

Improve your odds with Infosys Predictability



# Measuring Heuristic Evaluation Skills

Shanmugam Rajasekaran ([Shanmugam\\_R@infosys.com](mailto:Shanmugam_R@infosys.com))

Head, Experience Design

Shazeeye Kirmani ([Shazeeye\\_Kirmani@infosys.com](mailto:Shazeeye_Kirmani@infosys.com))

Senior Usability Engineer

May 27, 2006

# Contents

---

- Motivation
- Heuristic Evaluation
- Method
- Case Study: Indian Railways
  - Definition of categories and severity ratings
  - Scope of the evaluation
  - Benchmark
  - Results
- Applications

# Motivation

---

- Define and measure the heuristic evaluator's profile
- Identify specific training areas for developing heuristic evaluators
- Increase the quality of heuristic evaluation

# Heuristic Evaluation

- Jakob Nielsen's ten usability heuristics
  - Visibility of system status
  - Error prevention
  - Match between system and real world
  - User control and freedom
  - Help users recognize, diagnose and recover from errors
  - Recognition rather than recall
  - Consistency and standards
  - Flexibility and efficiency
  - Aesthetic and minimalist design
  - Help and documentation

Source: [http://www.useit.com/papers/heuristic/heuristic\\_list.html](http://www.useit.com/papers/heuristic/heuristic_list.html)

- 3 to 5 experts list the problems individually from which a consolidated report is obtained
- Severity rating:
  - Showstopper (3)
  - Major Issue (2)
  - Irritant (1)

	<a href="#">Passenger / PNR Status</a>	<a href="#">Trains between Imp. Stations</a>	<a href="#">Train/Fare Accommodation</a>
	<a href="#">Internet Reservation</a>		<a href="#">Tenders</a>
<a href="#">Upgradation Passenger Scheme / Chart</a>			<a href="#">Railway Retiring Room</a>

Indian Railways Online V  
Designed &

**Passenger Current Status**  
**Note:- Online PNR Status Enquiry will be available from 0400 hrs IST (GMT + 05:30) to**

**Problem:** There is no indication of user's current location  
**Category:** Interaction Design  
**Severity:** 2  
**Heuristic violated:** Visibility of system status

1. Identify the application which will be used to test evaluators' skills

2. Identify the categories to measure evaluators' skills  
E.g. Visual Design, Interaction Design, Information Architecture

3. Standardize the scope of the evaluation  
E.g. In amazon.com, we will evaluate only the ordering process

4. Standardize the time to do the evaluation

5. Standardize the evaluation format into problem, category and severity

6. Provide a knowledge transfer of the application

7. Execute the test

8. Use these tests to determine the benchmark; get a consensus on the problem list and severity ratings

9. Use this benchmark to measure the skills of each evaluator

# Case Study: Indian Railways

---

- 30 people were evaluated from a pool of UI designers
- Evaluation time was kept constant at 2 hours
- Scope of the evaluation was standardized
  - 38 screenshots
  - 10 main menu items
  - 13 sub menu items
- Feedback format was standardized into problem, category and severity
- Eight categories
  - Visual Design
  - Interaction Design
  - Navigation
  - Information Architecture
  - Labeling
  - Functionality
  - Content
  - Other (Accessibility, Branding, Web Guidelines)
- Three severity ratings
  - Showstopper (3)
  - Major Issue (2)
  - Irritant (1)

# Definition of Categories: Interaction Design (Page 1 of 8)

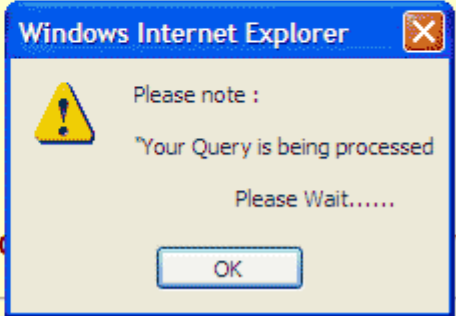
- Interaction between user and the interface
  - Error and information communication, incomplete feedback and status mechanisms, lack of user control, mismatch between interactions and user mental models

**Passenger Current Status**

**Note:- Online PNR Status Enquiry will be available from 0400 hrs IST (GMT + 05:30) to 2330 hrs.**

PNR Number :

**SMS at 7070,7886,676747,6677,8888 to**



**Problem:** The information message tells users to wait, but unless the OK button is clicked the user can't get his query results

**Category:** Interaction Design

**Severity:** 3

**Heuristic violated:** Error prevention

# Definition of Categories: Visual Design (Page 2 of 8)

- Appropriateness of page layout, color, font, animation, graphics, grouping
  - Visual inconsistency, absence of visual differentiation between links and labels, non-adherence to visual hierarchy principles, wrong placement of controls, ineffective use of real estate

	<a href="#">Passenger/ PNR Status</a>	<a href="#">Trains between Imp. Stations</a>	<a href="#">Train/Fare Accommodation</a>	<a href="#">Train Enquiry</a>	<a href="#">Weekly Availability Status at Stations</a>	<a href="#">Latest Announcements</a>	<a href="#">Ministry of Railways</a>	
	<a href="#">Internet Reservation</a>	<a href="#">Tenders</a>	<a href="#">Railway Retiring Room</a>	<a href="#">Rail SMS Services</a>	<a href="#">Helpful Links</a>	<a href="#">Summer Special Trains</a>		<a href="#">हिन्दी</a>
	<a href="#">Upgradation Passenger Scheme/Chart</a>				<a href="#">Home</a>	<input type="button" value="Reset"/>	<a href="#">Railway Websites</a>	

Indian Railways Online Website: <http://www.indianrail.gov.in>  
Designed & Hosted by [CRIS](#)

## Trains/Fare/Reservation Availability Status

Source Station Name :

Enter the few characters of Station, e.g.: "mum" for getting Station.

Destination Station Name :

Enter the few characters of

Class :

Select the class in which you wish to see all the classes,

**Problem:** Visual inconsistency; no visible alignment to grids; basic typographic principles not followed.

**Category:** Visual Design

**Severity:** 2

**Heuristic violated:** Consistency and standards

# Definition of Categories: Navigation (Page 3 of 8)

- Ease of flow across menus and within menus
  - Absence of optimal workflows, navigation aids for long pages, consistency, navigation to support forms, charts and related links

## **The Details of Derailment of Train No. 1108 (Bundelkhand Express) at Datia(Site of the Accident) on 03/10/2005**

- |  |                                  |
|--|----------------------------------|
| <b>1.Casualty</b>                        | <b>11 Persons</b>                |
| <b>2.Injured</b>                         | <b>37 Persons</b>                |
| <b>3.Ex gratia paid till 19:25 Hours</b> |                                  |
| • <b>Casualty</b>                        | <b>8 Persons paid@Rs.15000/-</b> |
| • <b>Grievous Injured</b>                | <b>10Persons paid@Rs 5000/-</b>  |
| • <b>Simple Injured</b>                  | <b>23 persons paid@Rs500/-</b>   |

**Problem:** The accident chart does not support navigation to the main site

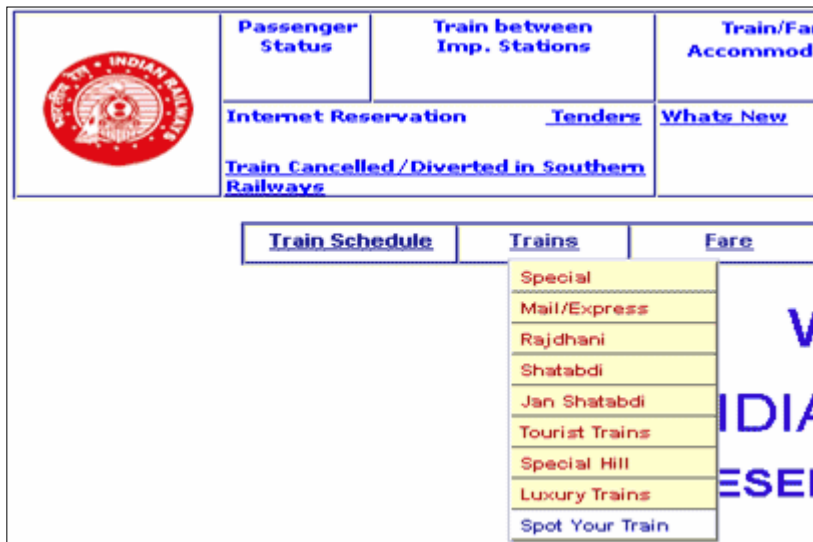
**Category:** Navigation


**Severity:** 2

**Heuristic violated:** Flexibility and efficiency

# Definition of Categories: Information Architecture (Page 4 of 8)

- Accurate categorization of information into appropriate menu groups
  - Absence of mutually exclusive groups, menus not prioritized, inappropriate menu width and depth



	Passenger Status	Train between Imp. Stations	Train/Fare Accomod:
	Internet Reservation	Tenders	Whats New
	Train Cancelled/Diverted in Southern Railways		
	Train Schedule	Trains	Fare
		Special	
		Mail/Express	
		Rajdhani	
		Shatabdi	
		Jan Shatabdi	
		Tourist Trains	
		Special Hill	
		Luxury Trains	
		Spot Your Train	

**Problem:** Information groups are not mutually exclusive. The groups “Trains between Imp. Stations”, “Train Schedule” and “Trains” overlap.

**Category:** Information Architecture

**Severity:** 3

**Heuristic violated:** Match between system and real world

# Definition of Categories: Labeling (Page 5 of 8)

- Appropriateness of labels as per user mental models
  - Inappropriate labels for menus, sub menus, links, controls, pages, paragraphs, categories, inconsistent labels (search, find, retrieve), presence of technical jargon and abbreviations

<a href="#">Availability Status at Stations</a>	<a href="#">Latest Announcements</a>	<a href="#">Ministry of Railways</a>
<a href="#">Home</a> ▼	<input type="button" value="Reset"/>	<a href="#">Railway websites</a>
<a href="#">mation</a>	<a href="#">About Us</a>	<a href="#">Vox Populi</a>

**Problem:** “Vox Populi” does not effectively communicate “Comments and Suggestions”

**Category:** Labeling

**Severity:** 2

**Heuristic violated:** Match between system and real world

# Definition of Categories: Content (Page 6 of 8)

- Completeness and accuracy of information on the interface to answer user queries
  - Missing content, wrong content, inappropriate presentation of information for scanability

## Details Of Derailment of Bundhelkhand Express Train on North Central Railway on 03-10-05

### IMPORTANT DETAILS REGARDING THE DERAILMENT OF TRAIN NUMBER 1108 BUNDELKHAND EXPRESS AT DATIA(MP) (SITE OF ACCIDENT) ON 03/10/2005

```
1|      RESERVATION CHART FOR AC 3-TIER SLEEPER          COACH : AS1      (CONCERT CHARTS)
|      COACH NO:                POSITION FROM ENGINE:
|  TRAIN : 1108  BUNDELKHAND EXP LEAVING VARANASI JN    ON 02 Oct 2005
|                        VARANASI JN    TO    GWALIOR
|
|*****
|S/B:QUOTA   :      NAME      :SEX:FROM: TO :   PNR   : MSG  :  TKT.NO. :PNDG:   NAME   :SEAT
|:CODE      :                :AND:   :   :   :   : STN  :      :AMT.: (IN HINDI) :BERTH
|NO.:      :*=VIP @=Tkt lost :AGE:   :   :   :   :      :Rs.  :      :NO.
|*****
| 1 | 1 |
| 2 | 2 |
```

**Problem:** Wrong chart (reservation chart) uploaded for the details of a particular train accident

**Category:** Content

**Severity:** 2

**Heuristic violated:** Error prevention

# Definition of Categories: Functionality (Page 7 of 8)

- Presence and appropriateness of all functions to serve all user goals
  - Missing functionalities, not clubbing functionalities, lack of optimal steps to execute a task, functionalities not designed as per user mental models

Indian Railways Online Website: <http://www.indianrail.gov.in> designed and hosted by CRIS.

## TRAINS BETWEEN A PAIR OF STATIONS

You Queried For

**Origin** NEW DELHI -[NDLS] **Destination** MUMBAI CENTRAL -[BCT] **Class** Sleeper

[Reservation Form](#)

Enter Your Journey

Date:

**Problem:** Does not club functionalities like “Get Availability”, “Get Fare” and “Get Schedule”

**Category:** Functionality

**Severity:** 2

**Heuristic violated:** Flexibility and efficiency

# Definition of Categories: Other (Page 8 of 8)

- Branding
- Web guidelines
- Accessibility guidelines

**Problem:** Brand image not represented consistently

**Category:** Other

**Severity:** 2

**Heuristic violated:** Consistency and standards

	Passenger/PNR Status	Trains between Imp. Stations	Train/Fare Accommodation	Train Enquiry	Weekly Availability Status at Stations	Latest Announcements
	Internet Reservation	Tenders	Railway Retiring Room	Rail SMS Services	Helpful Links	Summer Specials
	Upgradation Passenger Scheme / Chart				Home	Reset



# Definition of Severity Ratings

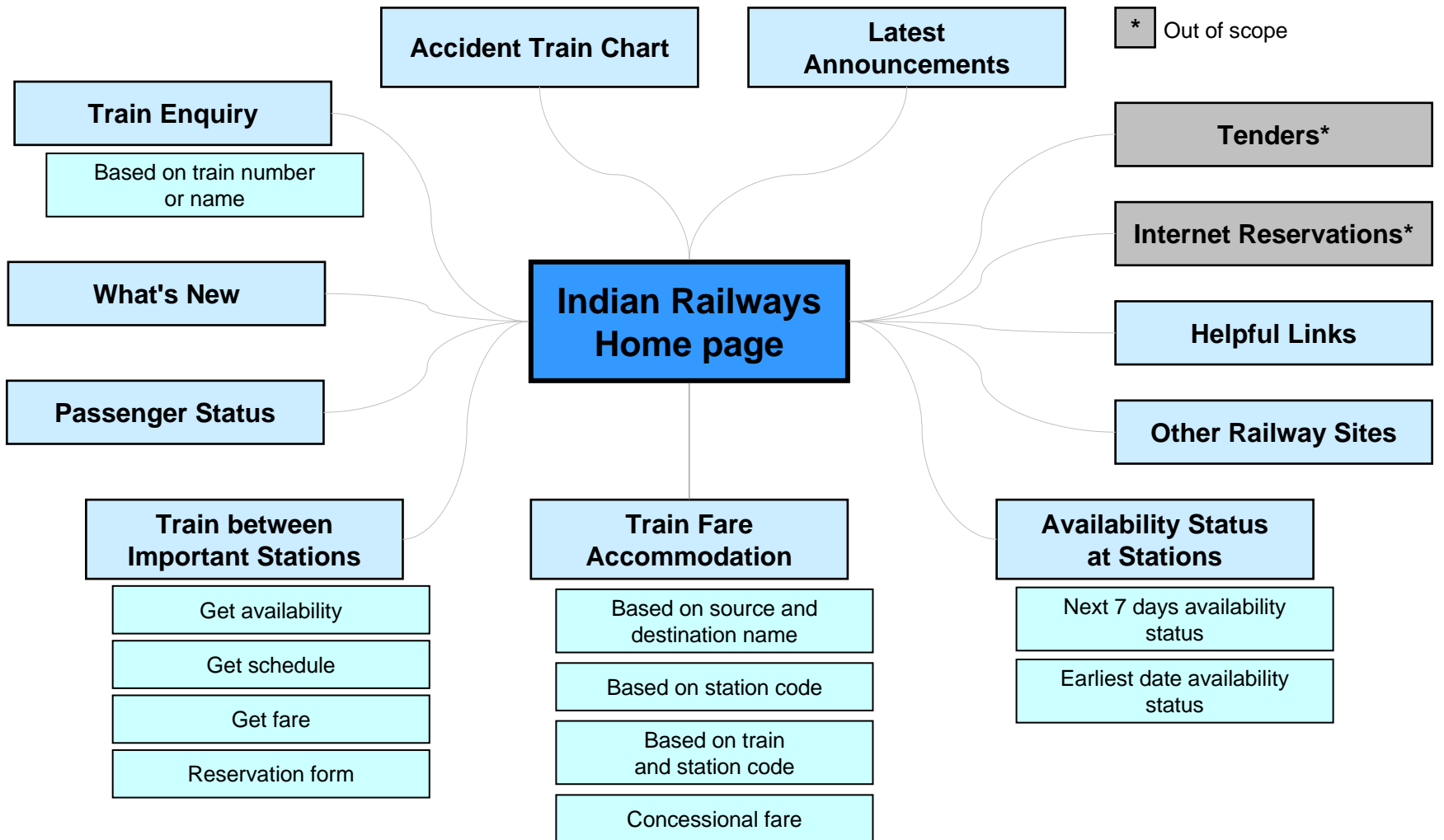
---

**Showstopper (3):** Catastrophic issues that prevent people from using the site effectively and hinder users from accomplishing their goals

**Major Issue (2):** Causes waste of time and increased learning / error rates

**Irritant (1):** Minor cosmetic or consistency issue, slows users down slightly and minimal violation of usability guidelines

# Scope of the Evaluation



# The Benchmark

Categories	Issues
Interaction Design	19
Visual Design	6
Navigation	22
Information Architecture	15
Labeling	15
Content	5
Functionality	6
Others	5
<b>Total Issues</b>	<b>93</b>

Inter-rater reliability	%
All 4 agreed	38%
3 of 4 agreed	60%
All 4 disagreed	2%

30 evaluators with heuristic evaluation experience ranging from 6 months to 5 years

# Results: Raw Data (Page 1 of 4)

## Breadth of Issues

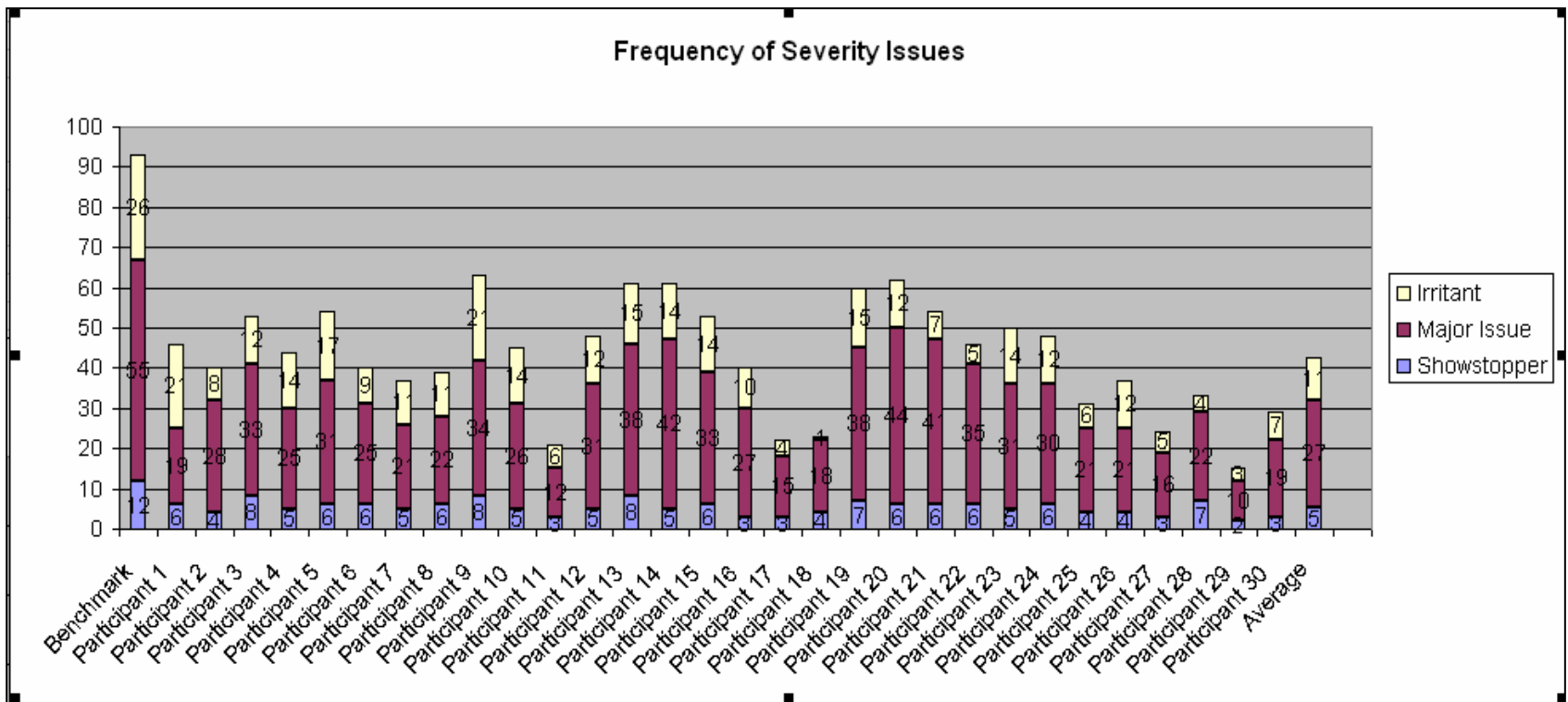
Evaluator Name	Total issues	Total number (unique issues) 93 issues	Interaction 19 issues	Visual 22 issues	Navigation 6 issues	IA 6 issues	Labeling 15 issues	Content 5 issues	Functionality 15 issues	Other 5 issues
Benchmark	212	93	19	22	6	6	15	5	15	5
Participant 1	56	46	12	12	2	5	6	3	4	2
Participant 2	64	40	10	10	2	4	5	3	4	2
Participant 3	63	53	15	9	2	4	5	4	11	3
Participant 4	56	44	8	17	2	2	7	3	4	1
Participant 5	92	54	13	16	4	5	3	4	7	2
Participant 6	54	40	9	11	4	5	4	2	2	3
Participant 7	54	37	6	11	3	3	7	2	2	3
Participant 8	72	39	10	13	3	2	4	2	2	3
Participant 9	132	63	10	19	5	6	8	3	7	5
Participant 10	64	45	11	17	2	2	3	3	6	1
Participant 11	27	21	6	6	3	2	1	1	2	0
Participant 12	126	44	10	17	3	3	5	1	5	0
Participant 13	160	61	7	17	6	5	8	6	7	5
Participant 14	135	61	13	19	6	4	5	4	7	3
Participant 15	74	53	10	14	3	4	7	5	7	3
Participant 16	74	40	6	9	2	3	7	4	6	3
Participant 17	28	22	5	9	3	1	2	0	1	1
Participant 18	23	23	8	7	1	3	2	1	0	1
Participant 19	142	60	14	20	5	5	5	3	5	3
Participant 20	75	62	14	14	3	5	7	5	11	3
Participant 21	97	54	6	19	5	5	6	3	7	3
Participant 22	53	46	11	10	4	6	7	5	2	1
Participant 23	50	50	12	11	4	4	6	4	5	4
Participant 24	74	48	16	9	2	2	6	2	8	3
Participant 25	53	31	10	10	4	1	1	2	2	1
Participant 26	88	37	8	16	2	3	2	1	3	2
Participant 27	40	24	11	6	1	0	3	2	0	1
Participant 28	43	33	7	12	3	1	2	5	3	0
Participant 29	18	15	3	4	2	0	2	2	2	0
Participant 30	50	29	7	8	1	1	7	1	1	3
Average	71	43	10	12	3	3	5	3	4	2

## Severity

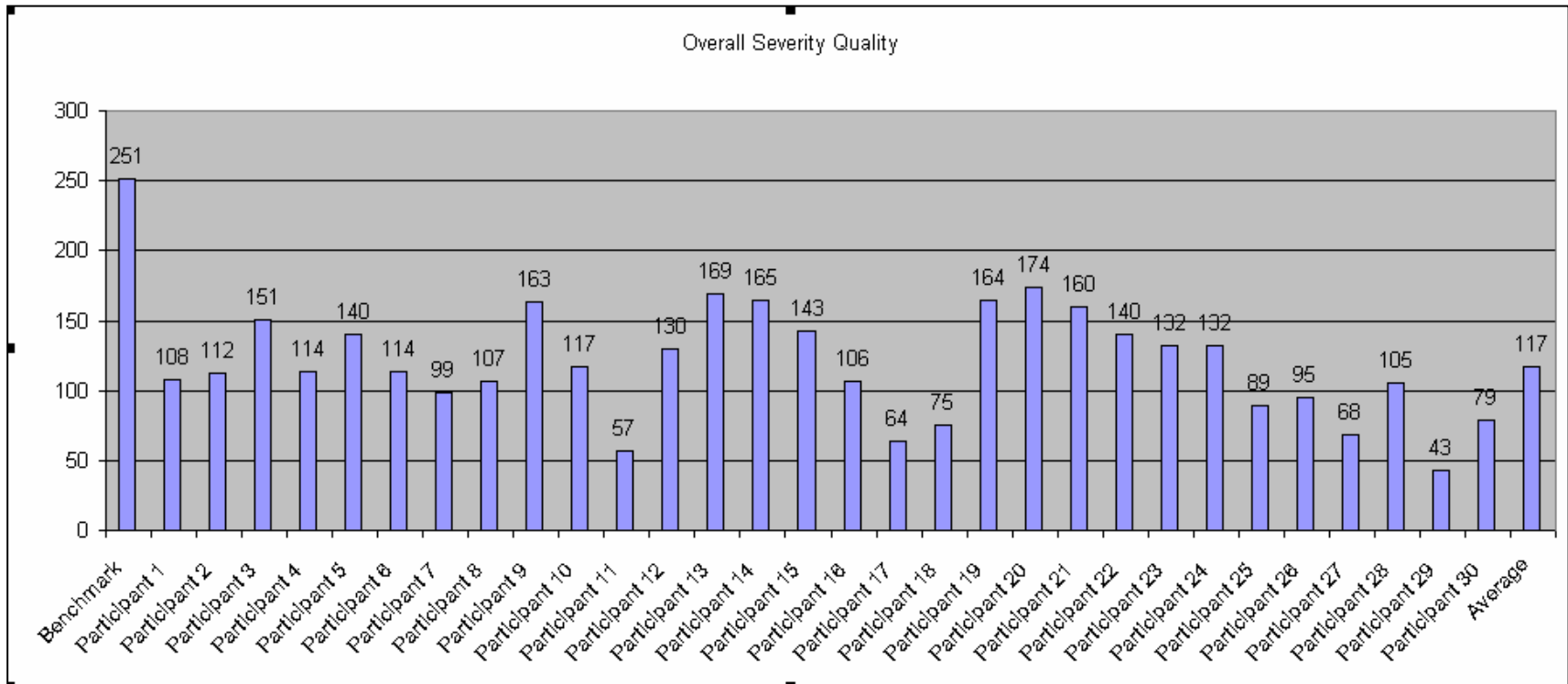
Evaluator Name	Showstopper 12 issues	Major 55 issues	Irritant 26 issues
Benchmark	12	55	26
Participant 1	6	19	21
Participant 2	4	28	8
Participant 3	8	33	12
Participant 4	5	25	14
Participant 5	6	31	17
Participant 6	6	25	9
Participant 7	5	21	11
Participant 8	6	22	11
Participant 9	8	34	21
Participant 10	5	26	14
Participant 11	3	12	6
Participant 12	5	31	12
Participant 13	8	38	15
Participant 14	5	42	14
Participant 15	6	33	14
Participant 16	3	27	10
Participant 17	3	15	4
Participant 18	4	18	1
Participant 19	7	38	15
Participant 20	6	44	12
Participant 21	6	41	7
Participant 22	6	35	5
Participant 23	5	31	14
Participant 24	6	30	12
Participant 25	4	21	6
Participant 26	4	21	12
Participant 27	3	16	5
Participant 28	7	22	4
Participant 29	2	10	3
Participant 30	3	19	7
Average	5	27	11

# Results: Analysis of Severity Issues (Page 2 of 4)

Frequency of Severity Issues



# Results: Analysis of Severity Quality (Page 3 of 4)



## Weightage factors

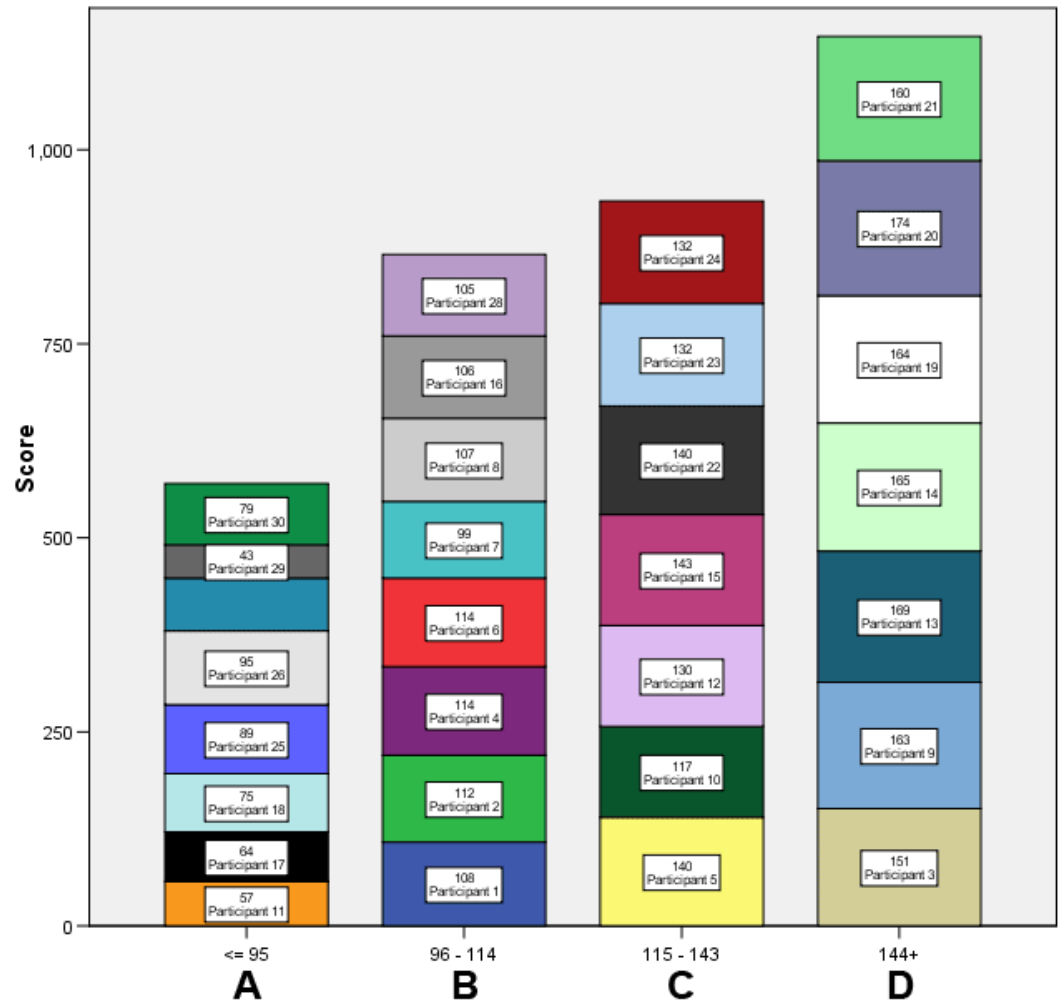
Showstopper = 5 points Major Issue = 3 points Irritant = 1 point

If Participant 1 identified 12 showstoppers, 55 major issues and 26 irritants, then,

Overall Severity Quality =  $12 \times 5 + 55 \times 3 + 26 \times 1 = 251$

# Results: Performance Bands of Evaluators (Page 4 of 4)

- A** Need to increase coverage of severity and breadth of issues significantly
- B** Need to increase coverage of severity and breadth of issues
- C** Average coverage of severity and breadth of issues.
- D** Best coverage of severity and breadth of issues.



# Application: Comparison (Page 1 of 3)

	<b>Collaborative HE</b>	<b>Individualist HE</b>
Application	Amazon.com	
Evaluators	3	
Expertise	Given Level	
Time	2 Hours	
Method	Collaborative	Individual
Result	125 issues 325 score	75 issues 180 score

# Applications (Page 2 of 3)

- Define the profile of an expert
- Compare experts' skills within an application (**A**)
  - Relative ratings of evaluation performance
- Compare expert skills across an application (**B**)
  - Is Participant 1 better at identifying issues on desktop applications or handhelds?

**A**

Evaluator Name	Total issues	Total number (unique issues) 93 issues	Interaction 19 issues	Visual 22 issues	Navigation 6 issues	IA 6 issues	Labeling 15 issues	Content 5 issues	Functionality 15 issues	Other 5 issues
Participant 1	56	46	12	12	2	5	6	3	4	2
Participant 2	64	40	10	10	2	4	5	3	4	2

**B**

Evaluator Name	Application	Benchmark	Showstopper	Major	Irritant	Participant's Score
Participant 1	Desktop	251	6	19	21	108
Participant 1	Handheld	145	5	15	18	88

# Applications (Page 3 of 3)

---

- Co-relating evaluation-expertise and UI complexity levels
- Recommending individual areas for improvement
  - Training improved the skills up to 250%
- Certification and tracking skill progress
- Assessing the quality of heuristic evaluation

## Improve your odds with Infosys Predictability



Thank You