1 Case Study: Evaluating a New Tool

This section reports on a case study where the heuristics were used by an organization to evaluate a test tool offering from a commercial vendor. This was the first case study to be completed and used v1.0 of the heuristics held on a GitHub repository. First we discuss the research environment, including diagnosis of what actions were required, and how research ethics were addressed with the participating organization. We describe the actions planned, scoping and actions taken. We describe the data collection, and the evaluation of the heuristics by the participants. Finally, we reflect on what was learned in the case study: perceived benefits of the heuristics, areas for improving the heuristics, and how to improve running of the case study. The improvements to the heuristics and the process are then taken into the next case study.

1.1 Research Environment

1.1.1 Codes

- Organization code [O1}
- Vendor code [V2}
- Tools being compared [T1} and [T2}
- Key contact: [P1]
- Diary and reflections author: [P1]

1.1.2 Organization

The author was invited to submit the heuristics for use by the organization [O1}. Contact [P1] asked the author for permission to use the heuristics, after taking part in an expert review of them and agreed to run as a case study. The case study took place in February 2024, with feedback to the author in April 2024.

[O1} is a multi-national company providing on-line web-based services in the entertainment and leisure industry. It has a multi-national customer-base. English is the business language of the organization. The company has a large engineering department, actively seek to improve the quality of their offering, and the efficiency of their engineering, including testing practices. They use a range of tools and automation to support testing.

Before the case study, [O1} was using a free open-source tool [T1} (OS) to support a specific aspect of its testing, accessibility of websites. The vendor, [V2}, supply another testing tool [T2} (COM) to [O1}, and proposed including the functionality of tool [T1} (OS) in its own offering. They asked [O1} to evaluate their updated tool with its additional functionality.

1.2 Diagnosis

1.2.1 Intervention, background, and purposes

The author met one-to-one with [P1] online twice to discuss and plan the case study. This included discussing the purpose of the case study and introducing the heuristics.

The vendor [V2} had requested that the organization [O1} evaluated their new tool function [T2} (COM) with a view to trialing and purchasing it to replace the existing free tool plug in [T1} (OS). The evaluation was to help organization [O1} understand whether to consider replacing their existing

tool/method for accessibility testing with the new vendor accessibility module. The evaluation would include a comparison for gaps, and differences (positive and negative) between tools.

The purpose of taking part in the case study for [P1] was to use the heuristics to aid the evaluation of tool [T2} (COM) against tool [T1} (OS), and decide whether changing tool would be beneficial.

The purpose of the case study for the author was to assess whether the heuristics helped in the evaluation process.

The author agreed with [P1] that the heuristics were to be applied by the participant to aid planning and execution of the evaluation. Timescale for the evaluation was February 2024, to be carried out in a series of meetings with the vendor and with in-house teams at [O1}, consisting in the latter case of people who potentially would use the new tool.

1.2.2 Current situation

The Design Engineering Team at [O1} build and test library components for the product engineering teams to use when building products. The designs are supplied by the UX Department. The Design Engineering team use a vendor supplied tool [T2} (COM), to manage and execute their tests, and in addition use a freeware accessibility checking plug-in. The plug-in with added code shows accessibility issues and is described by participant [P1] as a ``*nice plug in, it even has links to show you how to fix issues* '' This process, shown in Figure 1, has been used over the last three years.
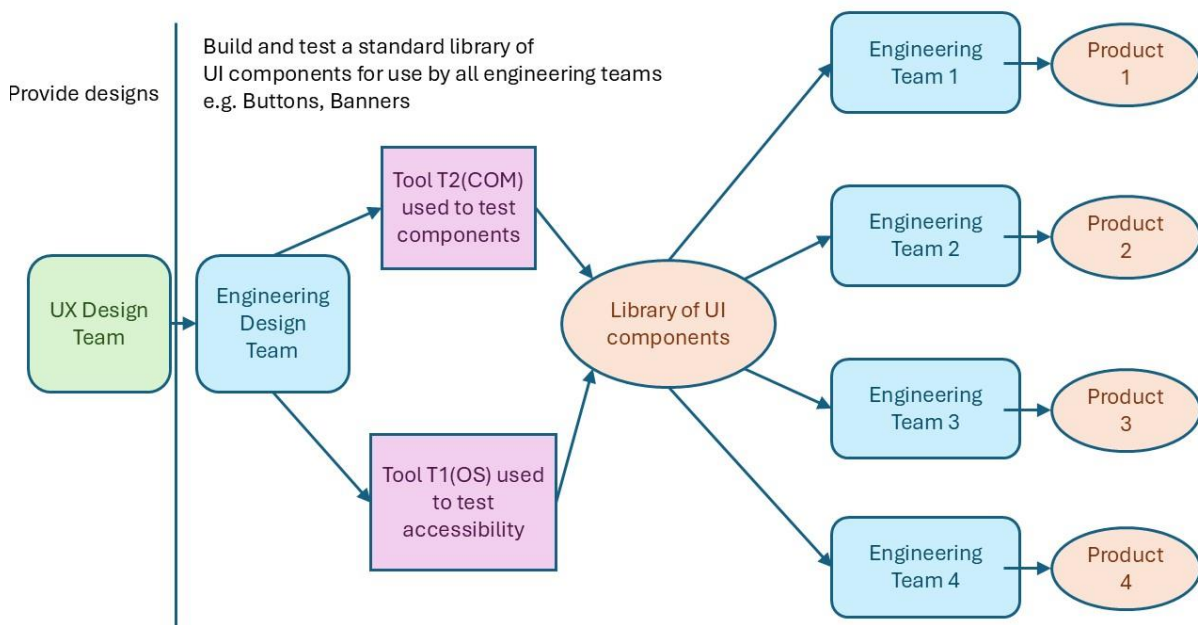

Figure 1 Design Engineering Team – current process and tools relevant to case study

1.2.3 New tool for evaluation

The vendor has added code to tool [T2} (COM), so that the plug-in is now integrated as part of its own dashboard, and would become a paid-for add on to the tool, entirely automated, and providing a report . Note that the plugin also produces a report. They are ``*pushing for [O1] to buy the new dashboard*'' [P1].
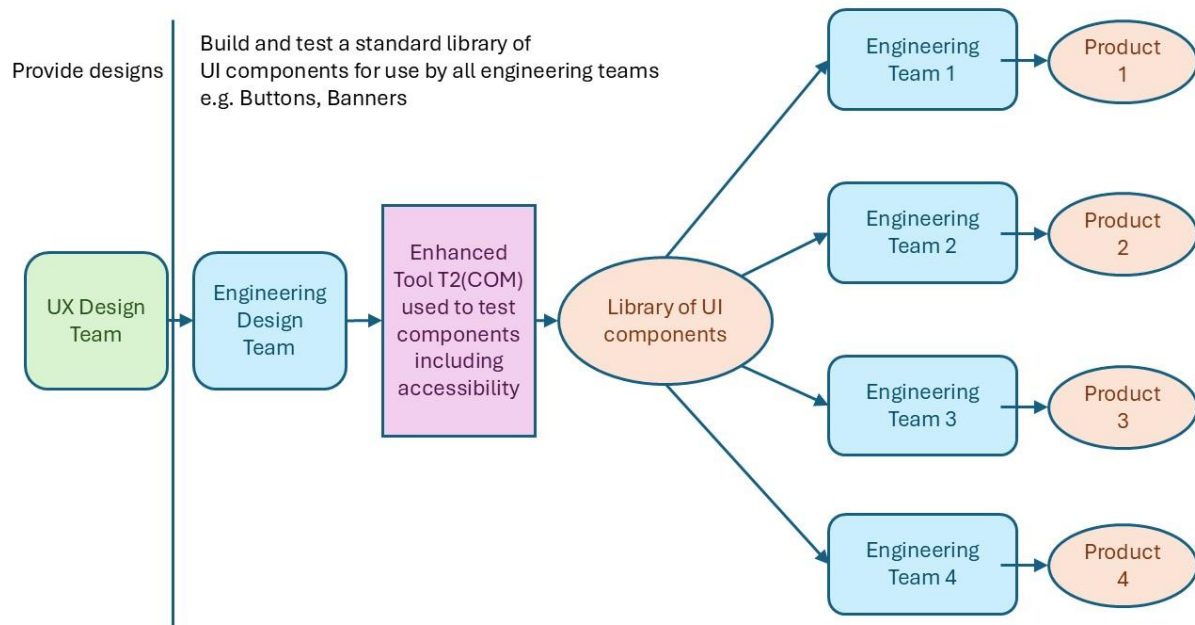
Figure 2 Vendor proposed process and tools relevant to case study

1.3 Action Planning

The planned sequence of events for the case study were:

1. Planning meetings: Researcher with [P1] to introduce the heuristics v1.0;
2. Use of Heuristics: [P1] ran the tool evaluation, using the heuristics and keeping a diary of events and usage of the heuristics, plus a score sheet;
3. Feedback and reflection meeting: Researcher with [P1] to review scores and diary entries, and reflect on the effect of the heuristics on the tool evaluation;
4. Researcher writes up report (this document)
5. Member checking: this report reviewed for fact checking by [P1]
6. Report refined, case study process refined, heuristics refined.

All meetings were online, none were video/audio recorded, notes taken during meeting. The planning meetings took place on 17 January 2024 and on 24 January 2024.

1.4 Action Taking

1.4.1 Scoping

The scope of the actions was use of the heuristics during the evaluation of the two tools. It was agreed that feedback would not be given to the researcher about specifics of the tools themselves, nor of the participant organization's software under test.

The list of heuristics for this case study are:

- H01. Why do we need this tool?
- H02. Who will use or be affected by the tool?
- H03. What previous experiences do people bring to the tool?
- H04. What communication needs or preferences do those people have?
- H05. What learning perspectives and goals do those people bring?

- H06. What learning preferences do those people have?
- H07. Where will the tool be used?
- H08. What workflows will the tool be part of?
- H09. What risks are associated with those workflows?
- H10. What work styles are acceptable in those workflows and teams?
- H11. When will the tool be used?
- H12. How long will the tool be used?

Each heuristic in the repository was linked to further information with questions, prompts, and activities to do.

1.4.2 Fieldwork

Participant [P1] ran the tool evaluation as a series of meetings/workshops, using the heuristics to aid planning and execution of the tool evaluation. Meetings were held with the new tool vendor, with testers doing the tool evaluation, and an additional meeting with the [O1}'s UX director. This last meeting arose out of insights from the tool evaluation.

Meetings were held during February, March and April 2024 (see Table 1). The initial meeting with the vendor was to decide whether to carry out the evaluation. [P1] used a limited number of the heuristics to guide this meeting. The next meetings, with the testers who would use the tool planned and executed the tool evaluation, using an increasing number of heuristics to drive the evaluation. There was an interim meeting with the vendor, to discuss progress. As a result of the tool evaluation to that point, [P1] had a meeting with [O1}'s UX director, to further evaluate both the tool and the workflows the tool was part of.

Date	At meeting	Purpose	Heuristics used by
21Feb2024	[P1] Vendor	Stage 1: decide whether to invest time evaluating the tool	[P1]
22Feb2024	[P1] Testers	Stage 2: Plan and set up tool evaluation	[P1] Testers
22Feb2024-25Feb2024	[P1] Testers	Stage 3a: Carry out tool evaluation: comparison of [T1} (OS) and [T2} (COM)	[P1] Testers
26Feb2024	[P1] Testers	Stage 3b: Review tool evaluation results	[P1] Testers
26Feb2024	[P1] Testers Vendor	Stage 3c: Discuss progress with the tool evaluation	[P1] Testers
04Mar2024	[P1] [O1} UX director	Stage 4: Additional meeting that arose out of the tool evaluation findings	[P1]
02Apr2024	[P1] Vendor	Stage 5: Final meeting with vendor to feedback on tool evaluation	Not used explicitly

Table 1 Meetings and other activities during the Tool Evaluation

Data collection was via diary entries kept by [P1] of events and insights including:

- Questions from participants and clarifications given;
- Changes made to questions and reasons for the changes;
- Assessment of each heuristic for:
  - usefulness (what thoughts did it prompt? What insights did it bring you to?);
  - understandability (how much help did you need to interpret the question?);

- flexibility (how did you interpret the question?);
- Notes of what activities were done to help answer the question;
- Order the questions were answered in;
- How questions related to other questions (e.g. revisiting question);
- Insights arising from the use of the heuristics; and
- Changes in actions from those insights.

1.4.3 Results

1.4.3.1 Stage 1: Using the heuristics to decide whether to carry out a full evaluation

Heuristics used:

- H01 Why do we need this tool?
- H02 Who will use or be affected by the tool?
- H07 Where will the tool be used?
- H08 What workflows will the tool be part of?

The purpose of the meeting was for organization [O1} to decide whether to evaluate the new tool – that is with the plug-in functionality integrated into the tool [T2} (COM) dashboard. Was it worth the effort of evaluating it?

The participant [P1] used the heuristics to plan and manage the meeting, and did not explicitly tell the vendor about them. The four heuristics selected very judged to be useful, understandable and flexible: Highest score of 5 for each factor was given for all four heuristics.

The heuristics the participant chose **not** to use were judged to be understandable and flexible, but not useful at this stage, so scored high for understandability and flexibility, but low for usefulness. Exception was H03 which was not scored.

1.4.3.2 Stage 2: Using the heuristics to plan and set up the tool evaluation

Heuristics used:

- H01. Why do we need this tool?
- H02. Who will use or be affected by the tool?
- H03. What previous experiences do people bring to the tool?
- H07. Where will the tool be used?
- H08. What workflows will the tool be part of?

Heuristic H03 was introduced at this stage to choose the team for the evaluation considering who had experience of the existing tool in use, and of the specific area of testing.

``*During an initial discussion with the tester who will be using the tool, the same questions that we used on the previous day were used. In addition, we also used H03 because this tester has used a similar tool before.''* [P1]

However, H03 was adapted:

``*With regards to H03, at this point and in context for this tool, not all of it was relevant. I think this is more to do with the nature of the tool we are evaluating. The tester who is helping the evaluation already uses a tool that is a plugin to a test automation tool. The tool that we are evaluating is an integration of this plugin into a dashboard. The conversation in this session was more around the existing processes using the existing plugin, and how will the new integration impact those.''* [P1]

The participant made the point that at this stage the heuristics were being used to make sure that the right people were involved in the evaluation, rather than for evaluating the tool itself.

The participant also noted that they may used heuristic H3 ``*almost subconsciously*'' when deciding who to ask to the meeting.

1.4.3.3 Stage 3: Using the heuristics to carry out the tool evaluation

Heuristics used:

- H01 Why do we need this tool?
- H02 Who will use or be affected by the tool?
- H06 What learning preferences do those people have?
- H07 Where will the tool be used?
- H08 What workflows will the tool be part of?

In this stage the [O1} team evaluated by comparing tool [T1} (OS) with [T2} (COM). The participant [P1] commented: ``*This phase was quite similar to the previous one, we did not really go through all the heuristics to make an evaluation*.'' [P1]

The evaluation, and the use of the heuristics gave rise to five insights (highlighting in bold by the author not the participant). Four were about the tool itself, and one was about the way the evaluation was carried out.:

1. The tool had such good perceived UX that the heuristics on learnability could be bypassed. The new tool was described by the participant as one that ``*doesn't require learning - just delivers a report*.'' The heuristics H04 and H05 were not overtly used:``[tool] *eliminates need for learning*'' with "**UX so beautiful you don't need to think about it**."
   - Insight: the heuristics may confirm good design as well as highlight design improvements.
2. Evaluating the tool using the heuristics on workflow made the participants question the current processes and behaviours around this testing: ``*The discussion with the tester was that the bugs reported from this plugin (whether integrated or not) were never acted upon anyway. We started to question* **whether we are using this tool at the right stage in the development process or not**.'' [P1]
   - Insight: the heuristics H07 and H08 enabled the evaluation team to look at the problems and challenges and question whether a tool would resolve them; Gaps in the process and behaviours were identified.
3. This led the team evaluating the tool to revisit the H01 heuristic and re-ask H01 "why are we using it?" because the problems uncovered by the tool were not given a high priority by the engineers:
   ``*At this point we started questioning the tech process that we use and who would benefit from the results of the tool.*
   In reality, at the moment it's the engineering teams who impact the design systems of our website who are doing accessibility testing. Whenever they get problems flagged by the tool, they do **log the critical ones, but they are never a high priority on the team's backlog***.*
   - Insight: revisiting the heuristic H01 ``Why?'' question in a more holistic way, noting that the engineers' view of issue severity and priority may not be the same as the UX Department's view, and that there are gaps in workflows.
4. *Which led the evaluation team to rethink H02 ``Who is the tool for?''*

> *The result of this meeting was that I will next talk to our UX Director to see whether this tool would be useful for them. As a result,* **we went back to questioning who this tool would be useful for**.'' [P1]

- Insight: revisiting the heuristic H02 ``Who?'' question in a more holistic way, to note that the engineers may not be the most useful primary users of the tool.

5. Additionally, the evaluation gave rise to some interesting changes in behavior depending on who was in the room. The feedback directly to the vendor was not the same as the feedback given in private: ``*Side note: I spotted that the way the tester talked to me about the tool was very different to how they talked about it with the vendor. In a personal discussion I was told "This is just the plugin we use, just integrated to the dashboard. We get no extra value". To the vendor, the same tester said "This is really nice work, well done guys" :D*'' [P1]

- Insight: this is a process point for the heuristics, which echoes findings in the researchers' studies S1-S6: when in a workshop people may be nervous of giving a reply that reflects what there would say in a private communication. Where the heuristics are used could affect their outcomes.

1.4.3.4 Stage 4: Using the heuristics for further tool evaluation based on insights from stage 3

Heuristics used:

- H01. Why do we need this tool?
- H02. Who will use or be affected by the tool?
- H03. What previous experiences do people bring to the tool?
- H04. What communication needs or preferences do those people have?
- H07. Where will the tool be used?
- H08. What workflows will the tool be part of?
- H10. What work styles are acceptable in those workflows and teams? [used indirectly]
- H11. When will the tool be used?
- H12. How long will the tool be used?

Based on the insights gained at the previous stage, the participant added an evaluation step: to engage the UX Department in the evaluation of the tool. This was new contact within the organization for the participant: ``*On this meeting I did not use the heuristics with the UX Director. I used them for myself but did not tell him about the study since it was my first time meeting him.*'' [P1] The heuristics were used as a checklist to guide the meeting.

``*Here I had an informal chat with our UX Director. I showed him the tool in the dashboard and he was quite keen on it. He mentioned that his UX teams have done accessibility training last year and they implement accessibility in their designs, but then when the Engineering department implements those designs,* **they never test for accessibility themselves to see that things the designs have been implemented as they should**. *We had a conversation about his team taking on training and doing work, for it to then* **not be put into use** *because the software developers don't follow those best practices.*'' [P1]

The work done on improvement in one team is disconnected from the work done in another team, which means that improvements are not propagated across the organization. Instead, improvements are started and lost further down the workflow.

Placing the tool in UX Department might enable them to identify where their accessibility requirements have not been implemented, when the engineers may not be fully aware of the importance of those requirements. However, a tool for UX Department needs to fit in as seamlessly as possible with their other work: ``***The UX Director is interested in this tool because it seamlessly runs accessibility tests for them without the software development teams having to intervene. Then they would***

be able to look into those results and put down specific requirements for engineering teams in the next iteration of development.'' [P1] In particular [P1] noted that there was no learning curve with the tool partly because of the good UX, and partly because the UX director is already familiar with the sort of outputs given by the tool.

The discussion ranged more widely across the heuristics, including into communication, hence the use of H04. The meaning and communication of the severity, and consequently priority of issues was the key area: ``*I wasn't sure here about whether I should put down H4. We did discuss the communication style of the tool. The tool lists down accessibility problems and sorts them by severity. This is nice, we did discuss that if the severity was better highlighted, it would be more useful because we are more interested in the critical problems.''* [P1]

This insight didn't answer the question about whether to adopt the new tool. The participant noted: ``*This does still raise the* **question about whether the new tool improves this or whether the old tool is good enough. In reality, it comes down to the cost of the new tool at the end of the day**.''* [P1]

The key insights gained from the evaluation using the heuristics was what the participant described as the ``**chasm**'' between UXD and Engineering. The participant is now thinking about "whole organization testing" and about how to improve the processes and communications. A resulting process and tool diagram might look like Figure 3, where the barrier between UXD and Engineering is removed or mitigated, and the UX teams use tool [T1} (OS) – or the equivalent dashboard within [T2} (COM) – to monitor accessibility , raise issues, and monitor their resolution, perhaps alongside Engineering. The Test Support Team will write a small additional piece of code which makes the plug-in run automatically every time the test suite is run, with a report on accessibility automatically arriving at the UX Department. The UX teams can then raise issues and triage for severity. This intervention will improve communication and triage of accessibility issues, while also saving the company financially as no fee to the external vendor is required.
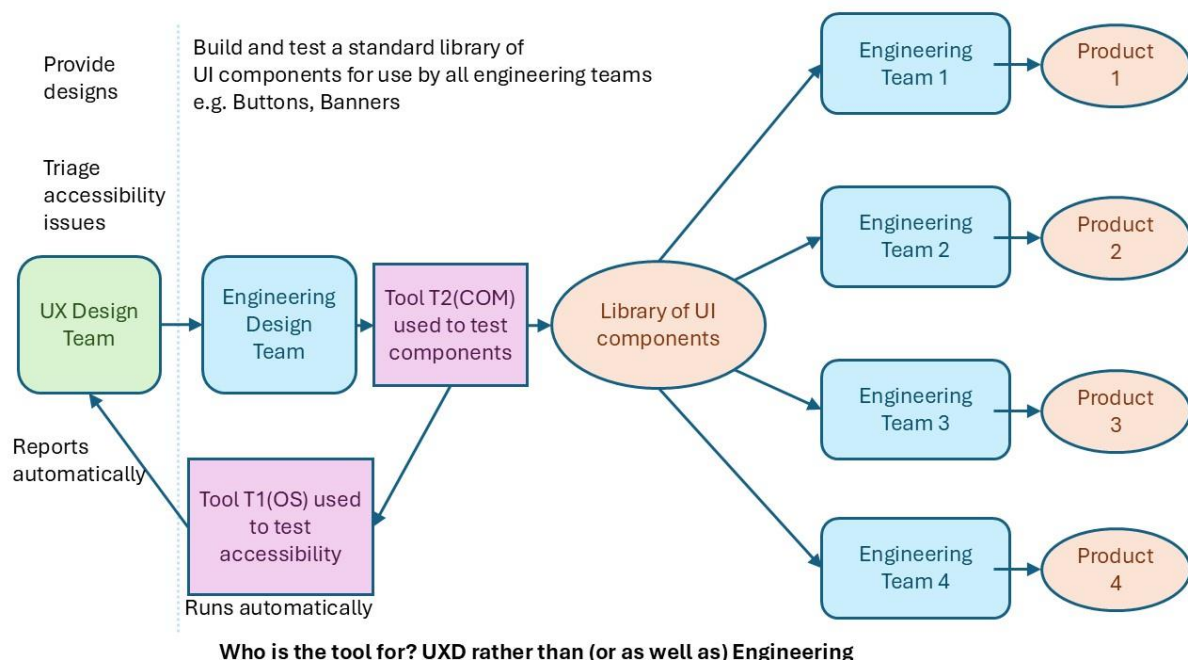

Who is the tool for? UXD rather than (or as well as) Engineering

Figure 3 Process discussion post insights from heuristics

1.4.3.5   Stage 5: Feedback to Vendor

The final stage of the evaluation of the tool was a feedback meeting. The heuristics were not used in this meeting, however in discussing the meeting with the participant we identified some potential changes to the heuristics.

The vendor used the meeting to find out about the progress with the evaluation, demo two new features of their tool, discuss the cost of the tool module that had been evaluated, and the legal implications of accessibility compliance.

The new features were relevant to the discussion with the vendor, and with the UX director: ``*During this meeting they showed me 2 new features they implemented since we started the trial. Coincidentally they were features that were helping what the UX Director and I were discussing regarding making the severity of the accessibility problems more visible.''* [P1]

A key point in the evaluation at this point was the financial aspect, and we realised in post-evaluation discussion that ROI is buried too far down in the heuristics for this type of evaluation process. It may be sensible to have a point on financial goals in H01, to catch that early; part of the discussion at stage 1 of the evaluation could have been ``Is this tool within our budget?'' which can be added to the prompts in the detail for H01. As the participant noted: ``*During this meeting, I then asked the vendor how much this tool will cost. It's a very steep price tag and it would essentially double the price of what we currently pay for the dashboard.''*

Although the vendor offered to help make a business case for the tool, based partly on legal requirements, and partly on business benefits, the participant noted that the organization is exempt from EU Disability Act 2025, because the products don't provide a necessary service. They also concluded that, although a business could be made for the tool, cost was an issue ``*The cost of the tool was the make or break here when it came to whether we will buy it or not.''*  [P1] while sorting out the process and communication gaps was more important, with a decision about the tool taken later:  ``*in reality, I would prefer putting my efforts elsewhere and fix how we work with testing overall''* [P1]

The conclusion of this tool evaluation for the participant was to work on improving test process and communication and to work with the UX Department to get them feedback on the results from the existing tool in a timely manner to allow severity and priority to be set by the UX teams. [P1] noted in a later conversation: ``*that saved 20,000 Euros!''* A later insight was from the engineering team on seeing these results, that accessibility could be built in earlier from the designs rather than waiting for issues to be raised.
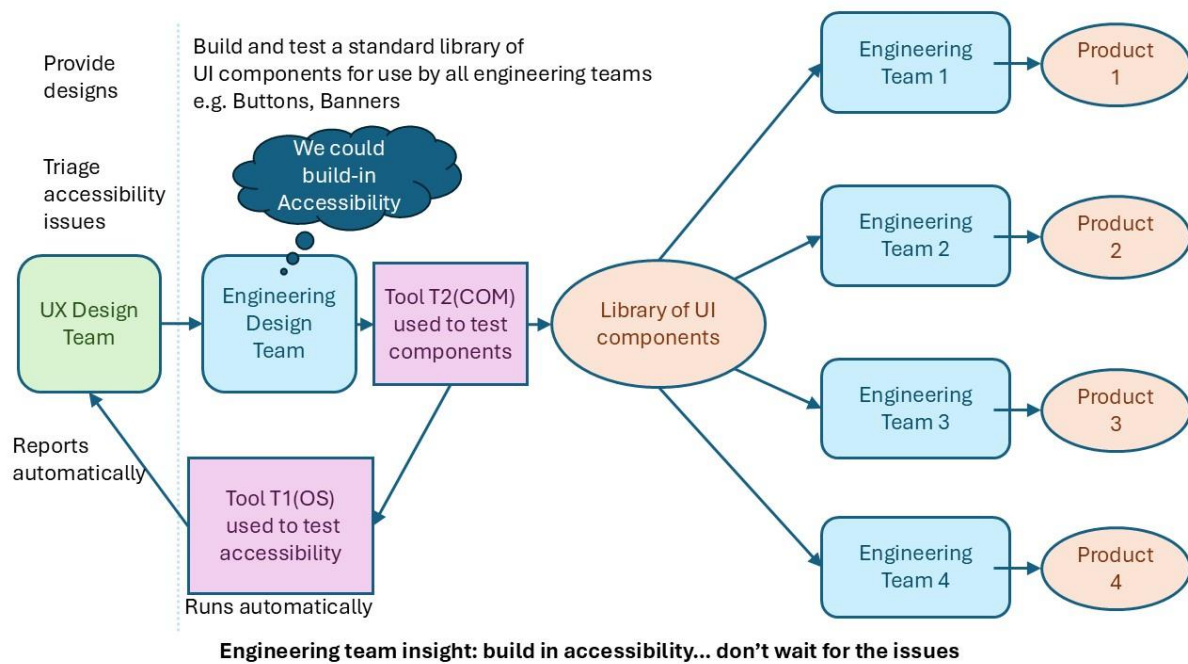
Figure 4 Later Engineering Team Insight: build in accessibility…don't wait for the issues

1.4.3.6   Overall feedback on the heuristics from the participant

The overall feedback on the heuristics from this case study was positive. The participant felt that the heuristics had enabled insights to be gained during the tool evaluation, and that they were natural to use: ``*I found that the conversation naturally flowed in the same order as the heuristics. This is quite impressive.*'' [P1]

1.5 Evaluating the Use of the Heuristics

In this section we evaluate the heuristics themselves, starting with summarizing their use and their scores, then notes on those that provided significant insights, then those not used at all, and finally identifying improvement to the heuristics, underlying prompts and explanations, and to the case study process.

1.5.1 Heuristics Summary Usage and Scores

Table 2 summarizes which heuristics were used at which stage of the tool evaluation. It is clear that a subset of the heuristics was useful in the earlier planning and preparation stages of the tool evaluation to facilitate discussion of whether to evaluate the tool at all, and who should take part in the evaluation. Note also that as insights happened some heuristics were revisited, and the previous answers reassessed, particularly H01, Why? and H02 Who?

	WHY?	WHO?					CONTEXT?						
Stage	H01	H02	H03	H04	H05	H06	H07	H08	H09	H10	H11	H12	Purpose of stage
evaluation stage 1	x	x					X	x					shall we evaluate?
evaluation stage 2	x	x	x				X	x					how should we do the evaluation?

evaluation stage 3	x	x	x				X	x					evaluation part 1
evaluation stage 4	x	x	x	x			X	x		x	x	X	evaluation part 2
evaluation stage 5													feedback to vendor

Table 2 Which heuristics used at different evaluation stages.

All the heuristics scored highly – 5 – for understandability and flexibility at all stages, so we have not tabulated those separately. The usefulness scores for the heuristics did vary, and these are summarized in Table 3. The heuristics marked with scores of 1 were either not useful in this context, for the participants, or were treated ``almost subconsciously'', or did not require a deep dive to answer: ``*Some heuristics were useful but I did not need a do deep dive into them. E.g. the learnability one was something I did consider but since there was no learning curve to the tool, I did not need to look into it further.''* [P1]

		usefulness			
	heuristic	evaluation stage 1 shall we evaluate?	evaluation stage 2 how and who involved?	evaluation stage 3 evaluation starts	evaluation stage 4 reinterrogating ``why'' and ``who?''
WHY?	H01	5	5	5	5
WHO?	H02	5	5	5	5
	H03	1	3	1	5
	H04	1	1	1	5
	H05	1	1	1	1
	H06	1	1	1	1
CONTEXT?	H07	5	5	5	5
	H08	5	5	5	5
	H09	1	1	1	1
	H10	1	1	1	4
	H11	1	1	1	5
	H12	1	1	1	3

Table 3 Usefulness of each heuristic at each stage (Participant assessment)

H03 scored 3 at stage 2 as it did not require much thought – it was implicitly obvious to the participant that someone experienced in using the current tool needed to be involved.

H10 scored 4 at stage 4 because work styles were not key to the discussion at this point.

H12 scored 3 at stage 4 because the ROI and financial questions were not asked soon enough in the process of going through the heuristics; perhaps if this had been asked at stage 1 the evaluation may not have taken place, although that would have meant the process insights were not gained.

1.5.2 Heuristics providing especially important insights

- H01. Why do we need this tool?
- H02. Who will use or be affected by the tool?

- H07. Where will the tool be used?
- H08. What workflows will the tool be part of?

As we saw in sections 1.4.3.3 and 1.4.3.4, this subset of the heuristics provided significant insights for the tool evaluation team, and redirected their efforts from tool acquisition to process and communication improvements.

1.5.3 Heuristics not used at all

- H05. What learning perspectives and goals do those people bring?
- H06. What learning preferences do those people have?
- H09. What risks are associated with those workflows?

These heuristics were not overtly used during the tool evaluation. The tool acquisition was considered to be low risk, in a low-impact area, so H09 was read but not used. H05 and H06 were not considered to need a deep dive, because the UX of the tool under evaluation removed the need to consider learnability. However, the participant noted that even with the ones not used, the heuristics were ``at the back of my brain'' and that "heuristics [are used] to prompt thought."

1.5.4 Next steps for the participant organization

There is a need to assess the test automation support provided in-house in organization [O1}, and [P1] is keen to use the heuristics to aid the in-house automation team. They are prepared to write up on diary form as another case study. We discussed how this context differs from the tool evaluation, and looked at how the heuristics H11, and H12, focus on quality attributes such as maintainability during changing requirements. The ideal would be for the team to predict which parts of the automated tests *should* fail because of requirements change and therefore which tests to change. The author demo'ed the page in the repository with the research outcomes on quality attributes.

1.6 Specifying Learning

No improvements to the top-level heuristic questions were identified. No changes were made to the questions, or to the order of the questions. No clarification questions were asked by the participant during the tool evaluation. Planning and starting the case study was easy, however this participant has an academic background and knows what is needed in a case study.

1.6.1 Improvements to heuristics

The detail below each question was useful, but the following improvements could be made:

- ROI and financials need to be earlier, especially for a tool evaluation;
- Put a prompt under H01 for financial goals;
- Cover this explicitly in the Tool Evaluation process explanation;
- The ROI question may need to be different for the tool designer/vendor and the tool evaluator/purchaser – cover that explicitly;
- Add steps to the evaluation suggested process, based on this case study.

1.6.2 Improvements to case study process

For the next case study, especially if it is over a longer time period, consider having a longer introduction to the heuristics repository. Include:

1. an initial review of the top-level questions
2. review process for their context
3. select questions if not using them all
4. show them specifically where to find the detail for those specific heuristics
5. have interim meeting to review progress and if they are looking at other heuristics introduce the detail for those at that point.

1.6.3 Reflections on the heuristics and their effect

Note that the participants reverted to their old thinking and had to be prompted towards the new thinking – the use of the research data in making that mind change

So the heuristics that arise as new ideas – 04-05-06 – need more pushing maybe?

Need to monitor in other case studies – that might be about taken for granted assumptions in this specific organization

Also add two extra columns to diary forms – did not use, and already thought about.