



0



1

## Pengantar

Bagaimana cara terbaik mengetahui usability sebuah antarmuka?

Bertanya terhadap user mengenai pengalaman (experience) dalam menggunakan antarmuka tersebut

Terdapat banyak jenis bentuk pertanyaan

- Pertanyaan terbuka / tertutup
- Rating
- Pilihan

Semua itu disebut dengan istilah **Self-reported Metrics**

2/70

2

## Pengantar

Istilah lain untuk mendefinisikan Self-reported Metrics

▪ **Subjective Data**

- Seringkali kita sulit mendapatkan data yang objektif
- Kemungkinan besar setiap partisipan akan memberikan data yang subjektif
- Tetapi dari sudut pandang usability specialist, hal itu tidak masalah

▪ **Preference Data**

- Seringkali preferensi mempengaruhi user dalam membuat suatu pilihan
- Hal tersebut terkadang diluar dugaan (sulit diperkirakan)

3/70

3

## Pengantar

Self-reported Metrics memberikan informasi penting tentang persepsi pengguna (user)

- Apa yang dipikirkan pengguna terhadap sistem?
- Apa yang dirasakan pengguna terhadap sistem?



4

## Collecting Self-Reported Data

Rating

- Cara paling efisien adalah dengan menggunakan sistem rating
- Jika menggunakan *open-ended questions* lebih sulit untuk dianalisa, jawaban harus dikategorikan



5/70

5

## Collecting Self-Reported Data

Rating

- **Likert Scales** adalah cara menyatakan level ‘kesetujuan’, bisa menggunakan 3,5,7 level

Mis: Likert scales 5 level

- sangat setuju
- setuju
- ragu-ragu
- tidak setuju
- sangat tidak setuju



6/70

6

## Collecting Self-Reported Data

Rating

- **Semantic Differential Scales** menggunakan skala yang serupa dengan likert scales dengan dua kata yang berlawanan untuk menunjukkan kecenderungan.

Biasanya menggunakan 5 atau 7 level skala

Mis:

Weak	○ ○ ○ ○ ○ ○	Strong
Beautiful	○ ○ ○ ○ ○ ○	Ugly
Hot	○ ○ ○ ○ ○ ○	Cold
Light	○ ○ ○ ○ ○ ○	Dark

- Terkadang sulit mencari lawan kata yang sesuai, contoh:

lapar >< kenyang ?	friendly >< not friendly ?
lapar >< tidak lapar ?	friendly >< unfriendly ?
lapar >< terisi ?	friendly >< hostile ?

Pemilihan lawan kata yang tidak sesuai, tentunya akan memberikan hasil yang berbeda

7/70

7

## Collecting Self-Reported Data

**When & How**

- Kapan waktu yang tepat untuk mengumpulkan data?
  - post-task** (ketika selesai mengerjakan setiap task)
  - post-session** (setelah menyelesaikan suatu sesi pengujian)
- Bagaimana cara mengumpulkan data?
  1. secara lisan
  2. menggunakan media kertas
  3. menggunakan online tools



8

## Post- Task Ratings

- Post-Task Ratings bertujuan untuk mengetahui task manakah yang menurut pengguna paling sulit.
- Cara paling mudah untuk mengetahui sulit atau tidaknya sebuah task adalah menggunakan sistem rating seperti likert scales
- After-Scenario Questionnaire:
  1. I am satisfied with the ease of completing the tasks in this scenario.
  2. I am satisfied with the amount of time it took to complete the tasks in this scenario.
  3. I am satisfied with the support information (online help, messages, documentation) when completing the tasks.

9/70

9

## Post- Task Ratings

**Analyzing Data**

- **Expectation Measure**

<p>1. Overall, I am satisfied with the ease of completing the tasks in this scenario □</p> <p>2. Overall, I am satisfied with the amount of time it took to complete the tasks in this scenario □</p> <p>3. Overall, I am satisfied with the support information (online-line help, messages, documentation) when completing the tasks □</p>	<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 10%;">1</td> <td style="width: 10%;">2</td> <td style="width: 10%;">3</td> <td style="width: 10%;">4</td> <td style="width: 10%;">5</td> <td style="width: 10%;">6</td> <td style="width: 10%;">7</td> </tr> <tr> <td>strongly disagree</td> <td>○ ○ ○ ○ ○ ○ ○</td> <td>strongly agree</td> </tr> <tr> <td>strongly disagree</td> <td>○ ○ ○ ○ ○ ○ ○</td> <td>strongly agree</td> </tr> <tr> <td>strongly disagree</td> <td>○ ○ ○ ○ ○ ○ ○</td> <td>strongly agree</td> </tr> </table>	1	2	3	4	5	6	7	strongly disagree	○ ○ ○ ○ ○ ○ ○	strongly agree	strongly disagree	○ ○ ○ ○ ○ ○ ○	strongly agree	strongly disagree	○ ○ ○ ○ ○ ○ ○	strongly agree
1	2	3	4	5	6	7											
strongly disagree	○ ○ ○ ○ ○ ○ ○	strongly agree															
strongly disagree	○ ○ ○ ○ ○ ○ ○	strongly agree															
strongly disagree	○ ○ ○ ○ ○ ○ ○	strongly agree															
1 2 3 4 5 6 7																	

**FIGURE 6.1**

The ASQ developed by Lewis (1991). This implementation is Gary Perlman's web interface, which can be found at <http://www.acm.org/perlman/question.cgi?form=ASQ>.

10/70

10

## Post- Task Ratings

**Analyzing Data**

- **Expectation Measure**

Average Expectation Rating (X)	Average Experience Rating (Y)	Quadrant
2.0	5.5	Promote it
2.5	6.0	Promote it
3.5	4.0	Big opportunity
4.0	5.0	Big opportunity
5.5	5.5	Don't touch it
6.0	2.0	Fix it fast
6.0	6.0	Don't touch it
6.5	2.5	Fix it fast
7.0	1.5	Fix it fast

**FIGURE 6.2**

Comparison of the average expectation ratings and average experience ratings for a set of tasks in a usability test. Which quadrants the tasks fall into can help you prioritize which tasks to focus on improving. Source: Adapted from Albert and Dixon (2003); used with permission.

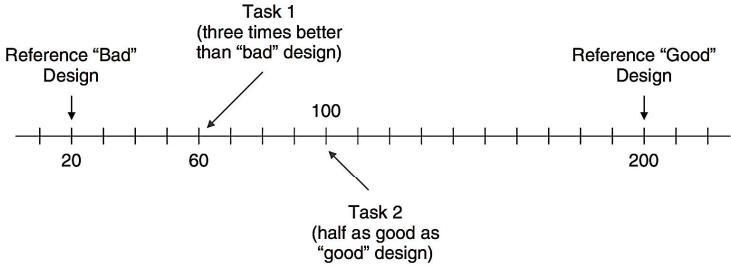
11/70

11

## Post- Task Ratings

**Analyzing Data**

- **Usability Magnitude Estimation**



**FIGURE 6.3**

Example of a participant's "usability ruler" and how it might be used for assessing the usability of the system for various tasks.

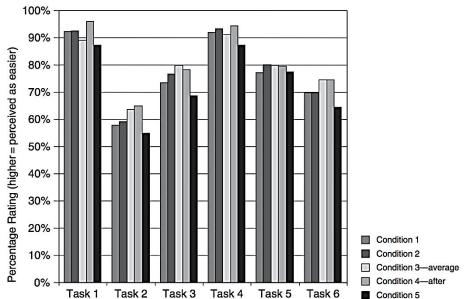
12/70

12

## Post- Task Ratings

**Analyzing Data**

- Todesco & Tullis (2006) mencoba membandingkan berbagai jenis self-reported metrics berbasis task
- Dari 5 kondisi yang dibandingkan, ternyata kelimanya memberi pola yang sama



**FIGURE 6.6**

Average subjective ratings split by task and condition. All five conditions (self-report techniques) yielded essentially the same pattern of results for the six tasks. Source: Adapted from Tedesco and Tullis (2006).

13/70

13

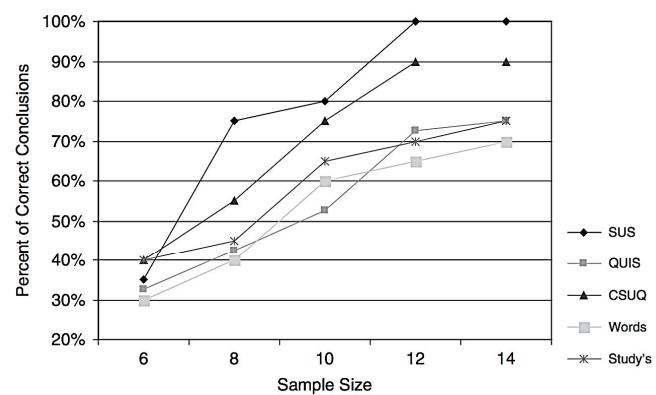
## Post- Session Ratings

- Post-Session Ratings bertujuan untuk menilai antarmuka sistem/piranti secara keseluruhan.
- Terdapat beberapa metode untuk melakukan post-session ratings:
  1. Aggregating Individual Task Ratings.
  2. System Usability Scale (SUS).
  3. Computer System Usability Questionnaire (CSUQ).
  4. Questionnaire for User Interface Satisfaction (QUIS)
  5. Usefulness, Satisfaction and Ease of Use Questionnaire (USE)
  6. Product Reaction Cards
  7. Website Analysis and Measurement Inventory (WAMMI)

14/70

14

## Post- Session Ratings

**FIGURE 6.13**

Data illustrating the accuracy of the results from random subsamples ranging from size 6 to size 14. This graph shows what percentage of the random samples yielded the same answer as the full dataset at the different sample sizes. Source: Adapted from Tullis and Stetson (2004).

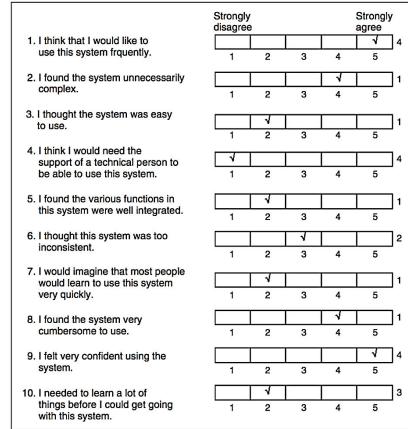
15/70

15

- Dibuat oleh John Brooke, berisi 10 statement yang terdiri dari 5 statement positif dan 5 statement negatif.

## Post- Session Ratings

System Usability Scale (SUS)



Total = 22      SUS Score =  $22 \times 2.5 = 55$

**FIGURE 6.8**

SUS, developed by John Brooke at Digital Equipment Corporation, and an example of scoring it. Source: From Brooke (1996).

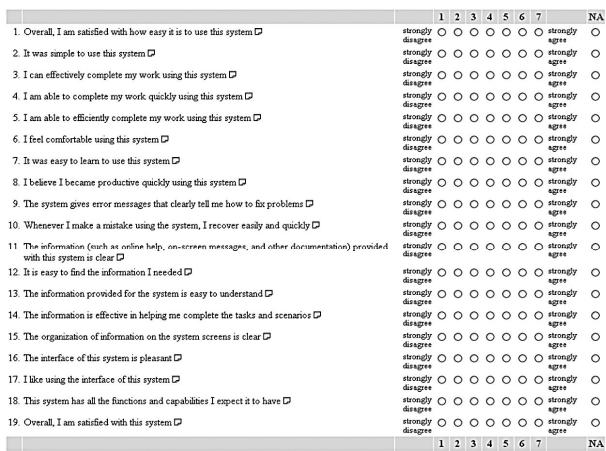
16/70

16

## Post- Session Ratings

### Computer System Usability Questionnaire (CSUQ)

- Hampir serupa dengan SUS, hanya terdapat beberapa modifikasi dan jumlah statement yang lebih banyak



**FIGURE 6.9**

The CSUQ. Source: Adapted from the work of Lewis (1995); used with permission.

17/70

17

## Post- Session Ratings

Questionnaire for User Interface Satisfaction (QUIS)

OVERALL REACTION TO THE SOFTWARE	0	1	2	3	4	5	6	7	8	9	NA
1. <input type="checkbox"/> terrible	<input type="radio"/>										
2. <input type="checkbox"/> difficult	<input type="radio"/>										
3. <input type="checkbox"/> frustrating	<input type="radio"/>										
4. <input type="checkbox"/> inadequate power	<input type="radio"/>										
5. <input type="checkbox"/> dull	<input type="radio"/>										
6. <input type="checkbox"/> rigid	<input type="radio"/>										
SCREEN	0	1	2	3	4	5	6	7	8	9	NA
7. Reading characters on the screen <input type="checkbox"/>	<input type="radio"/>										
8. Highlighting simplifies task <input type="checkbox"/>	<input type="radio"/>										
9. Organization of information <input type="checkbox"/>	<input type="radio"/>										
10. Sequence of screens <input type="checkbox"/>	<input type="radio"/>										
TERMINOLOGY AND SYSTEM INFORMATION	0	1	2	3	4	5	6	7	8	9	NA
11. Use of terms throughout system <input type="checkbox"/>	<input type="radio"/>										
12. Terminology related to task <input type="checkbox"/>	<input type="radio"/>										
13. Position of messages on screen <input type="checkbox"/>	<input type="radio"/>										
14. Prompts for input <input type="checkbox"/>	<input type="radio"/>										
15. Computer informs about its progress <input type="checkbox"/>	<input type="radio"/>										
16. Error messages <input type="checkbox"/>	<input type="radio"/>										
LEARNING	0	1	2	3	4	5	6	7	8	9	NA
17. Learning to operate the system <input type="checkbox"/>	<input type="radio"/>										
18. Exploring new features by trial and error <input type="checkbox"/>	<input type="radio"/>										
19. Remembering names and use of commands <input type="checkbox"/>	<input type="radio"/>										
20. Performing tasks is straightforward <input type="checkbox"/>	<input type="radio"/>										
21. Help messages on the screen <input type="checkbox"/>	<input type="radio"/>										
22. Supplemental reference materials <input type="checkbox"/>	<input type="radio"/>										
SYSTEM CAPABILITIES	0	1	2	3	4	5	6	7	8	9	NA
23. System speed <input type="checkbox"/>	<input type="radio"/>										
24. System reliability <input type="checkbox"/>	<input type="radio"/>										
25. System tends to be <input type="checkbox"/>	<input type="radio"/>										
26. Correcting your mistakes <input type="checkbox"/>	<input type="radio"/>										
27. Designed for all levels of users <input type="checkbox"/>	<input type="radio"/>										

18

## Post- Session Ratings

Usefulness, Satisfaction and Ease of Use Questionnaire (USE)

<ul style="list-style-type: none"> <li>▪ Dibuat oleh Arnie Lund</li> <li>▪ Terdiri dari 30 statement yang terbagi menjadi 4 kategori:</li> </ul> <ol style="list-style-type: none"> <li>1. Usefulness</li> <li>2. Satisfaction</li> <li>3. Ease of Use</li> <li>4. Learning</li> <li>5. Ease of Learning</li> </ol>	<p><b>Usefulness</b></p> <ul style="list-style-type: none"> <li>• It helps me be more effective.</li> <li>• It helps me be more productive.</li> <li>• It is useful.</li> <li>• It gives me more control over the activities in my life.</li> <li>• It makes the things I want to accomplish easier to get done.</li> <li>• It saves me time when I use it.</li> <li>• It <i>meets my needs</i>.</li> <li>• It does everything I would expect it to do.</li> </ul> <p><b>Ease of Use</b></p> <ul style="list-style-type: none"> <li>• It is easy to use.</li> <li>• It is simple to use.</li> <li>• It is user friendly.</li> <li>• It requires the fewest steps possible to accomplish what I want to do with it.</li> <li>• It is flexible.</li> <li>• <i>Using it is effortless.</i></li> <li>• <i>I can use it without written instructions.</i></li> <li>• <i>I don't notice any inconsistencies as I use it.</i></li> <li>• <i>Both occasional and regular users would like it.</i></li> <li>• <i>I can recover from mistakes quickly and easily.</i></li> <li>• <i>I can use it successfully every time.</i></li> </ul> <p><b>Ease of Learning</b></p> <ul style="list-style-type: none"> <li>• I learned to use it quickly.</li> <li>• I easily remember how to use it.</li> <li>• It is easy to learn to use it.</li> <li>• <i>I quickly became skillful with it.</i></li> </ul> <p><b>Satisfaction</b></p> <ul style="list-style-type: none"> <li>• I am satisfied with it.</li> <li>• I would recommend it to a friend.</li> <li>• It is fun to use.</li> <li>• It works the way I want it to work.</li> <li>• It is wonderful.</li> <li>• I feel I need to have it.</li> <li>• It is pleasant to use.</li> </ul>
---	--

FIGURE 6.11

The USE questionnaire. Source: From the work of Lund (2001); used with permission.

19/70

19

## Post- Session Ratings

**Product Reaction Cards**

- Dibuat oleh Joy Benedek & Trish Miner dari Microsoft
- Menggunakan pendekatan yang paling berbeda diantara metode yang lain
- User memilih kartu-kartu (dari 118 kartu yang disediakan). Setiap kartu berisi sebuah kata/statement yang sesuai dengan kondisi antarmuka.

**20/70**

20

## Post- Session Ratings

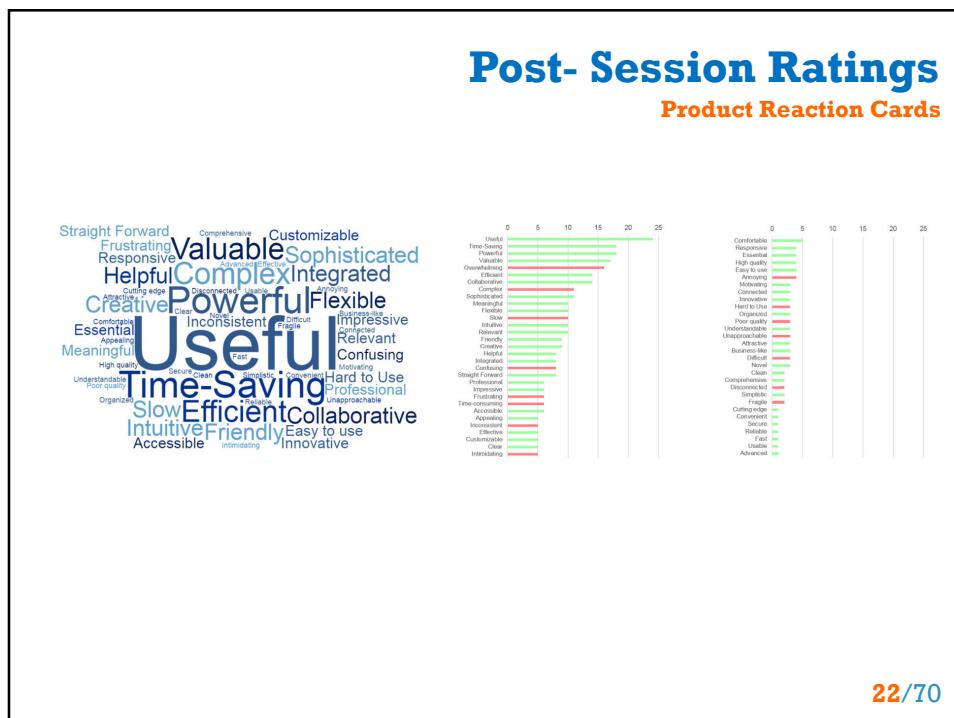
**Product Reaction Cards**

**The Complete Set of 118 Product Reaction Cards**

Accessible	Creative	Fast	Meaningful	Slow
Advanced	Customizable	Flexible	Motivating	Sophisticated
Annoying	Cutting edge	Fragile	Not secure	Stable
Appealing	Dated	Fresh	Not valuable	Sterile
Approachable	Desirable	Friendly	Novel	Stimulating
Attractive	Difficult	Frustrating	Old	Straightforward
Boring	Disconnected	Fun	Optimistic	Stressful
Business-like	Disruptive	Gets in the way	Ordinary	Time-consuming
Busy	Distracting	Hard to use	Organized	Time-saving
Calm	Dull	Helpful	Overbearing	Too technical
Clean	Easy to use	High quality	Overwhelming	Trustworthy
Clear	Effective	Impersonal	Patronizing	Unapproachable
Collaborative	Efficient	Impressive	Personal	Unattractive
Comfortable	Effortless	Incomprehensible	Poor quality	Uncontrollable
Compatible	Empowering	Inconsistent	Powerful	Unconventional
Compelling	Energetic	Ineffective	Predictable	Understandable
Complex	Engaging	Innovative	Professional	Undesirable
Comprehensive	Entertaining	Inspiring	Relevant	Unpredictable
Confident	Enthusiastic	Integrated	Reliable	Unrefined
Confusing	Essential	Intimidating	Responsive	Usable
Connected	Exceptional	Intuitive	Rigid	Useful
Consistent	Exciting	Inviting	Satisfying	Valuable
Controllable	Expected	Irrelevant	Secure	
Convenient	Familiar	Low maintenance	Simplistic	

**21/70**

21



22

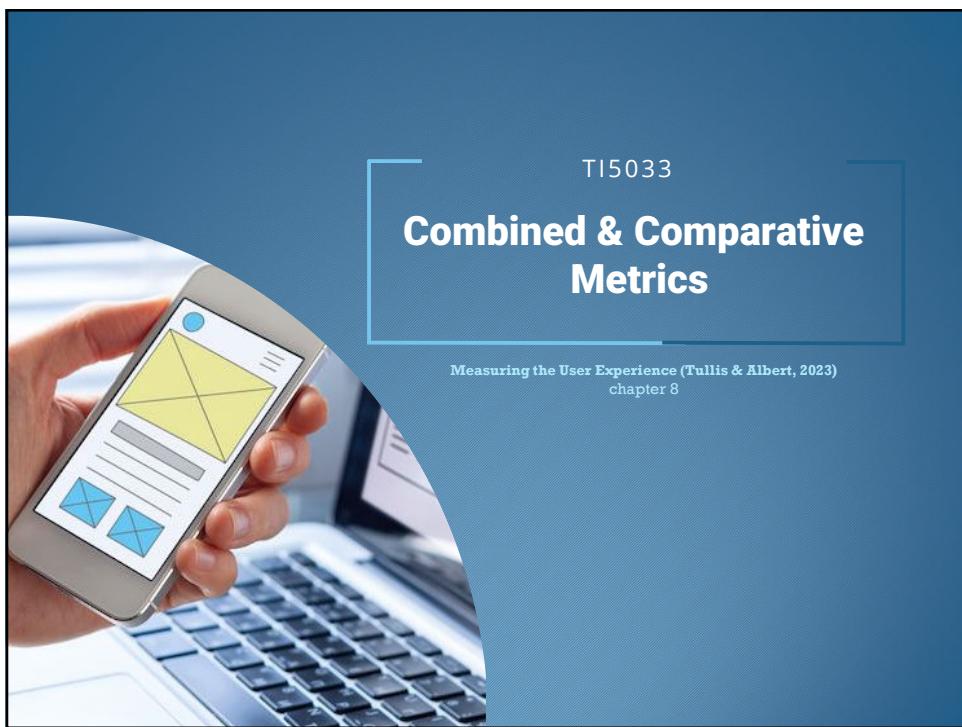
**Class Activity**

? ☰

### Membuat Kuesioner

1. Buatlah kelompok 2-3 orang
2. Buatlah suatu kuesioner untuk menguji kegunaan situs e-class UKDW, kemudahan penggunaannya dan kepuasan pengguna terhadap situs tersebut.
3. Kuesioner terdiri dari 10 pertanyaan *semantic differential scales*.
4. Mintalah 5 orang teman anda untuk mengisi kuesioner tersebut
5. Analisa dan berikan kesimpulan yang diperoleh

23



24

## Pengantar

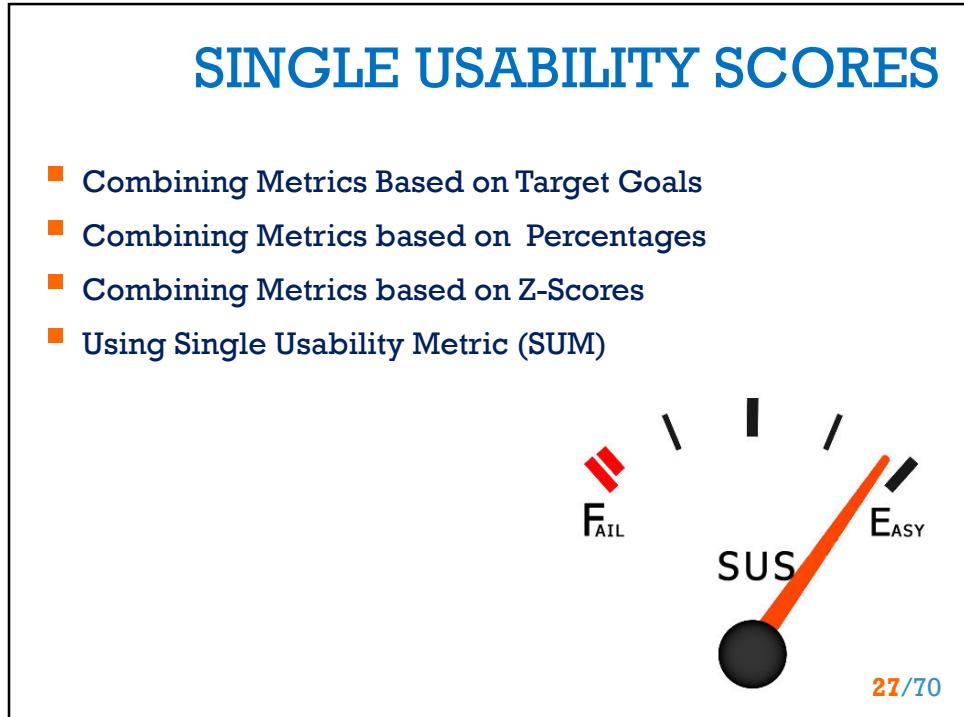
- Data usability ( task completion, time on task, self-reported) dapat digunakan untuk hasilkan ukuran baru yang mewakili keseluruhan ukuran → Combined dan comparative metrics
- Combined dan comparative metrics adalah nilai yang mudah untuk dipahami dan rangkuman dari semua ukuran
- Teknik : **single usability score** dan **perbandingan data usability dan data expert**



25



26



27

## COMBINING METRICS : TARGET GOALS

- Kombinasi ukuran untuk melihat ketercapaian sasaran
- Contoh : kombinasi ukuran task completion dan ukuran task time untuk melihat perfoma partisipan dalam gunakan produk.
- Apakah partisipan mencapai sasaran yang ditetapkan?  
Jika ya → 1, jika tidak → 0

28/70

28

## COMBINING METRICS : TARGET GOALS

Target Goal = TC min 80%, TT < 70 sec

Participant #	Task Completion	Task Time (secs)	Goal Met?
1	85%	68	1
2	70%	59	0
3	80%	79	0
4	75%	62	0
5	90%	72	0
6	80%	60	1
7	80%	56	1
8	95%	78	0
Average:	82%	67	38%

Table 8.1 Sample task completion and task time data from eight participants<sup>a</sup>.

<sup>a</sup>Also shown are averages for task completion and time and an indication of whether each participant met the objective of completing at least 80% of the tasks in no more than 70 seconds.

29/70

29

## HASIL PENGUJIAN

- Rata-rata Task Completion 82% - lebih dari tujuan
- Rata-rata Task time 67 detik – dianggap baik karena kurang dari batas (70 detik)
- Keduanya tidak perlihatkan keberhasilan tiap partisipan mencapai tujuan : minimal 80% tugas selesai dalam waktu kurang dari 70 detik
- Goal met/ketercapaian tujuan perlihatkan persentase partisipan yang sukses 38%

30/70

30

## COMBINING METRICS : PERCENTAGES

- Saat TIDAK ADA TARGET GOAL : Tiap skala/ukuran diubah dalam persentase, lalu hitung rata-rata
- Ukuran jumlah completed task dan subjective rating **nilai maksimumnya** adalah jumlah total tugas dan rating tertinggi
- Ukuran time : **nilai terbaiknya** adalah waktu tercepat dan nilai lain dibandingkan dengan waktu tercepat.

31/70

31

## Data Sampel Tes Usability 10 Partisipan

Participant #	Time Per Task (sec)	Tasks Completed (of 15)	Rating (0–4)
1	65	7	2.4
2	50	9	2.6
3	34	13	3.1
4	70	6	1.7
5	28	11	3.2
6	52	9	3.3
7	58	8	2.5
8	60	7	1.4
9	25	9	3.8
10	55	10	3.6

Table 8.2 Sample data from a usability test with 10 participants<sup>a</sup>.

<sup>a</sup>Time per task is the average time to complete each task, in seconds. Tasks completed are number of tasks (out of 15) that the user completed successfully. Rating is the average of a five-point task ease rating for each task, where higher is better.

32/70

32

## Contoh Perhitungan

- Partisipan #1
  - Time per task (sec) :  $(70-65)/(70-25) \times 100\% = 11\%$
  - Task Completed (of 15) :  $7/15 \times 100\% = 47\%$
  - Rating :  $2,4/4 \times 100\% = 60\%$
- Dimana ..
  - Time per task (sec) : 70 waktu terlama, 25 waktu tercepat dan 65 waktu yang diperoleh
  - Task completed : 15 adalah jumlah total tugas
  - Rating : 4 adalah nilai maksimum rating

33/70

33

### Tabel Transformasi Data ke Persentase

Participant #	Time	Tasks	Rating	Average
1	11%	47%	60%	39%
2	44%	60%	65%	56%
3	80%	87%	78%	81%
4	0%	40%	43%	28%
5	93%	73%	80%	82%
6	40%	60%	83%	61%
7	27%	53%	63%	48%
8	22%	47%	35%	35%
9	100%	60%	95%	85%
10	33%	67%	90%	63%

Table 8.3 Data from Table 8.2 transformed to percentages<sup>a</sup>.

<sup>a</sup>For task completion data, the score was divided by 15. For rating data, the score was divided by 4. For time data, the difference between the longest time (70) and the observed time was divided by the difference between longest (70) and shortest (25) times.

34/70

34

### Hasil

- Terendah 28%, tertinggi 85%
- Rata-rata keseluruhan 57.8% atau 58%
- Nilai 58% tidak mewakili 10 partisipan, karena itu gunakan 90% confidence interval :  $\pm 11\%$  sehingga confidence interval dari 47% sampai 69%
- 47% -----(-11)----- 58% -----(+11)----69%

		D7		f(x)	=CONFIDENCE(0.1,STDEV(B2:B11),COUNT(B2:B11))		
A	B	C	D	E	F	G	H
1 Participant	Average						
2	1	39					
3	2	56					
4	3	81					
5	4	28					
6	5	82					
7	6	61			90% confidence interval		
8	7	48			10.63385		
9	8	35					
10	9	85					
11	10	63					

35/70

35

## Perhitungan dengan Bobot

- Ukuran diberi bobot lebih karena dianggap penting dan menjadi pembeda
- Contoh :
  - Bobot time dan task (performa) > rating  
→ sasaran bisnis : kecepatan dan akurasi penggunaan produk
  - Bobot Rating (berasal dari Self-reported metrics) > performa  
→ sasaran bisnis adalah persepsi pengguna
- Bobot diselaraskan dengan business goal (sasaran bisnis) untuk produk
- Hasil cenderung mendekati bobot terbesar

**36/70**

36

## Dengan Bobot untuk Ukuran

Participant #	Time	Weight	Tasks	Weight	Rating	Weight	Weighted Average
1	38%	1	47%	1	60%	2	51%
2	50%	1	60%	1	65%	2	60%
3	74%	1	87%	1	78%	2	79%
4	36%	1	40%	1	43%	2	40%
5	89%	1	73%	1	80%	2	81%
6	48%	1	60%	1	83%	2	68%
7	43%	1	53%	1	63%	2	55%
8	42%	1	47%	1	35%	2	40%
9	100%	1	60%	1	95%	2	88%
10	45%	1	67%	1	90%	2	73%

Table 8.4 Calculation of weighted averages<sup>a</sup>.

<sup>a</sup>Each individual percentage is multiplied by its associated weight, these products are summed, and that sum is divided by the sum of the weights (4, in this example).

**37/70**

37

## COMBINING METRICS : PERCENTAGES (2)

Participant #	Tasks Completed (of 10)	# of Page Visits (min = 20)	Satisfaction Rating (0–6)	Usefulness Rating (0–6)	Tasks	Page Visits	Satisfaction	Usefulness	Average
1	8	32	4.7	3.9	80%	63%	78%	65%	71%
2	6	41	4.1	3.8	60%	49%	68%	63%	60%
3	7	51	3.4	3.7	70%	39%	57%	62%	57%
4	5	62	2.4	2.3	50%	32%	40%	38%	40%
5	9	31	5.2	4.2	90%	65%	87%	70%	78%
6	5	59	2.7	2.9	50%	34%	45%	48%	44%
7	10	24	5.1	4.8	100%	83%	85%	80%	87%
8	8	37	4.9	4.3	80%	54%	82%	72%	72%
9	7	65	3.1	2.5	70%	31%	52%	42%	49%

Table 8.5 Sample data from a usability test with nine participants<sup>3</sup>.

<sup>3</sup>Tasks completed are the number of tasks (out of 10) that the user completed successfully. Number of page visits is the total number of web pages that the user visited in attempting the tasks. (Typically, each revisit to the same page is counted as another visit.) The two ratings are average subjective ratings of satisfaction and usefulness, each on a seven-point scale (0–6).

38/70

38

## Perhitungan

- Task completed : total 10 tugas → nilai max 10 = 100%
- # Page visits : jumlah halaman yang dikunjungi untuk selesaikan tugas. Jumlah minimal : 20 (nilai terbaik) →  $20/\text{jumlah\_halaman} * 100\%$
- Subjective rating : satisfaction dan usefulness dari self-reported metrics, max : 6 → rating/6\*100%
- Average = jumlah persentase ukuran /4

39/70

39

## COMBINING METRICS : PERCENTAGES (3)

Participant #	Tasks Completed (of 10)	# of Errors	Satisfaction Rating (0–6)	Tasks	Accuracy	Satisfaction	Average
1	8	2	4.7	80%	60%	78%	73%
2	6	4	4.1	60%	20%	68%	49%
3	7	0	3.4	70%	100%	57%	76%
4	5	5	2.4	50%	0%	40%	30%
5	9	2	5.2	90%	60%	87%	79%
6	5	4	2.7	50%	20%	45%	38%
7	10	1	5.1	100%	80%	85%	88%
8	8	1	4.9	80%	80%	82%	81%
9	7	3	3.1	70%	40%	52%	54%
10	9	2	4.2	90%	60%	70%	73%
11	7	1	4.5	70%	80%	75%	75%
12	8	3	5.0	80%	40%	83%	68%

Table 8.6 Sample data from a usability test with 12 participants<sup>a</sup>.<sup>a</sup>Tasks completed are the number of tasks (out of 10) that the user completed successfully. Number of errors is the number of specific errors that the user made, such as data-entry errors. Satisfaction rating is on a scale of 0 to 6.

40/70

40

## Perhitungan

- Task completed : total 10 tugas → nilai max 10 = 100%
- #Error: jumlah error yang dilakukan, min =0  
 $\rightarrow (1 - (\text{jml\_error}/\text{jml\_error\_max})) * 100\%$
- Jml\_error\_max diperoleh dari data
- Subjective rating : satisfaction self-reported metrics, max : 6 → rating/6\*100%
- Average = jumlah persentase ukuran /3

41/70

41

## ATURAN

Tentukan nilai min –max yang mungkin terjadi

- Jika  $\min=0, \max=100 \rightarrow 0\% - 100\%$
- Jika  $\min=0, \max=\text{nilai\_max} \rightarrow \text{data}/\text{nilai\_max}*100\%$
- Jika  $\min=0, \max=? \rightarrow \text{nilai max} = \text{nilai data tertinggi}$
- Jika  $\min=?, \max=? \rightarrow (\text{data}_\text{max}-\text{data})/(\text{data}_\text{max}-\text{data}_\text{min})*100\%$

42/70

42

## DATA OUTLIER

Jika ada data yang mencolok perbedaannya

- Gunakan nilai mean dan standar deviasi
- Jika nilai data di luar 2 atau 3 kali std deviasi dari mean, maka nilai itu dianggap outlier, lalu tidak dianggap

43/70

43

## COMBINING METRICS : Z-SCORES

- Ubah nilai ke persentase gunakan z-score atau fungsi STANDARDIZE pada excel
- =STANDARDIZE(A1, AVERAGE(A:A), STDEV(A:A)) dengan data di kolom A.
- Tidak ada asumsi nilai Min dan Max dari data

	A	B	C	D	E	F	G
1	Time/Task		Task Complete		Rating		
2		65	0.98	7	-0.91	2.4	-0.46
3		50	0.02	9	0.05	2.6	-0.20
4		34	-1.01	13	1.97	3.1	0.43
5		70	1.30	6	-1.39	1.7	-1.35
6		28	-1.39	11	1.01	3.2	0.56
7		52	0.15	9	0.05	3.3	0.69
8		58	0.53	8	-0.43	2.5	-0.33
9		60	0.66	7	-0.91	1.4	-1.73
10		25	-1.59	9	0.05	3.8	1.32
11		55	0.34	10	0.53	3.6	1.07
12	Average	49.70		8.90		2.76	44/70
13	STD	15.57		2.08		0.79	

44

## Aturan

- Z score haruskan nilai ke arah yang sama, misal NILAI TERTINGGI ADALAH TERBAIK
- Untuk ubah arah kalikan data dengan -1
- Hasil z scores dan percentage dengan asumsi hampir sama.

45/70

45

## Aturan

- Z scores digunakan untuk bandingkan satu data set ke data set lain. Misalnya:
  - data usability test dari produk yang beda versi
  - data usability test yang sama dari beberapa grup pengguna
  - data usability test yang sama dari beberapa kondisi atau desain yang berbeda
- Sampel yang layak, misalnya 10 orang per kondisi

46/70

46

## Tabel Data – Z Scores

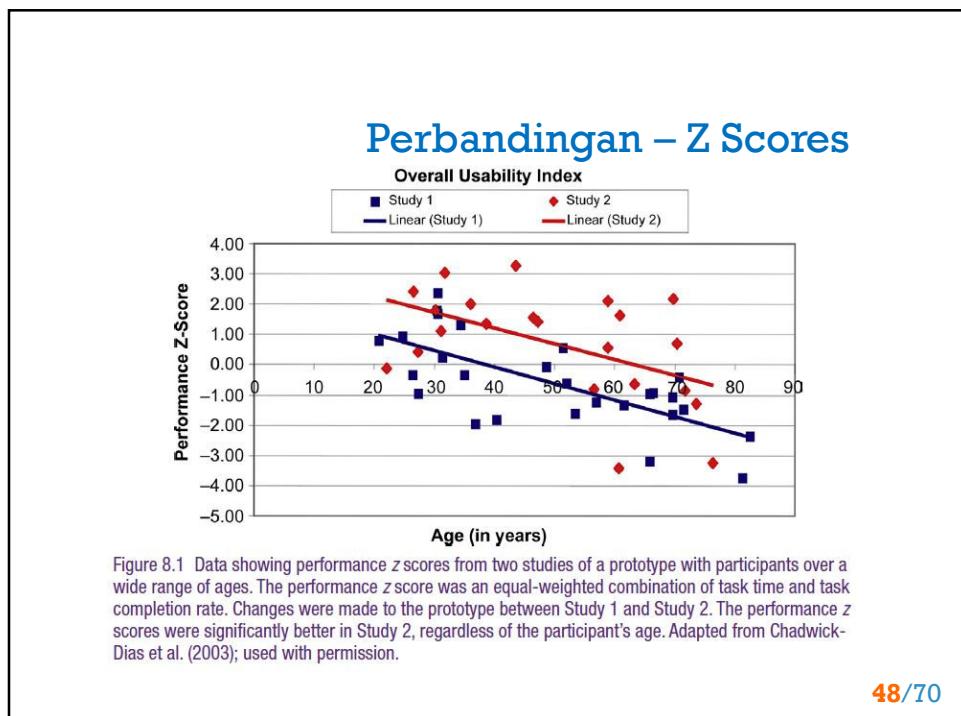
Participant #	Time Per Task (sec)	Tasks Completed (of 15)	Rating (0–4)	<i>z</i> Time*	<i>z</i> Time*(-1)	<i>z</i> Tasks	<i>z</i> Rating	Average
1	65	7	2.4	0.98	-0.98	-0.91	-0.46	-0.78
2	50	9	2.6	0.02	-0.02	0.05	-0.20	-0.06
3	34	13	3.1	-1.01	1.01	1.97	0.43	1.14
4	70	6	1.7	1.30	-1.30	-1.39	-1.35	-1.35
5	28	11	3.2	-1.39	1.39	1.01	0.56	0.99
6	52	9	3.3	0.15	-0.15	0.05	0.69	0.20
7	58	8	2.5	0.53	-0.53	-0.43	-0.33	-0.43
8	60	7	1.4	0.66	-0.66	-0.91	-1.73	-1.10
9	25	9	3.8	-1.59	1.59	0.05	1.32	0.98
10	55	10	3.6	0.34	-0.34	0.53	1.07	0.42
Mean				0.0	0.0	0.0	0.00	0.00
Standard deviation				1.0	1.0	1.0	1.00	0.90

Table 8.7 Sample data from Table 8.2 transformed using z scores\*.

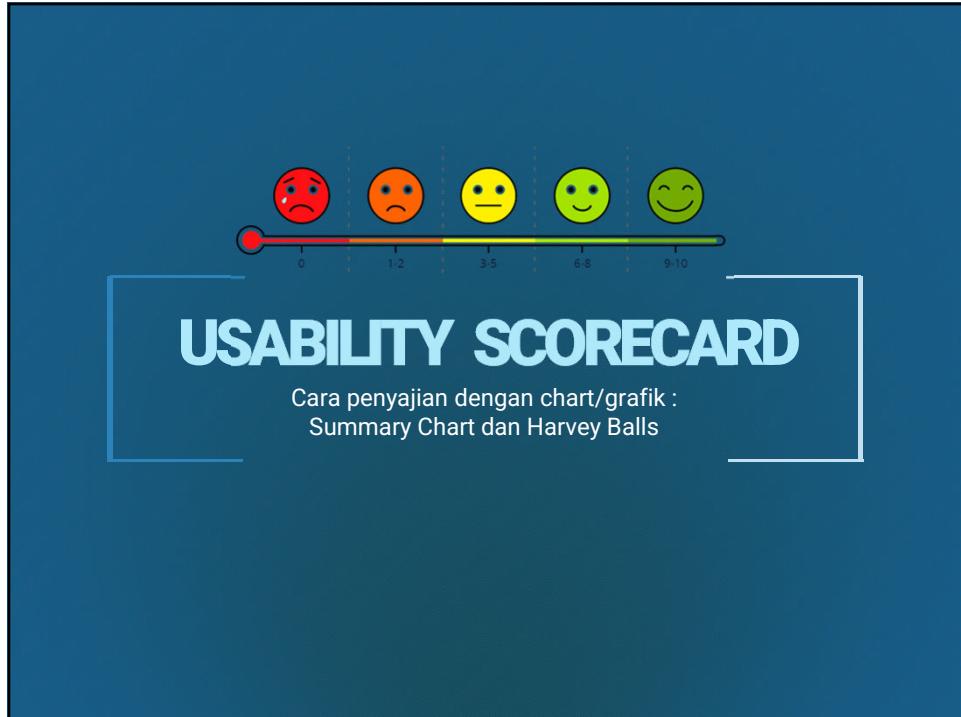
\*For each original score, the z score was determined by subtracting the mean of the score's distribution from it and then dividing by the standard deviation. This z score tells you how many standard deviations above or below the mean that score is. Since you need all the scales to have higher numbers better, the scale of the z scores of times is reversed by multiplying by (-1).

47/70

47



48



49

## Summary Chart

- Penyajian hasil pengukuran lewat Summary Chart
- Sasaran : sajikan data dari usability test sehingga trend dan aspek penting secara keseluruhan terdeteksi
- Contoh :
  - Tugas yang bermasalah untuk pengguna
  - Dua ukuran : task completion dan task ease rating

50/70

50

## SUMMARY CHART

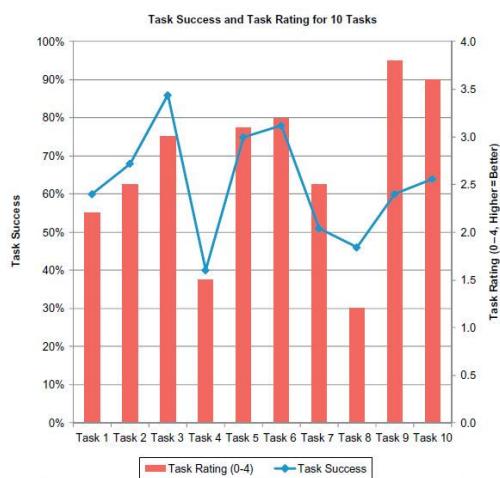


Figure 8.3 A sample combination column and line chart for 10 tasks. Task rating is shown via the columns and labeled on the right axis. Task success is shown via the lines and is labeled on the left axis.

51/70

51

## Penjelasan SUMMARY CHART

- Task 4 dan 8 : tugas yang paling bermasalah
- Ada perbedaan signifikan antar dua ukuran pada Task 9 dan 10 : pengguna kira mereka selesaikan tugas
- Tugas dengan nilai tinggi : Task 3,5 dan 6

52/70

52

## Summary Chart

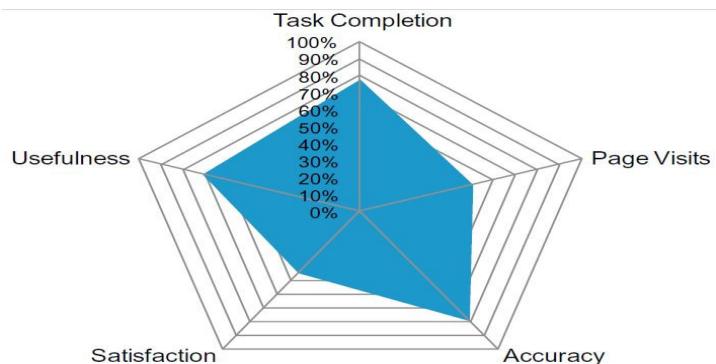


Figure 8.4 A sample radar chart summarizing task completion, page visits, accuracy (lack of errors), satisfaction rating, and usefulness rating from a usability test. Each has been transformed to a

53/70

53

## Penjelasan Radar Chart

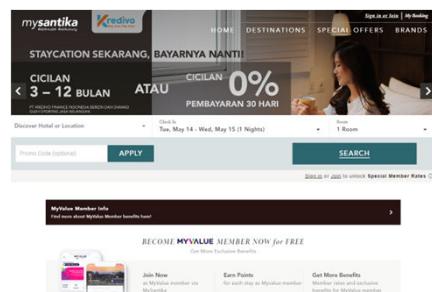
- Lima ukuran : Task completion, page visits, accuracy (lack of error), satisfaction, dan usefulness rating
- Walaupun task completion, accuracy, dan usefulness rating relatif tinggi, page visits dan satisfaction relatif rendah.

54/70

54

## Harvey Balls

- Dipopulerkan oleh Consumer Reports
- Presentasikan summary data untuk 3 atau lebih ukuran, tapi juga informasikan task-level information
- Contoh : evaluasi website hotel



55/70

55

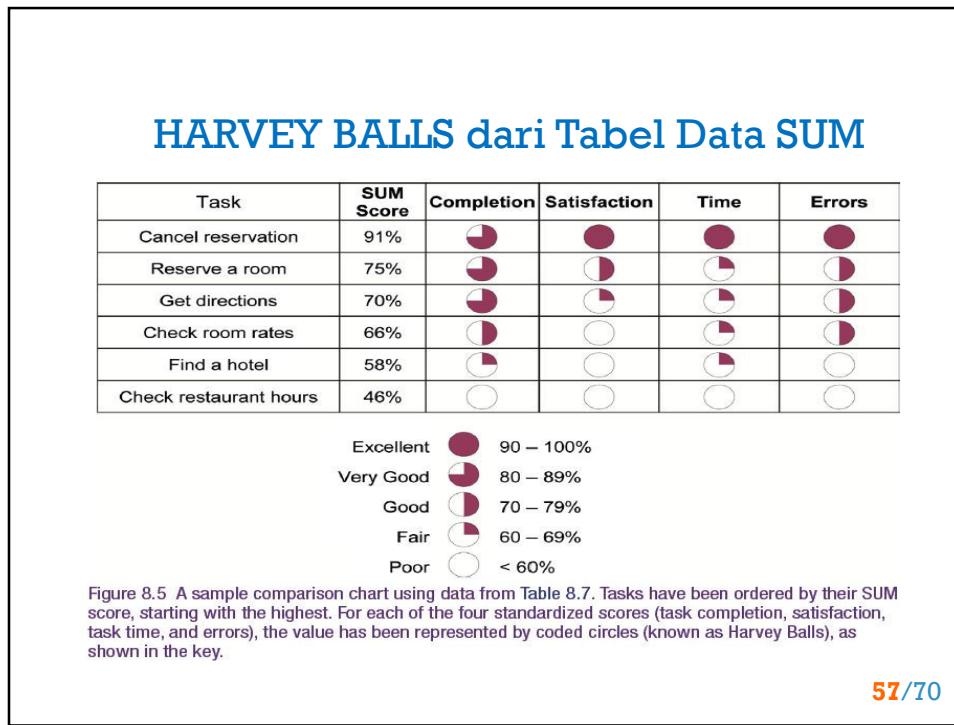
Task	SUM						
	Low	Mean	High	Completion	Satisfaction	Time	Errors
Reserve a room	62%	75%	77%	81%	74%	68%	76%
Find a hotel	38%	58%	61%	66%	45%	63%	59%
Check room rates	49%	66%	69%	74%	53%	63%	74%
Cancel reservation	89%	91%	99%	86%	91%	95%	92%
Check restaurant hours	22%	46%	68%	58%	45%	39%	43%
Get directions	56%	70%	73%	81%	62%	66%	71%
Overall SUM	53%	68%	68%				

Table 8.8 Sample standardized data from a usability test\*.  
\*After entering data for each participant and each task, these are the standard scores, overall SUM score and a confidence interval for it.

DIURUTKAN DARI MEAN YANG TERTINGGI

56/70

56



57

## COMPARISON TO GOAL & EXPERT PERFORMANCE

Cara untuk menyimpulkan data usability dengan membandingkannya terhadap external standard



58

### COMPARISON TO GOALS

- Hasil usability test dibandingkan dengan GOAL (sasaran, tujuan) yang ditetapkan sebelum tes
- Goal dapat ditetapkan untuk tiap task level( tiap ukuran) atau keseluruhan level

59/70

59

## CONTOH : TASK SPECIFIC GOAL

- Paling tidak 90% pengguna yang mewakili berhasil pesan kamar hotel yang sesuai
- Membuka akun online baru rata-rata kurang dari 8 menit
- Paling tidak 95% dari pengguna baru mampu beli produk online setelah memilih barang dalam waktu 5 menit

60/70

60

## CONTOH : OVERALL GOAL

- Pengguna berhasil menyelesaikan paling tidak 90% dari total jumlah tugas
- Pengguna berhasil selesaikan tugas-tugas dalam rata-rata 3 menit per tugas
- Pengguna menilai aplikasi rata-rata 80% menggunakan SUS rating

61/70

61

## Goals

- Usability goals diterapkan untuk task completion, time, accuracy dan satisfaction
- Goals harus dapat diukur
- Data yang ditetapkan harus mendukung pencapaian goal

62/70

62

## Sample Data of 8 tasks Target #page visits & Mean actual # page visits

	Target # of Page Visits	Actual # of Page Visits
Task 1	5	7.9
Task 2	8	9.3
Task 3	3	7.3
Task 4	10	11.5
Task 5	4	7
Task 6	6	6.9
Task 7	9	9.8
Task 8	7	10.2

Table 8.9 Sample data from eight tasks showing target number of page visits and mean of actual number of page visits.

63/70

63

### CHART ACTUAL VS TARGET

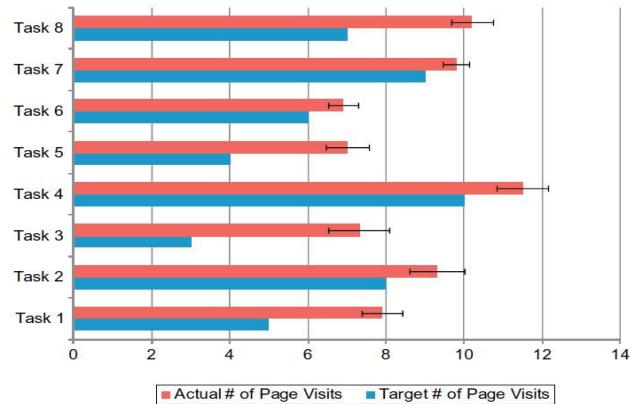


Figure 8.6 Target and actual number of page visits for each of eight tasks. Error bars represent the 90% confidence interval for the actual number of page visits.

**64/70**

64

### Page Visit Efficiency

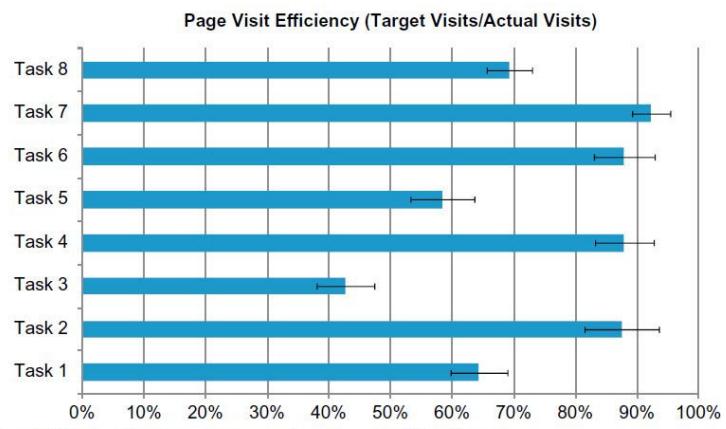


Figure 8.7 Ratio of target to actual page views for each of the eight tasks.

**65/70**

65

## COMPARISON TO EXPERT PERFORMANCE

- Bandingkan hasil usability test dengan performa ahli (expert) terhadap goal/sasaran/tujuan yang ditetapkan
- Beberapa ahli –Subject Matter Expert, menjalani usability test dan hasilnya, dalam bentuk rata-rata, digunakan untuk perbandingan

66/70

66

## COMPARISON TO EXPERT PERFORMANCE

- Bermanfaat untuk mengukur tingkat kesulitan beberapa tugas
- Tujuan perbandingan : melihat kedekatan hasil test antara pengguna non expert dan pengguna expert
- Biasanya digunakan pada pengukuran waktu

67/70

67

### Sample Data Actual vs Expert Time

Task	Actual Time	Expert Time	Expert/Actual
1	124	85	69%
2	101	50	50%
3	89	70	79%
4	184	97	53%
5	64	40	63%
6	215	140	65%
7	70	47	67%
8	143	92	64%
9	108	98	91%
10	92	60	65%

Table 8.10 Sample time data from 10 tasks in a usability test showing average actual time per task (in seconds), expert time per task, and ratio of expert to actual time.

68/70

68

### Ratio Chart

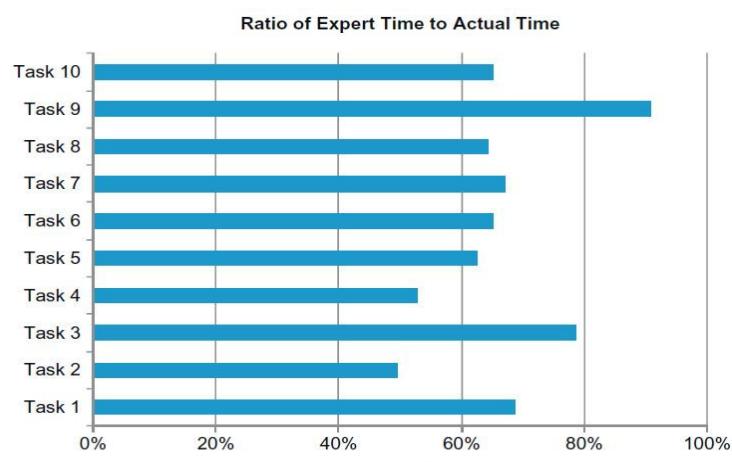


Figure 8.8 Graph of the ratio of expert to actual times from Table 8.10.

69/70

69



70