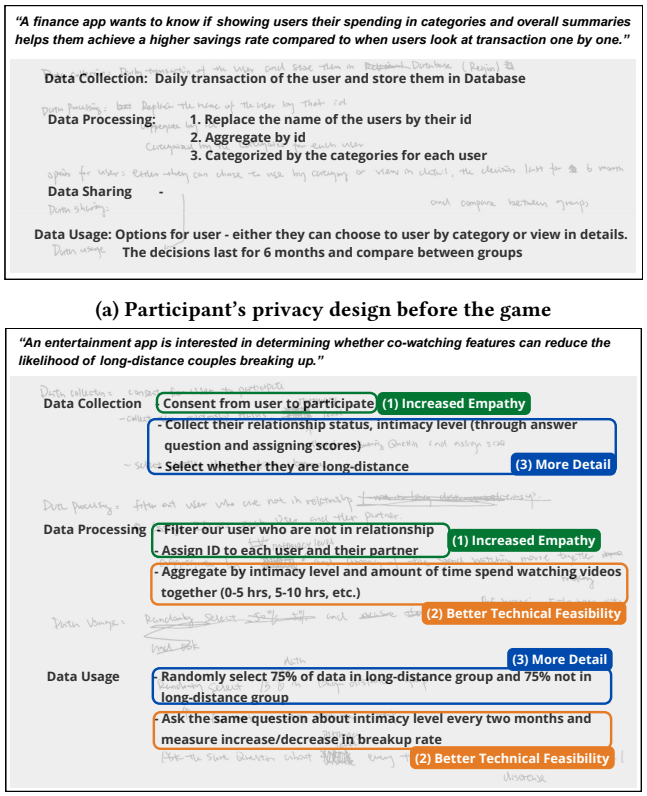
2.2 Privacy by Design

Previous research on privacy by design (PbD) focuses on helping privacy practitioners align their design decisions with compliance requirements and user expectations [1, 3, 12, 25, 64]. Tools such as compliance frameworks designed by the National Institute of Standards and Technology (NIST) and Privacy Impact Assessment (PIA), offer developers a checklist in complying with the General Data Protection Regulation (GDPR) and California Consumer Privacy Act (CCPA) regulations [53, 82]. More recent work translates PbD principles into design objectives [4, 29, 73] and develops tools for implementing privacy requirements [62, 87]. For instance, Hoepman's privacy design strategies map privacy regulations into eight system design strategies (e.g., minimize, hide, separate) [29]. Tools such as Privado.ai provide a code scanning platform that helps organizations automate GDPR compliance in their software [62].

In contrast to these PbD works, which focus on data practice, our work takes a human-centered perspective and focuses on training future data practitioners. Success in our work would improve practitioners' capability in designing privacy-sensitive data practices, benefiting the privacy of individuals whose data is processed by these practitioners.

2.3 Experiential & Game-Based Learning

Experiential learning, also known as active learning, has proven effective in helping students acquire practical skills and develop critical thinking and problem-solving abilities [37, 38, 52, 83]. In data science, this often comes through working on a capstone experience, or major project at the end of an educational program [5, 52, 63]. This is often a chance for students to utilize the skills they have learned to



(a) Participant's privacy design before the game

(b) Participant's privacy design after the game

Figure 2: Through Panopticon, participants' privacy designs improve in (1) empathy (e.g., addressing consent before collecting data), (2) technical feasibility (e.g., adopting an effective analysis method), and (3) level of detail (e.g., specifying the implementation of a new feature).

tackle real-world problems [52, 83]. Gamified learning has emerged as another form of experiential learning that engages students through interactive and immersive experiences [43]. Gamified security work help students understand security principles and improve performance in security assessments [23, 34, 51, 56, 66]. For example, CyberCIEGE immerses students in security decision-making by placing them in the role of information assurance decision-maker for some enterprise [78]. Control-Alt-Hack, a tabletop card game, engages players in white-hat hacking missions to raise awareness of cybersecurity [18, 19]. Panopticon builds on these principles by introducing privacy design concepts through an educational board game. Players actively engage in designing, critiquing, and iterating on privacy designs within a gamified framework.

3 Design Overview of Panopticon

The educational objective of Panopticon is to teach students the skills to design privacy-sensitive data practices (i.e., "design privacy"). Below, we present the key design concepts of Panopticon.

3.1 Applying Learning Science Principles

Previous research highlights several barriers faced by practitioners in designing privacy [13, 30, 42, 44]: (1) practitioners often struggle to articulate precise privacy-related design decisions [42, 44]; (2) practitioners frequently overlook important design decisions and fail to anticipate the potential implications of specific design choices [13, 44]; (3) many privacy-sensitive data practice designs are technically impractical [42].

We used two learning science principles, peer learning [10, 79] and formative feedback [31, 69], to design the game to help students overcome these barriers.

Panopticon employs the peer learning principle by encouraging players to evaluate others' privacy designs and provide feedback. Research has shown that peer learning increases students' possibilities to reflect and explore ideas when teachers are absent [10, 33, 45]. By critically assessing peers' designs, players must engage deeply with privacy concepts, which enhances their understanding and ability to articulate privacy-related design considerations. Further, a player may only be aware of a few privacy issues. By exposing players to their peers' perspectives, the game helps them identify overlooked design decisions and unexpected implications.

Panopticon also encourages players to provide **formative** privacy feedback on others' data practice designs throughout the design process, even when the final design is not ready. Research shows that learning is most effective when learners receive timely, specific, and actionable feedback, enabling them to understand their strengths, identify areas for improvement, and plan their next steps [54]. To support this, we set a tight time limit for the design tasks, prompting players to create only a rough sketch of their designs. This intentional constraint ensures that all designs remain in an unfinished state, making players more open to acknowledging and addressing the shortcomings in their work.

3.2 High-level Game Design

Panopticon draws inspiration from the classic economics-themed game Monopoly, but re-imagines Monopoly's financial system [57] as a data economy. While players in Monopoly alternate between the roles of landowners and tenants, players in Panopticon switch between digital service users and developers (Table 1).

Panopticon has two core game mechanisms: data points and trust scores. Data points function as currency - developers spend them to enter the market and earn them when users access their services, while users spend them to use services and earn them through effective critiques. Trust scores measure privacy design quality, starting at an initial value and decreasing when users identify privacy concerns through critiques.

When acting as developers, players must articulate their privacy design and data handling practices, starting with a baseline of consumer trust. However, if their practices raise privacy concerns, they gradually lose public trust and risk being expelled from the digital service market. Conversely, as users, players must spend data points to use other players' services. If they discover their data is at risk of misuse, they can critique the developer's practices, leading to the developer losing market share.

Table 1: Panopticon maps the real-estate concepts in Monopoly to concepts in modern digital marketplaces.

	Monopoly	Panopticon
Roles	Landlord/Tenant	Digital service developer/User
Assets	Real estates (lands, houses, and hotels)	Digital services
Currency	Monopoly money	Data points
Asset Acquisition	Purchase unowned lands by paying listed price	Pay a claim fee and sketch privacy designs to claim digital services
Landing on Others' Assets	Pay rent when landing on owned lands	Pay a subscription fee and critique the privacy design when landing on claimed digital services
Expanding Assets	Build houses/hotels on owned lands	Revising privacy designs
Special Action Spaces	Go, Community Chest, Chance, Go to Jail, Jail, Free Parking, Income Tax, Luxury Tax	Go, Data Reboot, Patent Piracy, Insightful Inquiry

3.3 Interactive Prototyping

We first developed a **paper-based game board** by adapting the Monopoly game board and associated rules (Table 1). We then conducted two interactive prototyping sessions [20] using the initial game board with five data science undergraduate students. Since the game was not fully developed then, the research team actively explained the mechanisms to the players during these sessions. In doing so, we can observe how students play the game when the final game design was not ready.

We made two key observations during the interactive prototyping sessions. First, students struggled to create well-defined data practice designs for others to critique. Frequently, a player would begin with vague descriptions of a data practice, making it difficult for the other player to provide meaningful critiques. Moreover, this vagueness made it challenging to validate whether the critiques were valid.

Second, selecting a data practice design scenario proved to be non-trivial. Initially, we allowed players to choose any data practice design scenario they were interested in. However, this approach led to two issues: (1) not all players were familiar with the contexts of the chosen data practices, and (2) some scenarios were too complex to design within the constraints of a game session.

3.4 Design Worksheet and Task Bank

To address the two challenges mentioned above, we developed a Design Worksheet to help players effectively think through and communicate data practice designs, and a Task Bank offers preselected data practice scenarios for players.

The initial design worksheet (Figure 3) includes two parts: a section that records key elements such as the player's name, research hypothesis, and trust score, and a flowchart of four data actions—*Data Collection, Processing, Sharing, and Usage*—with text boxes for documenting privacy design decisions.

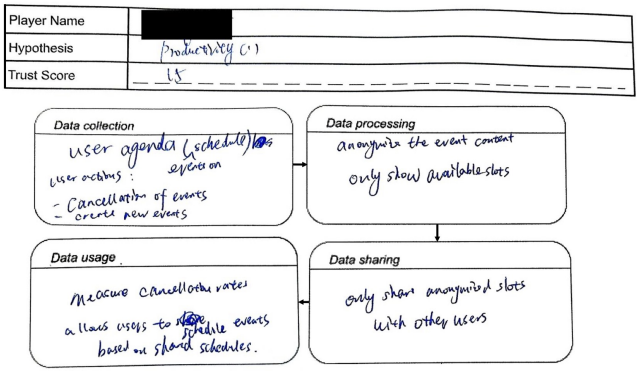


Figure 3: Initial design worksheet that tracks the privacy design's status (e.g., current trust score) and a dataflow.

This flowchart visually demonstrates how each data action builds upon previous actions, helping both the authors and the other players think through the interconnected nature of privacy design choices. Players complete this worksheet when claiming a digital service, documenting their decisions for each data action primitive. This structured format also aids players in articulating critiques of specific privacy design decisions. The flow chart formulation is inspired by Lean Privacy Review [35].

The question bank contains a set of data practice design scenarios inspired by real-world data practices (see complete scenarios in Appendix Table 8). We centered all scenarios around a standardized data science question to reduce the task complexity. An example scenario, drawing inspiration from OKCupid's controversial experiments [27], is:

Imagine that you are a data scientist at an online dating company, and your manager asks you to design an experiment to answer the following question: "Will increasing the visibility of shared interests and mutual connections on social networking apps enhance the likelihood of users finding compatible dating partners?" How would you design the experiment in a way that respects user privacy?

3.5 Initial Game Play Rules

We also articulated the basic rules for gameplay through the interactive prototyping session. All players start the game at the "GO!" space with A data points. Players will take turns rolling dice to move around the board, encountering two types of spaces: digital service spaces and special action spaces.

Players interact with digital service spaces in three ways:

- **Claiming Spaces:** When landing on an unclaimed space, players can establish ownership of the slot by paying B data points. The player must then draw a research question from the privacy design bank and create a privacy design within two minutes. Each new design starts with a C trust score.
- **Using Services:** Landing on an owned space requires the service owner to pay D data points. Players can then critique the owner's privacy design. Teachers will score the critiques on a scale from 0 to 5 points based on three criteria: specificity, justification, and

actionability [54], with a maximum possible score of 15 points. The critic receives (score × 10) data points from the owner, and the design's trust score decreases by the critique score.
- **Maintaining Ownership:** When a design's trust score drops to zero, the owner loses control of that service. To prevent this, owners can spend D data points to revise their design within one minute when landing on their own space.

Monopoly uses special action spaces to make the game more engaging, with each space having a rule (e.g., Go to Jail) that deviates from standard gameplay. We adapted the special action rules into our context (see Appendix Table 9).

Through these core mechanics, players experience iterative cycles of privacy design, peer critique, and design revision. The game continues until either a player bankrupts or the agreed session time is reached, with the winner being the player who accumulates the most data points.

4 Pilot Test and Iterations

We then conducted seven pilot test sessions and iteratively refined the game design based on our findings from the pilot studies.

4.1 Pilot Study

We recruited participants from our research lab and conducted seven pilot sessions with three players per session. Although Panopticon supports varying numbers of players, we selected the three-player format as it effectively balances the diversity of critiques with the logistical ease of coordinating and scheduling participants.

Every session begins with a tutorial explaining our game mechanics, followed by a 45-minute game (Figure 4). The pilot test sessions used a think-aloud protocol where participants verbalized their thought processes while creating privacy designs. We refined the game design iteratively based on participant feedback and then tested the modified game in a subsequent session. We paused the process until we observed evident improvements in participants' privacy designs after gameplay and confirmed that players found the game enjoyable. Through the process, we empirically determined the parameters for the game rules (see final parameters in Figure 1) and made the following key modifications.

4.2 Modified Game Board

We observed that the dice-rolling mechanism introduced randomness to peer learning opportunities. In one notable case, a player went through an entire session without landing on any digital service space. They did not have the chance to create, critique, or revise privacy designs.

We made two main changes to the game board to ensure students will have enough opportunities to interact with peers. First, we reduced the number of unique digital service spaces from 12 to 3. The final game board only has three digital services across three domains (Social, Health & Fitness, or Productivity), but each service can be triggered by 4 different slots (see Figure 1).

Second, we replaced the spaces that merely provided bonuses or penalties with action-required spaces (Table 9). Specifically, we invented the following action spaces to support learning objectives and mitigate randomness actively:

- **"Insightful Inquiry"** requires players to ask for elaboration on one of the existing designs as a charity space. The inquiry differs from the critique since it does not ask players to provide actionable insights. After receiving an inquiry, the design owner will have 1 minute to elaborate on their designs.
- **"Data Reboot"** enables players to revise any of their designs without cost. This ensures players can improve their designs even when dice rolls don't land them on their own spaces. Providing extra chances for players to refine their designs based on received feedback aims to promote iterative learning, preventing situations where poor initial designs persist due to a lack of revision opportunities.
- **"Patent Piracy"** allows players to claim any digital service space on the board. This addresses a common issue we observed during the pilot testing phase: some players never acquired spaces due to unlucky dice rolls. By providing an alternative path to space acquisition, we aim to let all players to practice privacy design and engage in peer critique.

4.3 Modified Design Worksheet

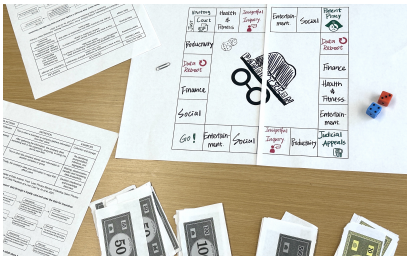
We noticed that feedback given by players during the pilot study was often vague, leaving those who received feedback unclear on how to revise. For instance, one common feedback is, "I think your design is collecting unnecessary user information," leaving recipients unclear on how to improve their designs.

The problem is twofold: the initial data practice designs are vague, and critics lack the skills to provide effective feedback. Our solution is inspired by prior research on crowdsourced design feedback [39, 54], highlighting that rubrics and examples effectively structure feedback by encouraging attention to deep and diverse criteria. Without such guidance, novices often focus on the first thing they notice, typically surface-level details [26].

We first added privacy design examples in the design worksheet to facilitate novice players' documentation of design choices at each stage (Figure 5). We then listed examples of different levels of critiques to let players have a better understanding of what feedback they are expected to give (Table 2).

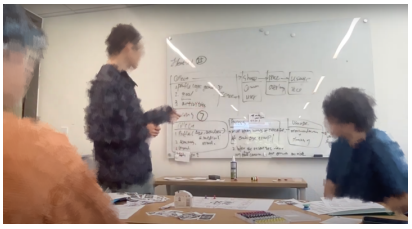
Table 2: Critique Examples. We provide the following example to show participants which critiques are preferred and which are not.

High quality Critiques	
Specific	*"The current data collection method risks exposing user privacy*
Justifiable	*because it logs detailed daily behaviors without anonymization.*
Actionable	*To mitigate this, implement data anonymization techniques before storage and restrict access to only aggregated data for analysis."*
Low quality Critiques	
Unspecific	*"This app might be invading users' privacy*
Unjustifiable	*<no justifiable description included>*
Unactionable	*Maybe look into this issue."*

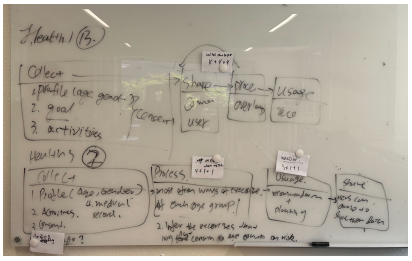


(a) The game board, banknotes, and worksheets used in the pilot studies.

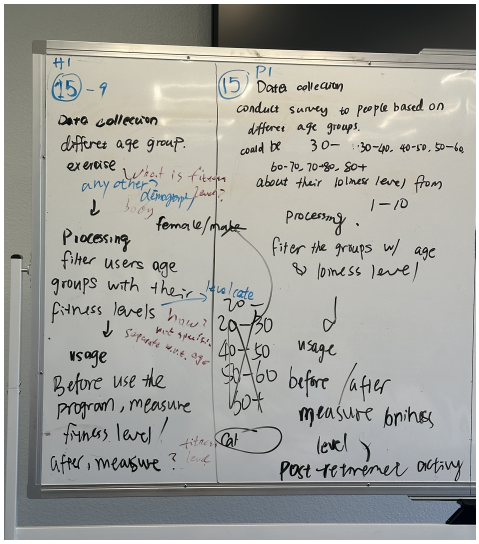
(b) Participants engaging in the game.



(c) A participant explaining privacy design to others during the game.



(d) Privacy design developed during a pilot game session.



(e) Participants propose their design on the whiteboard (black), others critique the design (red), then the design is further refined (blue).

Figure 4: Snapshots of the pilot game sessions. Through Panopticon, players engage in peer learning by proposing privacy designs, critiquing others' designs, and incorporating formative feedback to revise their own.

5 Evaluation

In this section, we present the design of our IRB-approved study and the quantitative and qualitative evaluation results of Panopticon.

5.1 Study Design

We based our evaluation design on Sheng et al.'s methodology for assessing educational security interventions [67]. We first presented participants with a data practice scenario and asked: "*As a data scientist, design a privacy-sensitive experiment to investigate this data practice scenario.*" After completing the first privacy design, participants underwent training, either through the Panopticon game or worksheet tutorials. Next, students will act as data scientists again to work on another data practice scenario. Finally, we asked them to complete an exit survey (Figure 6).

We developed two data practice scenarios: (1) whether co-watching features reduce long-distance relationship breakups (TikTok Scenario) and (2) whether categorized spending summaries increase savings rates (Finance App Scenario). We describe the two scenarios' exact descriptions in Table 8. To control for potential differences in scenario difficulty, we randomized their order - half the participants received the TikTok Scenario first, followed by the Finance App Scenario (Group TF), while the other half received the scenarios in the reverse order (Group FT).

We divided our participants into two groups:

- Game Group: Participants played Panopticon in groups of three for 40 minutes after watching a tutorial video. We chose the group size to enable peer learning while maintaining meaningful participation from each player. Two authors (one PhD student and one undergraduate student) took turns acting as the "teacher" who scores the critiques.
- Worksheet Tutorial Group: Participants spent at least 10 minutes studying our privacy design worksheet independently. We advised participants to begin the post-intervention task only when they felt their learning had reached a satisfactory level of saturation. This condition helps isolate the impact of game mechanics from the educational content.

We conducted all sessions using consistent recruiting, screening, and experiment administration procedures. We encouraged participants to think aloud while creating their designs, and we recorded their verbal explanations and written responses for analysis. The total experiment time was approximately one hour per session, including pre- and post-intervention designs, training, and completion of the exit survey.

5.2 Participant Recruitment and Demographics

We recruited 72 students from an American university through flyers posted around campus and advertisements on social platforms. We divided the game group (n=36) into 12 game sessions of 3 players each and the worksheet group (n=36) into 36 independent sessions. We compensated each participant in the game group with a $15 Amazon gift card and a $5 bonus for winning the game to provide an incentive for engagement and to reward achievement within the game. Participants in the worksheet group received $10.

We screened participants according to their majors' curriculum alignment with ACM Data Science Task Force's definition of Data Science [17]. We recruited participants with varying levels of

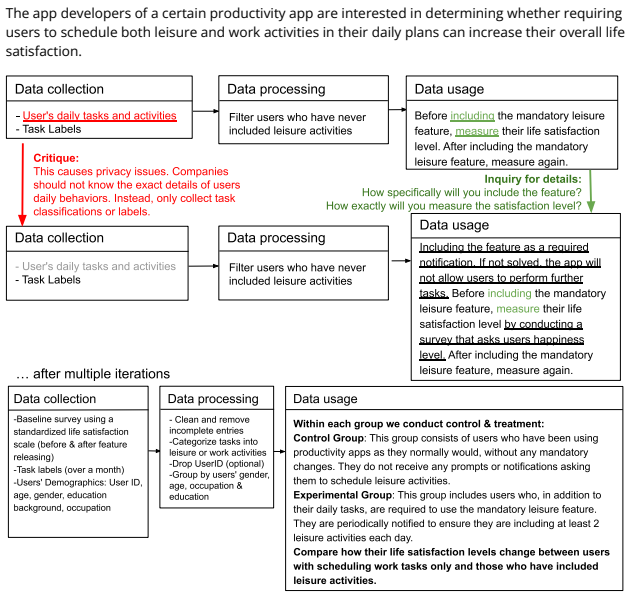


Figure 5: The final design worksheet (partial) with an example illustrating the iterative design process. The top section presents an initial design, starting with rough ideas. It then undergoes critique (red) and receives inquiries (green) from peers. Revising the design to address these issues results in the final version shown at the bottom.

proficiency in data science competencies (e.g., programming, data analysis, machine learning). These demographics help us understand how our game could help data science students with varying levels of prior domain knowledge. A summary of demographics is shown in Table 3.

Table 3: Participants demographics in each condition.

	Worksheet Group	Game Group
Gender		
Male	41.7%	28.0%
Female	58.3%	72.0%
Non-Binary	0.0%	0.0%
Education		
College Grad	8.3%	25.0%
College Undergrad	91.7%	75.0%
Major		
Physical Sciences	5.6%	0.0%
Biological Sciences	13.9%	8.3%
Social Sciences	16.7%	8.3%
Data Science	27.8%	33.3%
Engineering	27.8%	38.9%
Not Specified	8.3%	11.1%

5.3 Analysis Metrics and Method

In this section, we present the design of assessment rubrics, and the analysis method for our study.

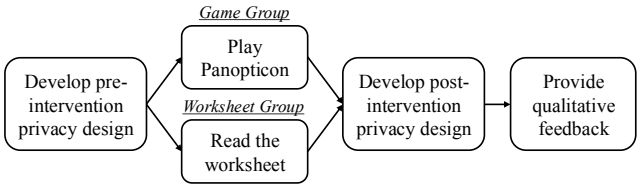


Figure 6: User study experiment design.

5.3.1 Design of Evaluation Metrics. Developing assessment rubrics to quantify learning outcomes for privacy education presented two key challenges: the lack of standardized evaluation methods and the need for consistent scoring across evaluators. Our initial rubric assessed privacy design competency using 1–5 scales across four criteria: data minimization, transparency, purpose specification, and data accuracy. However, pilot testing revealed significant limitations with this approach - even when experts identified similar improvements in designs, they often assigned different scores due to subjective scale interpretation.

To establish a more standardized assessment, we developed a rubric with binary scoring (0 or 1) across three key dimensions: *Design Detail*, *Technical Feasibility*, and *Stakeholder Respect*, covering the three barriers described in Section 3.1. Design Detail evaluates how thoroughly participants think through and articulate each data action, from collection to usage. Technical Feasibility assesses whether the proposed design can effectively answer the data practice scenario through appropriate data collection and analysis methods. Stakeholder Respect examines privacy protection measures, including regulatory compliance, user communication, security considerations, and stakeholder management.

We created specific yes/no/not applicable questions for each dimension through an iterative refinement process. Two authors first collaborated to develop initial assessment questions for each dimension. Using the draft rubric, these authors then independently evaluated all pre- and post-intervention designs. After completing the evaluations, the authors discussed potential improvements in the rubrics based on the scoring discrepancies. Then, the two authors re-evaluate all designs independently using the updated rubrics. We continuously refined the rubric criteria until they established consistent evaluation standards. We conducted three iterations in total.

This iterative process helped us transform subjective criteria into concrete questions. For example, rather than asking vague questions such as *"Is there transparency in this design?"*, we now ask concrete questions like *"Assuming the best data analysis method is used, can the obtained data address the data practice scenario? e.g., Is the data complete and accurate? Are there potential biases in the data collection? Does the data provide sufficient detail for analysis?"* Table 4 presents our complete assessment framework, detailing specific criteria and evaluation guidelines for each dimension.

To validate our rubric's reliability, we asked two independent privacy experts (two PhD students with 3 years experience in usable privacy research), blinded to the experimental conditions, to evaluate the 144 designs, 2 designs each from the 72 participants. For each dimension's assessment questions, raters provided categorical responses (0 = No, 1 = Yes, NA = Not Applicable). We

calculated agreement by comparing the alignment of categorical responses across all questions and designs. We considered two responses aligned only if both raters assigned the same categorical value. We assessed inter-rater reliability using Cohen's kappa coefficient, achieving [$\kappa = 0.84$]. A kappa score above 0.8 indicates strong agreement in fine-grained quality assessments [50]. Experts resolved evaluation disagreements through discussions after completing all independent ratings.

5.3.2 Analysis Method. We implemented a two-phase analytical approach to evaluate the efficacy of different interventions in improving participants' design skills and understanding of privacy designs. First, we conducted a linear mixed-effects regression analysis to assess the overall impact of interventions on privacy design quality, accounting for individual participant variations. For each design, we calculated the final score by averaging two independent expert ratings, where each rating represented the ratio of met criteria to total applicable criteria. The linear mixed-effects model employed these averaged privacy design scores as the dependent variable, treating intervention type (Game or Worksheet), demographic characteristics, educational background, prior domain knowledge, and scenario presentation order as fixed effects. Since each participant contributed two design scores, we controlled for individual differences using a random effect. For categorical variables, we applied dummy coding, with N-category factors represented by $N - 1$ dummy variables. We employed only a single dummy variable for gender (Male = 1, Female = 0) since no participant reported themselves as non-binary.

Second, we examined criterion-specific improvements within our three dimensions. For each criterion, we documented the pre- and post-intervention average scores for both Game and Worksheet groups, calculated the mean magnitude of improvement. Given that the pre-game scores were not normally distributed (Shapiro-Wilk test $p = 0.007$), we applied Mann-Whitney U tests to compare privacy design scores for each criterion before and after the intervention.

5.4 Quantitative Results

The linear mixed-effects regression analysis (Table 5) revealed that the game intervention had a significantly positive effect on privacy design scores ($b_{Game} = 0.207$, $p < 0.001$), while traditional worksheet-based education showed no significant effect ($b_{Worksheet} = 0.010$, $p = 0.067$). The intercept coefficient (0.402) represents the model's baseline prediction when using the reference group for each categorical variable (e.g., female, undergraduate). Fixed and random effects in the model accounted for 56% of the variance in privacy design scores (conditional R-squared = 0.56). Additionally, the order factor (TF vs. FT) showed no significant impact on privacy design scores, indicating that randomization effectively controlled for potential differences in the difficulty of the two scenarios.

To understand where these improvements occurred, we analyzed the impact across three key dimensions: Design Detail, Technical Feasibility, and Stakeholder Respect (Table 6). Within each dimension:

- In **Design Detail**, the game group showed significant improvements in three criteria. Most notably, Data Processing Articulation improved by 0.458 ($p < 0.001$) and Data Usage Articulation by 0.333 ($p < 0.001$). The worksheet group demonstrated statistically significant improvements only in Data Collection Articulation (improvement = 0.236, $p < 0.001$).
- For **Technical Feasibility**, both Data Accuracy (improvement = 0.222, $p < 0.05$) and Analysis Method Effectiveness (improvement = 0.264, $p < 0.01$) showed significant gains in the game group. The worksheet group's improvements were minimal (0.083 for both criteria) and not statistically significant.
- In **Stakeholder Respect**, participants demonstrated significant improvements in User Understanding (improvement = 0.278, p < 0.01) and Security Risk Consideration (improvement = 0.153, p < 0.05). While both groups showed negative trends in Stakeholder Consideration, the game group's smaller decline (-0.083 for the game group vs. -0.375 for the worksheet group) suggests that the interactive game session may help mitigate deterioration in stakeholder awareness.

When examining aggregate improvements across these dimensions (Figure 7), the game group consistently outperformed the worksheet group. The game group's improvement of 0.331 in Design Details (95% CI [0.196, 0.304]) more than twice the worksheet group's improvement of 0.138 (95% CI [-0.045, 0.309]). For Technical Feasibility, the game group achieved an improvement of 0.243 (95% CI [0.129, 0.374]), while the improvement in the worksheet group was not statistically significant. Although improvements in Respect for Stakeholders were more modest in the game group (improvement = 0.138, 95% CI [0.066, 0.220]), the worksheet group showed no significant improvement (-0.029, p > 0.05). The game group also maintained more consistent performance, reflected in their narrower confidence intervals.

Overall improvement across all dimensions revealed that the game group achieved a mean improvement of 0.236 (95% CI [0.176, 0.296]), representing a 354% increase compared to the worksheet group's 0.052 (95% CI [0.003, 0.098]). A Wilcoxon test confirmed highly significant improvements for the game group ($p = 4.40e{-}06$), while the worksheet group's improvements were not statistically significant ($p = 0.076$).

5.5 Qualitative Results

Our quantitative results demonstrate that Panopticon significantly improved participants' privacy design scores across all assessment rubrics dimensions. To understand how Panopticon achieved these improvements, we analyzed participants' pre- and post-intervention designs. We found that the integration of peer critique mechanisms and opportunities for iterative refinement guided participants to three key improvements: enhanced attention to detail, increased technical feasibility in designs, and more empathy towards stakeholders. These key improvements explain the quantitative disparities observed between interventions in Table 6.

Increased awareness of design details. Participants in the game group showed improvement in specifying detailed elements at each stage of privacy design (Figure 8). For example, one participant's pre-game design for the Finance app scenario merely stated *showing spending in category & overall sum*, omitting details

Table 4: Rubrics for privacy designs. The rubric assesses privacy designs across three key dimensions: Design Detail, Technical Feasibility, and Stakeholder Respect. Each dimension includes 2–4 specific questions, to which raters provide categorical responses (*Yes, No,* or *Not Applicable*) based on detailed descriptions, achieving 84% agreement.

Criterion	Description
Design Detail	
Did they articulate how to collect the data, if it involves data collection?	What data is being collected/stored? Who is the data observer? What is the data subject? Why is the company collecting/storing the data?
Did they articulate how to process the data, if it involves data processing?	What is the raw data? What is the derived data?
Did they articulate how to share the data, if it involves data sharing?	Why does the sender share the data? Who is the sender? Who is the recipient? Who will have access to the data for usage?
Did they articulate how they would use the data if it involved data usage?	What decisions or actions will the data inform? What is the potential impact of using the data on individuals or systems?
Technical Feasibility	
Assuming the best data analysis method is used, can the obtained data address the research question?	Is the data complete and accurate? Are there potential biases in the data collection? Does the data provide sufficient detail for analysis?
Assuming the best data is collected, can the chosen method answer the research question?	Does the method align with the research goals? Is the chosen method appropriate for the type of data? Are there assumptions made by the method that need to be addressed?
Stakeholder Respect	
Are the data collection, processing, sharing, and usage practices compliant with relevant laws and social norms?	Are the data collection, processing, sharing, and usage practices compliant with relevant laws and regulations (e.g., GDPR, HIPAA)? Do the potential benefits of the research outweigh the risks to the participants? Is the data collection limited to what is necessary for the project's stated objectives?
Did they ensure that users can clearly understand how their data will be collected, processed, and used in the design?	Is explicit consent obtained from users to collect and process their data? Will users fully be aware of how their data will be collected, processed, and used? Will data collection processes and purposes be clearly communicated to users?
Did they consider potential security risks such as hacking?	Are adequate security measures (e.g., anonymization, dropping sensitive info) in place to protect data from unauthorized access? How are potential security threats, such as hackers, accounted for? Is there a contingency plan in place to respond to data breaches?
Did they account for other stakeholders involved in data handling, if it involves other stakeholders?	Who are the other stakeholders involved in the process? How are their roles and responsibilities clearly defined and monitored? How are agreements with third parties enforced to ensure data protection?

like data collection frequency and specific categorization methods. Their post-game design for the TikTok scenario demonstrated significantly more detail by specifying precise data points like *Age, relationship lengths, gender* and establishing detailed experimental procedures with *Control Group: users without co-watching features* and *Experimental Group: group includes users with co-watching feature*.

Another participant started with a simple pre-game design for the TikTok scenario that only stated experimental actions with minimal detail (*two sets of long-distance couples* and *compare the break-up rate*). After playing our game, their design for the Finance app scenario incorporated specified more details across all stages: collecting specific user information like *user's daily spends* and users' *view history*; and implementing clear data processing steps to filter out users who *buy totally different things during the two periods*. We noted that this improvement in increased level of detail can be attributed to Panopticon's peer critique mechanism, as participants reported: "(B)y giving and receiving the critique, we get to realize the details we need."

In contrast, worksheet group participants showed weaker improvement in the level of detail. As seen in Figure 9a, their pre-worksheet response to the Finance app scenario simply stated *fetch the data with known information and population, identify useful data or label with importance level*, without specifying what constitutes *known information* or how to determine data importance. Their post-worksheet design for the TikTok scenario (Figure 9b) remained similarly vague with statements like*collect the basic information of the population accessing the survey*. We found worksheet participants, who did not receive peer feedback, continued to struggle with specifying concrete implementation steps.

Increased practicalities in designs. As some participants reflected, "In terms of data collection, before participating in the experiment, feasibility was basically not considered. After participating in the experiment, these aspects were taken into consideration." We noticed that the game group considered more practical

Table 5: Coefficients and p-values for the linear mixed-effects regression. A positive coefficient with a small p-value indicates that the factor increases the privacy design score. The game intervention significantly increases participants' privacy design scores by 0.207, whereas the worksheet intervention does not lead to a significant improvement.

Variable Name	Coefficient	P-value
Group: Game	**0.2069**	**< 0.001****
Group: Worksheet	**0.0999**	0.0672
Intercept	0.4019	< 0.001**
Gender: Male	0.0117	0.0796
Education: Master's Degree	0.1133	0.1054
Prior Knowledge	0.0083	0.6938
Order: TF	-0.0393	0.3986
Scene: T	-0.0351	0.2719
Major: Data Science	0.0066	0.9271
Major: Engineering	0.0023	0.9780
Major: Not specified	-0.1340	0.1441
Major: Physical Sciences	0.0477	0.6827
Major: Social Sciences	0.0340	0.7203
Random Effects	0.7686	0.0569

Conditional R^2 = 0.56

Note: * p < 0.05, ** p < 0.01

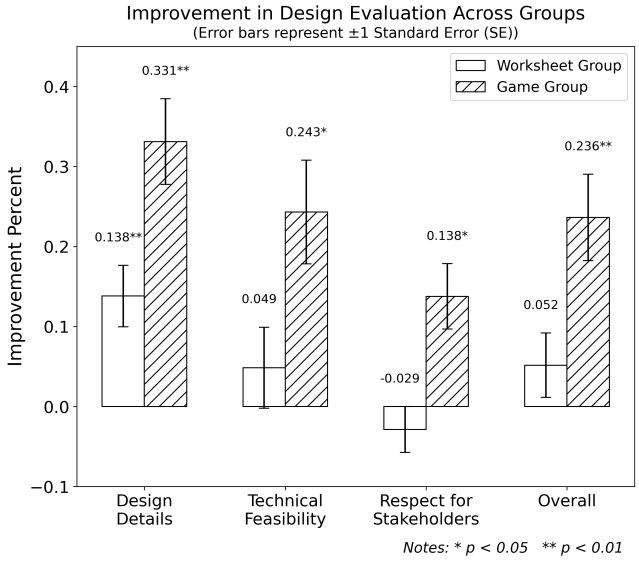


Figure 7: Participants in the game group showed greater improvement across all three dimensions compared to those in the worksheet group.

applications in their post-designs, that is, whether the obtained data and method can answer the data practice scenario (Figure 8). For instance, the pre-game design of one participant for the Finance app scenario showed a basic structure that didn't clearly establish how to measure the effectiveness of categorized spending (*either observe by units or divide by spending sectors by dollar value*). Their post-game design to the TikTok scenario included a clear experimental methodology by establishing control and treatment groups and specifying how to measure outcomes (*Compare how break-up rate change between the two groups*).

Similarly, in Figure 2a, the participant's pre-game design embedded a simplistic approach with *Options for user - either they can choose to user by category or view in details. The decisions last for 6 months and compare between groups* without specifying what data would enable this comparison or how to measure outcomes. Their post-game design (Figure 2b) demonstrated improved technical practicality by establishing a clear data collection strategy (*collect their relationship status, intimacy level through answering question and assigning scores*), specifying data processing steps (*aggregate by intimacy level and amount of time spending watching videos together (0–5 hrs, 5–10 hrs etc)*), and articulating how to assess the data practice scenario (*Ask the same question about intimacy level every two months and measure increase/decrease in breakup rate*).

We observed that the worksheet group continued to struggle with technical feasibility in their designs. In Figure 9a, their pre-worksheet design of the Finance app scenario included vague analytical goals like *discover the relationship between collected data and make a conclusion* without specifying what relationships to analyze or how to measure them. Their post-worksheet design, answering the TikTok scenario (Figure 9b), showed minimal improvement in accurately answering the data practice scenario. Their design maintained ambiguous data usage steps like *discover the findings by analyzing the relationship between data* and *control variables for different finding* without defining what specific relationships or variables would effectively answer the data practice scenario.

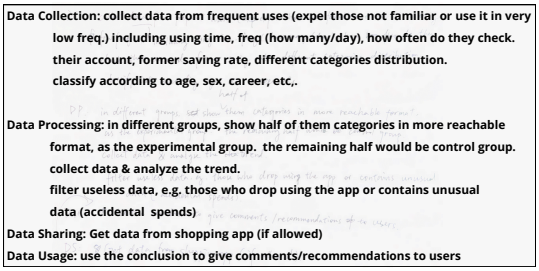
More empathy towards stakeholders. We found that the game group also increased awareness for user privacy and transparency in their post-game designs. Compared to the worksheet group, more participants in game group reported that "protecting user privacy and anonymization is essential during data processing phase." For example, in Figure 2a, a participant's pre-game design only focused on technical steps like *Replace the name of the users by their id.* After playing the game, their design (Figure 2b) demonstrated enhanced user empathy by: (1) Starting with *Consent from user to participate* before any data collection; (2) Adding privacy protection steps like *Filter out users who are not in relationship* and *Assign ID to each user and their partner*; (3) Clearly communicating data usage by specifying measurement intervals (*every two months*) and metrics (*intimacy level through answering question*). Another participant evolved from vague purpose statements in their pre-game design to clear specifications of how user data would be handled: *within each group, form control & treatment groups* and clearly disclosing the purpose *Compare how break-up rate change between the two groups.*

The worksheet group's designs remained technically focused without addressing stakeholder concerns. As shown in Figure 9c, merely stating *Remove sensitive information and share with public* without considering whether such sharing was necessary or justified. Their post-worksheet design for the TikTok scenario (Figure 9d) showed only surface-level progress in privacy considerations. While they maintained the idea of *Remove sensitive information and share data collected to public* and added *Form essays etc. to reach conclusion* as their purpose, they still didn't question the

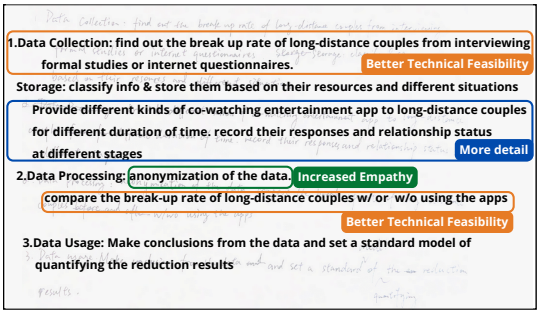
Table 6: Improvement in each assessment criterion. We calculate the mean improvement by averaging the differences in privacy design scores before and after the intervention and apply Mann-Whitney U tests for comparison. The game group shows significant improvement in 7 out of 10 criteria, whereas the worksheet group does not show significant improvement in any criterion.

Dimension	Assessment Criteria	Worksheet Group		Game Group	
		Improvement	p-value	Improvement	p-value
Design Detail	Data Collection Articulation	0.2361	0.0040**	0.2083	0.0172*
	Data Processing Articulation	0.1250	0.1167	0.4583	0.0000***
	Data Sharing Articulation	0.2000	0.2014	0.1607	0.2910
	Data Usage Articulation	0.0139	0.4404	0.3333	0.0004***
Technical Feasibility	Data Relevance	0.0972	0.1739	0.2222	0.0137*
	Analysis Method Effectiveness	0.000	0.4831	0.2639	0.0018**
Stakeholder Respect	Compliance with Laws	0.0139	0.4876	0.0139	0.2949
	User Understanding	0.0139	0.2773	0.2778	0.0038**
	Security Risk Consideration	-0.0972	0.8615	0.1528	0.0379*
	Stakeholder Consideration	-0.1250	0.6949	-0.0833	0.6743

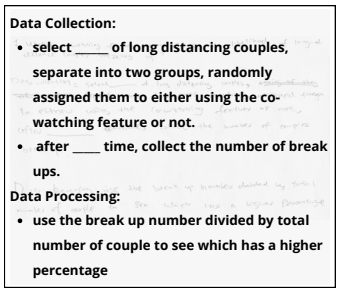
Note: * p < 0.05, ** p < 0.01, *** p < 0.001



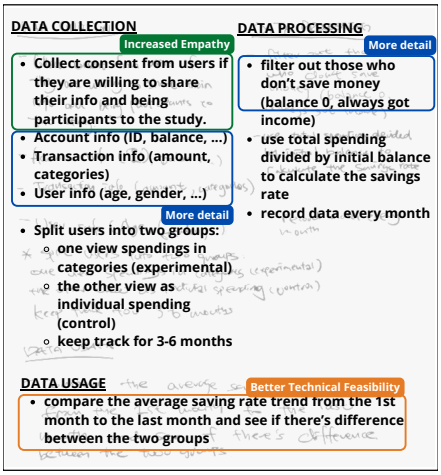
(a) P-G7's pre-game privacy design.



(b) P-G7's post-game privacy design.



(c) P-G30's pre-game privacy design.



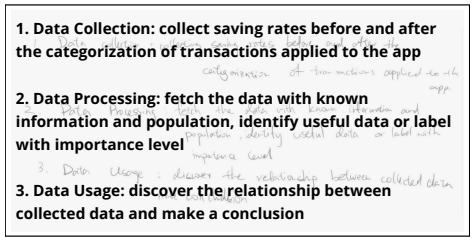
(d) P-G30's post-game privacy design.

Figure 8: Pre-/Post-game privacy designs. Participants in the game group improved by providing more details on data collection (P-G7, P-G30), processing (P-G7), and usage (P-G30), enhancing technical feasibility (P-G30), and demonstrating better respect for users by clearly defining how the data would be collected, processed, and used (P-G7). The game group shows greater improvement in privacy design skills compared to the worksheet group (Figure 9).

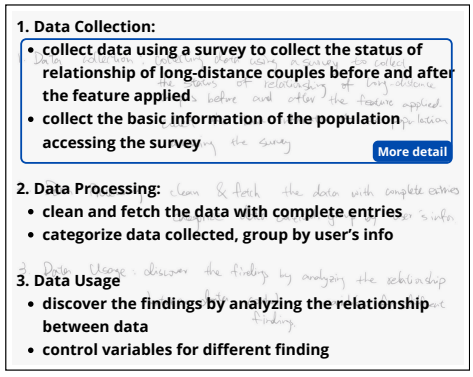
necessity of public data sharing or consider the potential privacy implications of sharing relationship data, even in anonymized form.

5.6 Participants Feedback

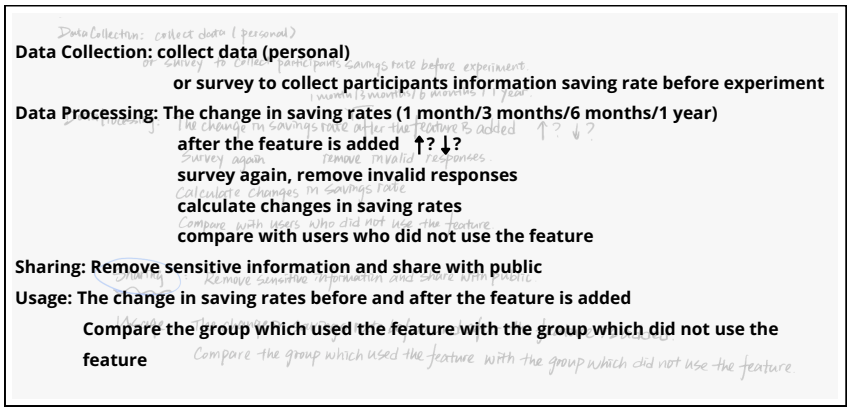
The post-study interviews revealed that Panopticon successfully maintained player engagement throughout the learning process. Multiple participants described the game as fun, which enhanced their learning experience. For example, P-G20 stated, "the game is really interesting" in the exit survey. P-G4 reflected that "learning how to do privacy designs with a board game makes me more willing to learn privacy designs." Other behavioral indicators also showcased how engaging the game was. Of the 36 players, 26 (73%) spontaneously asked if they can continue beyond the planned 40-minute time limit. By the end of our fourth session, P-G10, P-G11, and P-G12 insisted that they "want to play the game for a few more rounds." Beyond the game's entertainment value, post-study



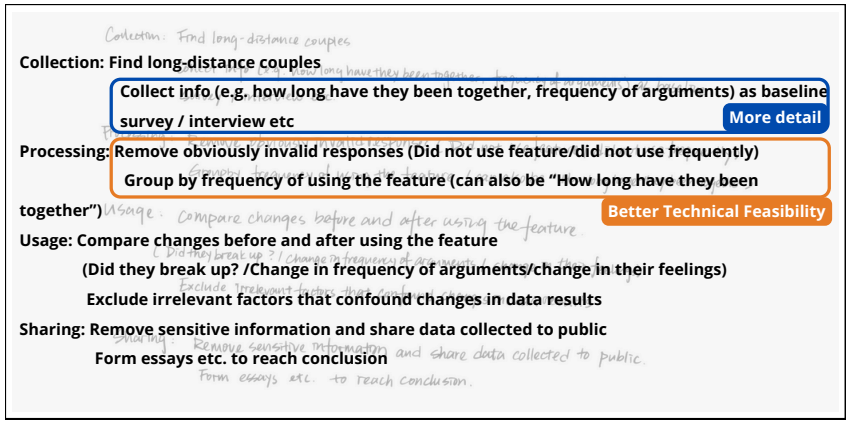
(a) P-W2's pre-worksheet privacy design.



(b) P-W2's post-worksheet privacy design.



(c) P-W5's pre-worksheet privacy design.



(d) P-W5's post-worksheet privacy design.

Figure 9: Pre-/Post-worksheet privacy designs. P-W2 shows improvement by providing more details on data processing, while P-W5 demonstrated the same level of privacy design skills before and after reading the worksheet. The improvement in the worksheet group is relatively marginal compared to that in the game group (Figure 8).

interviews also highlighted the game's ability to facilitate structured thinking under time constraints. As P-G35 noted, Panopticon "pushes me to think and capture ideas in a short period of time."

6 Discussion and Future Work

6.1 Potential Game Adaptions

The rules of Monopoly have many variations, as players often improvise to suit their needs and contexts [41]. Similarly, future educators may also adapt Panopticon in various ways.

The number of players in Panopticon. Our current evaluation focuses on the 3-player setting. However, Panopticon can potentially work well for different group sizes. Students may play the game in a solo setting [41], making the game like a rubber duck debugging session [85]. The main advantage of increasing the number of players is to increase the diversity of design critiques. However, one key challenge in multi-player settings (e.g., 6 players) is minimizing players' idle time. To mitigate this issue, we may allow players to critique and revise their designs asynchronously without blocking other players. Or we may allow players to critique the triggered data practice design collectively, fostering a collaborative learning environment.

When to use Panopticon? The ideal use case of Panopticon is an in-class activity in a middle-sized classroom, with a lecturer covering around 20 students. So, six groups would play the game simultaneously, and the lecturer could roam between groups to score the critique quality. A game session can last around 45 minutes, similar to our study design. Based on our study, participants, on average, will create 2 data practice designs, raise 4 critiques, and revise their designs 4 times.

Who can score critiques? One alternative design involves allowing players to score the critiques. For instance, in a three-player setup, the third player, who neither creates nor critiques the design, could take on the role of a judge. We did not choose this design, since we believe learning is more effective when supported by a trustworthy feedback mechanism. However, this player-as-judge design might be useful in a resource-constrained setting, e.g., a Massive open online course (MOOC) for privacy engineering.

6.2 Privacy Design Communication

We developed the design worksheet to facilitate effective communication of privacy design concepts. During the initial pilot tests, we observed that participants struggled to articulate the privacy design ideas they had in mind. Despite repeated reminders, some participants continued to focus on designing generic data science experiments rather than addressing privacy-related design decisions. Future research could explore alternative communication methods tailored specifically for privacy design.

6.3 Panopticon Playability

One potential approach to evaluate the playability of their games is through self-reported entertainment scores. However, previous research has shown that self-reported data can be biased toward experimenters' preferences [22]. We did not collect self-reported entertainment scores in our study. Instead, we looked for participants' behaviors and spontaneous feedback, which may suggest strong engagement with Panopticon. We observed that during game sessions, participants remained engaged even when waiting for others to develop privacy designs: they were either planning improvements to their own designs or preparing critiques of others' work. Additionally, 26 out of 36 (73%) players expressed interest in continuing beyond the planned 40-minute time limit.

6.4 Privacy-first Idealism in Panopticon

The privacy-first idealism in Panopticon does not perfectly reflect the real world. In reality, users adopt services for various reasons beyond privacy. Developers often must consider multiple factors in their design decisions, including regulatory constraints, conflicting stakeholder needs, and system integration challenges [7, 47].

We use this ideal setting to simplify the tasks since reflecting a comprehensive service design can take much longer than a class activity setting [11, 15, 74, 76, 81]. To minimize the dependencies on the other factors, we designed all the tasks as designing data science experiments rather than designing complete apps/services.

The current prototype of Panopticon allows students to practice privacy design in a controlled setting before engaging with more complex environments. This approach is similar to how language learners build foundational skills through structured exercises before engaging in real-world conversations [8]. Through Panopticon, we aim to help students develop a systematic and fundamental understanding of key privacy considerations. Future work could explore game scenarios that gradually introduce trade-offs, enabling students to transition from controlled settings to real-world privacy design challenges. For example, players could focus on a specific data practice in a workshop setting, iteratively refining its design.

6.5 Evaluating Privacy Designs

Evaluating privacy designs remains an open challenge due to the complexity of privacy trade-offs, the evolving nature of threats, and the diverse needs of stakeholders. Unlike security, where breaches provide clear failure points, privacy violations are often subtle, context-dependent, and difficult to quantify. Existing evaluation methods, such as expert reviews [53], user studies [28, 88], and vignette surveys [35], each have limitations in capturing real-world implications.

We employed a relatively lightweight method that allows teachers to evaluate privacy designs within the game. However, this approach has limitations, including the subjectivity of assessments and delays in the evaluation process. Future work could explore alternative methods, such as using large language models as a judge [89], to reduce evaluator subjectivity and latency.

6.6 Evaluating Learning Outcomes

Evaluating the learning outcomes of a new intervention is a well-documented challenge in education research [86]. The learning outcomes can be in multiple forms, such as knowledge gains, skill development, and conceptual understanding. The current prototype of Panopticon focuses on privacy design skill development in three aspects: (1) articulating precise design decisions, (2) anticipating the potential implications of design choices, and (3) balancing technical practicality with privacy requirements. We compared the game group with the individual worksheet group to understand the improvement in learning outcomes. Note that the worksheet approach is only one of the learning interventions today. Future work may look into other learning interventions, such as lecture-based teaching [70] and discussion-based learning [75].

6.7 Competitive or Collaborative Game

Our experiments with Panopticon revealed that players predominantly engaged in collaborative gameplay rather than directly competing against each other to win. Initially, we were concerned that players might collude to exploit the rules or deliberately provide malicious feedback, regardless of the quality of the privacy designs. However, our findings showed that nearly all participants prioritized critical thinking about privacy designs over gaining individual data points. As a result, the data points became a byproduct of their gameplay rather than the primary objective. Future research could explore more collaborative game designs [84] for privacy education. One example is a privacy escape game, where participants work together to analyze privacy risks in real-world scenarios within a time limit.

7 Conclusion

In this paper, we introduce Panopticon, an educational board game that helps data science students learn the skills of designing privacy-sensitive data practices (i.e., designing privacy) with fun. We used two learning science principles, peer learning, and formative feedback, to guide the game design. We evaluated the game through a user study with 36 players (i.e., 12 game sessions) and compared their learning outcomes to a control group (n=36) who learned privacy design through paper content. Our results suggest that Panopticon increased the learning outcomes by 354%, with significant improvements in all three dimensions.

Acknowledgments

The authors correct typos and grammatical errors using ChatGPT (GPT-4o) and Claude (Claude 3.5 Sonnet). We would like to thank our study participants whose involvements were essential to the

success of our research. We also appreciate the anonymous reviewers for their valuable feedback. This research was supported by the gift from Solix Technologies.

References

[1] 2019. ISO/IEC 27701:2019 — Security Techniques — Extension to ISO/IEC 27001 and ISO/IEC 27002 for Privacy Information Management — Requirements and Guidelines. https://www.iso.org/standard/71670.html Edition 1.

[2] D.-U. 202. 2024. Anonymity and Privacy-Responsible Data Science. https://dataresponsibly.github.io/rds24/assets/1011Privacy1017.pdf Accessed: 2024-11-18.

[3] Majed Alshammari and Andrew Simpson. 2017. Towards a Principled Approach for Engineering Privacy by Design. 161–177. https://doi.org/10.1007/978-3-319-67280-9_9

[4] Majed Alshammari and Andrew Simpson. 2017. Towards a principled approach for engineering privacy by design. In *Privacy Technologies and Policy: 5th Annual Privacy Forum, APF 2017, Vienna, Austria, June 7-8, 2017, Revised Selected Papers 5*. Springer, 161–177.

[5] Paul Anderson, James Bowring, Renée McCauley, George Pothering, and Christopher Starr. 2014. An undergraduate degree in data science: curriculum and a decade of implementation experience. In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education*. ACM, 145–150.

[6] American Library Association. 2010. Privacy. https://www.ala.org/advocacy/privacy Accessed: 2025-02-12.

[7] Sunil Babbar, Ravi Behara, and Edna White. 2002. Mapping product usability. *International Journal of Operations & Production Management* 22, 10 (2002), 1071–1089.

[8] Susan J Behrens, Judith A Parker, et al. 2010. *Language in the Real World*. Routledge. Oxon.

[9] Tamara Bonaci. 2022. *Introduction to Privacy Engineering*. https://peden.ece.uw.edu/pmp/wp-content/uploads/sites/2/2022/04/SPR22-Introduction-to-Privacy-Engineering-Bonaci.pdf Accessed: 2024-10-16.

[10] David Boud, Ruth Cohen, et al. 2014. *Peer learning in higher education: Learning from and with each other*. Routledge.

[11] Carnegie Mellon University. 2024. 17-631: Information Security: Privacy and Policy. https://www.cylab.cmu.edu/education/course-list/information-security-privacy-policy.html Accessed: 2024-11-18.

[12] Ann Cavoukian and Jeff Jonas. 2012. Privacy by design in the age of big data. (2012).

[13] Joana Cerejo and Miguel Carvalhais. 2023. Anticipation as a Tool for Designing the Future. In *International Conference on Design and Digital Communication*. Springer, 37–52.

[14] Merijke Coenraad, Anthony Pellicone, Diane Jass Ketelhut, Michel Cukier, Jan Plane, and David Weintrop. 2020. Experiencing cybersecurity one game at a time: A systematic review of cybersecurity digital games. *Simulation & Gaming* 51, 5 (2020), 586–611.

[15] Lorrie Faith Cranor. 2015. *Privacy engineering, privacy by design, and privacy governance*. https://cups.cs.cmu.edu/courses/pplt-fa15/slides/151117privacyengineering.pdf Accessed: 2024-10-16.

[16] Lorrie Faith Cranor and Norman Sadeh. 2013. Privacy engineering emerges as a hot new career. *IEEE Potentials* 32, 6 (2013), 7–9.

[17] Andrea Danyluk, Paul Leidig, Andrew McGettrick, Lillian Cassel, Maureen Doyle, Christian Servin, Karl Schmitt, and Andreas Stefik. 2021. Computing competencies for undergraduate data science programs: An ACM task force final report. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*. 1119–1120.

[18] Tamara Denning, Tadayoshi Kohno, and Adam Shostack. 2013. Control-Alt-Hack™: a card game for computer security outreach and education (abstract only). In *Proceeding of the 44th ACM Technical Symposium on Computer Science Education* (Denver, Colorado, USA) *(SIGCSE '13)*. Association for Computing Machinery, New York, NY, USA, 729. https://doi.org/10.1145/2445196.2445408

[19] Tamara Denning, Adam Lerner, Adam Shostack, and Tadayoshi Kohno. 2013. Control-Alt-Hack: the design and evaluation of a card game for computer security awareness and education. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*. 915–928.

[20] Anind K Dey, Timothy Sohn, Sara Streng, and Justin Kodama. 2006. iCAP: Interactive prototyping of context-aware applications. In *Pervasive Computing: 4th International Conference, PERVASIVE 2006, Dublin, Ireland, May 7-10, 2006. Proceedings 4*. Springer, 254–271.

[21] Danielle DuMerer. 2021. Technology for Public Policy. https://harris.uchicago.edu/files/2021-02/ppha38550_technologyforpublicpolicy_dumerer_spring_2021.pdf. Course Syllabus, Harris School of Public Policy, University of Chicago.

[22] Robert J Fisher and James E Katz. 2000. Social-desirability bias and the validity of self-reported values. *Psychology & marketing* 17, 2 (2000), 105–120.

[23] Mark Gondree and Zachary NJ Peterson. 2013. Valuing Security by Getting {[d0x3d!]}: Experiences with a Network Security Board Game. In *6th Workshop on Cyber Security Experimentation and Test (CSET 13)*.

[24] Google. 2024. Be Internet Awesome: Interland. https://beinternetawesome.withgoogle.com/en_us/interland/ Accessed: 2024-11-27.

[25] Weijia He, Nathan Reitinger, Atheer Almogbil, Yi-Shyuan Chiang, Timothy J Pierson, and David Kotz. 2024. Contextualizing Interpersonal Data Sharing in Smart Homes. *Proceedings on Privacy Enhancing Technologies* (2024).

[26] Catherine M Hicks, Vineet Pandey, C Ailie Fraser, and Scott Klemmer. 2016. Framing feedback: Choosing review environment features that support high quality peer assessment. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 458–469.

[27] Kashmir Hill. 2014. OkCupid Experiment Shows the Perils of Trusting Online Matchmaking. *Forbes* (2014). https://www.forbes.com/sites/kashmirhill/2014/07/28/okcupid-experiment-compatibility-deception/ Accessed: 2024-10-31.

[28] Lisa Janicke Hinchliffe. 2018. Privacy in User Research: Can You? *The Scholarly Kitchen*. https://scholarlykitchen.sspnet.org/2018/09/05/privacy-in-user-research-can-you/?utm_source=chatgpt.com

[29] Jaap-Henk Hoepman. 2014. Privacy design strategies. In *IFIP International Information Security Conference*. Springer, 446–459.

[30] Jason I Hong, Jennifer D Ng, Scott Lederer, and James A Landay. 2004. Privacy risk models for designing privacy-sensitive ubiquitous computing systems. In *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques*. 91–100.

[31] Alastair Irons and Sam Elkington. 2021. *Enhancing Learning through Formative Assessment and Feedback* (2nd ed.). Routledge, London. 248 pages. https://doi.org/10.4324/9781138610514

[32] Cynthia Irvine, Michael Thompson, and Karen Allen. 2005. The CyberCIEGE Game: An Immersive Educational Tool for Cybersecurity Training. In *Proceedings of the IFIP TC11 International Conference on Information Security*. Springer, 122–130. https://doi.org/10.1007/11594012_14 Accessed: 2024-11-27.

[33] C Kirabo Jackson and Elias Bruegmann. 2009. Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics* 1, 4 (2009), 85–108.

[34] Gokul Chettoor Jayakrishnan, Gangadhara Reddy Sirigireddy, Sukanya Vaddepalli, Vijayanand Banahatti, Sachin Premsukh Lodha, and Sankalp Suneel Pandit. 2020. Password: A serious game to promote password awareness and diversity in an enterprise. In *Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020)*. 1–18.

[35] Haojian Jin, Hong Shen, Mayank Jain, Swarun Kumar, and Jason I. Hong. 2021. Lean Privacy Review: Collecting Users' Privacy Concerns of Data Practices at a Low Cost. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. https://www.researcher-app.com/paper/8542433 Accessed: 2024-10-16.

[36] Dr. D. Keerthirathne. 2021. Peer Learning: An Overview. https://www.researchgate.net/profile/Dr-Keerthirathne/publication/355209445_Peer_Learning_an_Overview/links/616873f13851f9599407d660/Peer-Learning-an-Overview.pdf. Accessed on ResearchGate.

[37] Alice Y. Kolb and David A. Kolb. 2005. Learning Styles and Learning Spaces: Enhancing Experiential Learning in Higher Education. *Academy of Management Learning & Education* 4, 2 (2005), 193–212.

[38] David A. Kolb. 1984. *Experiential Learning: Experience as the Source of Learning and Development*. Prentice Hall.

[39] Markus Krause, Tom Garncarz, JiaoJiao Song, Elizabeth M Gerber, Brian P Bailey, and Steven P Dow. 2017. Critique style guide: Improving crowdsourced design feedback with a natural language model. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 4627–4639.

[40] Susan Landau. 2014. Educating engineers: teaching privacy in a world of open doors. *IEEE security & privacy* 12, 3 (2014), 66–70.

[41] Learn New Games. 2024. Solo Play: How to Play Monopoly by Yourself – Learn New Games. https://www.learnnewgames.com/solo-play-can-i-play-monopoly-by-myself/. (Accessed on 11/30/2024).

[42] Scott Lederer, Jason I Hong, Anind K Dey, and James A Landay. 2004. Personal privacy through understanding and action: five pitfalls for designers. *Personal and ubiquitous computing* 8 (2004), 440–454.

[43] Joey J. Lee and Jessica Hammer. 2011. Gamification in Education: What, How, Why Bother? *Academic Exchange Quarterly* 15, 2 (2011). https://cmu-ctp.github.io/assets/publications/pdfs/Lee-Hammer-Gamification-Education-2011.pdf

[44] Tony W Li, Arshia Arya, and Haojian Jin. 2024. Redesigning Privacy with User Feedback: The Case of Zoom Attendee Attention Tracking. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–14.

[45] Chia-Yu Liu and Hung-Ling Chen. 2020. Effects of peer learning on learning performance, motivation, and attitude. *International Journal of Education Economics and Development* 11, 4 (2020), 420–443.

[46] Ngar-Fun Liu and David Carless. 2006. Peer Feedback: The Learning Element of Peer Assessment. *Assessment & Evaluation in Higher Education* 31, 4 (2006), 457–468. https://web.edu.hku.hk/f/staff/412/2006_Peer-feedback-The-learning-element-of-peer-assessment.pdf Accessed: 2024-11-27.

[47] Lucy Ellen Lwakatare, Aiswarya Raj, Ivica Crnkovic, Jan Bosch, and Helena Holmström Olsson. 2020. Large-scale machine learning systems in real-world industrial settings: A review of challenges and solutions. *Information and software technology* 127 (2020), 106368.
[48] Sana Maqsood and Sonia Chiasson. 2021. Design, development, and evaluation of a cybersecurity, privacy, and digital literacy game for tweens. *ACM Transactions on Privacy and Security (TOPS)* 24, 4 (2021), 1–37.
[49] Philipp K Masur. 2020. How online privacy literacy supports self-data protection and self-determination in the age of information. *Media and Communication* 8, 2 (2020), 258–269.
[50] Mary L McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia medica* 22, 3 (2012), 276–282.
[51] Microsoft. 2025. Elevation of Privilege (EoP) Threat Modeling Card Game. https://www.microsoft.com/en-us/download/details.aspx?id=20303 Accessed: 2025-02-12.
[52] National Academies of Sciences, Engineering, and Medicine. 2018. *Data Science for Undergraduates: Opportunities and Options*. The National Academies Press, Washington, DC. https://doi.org/10.17226/25104
[53] National Institute of Standards and Technology 2020. *NIST Privacy Framework: A Tool for Improving Privacy Through Enterprise Risk Management*. National Institute of Standards and Technology, Gaithersburg, MD. https://www.nist.gov/system/files/documents/2020/01/16/NIST%20Privacy%20Framework_V1.0.pdf
[54] Tricia J Ngoon, C Ailie Fraser, Ariel S Weingarten, Mira Dontcheva, and Scott Klemmer. 2018. Interactive guidance techniques for improving creative feedback. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–11.
[55] Terms of Service; Didn't Read Project. 2024. Terms of Service; Didn't Read. https://tosdr.org Accessed: 2024-11-27.
[56] Marc Olano, Alan Sherman, Linda Oliva, Ryan Cox, Deborah Firestone, Oliver Kubik, Milind Patil, John Seymour, Isaac Sohn, and Donna Thomas. 2014. {SecurityEmpire}: Development and evaluation of a digital game to promote cybersecurity education. In *2014 USENIX Summit on Gaming, Games, and Gamification in Security Education (3GSE 14)*.
[57] Philip E. Orbanes. 2007. *Monopoly: The World's Most Famous Game and How It Got That Way*. Da Capo Press, Cambridge, MA.
[58] Yong Jin Park. 2013. Digital literacy and privacy behavior online. *Communication research* 40, 2 (2013), 215–236.
[59] Penn State University Libraries. 2024. Digital Shred: Privacy Literacy Toolkit. https://sites.psu.edu/digitalshred/ Accessed: 2024-11-21.
[60] Justin Petelka, Yixin Zou, and Florian Schaub. 2019. Put your warning where your link is: Improving and evaluating email phishing warnings. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–15.
[61] Christine Prince, Nessrine Omrani, and Francesco Schiavone. 2024. Online privacy literacy and users' information privacy empowerment: the case of GDPR in Europe. *Information Technology & People* 37, 8 (2024), 1–24.
[62] Privado.ai. 2024. Privado: Privacy Code Scanning Platform. https://www.privado.ai/ Accessed: 2025-02-13.
[63] Stephanie Rosenthal and Tingting (Rachel) Chung. 2020. A Data Science Major: Building Skills and Confidence. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education (SIGCSE '20)*. ACM, 178–184. https://doi.org/10.1145/3328778.3366791
[64] Marco Saltarella, Giuseppe Desolda, Rosa Lanzilotti, and Vita Santa Barletta. 2024. Translating privacy design principles into human-centered Software Lifecycle: A literature review. *International Journal of Human–Computer Interaction* 40, 17 (2024), 4465–4483.
[65] San Jose Public Library. 2024. Virtual Privacy Lab. https://www.sjpl.org/privacy/ Accessed: 2024-11-21.
[66] Xiaoxin Shen, Eman Alashwali, and Lorrie Faith Cranor. 2024. What Do Privacy Advertisements Communicate to Consumers? *arXiv preprint arXiv:2405.13857* (2024).
[67] Steve Sheng, Bryant Magnien, Ponnurangam Kumaraguru, Alessandro Acquisti, Lorrie Faith Cranor, Jason Hong, and Elizabeth Nunge. 2007. Anti-Phishing Phil: The Design and Evaluation of a Game That Teaches People Not to Fall for Phish. In *Proceedings of the 3rd Symposium on Usable Privacy and Security (SOUPS '07)*. ACM, Pittsburgh, Pennsylvania, USA, 88–99. https://doi.org/10.1145/1280680.1280692
[68] Valerie J. Shute. 2008. Focus on Formative Feedback. *Review of Educational Research* 78, 1 (2008), 153–189. https://journals.sagepub.com/doi/10.3102/0034654307313795
[69] Valerie J Shute. 2008. Focus on formative feedback. *Review of educational research* 78, 1 (2008), 153–189.
[70] Emily M Smith and NG Holmes. 2017. Seeing the real world: Comparing learning from verification labs and traditional or enhanced lecture demonstrations. *arXiv preprint arXiv:1712.03174* (2017).
[71] CUPS Staff. 2006. Computer Science and Society. https://cups.cs.cmu.edu/courses/compsoc-sp06/. Course Schedule, Carnegie Mellon University.
[72] CSE484 Staff. 2009. Introduction to Computer Security. https://courses.cs.washington.edu/courses/cse484/09wi/lectures/Lecture01-2009.pdf. Lecture Notes for CSE 484, University of Washington.
[73] FTC Staff. 2011. Protecting consumer privacy in an era of rapid change–a proposed framework for businesses and policymakers. *Journal of Privacy and Confidentiality* 3, 1 (2011).
[74] William Stallings. 2019. *Information privacy engineering and privacy by design: Understanding privacy threats, technology, and regulations based on standards and best practices*. Addison-Wesley Professional.
[75] Tova Stenlund, Fredrik U Jönsson, and Bert Jonsson. 2017. Group discussions and test-enhanced learning: individual learning outcomes and personality characteristics. *Educational Psychology* 37, 2 (2017), 145–156.
[76] John Sweller. 2011. Cognitive load theory. In *Psychology of learning and motivation*. Vol. 55. Elsevier, 37–76.
[77] Tactical Tech. 2024. Data Detox Kit. https://tacticaltech.org/projects/data-detox-kit/ Accessed: 2024-11-21.
[78] Michael Thompson and Cynthia Irvine. 2011. Active learning with the {CyberCIEGE} video game. In *4th workshop on cyber security experimentation and test (CSET 11)*.
[79] Keith J Topping. 2005. Trends in peer learning. *Educational psychology* 25, 6 (2005), 631–645.
[80] Sabine Trepte, Doris Teutsch, Philipp K. Masur, Carolin Eicher, Mona Fischer, Alisa Hennhöfer, and Fabienne Lind. 2014. Do People Know About Privacy and Data Protection Strategies? Towards the "Online Privacy Literacy Scale" (OPLIS). In *Reforming European Data Protection Law*, Serge Gutwirth, Ronald Leenes, Paul de Hert, and Yves Poullet (Eds.). Law, Governance and Technology Series, Vol. 20. Springer, 333–365. https://doi.org/10.1007/978-94-017-9385-8_14
[81] University of Washington. 2024. CSE 564: Computer Security. https://courses.cs.washington.edu/courses/cse564/24wi/syllabus/ Accessed: 2024-11-18.
[82] U.S. Securities and Exchange Commission, Office of Information Technology. 2007. *Privacy Impact Assessment (PIA) Guide*. U.S. Securities and Exchange Commission, Alexandria, VA. https://www.sec.gov/about/privacy/piaguide.pdf Revised version.
[83] Richard D. De Veaux, Mahesh Agarwal, Maia Averett, Benjamin S. Baumer, Andrew Bray, Thomas C. Bressoud, Lance Bryant, Lei Z. Cheng, Amanda Francis, Robert Gould, et al. 2017. Curriculum guidelines for undergraduate programs in data science. *Annual Review of Statistics and Its Application* 4 (2017), 15–30.
[84] Wendy Wang, Yu Tao, Kai Wang, Dominik Jedruszczak, and Ben Knutson. 2016. Leveraging Crowd for Game-based Learning: A Case Study of Privacy Education Game Design and Evaluation by Crowdsourcing. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 1116–1121.
[85] Wikipedia. 2024. Rubber duck debugging - Wikipedia. https://en.wikipedia.org/wiki/Rubber_duck_debugging. (Accessed on 11/30/2024).
[86] Michael Wilkes and John Bligh. 1999. Evaluating educational interventions. *BMJ* 318, 7186 (1999), 1134–1137. https://doi.org/10.1136/bmj.318.7186.1134
[87] Kim Wuyts, Laurens Sion, and Wouter Joosen. 2020. Linddun go: A lightweight approach to privacy threat modeling. In *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 302–309.
[88] Eric Zeng and Franziska Roesner. 2019. Understanding and improving security and privacy in {multi-user} smart homes: A design exploration and {in-home} user study. In *28th USENIX Security Symposium (USENIX Security 19)*. 159–176.
[89] Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric Xing, et al. 2023. Judging llm-as-a-judge with mt-bench and chatbot arena. *Advances in Neural Information Processing Systems* 36 (2023), 46595–46623.

A Task Bank

See Table 7.

B Data Practice Scenarios

See Table 8.

C Special Action Spaces

See Table 9.

Table 7: We designed 2 data practice scenarios for each digital service type. Each prompt contains a Context, a Challenge, and a Hypothesis. These data practice scenarios help players practice communication of privacy design ideas.

Social
In the increasingly digital world of romance, dating apps have become a primary means for people to connect with potential partners. A popular dating app currently features a matching page that displays user profiles with basic information such as photos, age, occupation, and a short bio. Users can swipe right to express interest or left to pass. In an effort to gain a competitive edge and differentiate itself, the app is exploring new ways to enhance its matching system and user experience. A social networking app is interested in knowing **whether displaying mutual connections on the matching page could increase the likelihood of successful matches**.
Social media platforms offer a wide variety of text-based content, from short-form posts and status updates to longer articles and blog entries. These apps typically include features such as infinite scrolling, personalized content feeds, like and share buttons, and comment sections. An entertainment app is now interested in determining **whether content with mostly positive connotations (such as success stories, acts of kindness, or motivational quotes) or content with mostly negative connotations (such as critical news updates, societal issues, or personal struggles) is more effective in increasing users' engagement time**.
Health & Fitness
Many fitness apps offer features designed primarily for younger, active users. These include personalized high-intensity workout recommendations based on physical attributes and goals, extensive libraries of vigorous exercise videos, and customizable fitness plans emphasizing rapid progression. Users can track detailed progress and integrate data from wearable devices for comprehensive health monitoring. In an effort to expand its user base and promote inclusive fitness solutions, the app is exploring new ways to cater to diverse age groups and fitness levels. A fitness app is interested in exploring **whether the integration of age-specific fitness programs could encourage older adults to engage in more regular exercise routines**.
Fitness apps traditionally focus on longer, dedicated workout sessions, offering personalized high-intensity routines, exercise videos, and tracking progression. These typically range from 30 minutes to an hour, often making daily exercise challenging for many users. However, in an effort to address the growing health concerns associated with sedentary behavior, the app is exploring innovative ways to promote physical activity throughout the day, beyond just these longer workout sessions. A fitness app is interested in exploring **whether integrating features that schedule fitness breaks (short periods of physical activity incorporated during working hours) can improve users' overall productivity at work**.
Productivity
Productivity apps typically focus on features like task management, goal setting, and time tracking, primarily catering to working professionals or students. These apps often include calendar integration, to-do lists, and project planning tools. However, as the global population ages, a significant challenge has emerged: loneliness and social isolation among older adults. Recognizing this growing concern and the changing demographics, a productivity app is interested in exploring **whether incorporating recommendations for local volunteer activities can make older adults more active and engaged post-retirement**.
Productivity apps commonly offer features like to-do lists, calendar integration, and goal tracking. These apps typically allow users to create tasks, set deadlines, and receive reminders. Some include more advanced features such as time tracking, project management tools, and focus timers. Despite the availability of various productivity tools, students often report that procrastination remains a barrier to their academic success. In light of this challenge, a productivity app is interested in exploring **whether incorporating mandatory study scheduling features can improve students' academic performance**.

Table 8: Scenarios used to evaluate participants' privacy design skills.

Scenario	Scenario Description
TikTok Scenario	Short-form video platforms like TikTok typically offer individual viewing experiences, with features like personalized content recommendations, video playback, and social sharing options. Long-distance couples frequently seek ways to build memories together, even when they cannot be physically present in the same location. This desire for shared experiences despite geographical separation presents a unique opportunity for entertainment apps. An entertainment app is interested in determining whether co-watching features can reduce the likelihood of long-distance couples breaking up.
Finance App Scenario	Peer-to-peer transaction apps like Venmo and Zelle have become essential tools for money transfers and expense tracking. These apps provide users with comprehensive records of their financial transactions, offering unprecedented visibility into personal spending habits. Recently, many have introduced automatic categorization of transactions, allowing users to see an overview of their spending by category before diving into specific details. A finance app wants to know if showing users their spending in categories and overall summaries helps them achieve a higher savings rate compared to when users look at each transaction one by one.

Table 9: Special action space exploration. We designed multiple sets of special action spaces throughout the design process. Ultimately, we kept those that introduced extra opportunities for players to initiate or revise privacy designs and interact with peers.

Special Action	Description	Status
Insightful Inquiry	Ask other players for elaboration on one of the existing designs.	Kept
Data Reboot	Revise any of the owned designs without cost.	Kept
Patent Piracy	Claim any claimed/unclaimed digital service space on the board.	Kept
Privacy Violation	Skip one turn	Discarded
Innovation Hub	Receive 200 bonus data points.	Discarded
Research Grant	Draw a card to determine the amount of additional funding you receive (0-250 data points).	Discarded
Regulatory Compliance	Draw a card determining penalties (0-250 data points) for compliance issues.	Discarded
Regulatory Enforcement	Move to Regulatory Compliance space and skip one turn.	Discarded
Free Consultation	Choose another player to give them advice on one of their data practices.	Discarded