Authentic Testing White Paper

Job-Task Analysis for Performance Testing

Wallace Judd, PhD

Performance Test Development Series 02

Job-Task Analysis for Performance Testing

Wallace Judd, PhD

Performance Test Development Series 02

Job-Task Analysis for Performance Testing

Scope & Prerequisites

One of the first decisions to be made by the SME team is to define the scope of the exam. What specific content is to be evaluated? One method that facilitates this is to enumerate skills the test will address, and to clarify additional skills that are prerequisite and will be assumed already mastered by candidates; and to list skills which are beyond those expected of candidates.

Another way of arriving at the scope is to characterize the audience for the test, and the appropriate uses of the test.

Tools

If the test is a software test, then operating systems, software version and configuration must be detailed. The best way to define these is to create an initial specification, and then revise it as specific tasks come up, so in conclusion the specification is clear for all items.

Additional resources available or prohibited must be specified as well. A group doing a JTA for a software development tool surprisingly asked that the internet be available to candidates without restriction, since that was critical to their daily development efforts. After pointing out the security issues with that we compromised on enforcing a white list of internet sources which would be available to candidates.

If the test is a software test, then operating systems, software version and configuration must be detailed. The best way to define these is to create an initial specification, and then revise it as specific tasks come up, so in conclusion the specification is clear for all items.

Three Types

There are three types of JTAs. The simplest is for skills defined by another organization—an external JTA. A second one is for a new product or market. The third is for a product that has a minimum of several hundred users who have used it extensively.

External JTA

An external JTA is adopted from another entity, perhaps a governmental body, a professional association, educational institution or influential company.

A simple example is a skill set for construction hazards inspection. The following is a list of the Top 10 OSHA violations for 2020.

- Fall Protection
- Hazard Communication Standard
- Respiratory Protection
- Scaffolding
- Ladders
- Lockout/Tagout
- Powered Industrial Trucks
- Fall Protection–Training Requirements
- Eye and Face Protection
- Machinery and Machine Guarding

https://www.osha.gov/top10citedstandards

Note that these are in order of frequency nationally, which may not reflect the frequency with which they're observed in a specific construction industry. Consequently, the Scheme Committee would have to weight them by importance for a specific context.

A more complex sample is drawn from a study by a driver education company that conducted a two-year study of driving skills and

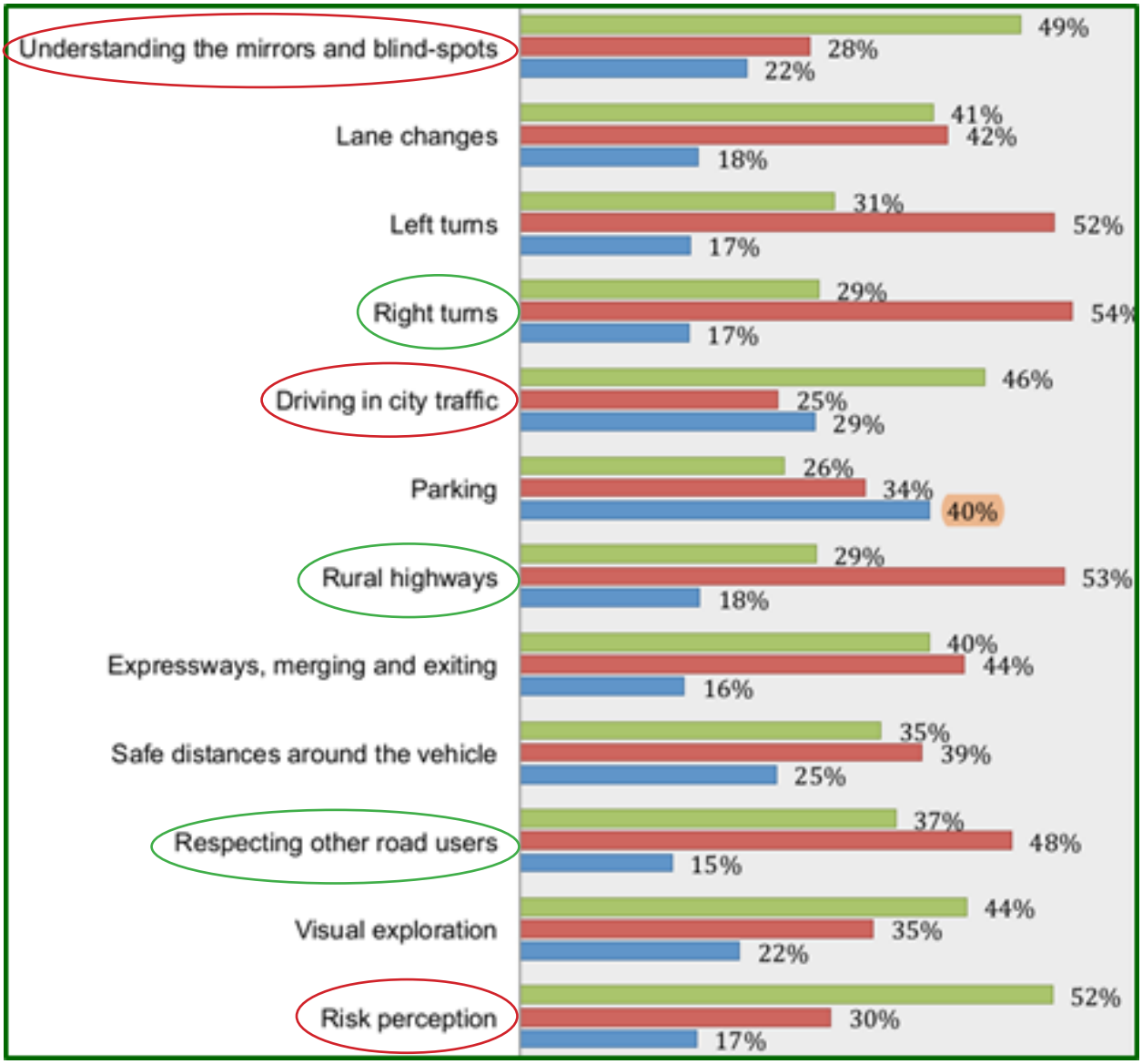

Figure 1. A two-year study compared student drivers' reactions to instruction on simulators and on the road. The top bars shows students who thought the simulation was better at teaching these topics; the second bar is students who thought on-the-road instruction was better.

Figure 2. Subject-matter experts have used sticky notes to jot down tasks associated with different domains and subdomains of professional practice.

the reaction of student drivers to education, either on simulators or in person.

The red circles indicate topics in which students preferred simulations. The green circles point to topics in which students preferred on-the-road instruction.

From these research results, the Scheme Committee may decide that the red-circled topics should be emphasized, while the green-circled topics can be de-emphasized. The rationale for the choices should be made explicit in the JTA report. https://www.researchgate.net/publication/319115341

An external JTA of intimidating proportions might be the flight requirements for a private pilot published in the 528-page *Pilot's Handbook of Aeronautical Knowledge* by the FAA (Federal Aviation Administration).

Whoever defines the set, it must be adopted and approved by your organization's Scheme Committee – the body that defines the scope and content of the exam.

Weighting

No matter the source, the contents should be evaluated and weighted by the Scheme Committee, with a brief rationale for weighting each topic.

Review

The JTA must document a review procedure to make certain that whenever the external source is updated or revised, the new version is considered for reevaluation and weighting.

New Content: Performance Testing

This JTA requires an assembly of 6–9 SMEs who were on the development team, in marketing or design, or who have beta tested the product with users. The SME teams creates a list of domains which describe how the product is used, then elaborate tasks under each domain.

To the left is a set of Post-It notes for one domain, "Storage Management." The green notes are topics. The pink Post-Its are specific tasks relevant to the topic beside them. Note that there can be more than one task beside specific tasks.

The tasks can be categorized by each SME by difficulty (easy, medium, or hard to use) and by frequency of use (infrequently, fairly often, often, daily). Awarding points to tasks at each level and adding them together yields a way to estimate their relative importance. (Some recommend

Figure 3. After hopping around the conference room sorting tasks captured on sticky notes, subject-matter experts take a moment to smile about their work product

multiplying difficulty by frequency to increase dispersion. In terms of rating it usually doesn't make a major difference.)

Average importance across all SMEs to produce a mean importance for each task, then sum the average importance for all tasks in each domain. This produces an overall task rating for each domain. Sum the domains and divide each domain by the total to produce a relative weighting for every domain. The domain weightings are refined by group consensus in another round.

The picture on the previous page shows an SME team conducting a JTA, with Post-It notes on the widows describing tasks & domains. This was at that time the exclusive set of experts for the specific product we were conducting a JTA for.

Your JTA could be over at this point.

If, however, you want to confirm or refine your domain weightings and you can assemble enough initial users to yield viable results, you can conduct a user survey of the tasks as described below and reconcile the results with the SME domain weightings.

Existing Content

This method assumes that you have an audience of users experienced enough to rate the tasks addressed by the content.

Have the SME group define tasks which address the content. Assemble the tasks into domains with appropriate prompts on a survey.

Conduct a user survey of the tasks, asking for respondents to weight each task's difficulty and frequency. Roll up the results as described above, and use this as a preliminary estimate of domain weightings.

After receiving the analysis, ask the SME team to rate the tasks independently as well and summarize their results. The survey estimates can then be reconciled with the SME domain weightings either by the SME group itself, or by a scheme committee.

Extension to Multiple Choice

Multiple choice exams are more difficult to evaluate, simply because knowledge is more vague than specific tasks.

Take for example the knowledge requirement that a candidate "Be able to analyze system log files." This knowledge could require that they be able to list who was on the system at a specific time (an easy task). The knowledge could mean that the candidate be able to analyze who is using the printer extensively (medium task). At its highest level, it could mean for the candidate to analyze the log file to see who surreptitiously entered the system at what time and where they went and which files they modified (very hard task). Note that in the JTA analysis, any of these three are valid interpretations of the knowledge requirement "Analyze system log files."

In my experience, it is more appropriate to survey specific tasks, rank the domains by tasks, then document the knowledge necessary to conduct the task successfully.

Response Coding

This analysis assumes that the survey responses were recorded as text. The text can be translated into numeric values as shown in the two tables below. Importance was not coded 5, 4, 3, 2, 1 because that scaling tends to group all responses together. The scale in the table at left is an example of how Importance could be coded.

Importance	Weight
Critical	10
Very Important	7
Important	4
Somewhat Important	2
Not Important	0

Frequency	Weight
Daily	9
Weekly	5
Monthly	3
Rarely	1
Never	0

Figure 4. Converting survey responses to numerical values allows you to rank tasks in terms of criticality.

Likewise, Frequency can be coded on a scale shown in the table at right, which again serves to spread responses for analysis.

The resulting codings from these tables can be used to rate topics and domains. For each topic the product of all Frequency and Importance ratings are added, then averaged. These mean Frequency + Importance ratings will be used to rank topics overall. (Again, you may multiply instead of add.)

JTA Core Problem

The core problem with JTA procedures is that there does not appear to be a concise way of estimating whether the JTA accurately describes the performance of practitioners.

Having a clearly defined scope helps JTA developers. Having a clear definition of the minimally qualified candidate also helps. But it may be that the problem of JTA precision may be inherent in the fact that different user groups use the content knowledge in different ways.

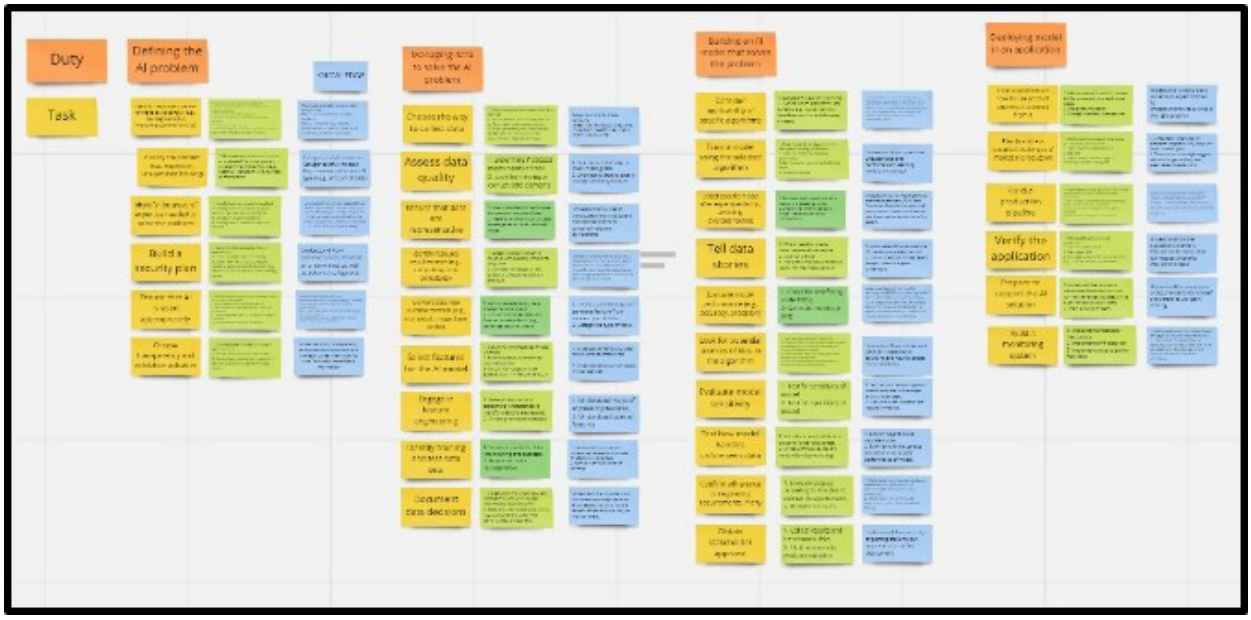

Figure 5. In a virtual environment, miro.com offers some of the functionality of walls and sticky notes.

Ideal JTA

The ideal JTA would record with precision and over a relatively long time how practitioners used the subject tool or knowledge in their daily practice. This record could then be parsed to see which skills were used in specific contexts. The enterprise using the tool or knowledge could then rate the frequency with which those contexts arose in users of specific roles, and weight the domains of the test appropriately. This is the ultimate JTA environment.

Online Tool

One online tool named Miro allows your SME team to construct the Post-It Notes online and collaboratively.

The picture to the right shows domains in orange, tasks in yellow, checking or scoring notes in green, and knowledge underlying each task in blue.

Work Product

The results of a JTA include:

- Scope Statement
- Software Specification and Execution Environment
- Help Resources Available during the exam
- Training Resources Available prior to the exam
- Public-Facing Blueprint with Main Topics, Domain Weightings, & optional subtopics
- Internal Blueprint with Main Topics, Weightings, Tasks, & Checking Process

3349 Monroe Avenue, Suite 336
Rochester, N.Y. 14618-5513
www.AuthenticTesting.com
+1-703-777-7321