# Usability Issues from Usability Testing

## Quiz App

### Category View

A1 Users did not immediately recognize what the view is about, requiring them to read the categories to understand it.

A2 Users were confused by the sorting of the categories, as they could not identify a clear sorting rule, such as alphabetical order.

A3 The light/dark mode toggle was misleading; users expected the icon to change from a sun to a moon when toggled.

A4 Difficulty was encountered in finding the "Geography" category due to the lengthy list without a search function.

### Setup View

A5 The design was perceived as chaotic.

A6 Uncertainty regarding "Any" as a difficulty level and its emoji.

A7 Delay after pressing the "Start" button led to multiple presses.

A8 The back button was misleadingly labeled as "Back" instead of something more descriptive like "Change Topic", causing navigation confusion for some users.

### Quiz View

A9 Some users were unsure which answer they had logged in and whether it was correct. This is due to the sudden change of the screen, accompanied by green, red, and yellow indicators, along with multiple symbols appearing.

A10 The cross and checkmark symbol caused confusion among users.

A11 Text changed from one line to two lines after displaying the cross / checkmark symbol.

A12 When "Any" difficulty was chosen, users wanted to know the difficulty level of questions shown to them.

A13 Current Points were not visible to the user, leading to uncertainty about their score.

A14 Users were unable to correct their answers when they accidentally pressed the wrong button.

**Score View**

A15 Users who pressed "Try again" to answer different questions on the same topic were surprised to receive the same questions again.

A16 Pressing "Back" led users to the setup screen instead of the categories screen, which was not as expected.

# To-Do App

**List View**

A17 When trying to tap the "Add To-Do Task" button was often missed due to its small size and placement at the top of the screen.

A18 Task color coding was unclear, particularly the use of blue instead of orange, to follow the expected traffic light colors of red, orange, green for high, medium, and low priorities.

A19 Task sorting by priority was not clear until tasks of all priority levels were displayed.

A20 Accidentally marking tasks as completed instead of opening the edit view.

A21 Some users cleared tasks individually by swiping, overlooking the option to use the trash bin for clearing all tasks at once.

A22 Some users were unable to find the swipe-to-delete feature for individual tasks, instead trying a long press or searching for an edit mode button.

A23 Uncertainty about the disappearance of completed tasks (e.g., if they disappear automatically after a certain time).

**Task Detail View**

A24 Tendency to create a single task for all activities (e.g., "Daily Tasks") due to lack of subtasks or examples, leading to overloaded task descriptions.

A25 It was not clear to all users about the optional nature of the task description field.

A26 Lack of clarity on why reminders were marked as invalid when accidentally set to past dates.

A27 Desire for a delete option within the task editing interface.

# Usability Issues from Expert Review

## Quiz App

**Category View**

B1 The app defaults to light mode, irrespective of the device's preferences.

B2 The yellow and white color scheme of the app make text hard to read.

B3 The spacing between the list and heading is too narrow to effectively create a hierarchy.

B4 The title "Trivia Game" does not clearly indicate the screens purpose, forcing users to explore the content for clarity.

B5 The order of the categories appears arbitrary.

B6 Additional explanations for some categories might be beneficial, such as clarifying that "General Knowledge" includes a mix of questions from all categories.

**Setup View**

B7 The previously selected category and its corresponding title on the setup screen sometimes do not match (e.g., "Video Games" vs. "Entertainment: Video Games").

B8 The back button's blue accent clashes with the app's yellow accent color.

B9 There is an overload of content on a single view.

B10 Inconsistent grid layouts cause visual confusion.

B11 The lack of visual guides makes the screen appear disorganized, with buttons missing defined tap areas.

B12 Texts are inconsistent in their use of capitalization.

B13 It is unclear whether the "Any" option randomizes difficulty per question or for the entire quiz.

B14 The "Multiple Choice" text wrapping is visually unappealing.

B15 The start button is slimmer and has a shadow, leading to a lack of consistency in button design.

B16 Saving the last setup configuration could simplify repeated use.

**Quiz View**

B17 There is no option to exit a game in progress.

B18 The progress label ("1/10") could be confusing as there is no textual description of what the numbers represent.

B19 The design of the answer buttons is not consistent with the rest of the button designs.

B20 Once an answer is selected, it is instantly submitted, with no option to revise it.

B21 The presentation logic of the symbols for correct and incorrect answer (checkmark and cross) can be confusing.

B22 When submitting an incorrect answer, the black text on the red background of the answer button creates a contrast that is difficult to read.

**Score View**

B23 The visual hierarchy is unclear, making it hard to distinguish the importance of different elements.

B24 The alignment of the "10/10" label to the progress indicator is not centered.

B25 The "10/10" label and progress indicator are not relevant on the result screen.

B26 Text elements are overly large and clustered at the top, lacking a balanced distribution.

B27 The positioning of buttons towards the top of the screen makes them less accessible.

B28 The design does not effectively utilize secondary buttons to indicate less important actions.

B29 The "Try Again" feature restarts with the game with the same questions, which could be clearer in its description.

# To-Do App

**List View**

B30 The app's interchangeable use of "Tasks", "Things", and "Items" creates terminology inconsistency.

B31 The wording of the headline "Things to Do" can confuse, especially in the empty state.

B32 During the empty state, showing a disabled trashcan is confusing.

B33 There is a lack of clear visual hierarchy due to the font sizes in the empty state.

B34 Inconsistent icon sizes detract from the app's visual coherence.

B35 Despite being a key function, the add task button is small and positioned in the toolbar, away from easy thumb access.

B36 The task's sorting is unclear and unexplained.

B37 The color coding for task priority does not follow the expected traffic light scheme.

B38 Completed tasks do not visually stand out enough from pending tasks.

B39 Once tasks are completed, their details become inaccessible.

B40 The search feature is hidden behind a swipe-down gesture.

B41 The search feature's empty state has no informative content.

B42 The absence of a confirmation step for swipe-to-delete actions could result in accidental task removal.

B43 An expected "select to delete" functionality is missing.

**Task Detail View**

B44 The view's title "Edit Task" when adding a task is misleading.

B45 The back button's label "Things to Do" is unnecessarily long and not straight-forward.

B46 The word "Title" is redundantly used both as the label and the placeholder for the title field.

B47 Unlike the other fields, the description text field contains no placeholder.

B48 The description field's dynamic resizing is unconventional and might not guide user input as effectively as a fixed size would.

B49 The ability to dismiss the keyboard by tapping on the background is missing.

B50 The arrow icons in the priority field do not clearly indicate a dropdown menu.

B51 The use of gray color for the text in the priority field misleadingly suggests that it is disabled or a placeholder.

B52 The word "Priority" is redundantly used both as the label and the placeholder for the priority field.

B53 The "Remove Reminder" button is too prominent and could be mistakenly perceived as a save button.

B54 The alert message for reminders set in the past does not clearly explain the problem.

B55 Alerts for invalid reminder dates should be given immediately when an invalid date is selected, not upon saving.

B56 The action for saving should be labeled as "Save" instead of "Done".

B57 It is unclear that the title field is required for task creation and that the "Done" toolbar button is grayed out until it is filled in.

B58 A delete option within the task detail view is missing.

# Usability Issues from UX-LLM

## Quiz App

**Category View**

C1 Insufficient contrast between text and background color: The yellow background with white text on the category buttons may not provide enough contrast for users with visual impairments or when viewing in bright light conditions.

C2 Inconsistent use of dark mode: The toggle for light and dark mode changes the overall UI theme but retains a yellow background for category buttons, which could lead to a disjointed user experience in dark mode.

C3 No visual feedback on button press: The category buttons do not appear to have any visual feedback when tapped, which could leave users uncertain whether their input has been registered.

C4 Lack of category description: The buttons only display the category name without any description, which might not be enough information for users to decide which category to choose, especially if the category names are not self-explanatory.

C5 Fixed button height: The category buttons have a fixed height, which might not accommodate category names of varying lengths or localization into other languages where the text might be longer.

C6 Potential for hidden navigation bar elements: The sun icon for toggling light/dark mode might be hard to find for users, as it blends with the navigation bar elements and does not communicate its function clearly without text or a more descriptive icon.

C7 No indication of loading state: When the app is fetching categories, there is no visual indication to the user that data is being loaded, which could lead to confusion if the process takes longer than expected.

**Setup View**

C8 Inconsistent option button layout: The "True/False" option button for the "Question Type" is not aligned with the other buttons, which may confuse users about whether it is selectable or part of another group.

C9 Lack of visual feedback on user interactions: There is no indication that a user's selection has been registered aside from the color change, which might not be enough for users with visual impairments or those not perceptive to subtle changes.

C10 Unclear error messaging position: The error message appears at the bottom of the view, which might not be noticed by users if they do not scroll down or if there is no indication that an error has occurred.

C11 Ambiguous loading state: The loading indicator is placed on top of the "Start!" button when fetching questions, which might make users think that the button is disabled or that the app has frozen, rather than understanding that a process is ongoing.

C12 The "Start!" button's text color is similar to the background color, which may reduce contrast and readability, potentially making it difficult for users to identify the call-to-action button.

C13 The app view lacks a clear indication of the current step or progress within the quiz configuration process, which might leave users uncertain about how far along they are in the setup.

C14 No clear visual hierarchy between different sections (Difficulty, Question Type, Number of Questions), which can make the view appear as one large block of content, potentially overwhelming users.

**Quiz View**

C15 Inconsistent feedback on selection: The visual feedback for correct and incorrect answers is inconsistent. The correct answer is only indicated by the color green with a checkmark, while the incorrect answer is indicated by the color red with an "X" and also shifts to the left, which may be confusing for the user.

C16 Lack of clear navigation cues: There is no clear indication of how to proceed to the next question after a selection is made, which could lead to user confusion. The "Next" button appears only after a selection, and it may not be immediately obvious to the user.

C17 Inadequate touch targets: The "Next" button may have an inadequate touch target size, which could make it difficult for users to tap accurately, especially on devices with smaller screens.

C18 Limited question visibility: The question text does not have a maximum line limit, which may cause issues if the question is too long, leading to the text being cut off or the layout looking cluttered.

C19 Ambiguous progress bar: The progress bar does not have a label or any indication of what it represents, which could lead to confusion about the user's progress within the quiz.

C20 Color contrast issues: The colors used for the "Next" button (yellow text on a yellow background) have low contrast, making it difficult to read, especially for users with visual impairments.

C21 No indication of the selected answer: There is no persistent indication of which answer was selected by the user after the answer is revealed, which can be problematic if the user wants to review their choice before moving on.

C22 Lack of confirmation before moving to the next question: There is no mechanism to confirm the user's choice before automatically advancing to the next question, which may lead to accidental selections.

**Score View**

C23 The color contrast between the yellow buttons and the white text may not be sufficient for all users, particularly those with visual impairments, which can lead to readability issues.

C24 Lack of clear visual hierarchy could make it difficult for users to quickly understand the results; the results text and the buttons seem to have similar visual prominence.

C25 The buttons' labels "Try again!" and "Back" may not be descriptive enough to indicate their function clearly, which can lead to confusion about what action will be taken.

C26 There appears to be no indication of a correct answer review option, which might be beneficial for users who want to learn from their mistakes.

C27 The feedback provided to the user does not appear to be personalized beyond the number of correct answers; more detailed feedback could enhance the learning experience.

C28 There is no visual feedback on user interactions; for example, the buttons do not change appearance when tapped, which could leave users uncertain if their input has been registered.

C29 The spacing between the "Try again!" and "Back" buttons may not be enough to prevent accidental presses, which could lead to frustration.

C30 The fixed width and height of the buttons may not accommodate longer text for localization, potentially resulting in text clipping or layout issues in different languages.

## To-Do App

**List View**

C31 Inconsistent visual feedback on task completion: The checkboxes provide a visual cue when a task is completed, but the accompanying text does not change its appearance, which could be improved by altering the text style to indicate completion more clearly.

C32 Lack of clear separation between tasks: The tasks are listed without clear dividers or spacing, which can make it difficult to distinguish individual tasks in longer lists.

C33 Ambiguous delete functionality: The delete icon's function in the navigation bar is not clear, and could be improved with a text label or a more descriptive icon.

C34 Missing visual hierarchy in list items: The task title and due date are styled similarly, making it hard to identify the most important information quickly.

C35 No indication of swipe actions: If swipe actions are available, they should be indicated visually to educate users.

C36 Redundant navigation links for completed tasks: Navigation elements for completed tasks could mislead users, and should be removed or visually distinguished.

C37 Unclear priority indication: The meaning of colored checkboxes as priority indicators is not obvious without a legend or explanation.

C38 Overloaded navigation bar: Having add and delete actions in the same place can be overwhelming; separating these actions can reduce cognitive load.

C39 Inconsistent interaction patterns: The use of toggles for marking tasks as completed may not align with other interaction patterns in the app.

C40 Lack of feedback when adding or removing tasks: No visual or auditory feedback is provided, which could leave users unsure of the action's success.

**Item Detail View**

C41 Incomplete visual feedback for priority selection: No clear indicator like a checkmark or highlight to show the selected priority level when the picker is collapsed.

C42 Inconsistent navigation title: The title says "Edit task" which might be misleading for new tasks; it should change based on the action context.

C43 No visual separation between date picker and remove button: These elements are placed next to each other without clear visual separation, possibly causing confusion.

C44 No indication of mandatory fields: Fields like the title, which are required, are not marked, potentially leading to incomplete task saving attempts.

C45 Lack of confirmation or undo functionality for removing reminders: This can lead to accidental data loss.

C46 No clear error messaging for invalid due dates: The alert shown does not specify why the date is invalid or guide the user to correct it.

C47 Inconsistent button states: The "Done" button is disabled if the title is empty, but no similar constraints exist for other critical fields.

C48 No feedback on successful task creation or update: There is no visual confirmation that the task has been successfully saved.

C49 Limited accessibility support: UI elements lack descriptive labels for accessibility tools, hindering usability for visually impaired users.

| ID | E1 | E2 | ID | E1 | E2 | ID | E1 | E2 | ID | E1 | E2 |
|----|----|----|----|----|----|----|----|----|----|----|----|
| C1 | A | A | C14 | A | A | C27 | B | A | C40 | B | B |
| C2 | C | B | C15 | A | A | C28 | D | D | C41 | B | A |
| C3 | D | D | C16 | A | A | C29 | B | B | C42 | A | A |
| C4 | A | A | C17 | C | B | C30 | B | B | C43 | A | A |
| C5 | B | B | C18 | A | B | C31 | D | D | C44 | A | A |
| C6 | A | B | C19 | B | A | C32 | B | A | C45 | A | A |
| C7 | A | A | C20 | A | A | C33 | A | A | C46 | A | A |
| C8 | A | A | C21 | A | C | C34 | D | D | C47 | A | A |
| C9 | B | A | C22 | A | A | C35 | A | A | C48 | B | C |
| C10 | C | A | C23 | A | A | C36 | A | A | C49 | C | A |
| C11 | A | B | C24 | A | A | C37 | A | A | | | |
| C12 | A | A | C25 | C | A | C38 | B | B | | | |
| C13 | A | B | C26 | B | A | C39 | B | B | | | |

**Legend:** ID = Usability Issue ID, E1 = UX Expert 1, E2 = UX Expert 2,
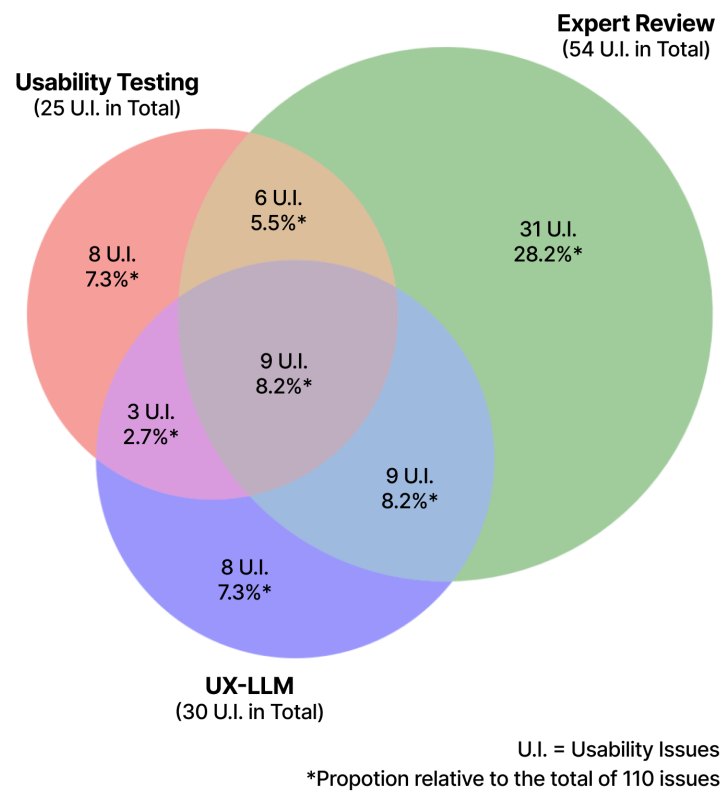A = Usability Issue, B = No Usability Issue, C = Uncertain, D = Incorrect/Irrelevant Statement.

Usability issues of UX-LLM annotated with assessments from UX experts.
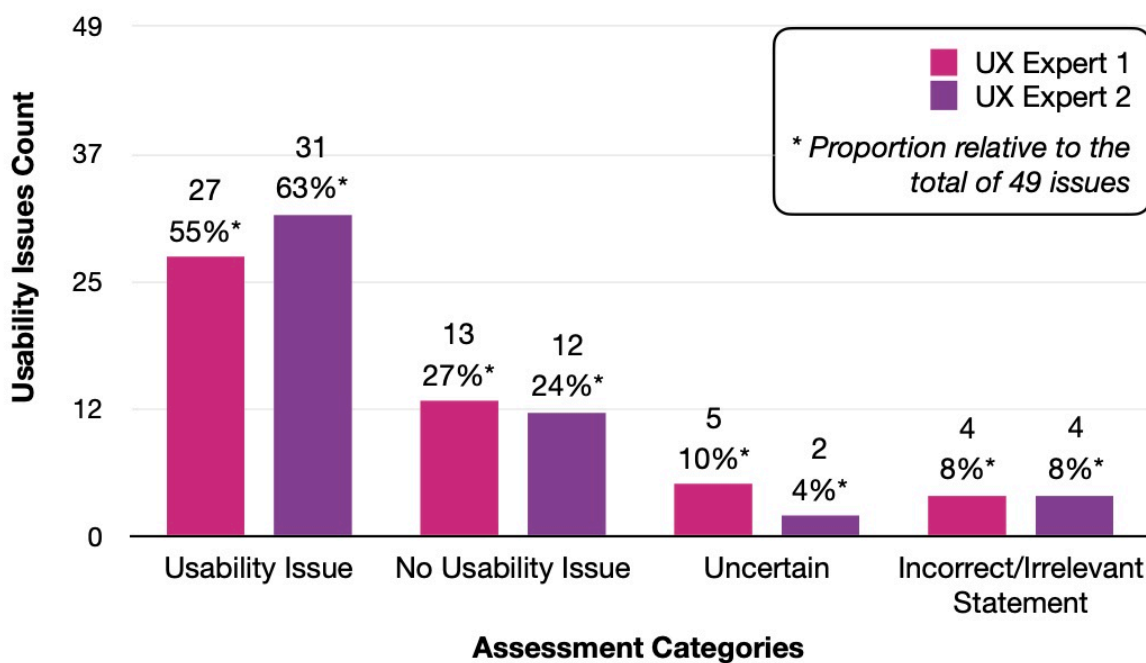
Visualization of UX experts' assessment regarding usability issues identified by UX-LLM.

| View | Usability Testing | Expert Review | UX-LLM |
|---|---|---|---|
| Category View (Quiz App) | A1 | B4 | / |
| | A2 | B5 | / |
| | A3 | / | C6 |
| | / | B2 | C1, C12, C20, C23 |
| | / | B6 | C4 |
| Setup View (Quiz App) | A5 | B9, B10, B11, B14 | C8, C14 |
| | A6 | B13 | / |
| | A7 | / | C11 |
| Quiz View (Quiz App) | A9, A10 | B21 | C15, C21 |
| | A14 | B20 | C22 |
| | / | B18 | C19 |
| Score View (Quiz App) | A15, A16 | B29 | C25 |
| | / | B23, B26 | C24 |
| List View (To-Do App) | A17 | B35 | / |
| | A18 | B37 | C37 |
| | A19 | B36 | / |
| | A21 | / | C33 |
| | A22 | B43 | C35 |
| | / | B38 | C32 |
| | / | B39 | C36 |
| Task Detail View (To-Do App) | A25 | B57 | C44, C47 |
| | A26 | B54 | C46 |
| | A27 | B58 | / |
| | / | B44 | C42 |
| | / | B51 | C41 |
| | / | B53 | C43 |

Overlap of usability issues from reference apps identified through: usability testing, expert review, and UX-LLM.

**Expert Review**
(54 U.I. in Total)

**Usability Testing**
(25 U.I. in Total)

6 U.I.
5.5%*

8 U.I.
7.3%*

31 U.I.
28.2%*

9 U.I.
8.2%*

3 U.I.
2.7%*

9 U.I.
8.2%*

8 U.I.
7.3%*

**UX-LLM**
(30 U.I. in Total)

U.I. = Usability Issues
*Propotion relative to the total of 110 issues

Venn Diagram showing the overlap of usability issues identified by usability testing, expert review, and UX-LLM.

$$Precision_{E1} = \frac{TP_{E1}}{TP_{E1} + FP_{E1}} = \frac{27}{27 + 17} \approx 0.61$$

$$Recall_{E1} = \frac{TP_{E1}}{TP_{E1} + FN} = \frac{27}{27 + 51} \approx 0.35$$

$$Precision_{E2} = \frac{TP_{E2}}{TP_{E2} + FP_{E2}} = \frac{31}{31 + 16} \approx 0.66$$

$$Recall_{E2} = \frac{TP_{E2}}{TP_{E2} + FN} = \frac{31}{31 + 51} \approx 0.38$$