Case Study: Usability Testing of a Medical Lab Device

Jeremiah Parry-Hill hello@jparryhill.com

Opportunity Brief

The VITROS 5600 is a machine manufactured by Ortho Clinical Diagnostics. It is framed as the workhorse of medical fluid sample testing. Because its design undergoes heavy regulatory oversight on a 3-year release cycle, it was important to the client to justify improvements through an intentional process. The client's stated goal was to have a set of observations and recommendations prepared in time for the device's next redesign approval.

Figure 1: The main interface screen of the VITROS 5000.

Users

The client, a product manager at OCD, provided a background for the user profile as follows. People who work in laboratory testing have a 4-year medical technician degree. The workforce tends to skew older, as wage compression and other factors have discouraged younger technicians from entering the profession.

The nature of the work involves multiple technicians operating the machine at various overlapping times. Rather than a “lone user, single process flow,” it’s not uncommon for urgent tests to create an interrupt-driven workflow with frequent state changes. The work environment is crowded, enclosed, and busy.

For the purposes of this test plan, we recruited two sets of users: a group of students in a local Laboratory Science Technology program, and a convenience sample of employees from the client’s manufacturing facility.

Project team and my role

Our usability team consisted of four consultants. Our skills included interviewing, data analysis, testing protocols, and facilitation. My role was to be the primary client-facing contact, to lead user recruitment, and to facilitate user tests. I contributed to writing the testing protocols, analyzing quantitative and qualitative data, writing the final report, and presenting the findings and recommendations to the client.

Process

At a high level, our process involved:

- A heuristic evaluation of the current state of the system. The client provided a virtual machine version of the software used for training.

- An initial client meeting to capture goals and background information and set expectations around project scope. It was helpful to have already completed the heuristic evaluation to provide context to this conversation.
- A site visit to observe the device in use with an expert user. The lead designer of the system showed us how he expects the system to be used. This was a useful dual-purpose interview that gave us context for the system's intended implementation model.
- Our team then identified real user demographics to inform recruitment of a relevant user sample. This allowed us to target a baseline minimum of clinical lab experience.
- We identified use case scenarios that would reflect the type of tasks and rhythm of interruptions that accurately represent real-world scenarios.
- We recruited users through a screener survey.
- We ran in-person sessions with selected users. Half the sessions took place in our usability testing lab, with the remainder onsite at the client's offices.
- Users were instructed to carry out certain tasks, and encouraged to self-narrate their thought process.
- Users consented to being recorded. We used Morae to capture footage of the users' remarks and nonverbal reactions as an overlay on the screen recording.
- We analyzed the results and presented insights and recommendations to the client.

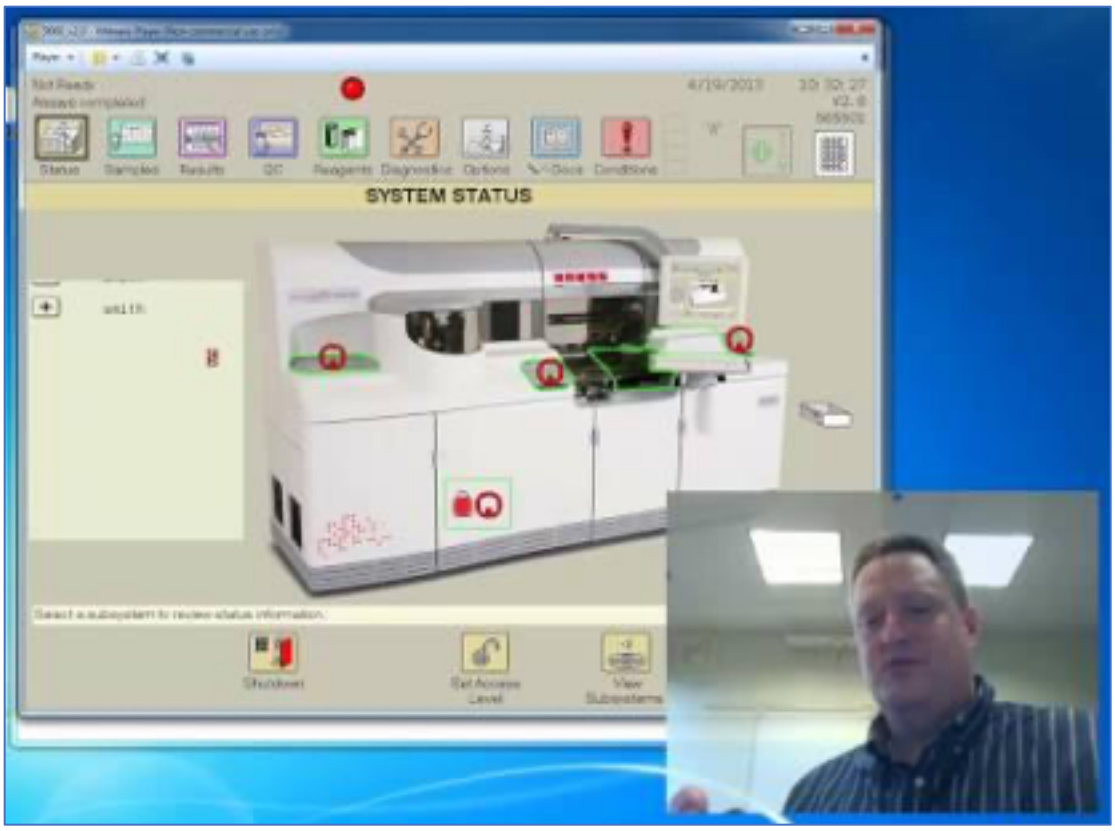

Figure 2: A user reacts to the main interface screen. "Where would you begin?"

Decision points and decision-making process

Our team was able to use discussion, grounded in the goals of the project, to navigate decisions. Key moments included:

- How many users to test with? The classic Jakob Nielsen answer of "5 users is usually enough" guided our decision here, in conjunction with the project budget. We could offer a modest compensation to 5 clinical tech students, and we obtained 4 more bonus users from the client's staff at no additional cost because they participated on their own working time.
- What to test? This was the area that offered the most creativity to our team as we discussed the sequence of activities and the optimal balance of things we could ask

users to do that would show us relevant behaviors before the user became fatigued or time ran out. In the end, we asked users to perform three tasks: two types of lookup (find a patient record, and find all records that match criteria) and a simple select-test-type-and-begin-testing task.

- What type of data stories matter? Half of our team, myself included, believe deep insights come from the qualitative experiential aspects of user interactions. Still, we acknowledged that quantitative analysis (time on task, error rates, etc.) can provide additional depth to the insights and satisfy the client who is comforted by hard numbers to justify decisions. We empowered those team members who believed most strongly in the case for quantitative analysis to take ownership of those parts of the study design and data stories. In the end, this resulted in a richer mixed-methods package of insights for the client.

Insights

The principal finding was that users expected to find a generic Search bar for any lookup task. The existing navigation model and terminology used in the interface only generated confusion. Although a Help function was provided, it was rarely used; “I don’t usually go for the Help button because it’s usually not helpful.”

Our recommendations to the client, based on patterns observed in testing, included:

1. Add a search function to enable text-based lookup of records.

2. We observed several cases of repetitive labor when settings were not retained between sessions. We recommend retaining low-risk settings between instances of tasks and either clearing high-risk settings or prompting users to verify their reuse.

3. Allow users to drill into expanded views of records by clicking anywhere on the line item (in the current interface, the expand button is an overlooked plus-sign).
4. Alphabetize, or otherwise organize, the list of test types and allow expert users to customize the order based on frequency of use.

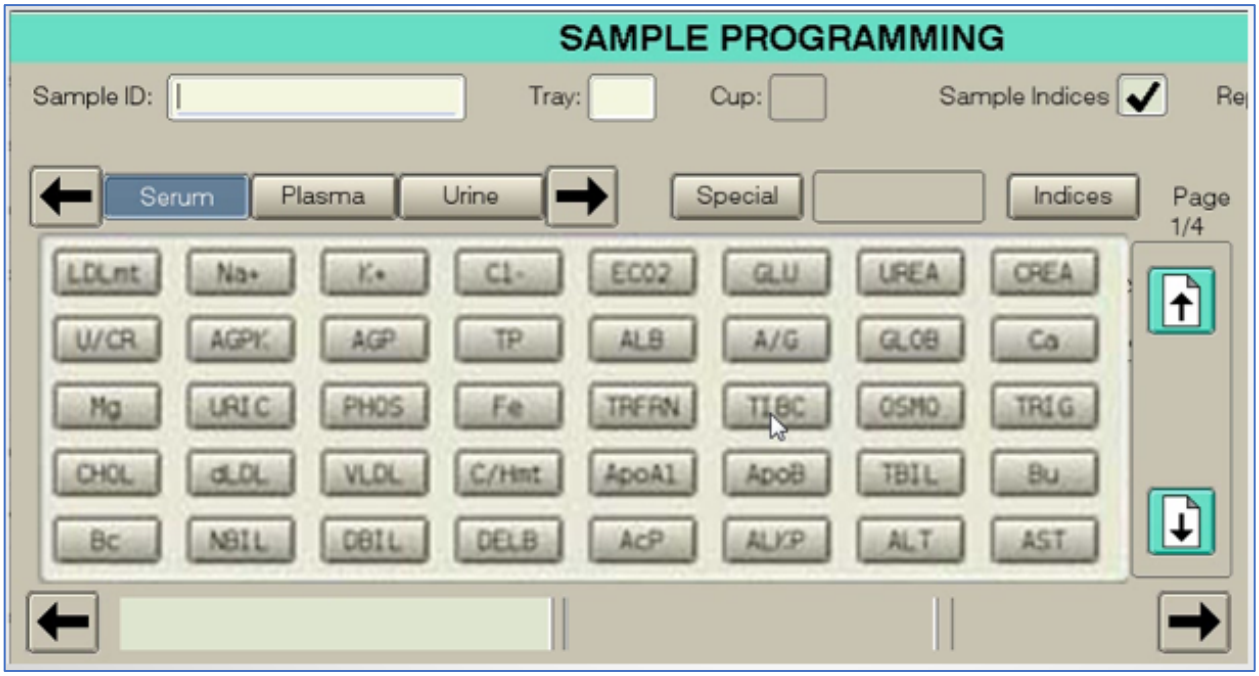

Figure 3: The list of available fluid sample test types is presented in no apparent order.

Constraints and limitations of the project

This project was an intentionally brief engagement. The client, a product manager, intended to use the success of this effort to build the case for staffing up an in-house UX research team. Given that scope, our consulting team focused on generating quick high-impact insights to demonstrate the value of research activities.

With our sample of novice users, the insights we surfaced tended toward superficial first impressions of the system and its learnability. Such results were well-aligned with the client's immediate goal to make the product show well in sales and

training contexts, but if any further consideration of the system's design were to take place, we would have to recruit expert users and consider deeper contextual inquiry.

Future work

There's an immediate opportunity to step back and reconsider the information architecture with an eye toward redesign. We identified several areas in the system interfaces where the labels and models were inconsistent or unclear.

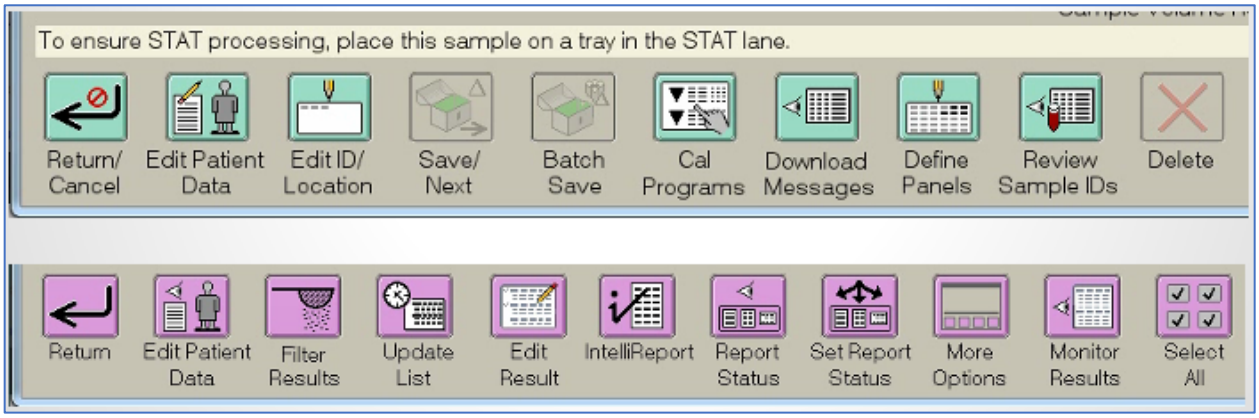

Figure 4: Inconsistent navigational models used in different parts of the system.

Based on our heuristic analysis, we came away with the impression that the current interface is arranged from a "machine-centered" perspective. The machine challenges the user to understand how it perceives a patient records and lab samples. A redesign effort should start by placing the needs of the lab technician, with all their interruptions and task flows, at the center of the system's design paradigm.