Assignee
Part II

David Wu

Contents

1. Overview 4
  1.1. Introduction 4
  1.2. Philosophy 4
  1.3. Methodology 4
    1.3.1. Evaluation Survey 4
    1.3.2. Technical Testing 4
  1.4. Resources 4
    1.4.1. Repository 5
    1.4.2. Prebuilt Binaries 5
2. Evaluation Survey 6
  2.1. Construction 6
    2.1.1. Respond Comfort 6
    2.1.2. Structural Design 7
    2.1.3. Accessibility 8
    2.1.4. Presentation 8
    2.1.5. Implementation 9
    2.1.6. Auxiliary 9
  2.2. Distribution 9
    2.2.1. Embedded Form 9
    2.2.2. Embedded Link 10
    2.2.3. Other Methods 10
    2.2.4. Population Sample 10
  2.3. Results Analysis 11
    2.3.1. Preprocessing 11
    2.3.2. Quantitative Analysis 11
      2.3.2.1. Demographics 11
      2.3.2.2. Descriptive Statistics 13
      2.3.2.3. Correlation 15
      2.3.2.4. Consistency 18
    2.3.3. Qualitative Analysis 19
      2.3.3.1. Functionalities 19
      2.3.3.2. Pain Points 19
      2.3.3.3. Improvements 19
      2.3.3.4. No-Responses 20
      2.3.3.5. Synthesis 20
  2.4. Improvements 20
    2.4.1. Open Dyslexic 20
    2.4.2. Invitation Link 21
    2.4.3. Planned 21
3. Technical Testing 22
  3.1. Environment 22
  3.2. Functional 22
    3.2.1. Framework 22

3.2.2. Coverage ... 22
3.2.2.1. Unit Tests ... 22
3.2.2.2. Integration Tests ... 22
3.2.2.3. E2E Tests ... 23
3.3. Usability ... 23
3.4. Interface ... 23
3.5. Database ... 24
3.6. Performance ... 24
3.7. Compatibility ... 25
3.8. Security ... 26
4. Credits ... 27
4.1. Common ... 27
4.2. Survey Build ... 27
4.3. Survey Analysis ... 27
4.4. Technical Testing ... 27
4.5. Reference ... 27
4.6. Report ... 27
5. Final Remarks ... 28

1. Overview

1.1. Introduction

This report details the testing and evaluation of Assignee, a full-stack web application developed for the ICT SBA task of implementing an assignment management system. The assessment integrates empirical user experience (UX) research with rigorous technical testing to provide a holistic view of readiness and areas for improvement.

Within this report, we:

- Empirically evaluate UX across 7 key dimensions: Useful, Usable, Accessible, Findable, Credible, Valuable, and Desirable.
- Rigorously test robustness across 7 key areas: Functional, Usability, Interface, Database, Performance, Compatibility, and Security.
- Synthesize findings from both streams into a unified, actionable analysis.
- Provide prioritized, evidence-based recommendations for fixes and improvements.

1.2. Philosophy

We employed a mixed-methods approach:

- Quantitative: Statistical analysis of structured survey data.
- Qualitative: Analysis of open-ended user feedback and heuristic evaluation.
- Empirical: Objective technical testing using automated tools and manual procedures.

1.3. Methodology

1.3.1. Evaluation Survey

A digital survey was administered to a target group of users to measure subjective experience. The instrument used a 5-point Likert scale to quantify perceptions across the 7 core constructs: Useful, Usable, Accessible, Findable, Credible, Valuable, and Desirable.

Demographic data was optionally collected for cohort analysis. Results were analyzed using descriptive statistics to identify significant differences between user groups.

1.3.2. Technical Testing

Technical quality was assessed through structured, multi-faceted testing.

Functionality was validated via automated test cases. Usability was analyzed with user logging and Lighthouse. Interface was tested against usage heuristics. Database integrity and query performance were analyzed. Performance was measured using automated tools and Lighthouse. Compatibility was verified against a defined matrix. Security posture was assessed through automated vulnerability scanning.

1.4. Resources

The following resources are provided for SBA invigilators' reference and validation purposes. The version of Assignee that is tested and evaluated is a slightly updated version of the one presented in Part I, which the report was also updated:

Report Part I° (https://github.com/wavim/assignee/blob/main/report/report-p1.pdf)

1.4.1. Repository

The complete project is hosted in a repository, accessible at

Repository° (https://github.com/wavim/assignee)

for inspection.

The version of Assignee that is tested and evaluated is entirely based off the commit with hash `3ab4cc5f5d2c9357c964e3d5191fb947429238f6`.

1.4.2. Prebuilt Binaries

To accommodate environments without development dependencies, prebuilt archives are available in

Assignee Releases° (https://github.com/wavim/assignee/releases)

for invigilators.

The version of Assignee that is tested and evaluated is tagged `v1`, available for x86-64 Windows.

To execute the application:

1. Extract the archive to a preferred location
2. Run the prebuilt binary `app.exe` and follow prompts

Database records persist in the `app.db` file.

2. Evaluation Survey

Backing UX analysis is a digital survey to efficiently gather precise data, providing detailed, accurate, and easily analyzable information faster and more comprehensively than traditional methods.

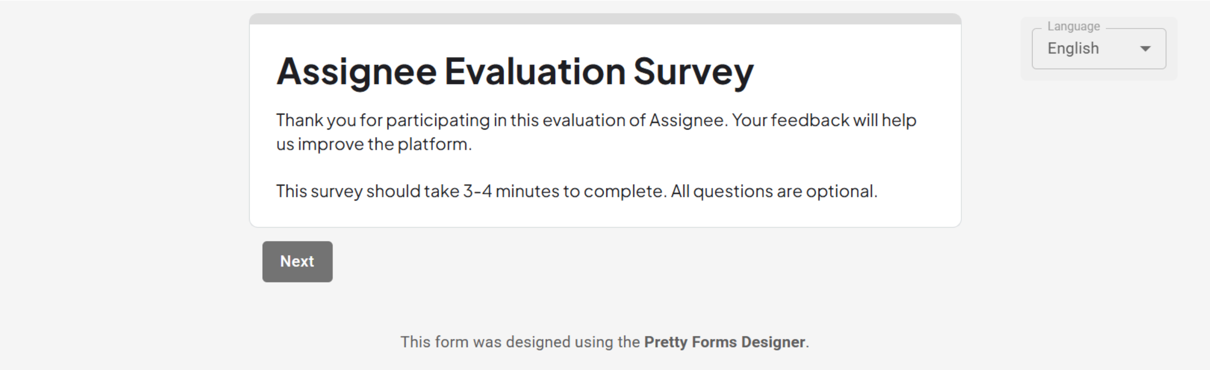
Within this chapter, we detail the construction, distribution, and analysis of the user experience survey. Do note that many rationales and content of the survey is detailed in the analysis.

2.1. Construction

The survey instrument was meticulously designed according to a respondent-centric philosophy that prioritizes accessibility, reduced cognitive burden, and ethical data collection, while simultaneously ensuring the gathering of robust, analyzable data to meet the evaluation's objectives.

2.1.1. Respond Comfort

The core design principle was to minimize barriers to participation and decision fatigue. All questions were made optional, eliminating any sense of coercion and allowing participants to engage only with items they felt comfortable answering. To respect participants' time and mental effort, the survey was designed to be completable within 3–4 minutes. This brevity targets a higher completion rate and reduces survey abandonment, directly enhancing data quality and volume for analysis.



Assignee Evaluation Survey

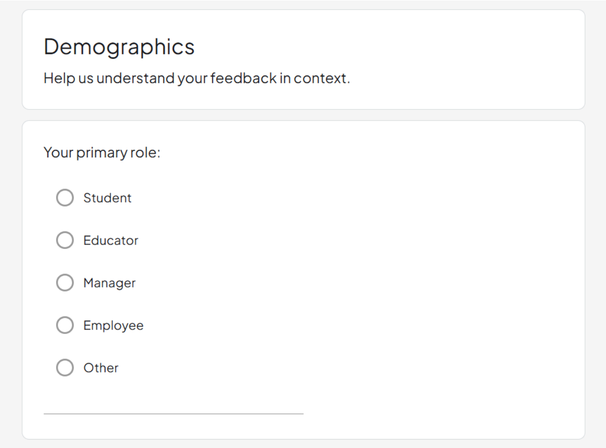
Thank you for participating in this evaluation of Assignee. Your feedback will help us improve the platform.

This survey should take 3-4 minutes to complete. All questions are optional.

Language: English

This form was designed using the **Pretty Forms Designer**.

Figure 1: Survey Introduction



Demographics

Help us understand your feedback in context.

Your primary role:

○ Student

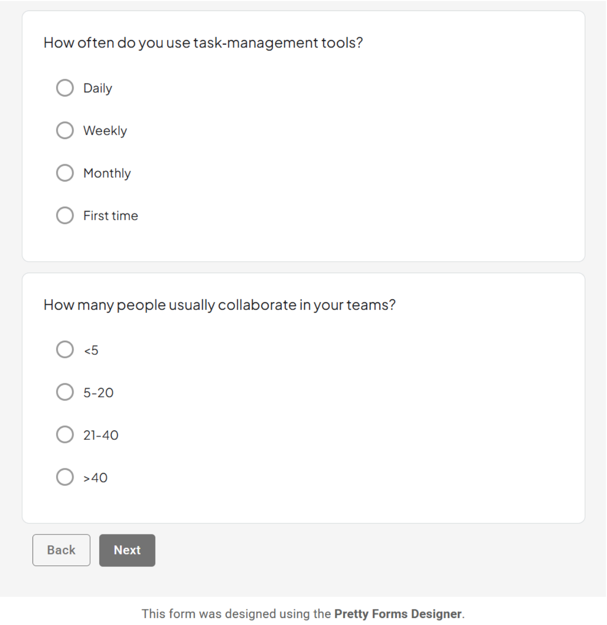
○ Educator

○ Manager

○ Employee

○ Other

Figure 2: Demographics (1)



How often do you use task-management tools?

○ Daily

○ Weekly

○ Monthly

○ First time

How many people usually collaborate in your teams?

○ <5

○ 5-20

○ 21-40

○ >40

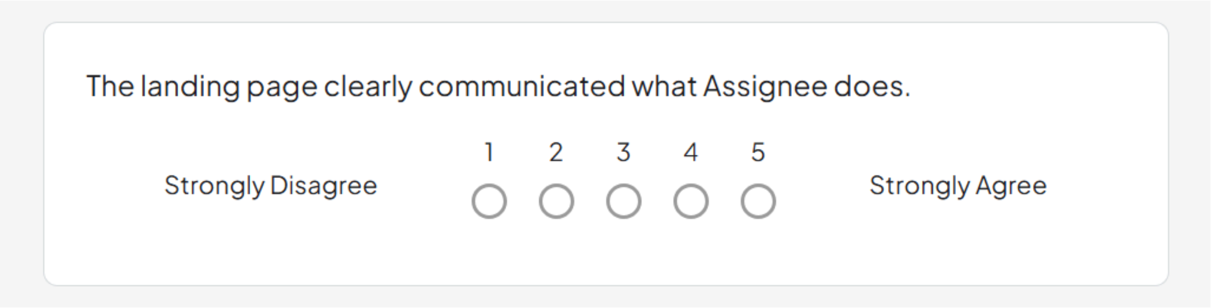
This form was designed using the **Pretty Forms Designer**.

Figure 3: Demographics (2)

2.1.2. Structural Design

The questionnaire employed a hybrid quantitative-qualitative structure to provide both measurable metrics and rich contextual insights. This balances the effects of a shorter survey.

The seven key UX constructs: Useful, Usable, Accessible, Findable, Credible, Valuable, and Desirable were each measured using standard 5-point Likert scale questions (e.g. The available accessibility options were sufficient to my needs: Strongly Disagree - Strongly Agree) in their dedicated page. The scale was chosen over e.g. 7-point scale to simplify the cognitive task for respondents, enhancing reliability, while still providing sufficient gradation for meaningful statistical analysis, including descriptive statistics and variance.



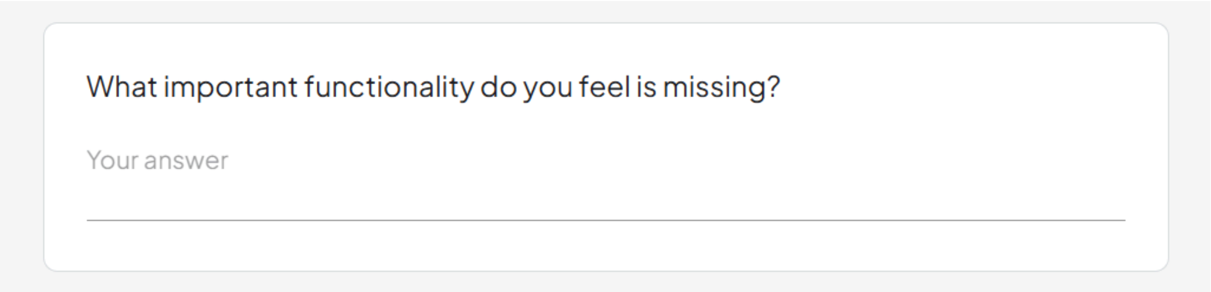
The landing page clearly communicated what Assignee does.

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

Figure 4: Likert 5-Point Scale

Following core sections, open-ended fields were included (e.g. What important functionality do you feel is missing?). This design provides crucial context for the quantitative scores, helping to explain why a score was given and surfacing unanticipated issues or praises that predefined questions might miss.

By following this structure, (i.e. 3 Likert scale + 1 open-ended question) for each separate page concerning each UX area, this reduces decision fatigue and form abandonment. It is also more modularized for clearer input and analysis.



What important functionality do you feel is missing?

Your answer

Figure 5: Open Ended Question

The consistent use of a single, simple question format (Likert scale) minimizes the mental effort required to parse new instructions, reducing decision fatigue and increasing the accuracy of responses.

2.1.3. Accessibility

The design actively promotes inclusivity and ethical practice.

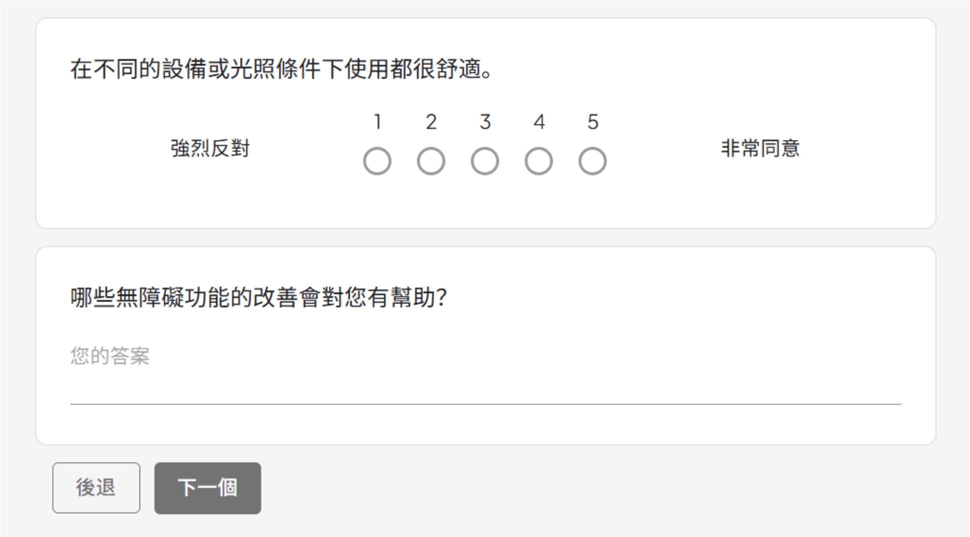
By strictly avoiding complex question types (e.g. image click-maps, complex branching), the survey remains compatible with screen readers and other assistive technologies. Furthermore, this simplicity allows the survey to be easily printed and distributed offline as a paper form, ensuring participation from users with special needs.

2.1.4. Presentation

Professional presentation was used to foster trust and engagement.

The survey was visually themed to match Assignee's color palette and typography. This creates a seamless, professional experience that reinforces the survey's legitimacy and connects the feedback directly to the product.

The instrument was designed with multi-language support in mind, ensuring clarity and comfort for a diverse user base, which in turn improves response accuracy and participation rates across different groups.



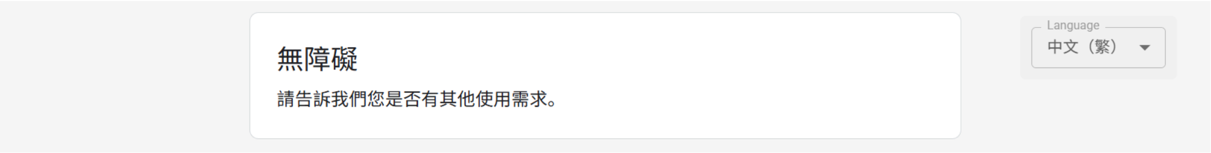
在不同的設備或光照條件下使用都很舒適。

	1	2	3	4	5	
強烈反對	○	○	○	○	○	非常同意

哪些無障礙功能的改善會對您有幫助?

您的答案

Figure 6: Chinese Translation Available



無障礙

請告訴我們您是否有其他使用需求。

Language: 中文（繁）

Figure 7: Translation Toggle

2.1.5. Implementation

The survey was implemented using Google Forms, augmented with the Pretty Forms Designer plugin. This combination provided an optimal, resource-efficient platform that aligned with the survey's respondent-centric and analytical goals:

- Intuitive interface allowed for rapid development and iteration.
- Free and instantly accessible via any web browser, ensuring no barriers to participation.
- Responses were collected in real-time into a Google Sheet, eliminating manual data entry and creating a direct feed for statistical analysis.

Furthermore, respondent emails are not collected, reassuring them to complete the survey without feeling forced.

2.1.6. Auxiliary

Standardized frameworks like the Net Promoter Score (NPS), or the System Usability Scale (SUS) were considered but ultimately not selected. While valuable for benchmarking, these tools provide generalized scores (e.g. loyalty) that are less effective for diagnosing specific, actionable insights in a bespoke application.

Instead, concrete questions on UX areas allows us to move beyond a single number and pinpoint precisely which aspects of the user experience are strengths or require intervention, delivering the targeted analysis necessary for effective, iterative improvement in a development context.

This construction demonstrates a deliberate focus on obtaining high-quality, actionable data by first respecting the participant's experience, ensuring that the insights gathered are founded on comfortable, voluntary, and clear feedback.

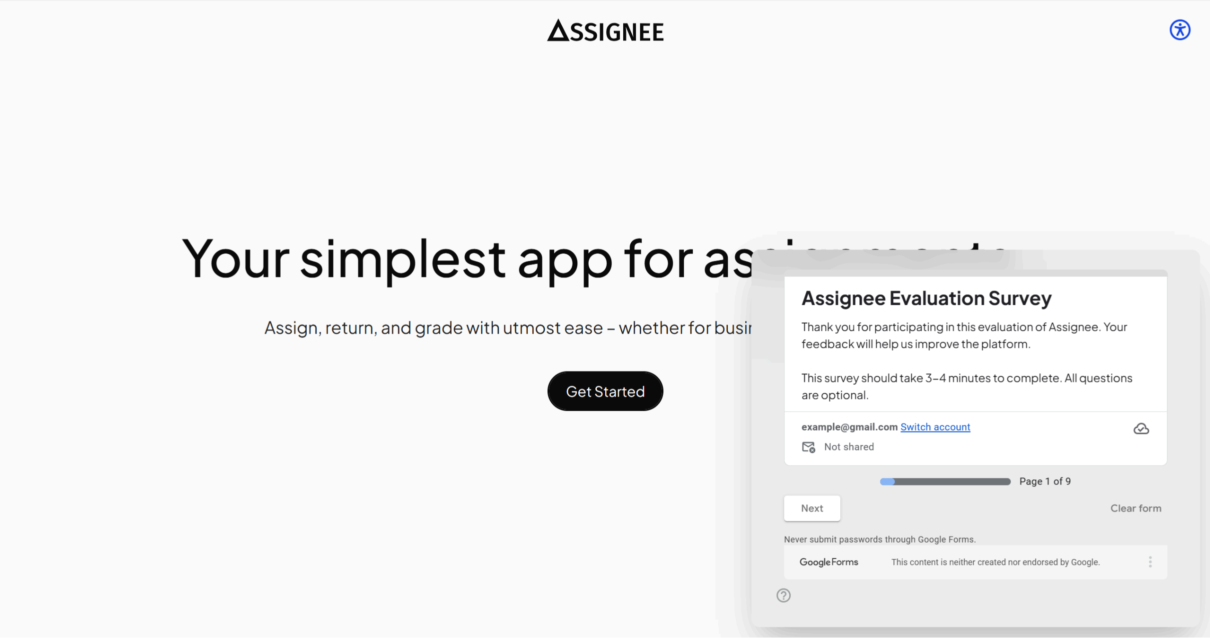
2.2. Distribution

The survey was strategically distributed to gather targeted, relevant feedback from actual or potential users of the Assignee application.

2.2.1. Embedded Form

To seamlessly integrate the feedback mechanism into user environments and maximize engagement, the Google Form survey was directly embedded within the Assignee website, which is adapted to the Assignee modal theme seamlessly.

The survey was embedded using an HTML `<iframe>` element, with Google taking care of its security concerns. Presenting the form as a native component of the website rather than an external link creates a frictionless transition for users from exploring the application to providing feedback.



Your simplest app for as

Assign, return, and grade with utmost ease – whether for busi

Assignee Evaluation Survey

Thank you for participating in this evaluation of Assignee. Your feedback will help us improve the platform.

This survey should take 3–4 minutes to complete. All questions are optional.

example@gmail.com Switch account

Not shared

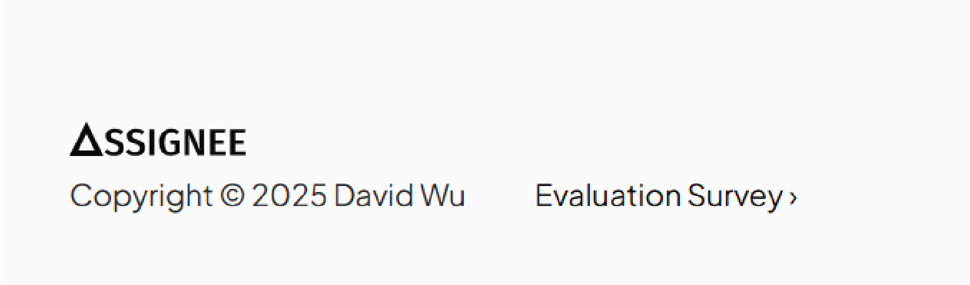
Never submit passwords through Google Forms.

This content is neither created nor endorsed by Google.

Figure 8: Embedded Survey

2.2.2. Embedded Link

The survey link was also embedded within Assignee. This ensured that participants could respond to the survey later, without having to fill in an in-page embedded form.



Copyright © 2025 David Wu

Evaluation Survey ›

Figure 9: Footer Survey Link

2.2.3. Other Methods

To broaden reach, the link was also shared with my family, friends, and mentors. They tend to explore Assignee more in depth and have more time to give an accurate evaluation.

Additionally, we could use registration emails to send survey invitations, but since this is not included in a Term of Agreement, it might not be the best option.

2.2.4. Population Sample

Distribution aimed for a convenience sample of the core user base. While not statistically random, this approach efficiently gathers insights from the most relevant audience for a formative evaluation.

The survey was active for a month to capture a sufficient volume of responses. Participation was entirely voluntary. No personally identifiable information was collected, ensuring respondent anonymity.

This multipoint distribution strategy was designed to facilitate easy access for users while directly targeting the population whose experience is most critical to evaluate.

2.3. Results Analysis

This section presents the analysis of the survey data. The analysis follows a two-stage approach: first, quantitative analysis of Likert scale items mapped to the 7 UX constructs, followed by thematic analysis of open-ended qualitative feedback.

The results of the Google Form survey is linked to an Excel file, allowing sophisticated graphing and analysis with Excel functions. There are 31 responses when the survey ended.

2.3.1. Preprocessing

Prior to analysis, potential outliers were identified by flagging responses falling $2\sigma_s$ (sample standard deviations) from the mean. In line with best practices for preserving data integrity and acknowledging critical feedback, no responses were removed.

However, to ensure our summary statistics were representative, we present both the mean and the more robust median for key constructs. The analysis of qualitative comments from these outliers was included to understand the root causes of their negative experience.

Furthermore, open-ended responses are slightly edited to ensure a consistent feedback format.

2.3.2. Quantitative Analysis

The 5-point Likert scale items were grouped and analyzed within the 7 UX constructs.

Note that unless otherwise specified, σ_s stands for sample standard deviation with Bessel's correction applied, instead of population standard deviation. It is used instead of s to be better distinguished from words.

2.3.2.1. Demographics

Demographic profiles allow us to interpret UX evaluation results with much greater nuance.

This shows a strong focus on the application's core educational use case:

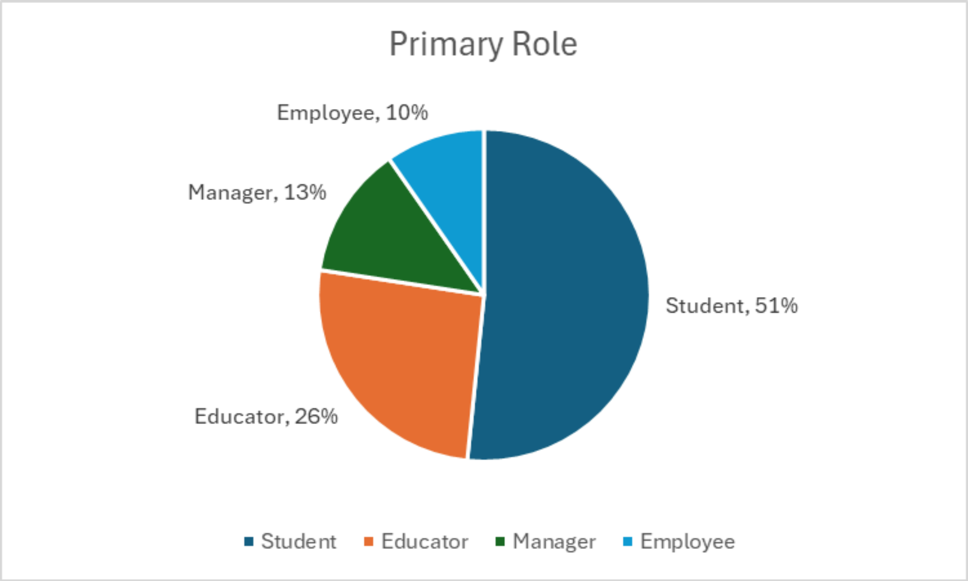


Figure 10: Primary Role

The sample is heavily weighted toward the primary academic roles of Student and Educator (77% combined). Feedback from these groups is most critical for evaluating Assignee's fit for its intended educational environment.

This reveals a user base with generally high familiarity with similar tools, setting a higher expectation bar:

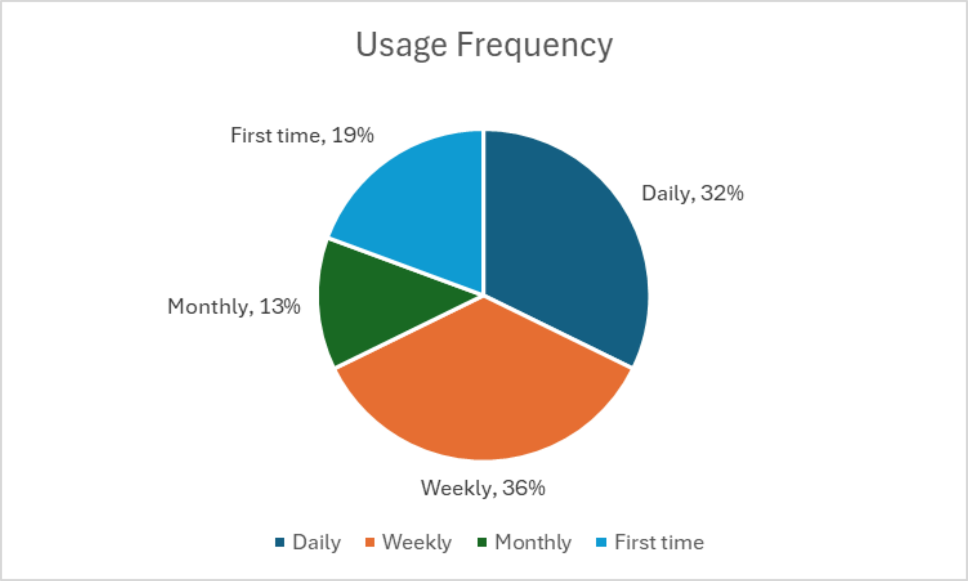


Figure 11: Usage Frequency

A significant majority (68%) are regular users (Daily/Weekly). Their feedback is informed by experience with other platforms. The 19% who are first time users are vital for identifying onboarding and initial usability hurdles.

This indicates the scale of collaboration Assignee needs to support effectively:

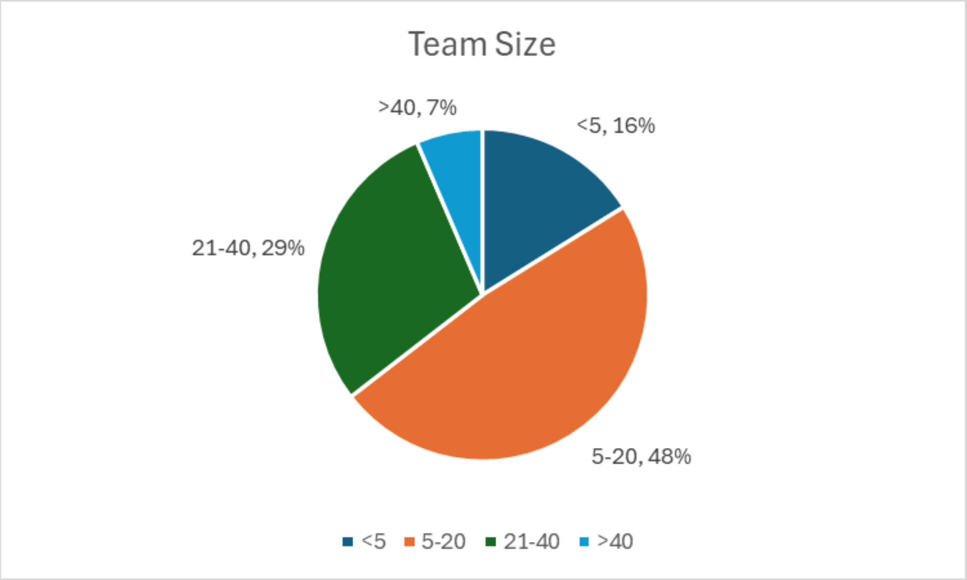


Figure 12: Team Size

Collaboration is primarily small to medium-sized groups (77% in teams of 5-40). This validates the core team model. A few >40 cases highlights a potential edge case for scalability and permissions management.

2.3.2.2. Descriptive Statistics

Mean (μ `=AVERAGE(:)` $\in [1,5]$), median ($\tilde{x}$ `=MEDIAN(:)` $\in [1,5]$), sample s.t.d. (σ_s `=STDEV.S(:)`), and skewness (γ_1 `=SKEW(:)`) were calculated for each construct to gauge central tendency and response dispersion. Mode is ignored because the sample size is not large enough for reference.

Excel dragging fill handle is utilized to compute statistics efficiently.

USEFUL	μ	$\tilde{x}$	σ_s	γ_1
The landing page clearly communicated what Assignee does.	4.03	4	0.84	−1.53
The team system (roles, workflow) supports my work needs.	4.45	5	0.85	−1.41
Assignee provides the key features I need to manage tasks and work.	4.23	4	0.88	−0.79

The high medians ($\tilde{x} = 4 \lor 5$) confirm strong perceived usefulness. The team system ($\mu = 4.45$) is the standout. The landing page ($\mu = 4.03$) has the lowest mean and most negative skew ($\gamma_1 = -1.53$) in this group, which means most ratings are high (clustered at $4 \lor 5$), but a small group of low scores are pulling the mean down and creating that strong negative skew. This aligns with its higher standard deviation ($\sigma_s = 0.84$), showing more disagreement. This could be due to the lack of descriptive figures and images on the landing page.

USABLE	μ	$\tilde{x}$	σ_s	γ_1
The sign-in/sign-up process was clear and smooth.	4.13	4	0.92	−1.36
Creating, joining, and managing teams was easy to understand.	4.52	5	0.77	−1.70

Creating, editing, and completing tasks was straightforward.	4.55	5	0.78	−1.86

Excellent core usability. Very high medians ($\tilde{x} = 5$) for team and task management, with highly negative skew ($\gamma_1 \approx -1.78$), indicating near-universal ease of use. The sign-in process has a slightly less negative skew and higher standard deviation ($\sigma_s = 0.92$), confirming it as the main, but still minor, source of inconsistent experience. This could be due to dynamic password requirements that could confuse low-tech users.

Accessible	μ	$\tilde{x}$	σ_s	γ_1
Text, icons, and spacing were clear and easy to read.	4.65	5	0.84	−2.49
The available accessibility options were sufficient to my needs.	4.48	5	0.92	−1.87
Usage was comfortable under different device or lighting conditions.	4.61	5	0.84	−2.34

Superlative scores. Perfect medians ($\tilde{x} = 5$), very high means ($\mu \approx 4.6$), and extremely negative skew (2 out of 3 with $\gamma_1 < -2.3$). This combination means an overwhelming majority rated these aspects as 5, with almost no spread ($\sigma_s \approx 0.87$). This is the model of a high-performing, universally agreed-upon feature set. The available accessibility options ($\mu = 4.48$) has the lowest mean and highest standard deviation, which could be related to open-ended feature requests.

Findable	μ	$\tilde{x}$	σ_s	γ_1
Navigation made it easy to locate functionalities.	4.55	5	0.87	−2.80
Search and filtering returned my expected results.	4.26	5	1.06	−1.61
Finding teams, members, and tasks was effortless.	4.29	5	1.10	−1.59

Marvelous medians ($\tilde{x} = 5$). Navigation is excellent with the highest mean and very negative skew. For search and finding items, the means are lower ($\mu \approx 4.28$) and skew is less negative, but crucially, the standard deviations are the largest in the set ($\sigma_s \approx 1.08$). However, this is likely related to the lower respond rates of these questions, since all ratings are optional and many people might have never tried using searching features.

Credible	μ	$\tilde{x}$	σ_s	γ_1
I felt confident that my data was secure.	4.71	5	0.82	−3.59
The site performed reliably with minimal errors.	4.71	5	0.82	−3.59
The overall experience felt trustworthy and professional.	4.61	5	0.88	−2.86

The construct demonstrates exceptional statistical strength, with perfect medians ($\tilde{x} = 5$), near-perfect means ($\mu \approx 4.68$), extremely low standard deviations ($\sigma_s \approx 0.84$), and the most negative skewness ($\gamma_1 \approx 3.35$) indicating high consensus. While survey responses on credibility can be subject to positive bias, this consistent data suggests a foundational sense of trust among users. This subjective perception is objectively corroborated by technical testing, which found the application to be reliable and free of critical bugs, confirming that the high credibility scores are well-founded.

VALUABLE	μ	$\tilde{x}$	σ_s	γ_1
Assignee helped me stay organized or complete work faster.	4.07	4	1.00	−1.07
Assignee feels worth my time and effort to use.	4.13	4	0.99	−1.15
I would recommend Assignee to others in my role.	3.90	4	0.98	−0.71

This is the most critical area for improvement. It has the lowest means ($\mu \approx 4.00$) and some of the least negative skews ($\gamma_1 > -1.15$) in the entire survey. The medians are all 4, none 5. The combination tells a clear fact: opinions are clustered lower and are more spread out ($\sigma_s > 0.98$), indicating weaker and less unanimous perception of value. The recommendation question is the absolute lowest mean, but could be related to the lower maturity and lack of sophisticated features of Assignee when compared to mainstream task-management tools, which are developed with much more time and effort.

DESIRABLE	μ	$\tilde{x}$	σ_s	γ_1
The interface design (layout, colors, font) felt visually appealing.	4.65	5	0.84	−2.49
The overall experience felt pleasant and engaging.	4.23	4	0.99	−1.59
Assignee is better compared to alternative task tools.	3.52	4	0.85	−0.57

Strong visual appeal with very high mean ($\mu = 4.65$), low standard deviation ($\sigma_s = 0.84$), and very negative skew ($\gamma_1 = -2.49$). The competitive comparison item ($\mu = 3.52$) is the lowest score in the survey, with the least negative skew ($\gamma_1 = -0.57$). This statistical profile is anticipated considering the relative naivety of Assignee in comparison to feature complete mainstream alternatives.

2.3.2.3. Correlation

The mean scores for each UX construct, segmented by user role, usage frequency, and team size, reveal significant variations in how Assignee is perceived by different groups. This analysis highlights where the application excels and pinpoints specific demographics that may require tailored improvements.

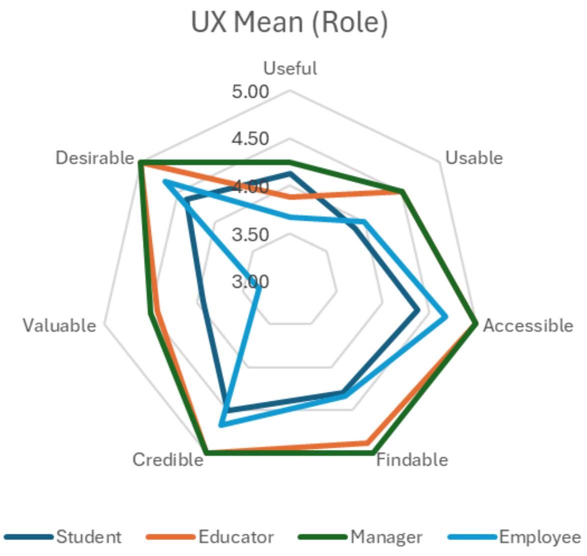


Figure 13: UX Mean by Primary Role

The experience varies considerably between roles, with Educators and Managers reporting consistently superior experiences across all 7 constructs compared to Students and Employees.

Educators and Managers (Power Users): These groups give the highest scores, particularly for Accessible, Credible, and Desirable (all at 5.00). This indicates the application successfully meets the needs of users who manage workflows and teams, fostering high trust, usability, and satisfaction.

Students (Primary User Base): As the largest respondent group, students show low scores in Usable (3.88) and Useful (4.13). This critical finding suggests the core task-management interface and perceived utility for students are not yet fully optimized and represent the most significant opportunity for improvement.

Employees (Small Sample): This group mirrors the student pattern of lower perceived scores: Valuable (3.33) and Useful (3.67), indicating a potential gap in demonstrating immediate, tangible value for general professional use outside a strict educational context.

However, the presence of outliers and the size of samples must also be considered when evaluating mean, which would become evident soon.

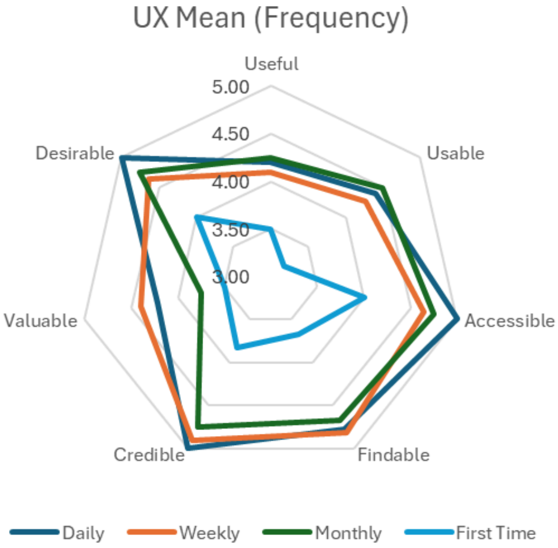


Figure 14: UX Mean by Usage Frequency

Scores show a strong positive correlation with user familiarity, revealing onboarding as a key hurdle that impacts initial perception.

Daily/Weekly Users (Established Users): Users with regular exposure report high, consistent scores. Daily users, in particular, rate Accessible, Credible, and Desirable at a perfect 5.00, suggesting that sustained use builds strong trust and appreciation for the design.

Monthly Users (Small Sample): This group mirrors the established user pattern of high perceived scores in similar areas, indicating the potential of Assignee in long term usage contexts.

First-Time Users (Small Sample): This group reports low scores across every single construct. The most pronounced gaps are in Usable (3.17), indicating that the initial interaction feels less intuitive. This validates qualitative feedback about onboarding confusion and underscores the need to smooth the first-time user journey.

The outliers from first-time users influence previous judgments on user role correlation, since certain groups are low in sample, thus the mean UX could be relatively low compared to the whole population. This must not be ignored, and we should view all analysis with synthesis in mind.

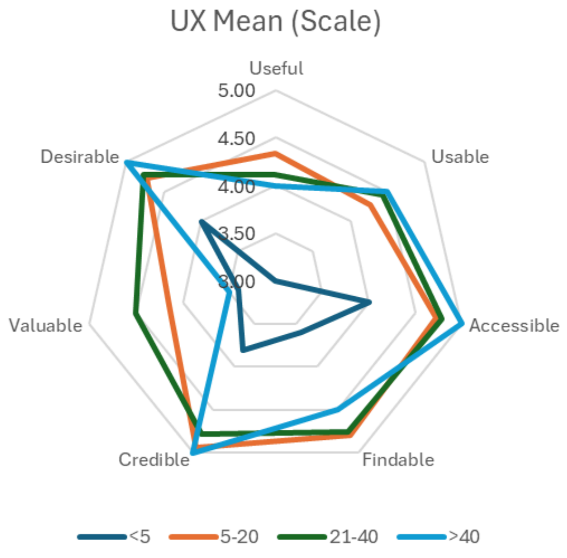


Figure 15: UX Mean by Team Size

The application demonstrates a clear sweet spot for collaborative scale, performing best with mid-sized teams.

Teams of 5-20 and 21-40 (Core Audience): These groups, representing the majority of respondents (e.g. classroom users), report high and very similar scores across all constructs, especially Credibility (≈ 4.86). This confirms Assignee's core architecture is well-suited for standard collaborative group work, especially in academic settings.

Very Small Teams (<5, Small Sample): Users in very small teams report markedly lower scores, with Useful (3.00) and Usable (3.00) being critical lows. This could suggest features may feel overly complex or unnecessary for minimal collaboration, but most likely to be the result of blatant outliers and a very small sample size.

Large Teams (>40, Small Sample): While based on limited data, scores remain high except for a notable dip in Valuable (3.50). This may indicate that for very large groups, advanced administrative or reporting features, which enhance value for managers, may be lacking at this stage.

2.3.2.4. Consistency

To validate the internal consistency and reliability of the Likert scale questions in the survey, Cronbach's alpha is computed with statistics.

The coefficient is computed as follows:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k} \sigma_{y_i}^2}{\sigma_X^2}\right) \in [0,1] \qquad (1)$$

where k represents the number of questions in the survey, i.e. 31.

It turns out that $\alpha = 0.97$ in our case. Since the coefficient is considered to be better when it is close to 1 in general (baseline > 0.70), it verifies the validity of the survey results.

2.3.3. Qualitative Analysis

The open-ended survey responses provide crucial context for the quantitative scores, revealing specific user pain points, desires, and ideas for improvement. The following thematic analysis synthesizes this feedback to guide strategic development.

2.3.3.1. Functionalities

Responses to ‘missing functionality’ and ‘add value’ converge into several clear themes, indicating where users perceive gaps in Assignee’s offering.

- Enhance Communication: The most frequent request cluster centers on improving team interaction.

Users explicitly asked for Announcements, Team Posts, and Group Chat. This indicates that while core task management works, the surrounding collaborative ecosystem feels underdeveloped. This is explained in Report Part I.

- Rich File Content Management: A strong need exists for better handling of attachments and content.

Requests for Multiple Attachments, In-app File Viewing and Editing suggest that the current file system is seen as a limitation to seamless workflow.

- Specialized Advanced Features: A segment of power users requested advanced tools that suit specific contexts.

Users asked for Student Grading Hub, LaTeX Equation Support, and features for managing ‘larger scale’ projects. These represent opportunities for high-impact, niche differentiation. This requires major changes in Assignee’s current philosophy of being general purpose.

2.3.3.2. Pain Points

Feedback on confusion was highly concentrated, revealing specific and critical friction points.

- The Authentication Bottleneck: An overwhelming majority of confusion-related comments centered on account creation.

Specific pain points included unclear ‘password requirements’ confusion over the ‘password creation’ process. This provides definitive evidence that the initial user sign-up is the dominant, however minor, usability barrier and directly impacts first-time user experience scores.

2.3.3.3. Improvements

General improvement ideas align with and expand upon the themes above, pointing towards maturation and polish.

In general, users requested ‘more mature features’ like ‘Notifications’, ‘Tags’, ‘Invitation Links’, and ‘More Role Hierarchy’. This suggests that while the core is functional, the product feels like an early version lacking the polish and depth expected of a primary tool.

Suggestions like ‘More Locales’ and ‘Dyslexia Font’ point to desired enhancements in user accessibility.

2.3.3.4. No-Responses

The lack of responses to 'What was hardest to find?' and 'Have you encountered any bugs?' is itself a finding. It indicates that discoverability issues are subtle, and that users did not encounter obvious, memorable bugs during their evaluation period.

On finding items, it is worth noting that many dashboards in Assignee are sorted in alphabetical order, if not chronological order. This ensures consistent UI states and allows quick lookup once the user got used to the layout.

2.3.3.5. Synthesis

The qualitative data tells a coherent story that directly explains the quantitative scores:

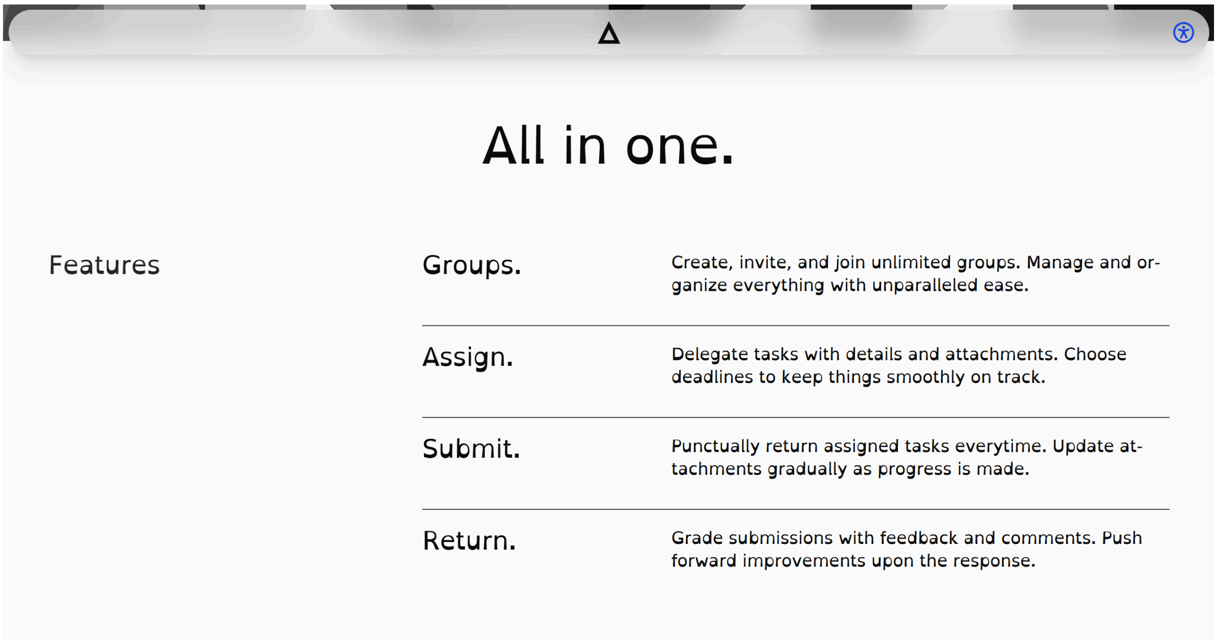
- The minor friction in onboarding (password confusion) is the likely primary driver of the lower scores from first-time users.
- The demand for communication features (chat, posts) and advanced tools is the key to improving perceived usefulness and valuableness.
- Requests for mature features and clarity align with the need to boost credibility and desirability against established alternatives.

2.4. Improvements

The following are derived directly from the synthesis of quantitative and qualitative feedback.

2.4.1. Open Dyslexic

To be more inclusive in typography, the Open Dyslexic font is added as a user-selectable option in accessibility settings.



All in one.

Features

Groups. Create, invite, and join unlimited groups. Manage and organize everything with unparalleled ease.

Assign. Delegate tasks with details and attachments. Choose deadlines to keep things smoothly on track.

Submit. Punctually return assigned tasks everytime. Update attachments gradually as progress is made.

Return. Grade submissions with feedback and comments. Push forward improvements upon the response.

Figure 16: Open Dyslexic Font Support

The unique design of characters increases legibility and allows users with dyslexia to read faster with ease.

2.4.2. Invitation Link

A prominent ‘Copy Invite Link’ button to the team management dashboard to ease group invitation.

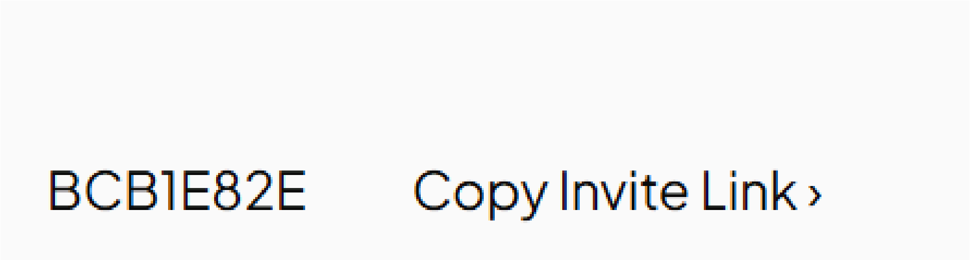


Figure 17: Open Dyslexic Font Support

The link would include a query string with the invitation code included, eliminating the need for manual input.

2.4.3. Planned

The following validated suggestions have been documented and prioritized for future development cycles. Urgency and effort matrices are considered in planning future work.

1. High Priority (Core Value)
   - Internationalization: Translate manually for more locales;
   - Team Communication Features: Announcements and Team Posts;
2. Medium Priority (Features)
   - Notification System: Adding user notifications for tasks and messages;
   - Advanced Role Management: Exploring more granular permissions;
3. Long-Term (Strategic)
   - Enhanced File Handling: Enabling multi-file uploads, in-app previews and edits;
   - Specialized Tools: Investigate paths for a Student Grading Hub or LaTeX support;

A user tutorial or alike is feasible but unnecessary at this stage for its simplicity of design and minimality of features and views. Nevertheless, it could be arranged for future improvements given long-term updates implement more features.

This analysis demonstrates that user feedback could provide a logical and actionable blueprint for evolving Assignee from a functional prototype into a robust, user-centric platform.

3. Technical Testing

This section details the empirical evaluation of Assignee's technical robustness across 7 critical domains. The testing employs both automated tools and manual procedures to validate functionality, usability, interface, database, performance, compatibility, and security.

3.1. Environment

The tests were performed on my IdeaPad 5i Pro 2021 (x86-64) Windows 10/11 Pro, with a complete development stack. Node.js 22 LTS is used as runtime, specific dependency versions are available in the repository `package.json` files.

3.2. Functional

To verify that all core features work as specified without error, a suite of automated and manual test cases was executed, covering positive, negative, and edge-case scenarios for different user contexts.

3.2.1. Framework

To ensure code reliability and prevent regressions, automated testing was implemented using the Jest framework.

Jest was selected as the primary testing framework for Assignee due to its built-in code coverage reporting, and excellent support for both unit and integration testing. The clear, descriptive assertion library aligns with the goal of writing tests that serve as documentation.

3.2.2. Coverage

A three-level testing strategy was employed for maximum system robustness.

3.2.2.1. Unit Tests

Isolate and test individual functions and components, such as utilities.

```
// Cryptography Utils
test("Hashes match", () => {
  const { hash, salt } = chash("test");
  expect(match("test", hash, salt)).toBe(true);
});
```

This verifies that the smallest, isolated parts of code (functions, methods) work correctly and meet their design, catches bugs early, and ensures reliability.

3.2.2.2. Integration Tests

Verify interactions between modules, such as API endpoints and database models.

```
// Retrieve Tasks API Route
test('GET /api/tasks', async () => {
  // Frontend AXIOS Call
  const res = await api.get('/api/tasks');
  // Backend Evokes DBMS
  expect(res.data).toBe(...);
});
```

This verifies that different modules and components work together correctly and seamlessly as a cohesive whole, the system functions reliably and meets overall requirements.

3.2.2.3. E2E Tests

Simulate complete user flows for critical paths e.g. registration and task creation.

Functional or End-to-End (E2E) tests simulate complete user scenarios. It focuses on user journeys rather than isolated functions, ensuring data flows correctly across layers and confirming a reliable user experience.

Browser simulation frameworks like Puppeteer and Selenium could serve the purpose well, but the implementation could require a lot of effort. In those cases, manual testing would be unavoidable.

3.3. Usability

To assess the efficiency, learnability, and satisfaction of the user interface in practice.

Supplement the UX survey with targeted task-based scenarios, conducted with a selected group of users. Evaluate the 5Es (Effective, Efficient, Engaging, Error Tolerant, Easy to Learn) factors via logging and analysis.

However, the evaluation survey alone usually suffice, and further investigations would cost too much effort. Only if absolutely necessary should this be performed, since analyzing logs could be tedious and difficult.

Google Lighthouse snapshots are used to ensure all accessibility measures are implemented.

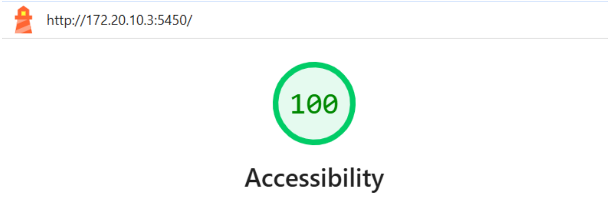


Figure 18: Lighthouse Accessibility Score

3.4. Interface

To verify the correctness, reliability, and robustness of the shared contracts and data exchanges between different layers (frontend and backend).

Unit testing of Zod schemas are performed to ensure they correctly accept valid data and reject invalid data with appropriate errors.

```
// Zod Interface Validation
test("Invalid payload", () => {
  const req = { ..., badreq: true };
  const res = SigninRequest.safeParse(req.body);

  expect(res.success).toBe(false);
  expect(res.error).toBe(...);
});
```

Testing that validation errors from Zod schemas are properly propagated and transformed into consistent, user-friendly API error responses.

The approach using Zod schemas established a strong, type-safe interface layer that effectively prevented invalid data from propagating into the core application logic. This resulted in robust API behavior and consistent error handling.

API endpoints must adhere to their defined request-response contracts, including status codes, data shapes, and error formats.

```
// Retrieve Tasks Payload
test("GET /api/tasks", () => {
  const res = await api.get('/api/tasks');
  const { success } = GetTasksResults.safeParse(res.data);

  expect(res.success).toBe(true);
});
```

These unit testing on interfaces ensures reliable communication within the system.

3.5. Database

To ensure data integrity, reliability, and performance of all database operations under load.

Testing included SQL query analysis, and constraint validation. All create, read, update, and delete operations must maintain referential integrity. Key queries (e.g. authentication rotation) must be analyzed to maximize concurrency, therefore performing better under spikes:

```
// Benchmark Spikes
for (let i = 0; i < 10_000; i++) {
  // Perform DBMS Operations
}
// Check For Corruption
```

All critical queries could execute within acceptable thresholds to provide a seamless and smooth experience to users. The database schema is sound and no integrity violations are found.

However, it is found that massive simultaneous updates could make the database struggle. SQLite is known for its lack of concurrency features, and most queues are implemented on the application side instead (e.g. Prisma), which leads to deteriorated performance when heavy operations like file upload is performed simultaneously.

The fix is simply to replace SQLite (only chosen for packaging) with an alternative mainstream DBMS, and to set file size limits on attachments e.g. 5MB.

3.6. Performance

To measure and validate the responsiveness, stability, and scalability of the application.

Frontend metrics were gathered via Google Lighthouse.

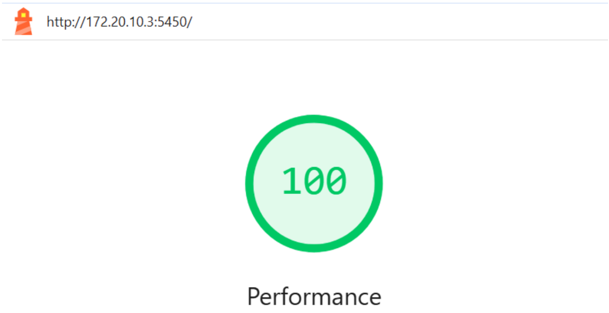
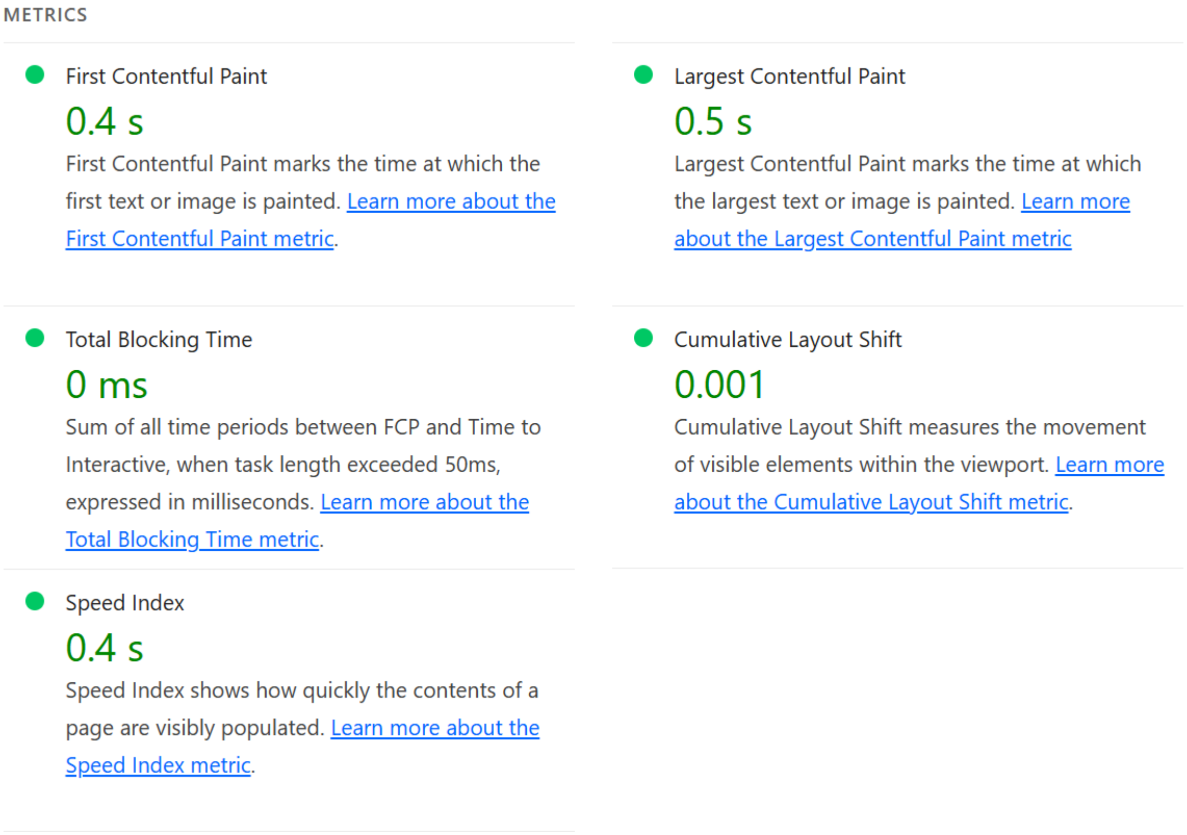


Figure 19: Lighthouse Performance Score



METRICS

● First Contentful Paint

0.4 s

First Contentful Paint marks the time at which the first text or image is painted. Learn more about the First Contentful Paint metric.

● Largest Contentful Paint

0.5 s

Largest Contentful Paint marks the time at which the largest text or image is painted. Learn more about the Largest Contentful Paint metric

● Total Blocking Time

0 ms

Sum of all time periods between FCP and Time to Interactive, when task length exceeded 50ms, expressed in milliseconds. Learn more about the Total Blocking Time metric.

● Cumulative Layout Shift

0.001

Cumulative Layout Shift measures the movement of visible elements within the viewport. Learn more about the Cumulative Layout Shift metric.

● Speed Index

0.4 s

Speed Index shows how quickly the contents of a page are visibly populated. Learn more about the Speed Index metric.

Figure 20: Lighthouse Performance Metrics

Largest Contentful Paint dominates one's first-impressions, and decides if visitors would like to stay for content.

Backend load testing was simulated using Axios, and benchmarked with BenchJS' asynchronous testing suite. The procedure is detailed in the last section. The tests help maintain the scalability of the application.

3.7. Compatibility

To ensure consistent functionality and appearance across various devices, operating systems, browsers, and viewports.

The application was tested on a matrix of client environment combinations in order to simulate varying real-life settings:

Device	Type	System	Browser	Resolution	Network
iPhone XR	Mobile	iOS 18+	Safari 16.4+ (2023-04-11)	414 × 896	WiFi/4G
iPad Air 4	Tablet	iPadOS 18/26+	Safari 16.4+ (2023-04-11)	1180 × 820	WiFi/4G
IdeaPad 5i Pro 2021 (x86-64)	Desktop	Windows 10/11+ Pro	Chromium 111+ (2023-05-09)	Responsive (simulated)	WiFi/4G (simulated throttling)
IdeaPad 5i Pro 2021 (x86-64)	Desktop	Linux Mint 21+ LTS (Debian APT)	Firefox 114+ (2023-06-06)	Responsive (simulated)	WiFi/4G (simulated throttling)
IdeaPad 5i Pro 2021 (x86-64)	Desktop	Linux Fedora Adams 42+ (RedHat DNF)	Firefox 114+ (2023-06-06)	Responsive (simulated)	WiFi/4G (simulated throttling)

- Browsers on iOS are all Safari Webview based;
- Both Chrome and Edge are all Chromium based;

The bowser versions supported all provide ‘Baseline: Widely Available’ features, as defined by the WebDX Community Group.

Layouts adjusted correctly. Touch targets were adequately sized. Assignee displays and functions correctly across all modern browser environments, ensuring broad accessibility.

3.8. Security

To identify vulnerabilities that could compromise data confidentiality, integrity, or availability.

Lighthouse is once again utilized to automatically scan for potential vulnerabilities in Headers (e.g. expiry, cache) and XHR etc.

A possible improvement is to adopt HTTPS (SSL encrypted) connection instead of HTTP. However, HTTPS requires signing and is not practical for a school project.

4. Credits

This is a continued list of credits in additional to ones mentioned in Report Part I. Important tools are listed again.

4.1. Common

- Git (https://git-scm.com)
- GitHub (https://github.com)
- VSCode (https://code.visualstudio.com)
- Catppuccin (https://github.com/catppuccin/vscode)

4.2. Survey Build

- Logic Form (https://logicform.io)
- Google Forms (https://docs.google.com/forms)
- Pretty Forms Designer (https://www.addxt.com/pretty-forms-designer)

4.3. Survey Analysis

- Google Sheets (https://docs.google.com/spreadsheets)
- Microsoft Excel (https://www.microsoft.com/en-us/microsoft-365/excel)
- Wikipedia (https://www.wikipedia.org)

4.4. Technical Testing

- Jest (https://jestjs.io)
- Node.js (https://nodejs.org)
- Zod (https://github.com/colinhacks/zod)

4.5. Reference

- Browser Stack (https://www.browserstack.com)
- WebDX Baseline (https://web-platform-dx.github.io/web-features)

4.6. Report

- Typst (https://typst.app)
- ILM (https://typst.app/universe/package/ilm)
- Tiny Mist (https://github.com/Myriad-Dreamin/tinymist)
- PDF Viewer (https://github.com/tomoki1207/vscode-pdfviewer)
- DeepSeek (https://chat.deepseek.com)
- LTeX+ (https://github.com/ltex-plus/vscode-ltex-plus)

While only a selection is mentioned here, I am deeply grateful for the entire ecosystem that made Assignee possible.

5. Final Remarks

This report reflects original research, analysis, and insights. While generative AI tools were selectively used to enhance the clarity of this report, all core ideas, findings, and conclusions remain the product of human effort.

In the AI-augmented era, I have taken care to uphold academic and professional integrity throughout this work, by acknowledging how to use AI tools responsibly.

As I wrap up this report, I extend my thanks to the many tools, contributors, and collaborators who made Assignee a reality. I want to affirm my commitment to respecting intellectual property and licensing rights. In an era of boundless digital collaboration, I believe progress thrives when credit is given fairly, and innovation is built responsibly.