



# Tips and Tricks for Measuring the User Experience

---

**Tom Tullis and Bill Albert  
Fidelity Investments**

**Usability and User Experience 2008**

**UPA-Boston's Seventh Annual Mini UPA Conference**

**May 28, 2008**



# Who we are

---

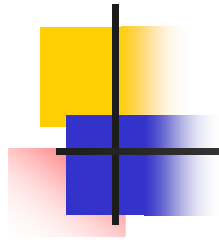
- Tom Tullis  
Senior Vice President, User Insight, Fidelity Investments  
[Tom.Tullis@fmr.com](mailto:Tom.Tullis@fmr.com)
- Bill Albert  
Director, User Insight, Fidelity Investments  
[Bill.Albert@fmr.com](mailto:Bill.Albert@fmr.com)



# Our Top 10 List

---

1. Know your data
2. Show your confidence (intervals)
3. Deal with binary success data (appropriately)
4. Compare means
5. Consider using expectation measures
6. Use the System Usability Scale (SUS)
7. Show frequency distributions
8. Combine different metrics
9. Use appropriate tools
10. Present data appropriately



Tip 1: Know your data  
(and what you can & can't do with it)



# Independent & Dependent Variables

---

- Independent variables: The things you manipulate or control for, e.g.,
  - Characteristics of participants (e.g., age, sex, relevant experience)
  - Different designs or prototypes being tested
  - Tasks
- Dependent variables: The things you measure, e.g.,
  - Task Success
  - Task Time
  - SUS score
  - etc.



# Types of Data

---

- Nominal (aka Categorical)
  - e.g., Male, Female; Design A, Design B.
- Ordinal
  - e.g., Rank ordering of 4 designs tested from Most Visually Appealing to Least Visually Appealing.
- Interval
  - e.g., 7-point scale of agreement: “This design is visually appealing. Strongly Disagree . . . Strongly Agree”
- Ratio
  - e.g., Time, Task Success %



# Ordinal vs. Interval Rating Scales

- Are these two scales different?

Poor    Fair    Good    Excellent  
Poor                  Excellent

Top scale is ordinal. You should only calculate frequencies of each response.

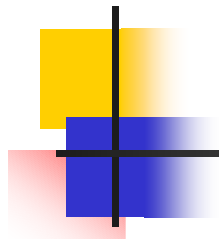
Bottom scale can be considered interval. You can also calculate means.



# Statistics for each Data Type

**Table 2.3** Choosing the Right Statistics for Different Data Types and Usability Metrics

Data Type	Common Metrics	Statistical Procedures
Nominal (categories)	Task success (binary), errors (binary), top-2 box scores	Frequencies, crosstabs, Chi-square
Ordinal (ranks)	Severity ratings, rankings (designs)	Frequencies, crosstabs, chi-square, Wilcoxon rank sum tests, Spearman rank correlation
Interval	Likert scale data, SUS scores	All descriptive statistics, <i>t</i> -tests, ANOVAs, correlation, regression analysis
Ratio	Completion time, time (visual attention), average task success (aggregated)	All descriptive statistics (including geometric means), <i>t</i> -tests, ANOVAs, correlation, regression analysis



Tip 2: Show your confidence  
(intervals)



# Confidence Intervals

---

- Assume this was your time data for a study with 5 participants:

	<b>Task Time (secs)</b>
P1	43
P2	53
P3	59
P4	68
P5	49

Does that make a difference in your answer?

# Calculating Confidence Intervals

=CONFIDENCE(<alpha>,<std dev>,<n>)

	Task Time (secs)
P1	43
P2	53
P3	59
P4	68
P5	49
<b>Mean:</b>	<b>54.4</b>
<b>Std Dev:</b>	<b>9.6</b>
<b>N:</b>	<b>5</b>
<b>95% Confidence Interval:</b>	<b>8.4</b>

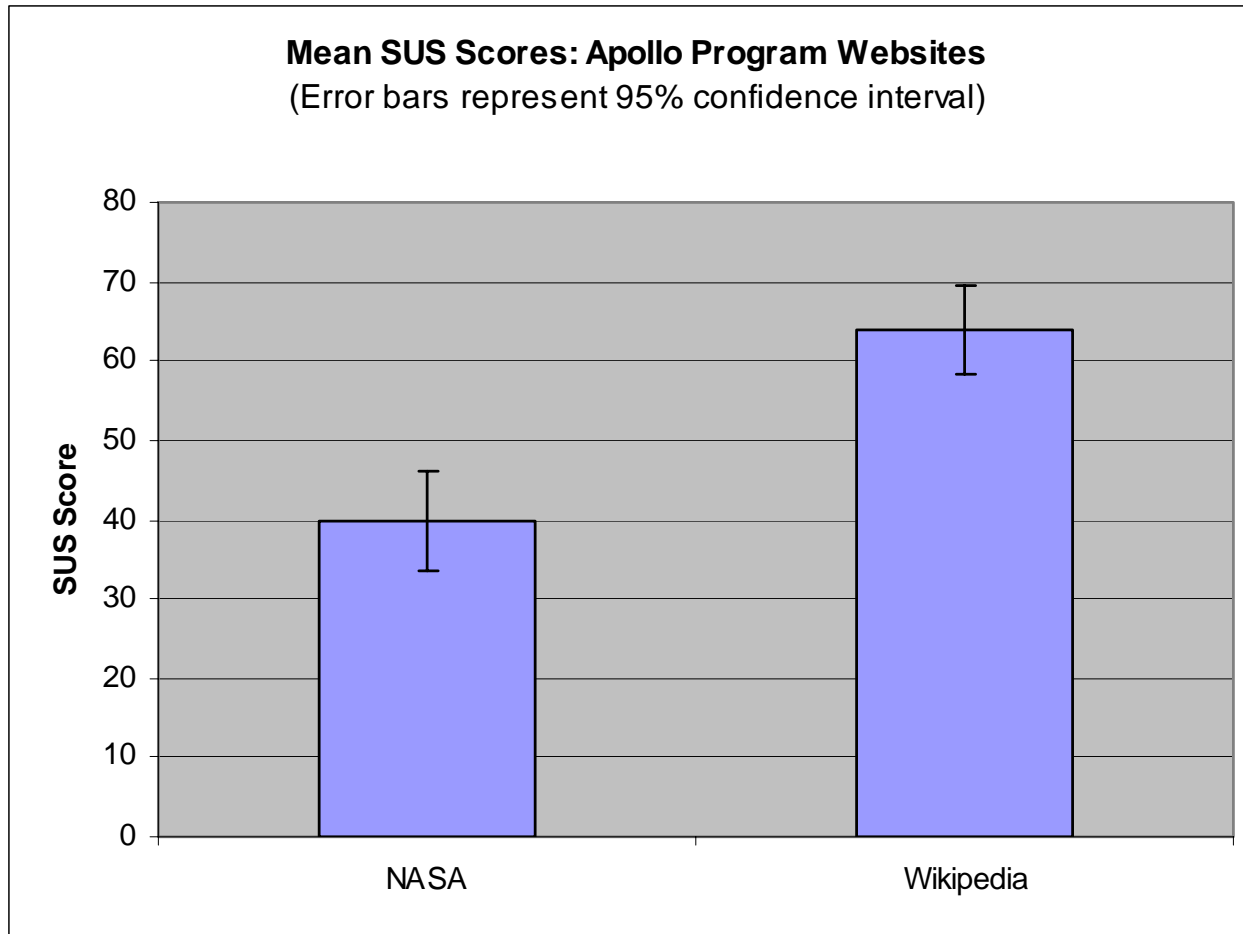
<alpha> is normally .05 (for a 95% confidence interval)

<std dev> is the standard deviation of the set of numbers (9.6 in this example)

<n> is how many numbers are in the set (5 in this example)

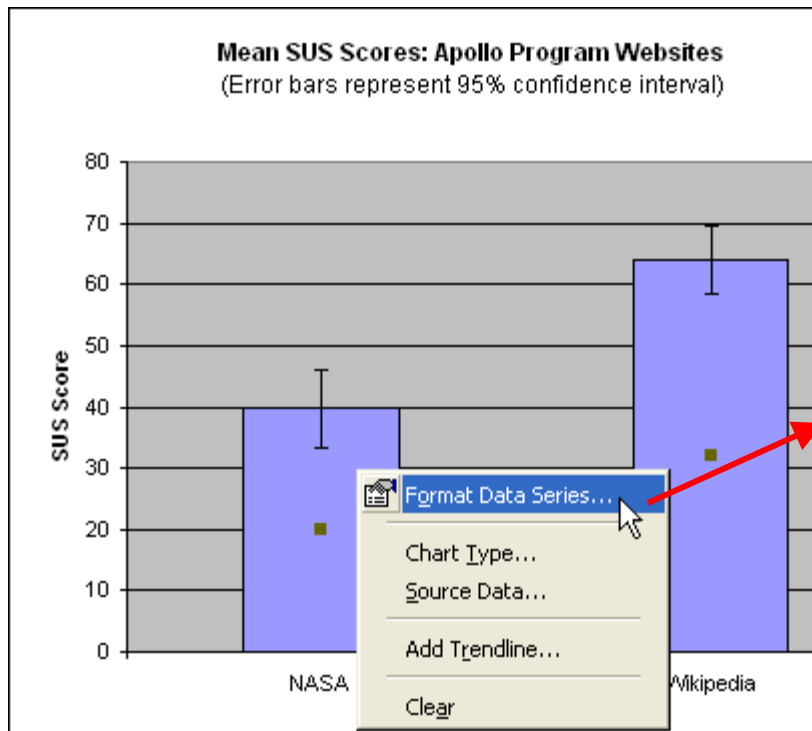
Excel example

# Show Error Bars



[Excel example](#)

# How to Show Error Bars



**Format Data Series**

Patterns | Axis | **Y Error Bars** | Data Labels | Series Order | Options

Display

Both  Plus  Minus  None

Error amount

Fixed value: 10

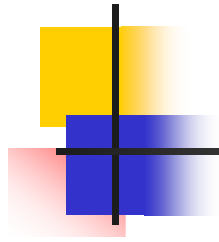
Percentage: 5 %

Standard deviation(s): 1

Standard error

Custom: + =Sheet1!\$E\$5:\$F\$5  
- =Sheet1!\$E\$5:\$F\$5

OK Cancel



Tip 3: Deal with Binary Success  
Data (appropriately)



# Binary Success

- Pass/fail (or other binary criteria)
- 1's (success) and 0's (failure)

Participant	Task 1	Task 2	Task 3	Task 4	Task 5
P1	1	0	1	0	0
P2	1	0	1	0	1
P3	1	1	1	1	1
P4	1	1	1	1	1
P5	0	0	1	1	1
P6	1	0	0	1	1
P7	0	1	1	1	1
P8	0	0	1	1	0
P9	1	0	1	0	1
P10	1	1	1	1	1
P11	0	1	1	1	1
P12	1	0	1	1	1
<b>Average</b>	<b>67%</b>	<b>42%</b>	<b>92%</b>	<b>75%</b>	<b>83%</b>
<b>Confidence</b>					
<b>Interval (95%)</b>	<b>28%</b>	<b>22%</b>	<b>29%</b>	<b>29%</b>	<b>29%</b>



# Confidence Interval for Task Success

---

- When you look at task success data across participants for a single task the data is commonly binary:
  - Each participant either passed or failed on the task.
- In this situation, you need to calculate the confidence interval using the binomial distribution.

# Example

	Task 1
P1	1
P2	1
P3	1
P4	0
P5	1
P6	1
P7	0
P8	1

1=success,  
0=failure.  
So, 6/8  
succeeded,  
or 75%.

Easiest way to calculate confidence interval is using Jeff Sauro's web calculator:  
<http://www.measuringusability.com/wald.htm>

**Measuring Usability**  
Quantitative Usability, Statistics & Six Sigma by Jeff Sauro

Home Usability Statistics Six Sigma

### Confidence Interval Calculator for a Completion Rate

by Jeff Sauro | October 1, 2005 :: [2 Related Articles](#):: [2 Related Questions](#)

Use this calculator to calculate a confidence interval and best point estimate for an observed completion rate. This calculator provides the [Adjusted Wald](#), [Exact](#), [Score](#) and [Wald](#) intervals.

Input Table		Results Table				
Passed	Total Tested	Confidence Intervals			Point Estimates	
6	8	Low	High	Margin of Error <sup>*</sup>	Best Estimate	
Calculate		Adj. Wald	0.4009	0.9369	0.2680	0.7000
Confidence Level: 95%		Exact	0.3491	0.9681	0.3095	MLE 0.7500
Likely Population Completion Rate		Score	0.4093	0.9285	0.2596	LaPlace 0.7000
Unknown		Wald	0.4499	1.0501	0.3001	Jeffrey's 0.7222
					Using Alpha: .05	Wilson 0.6689

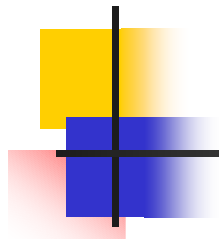
# Chi-square

- Allows you to compare actual and expected frequencies for categorical data.

`=CHITEST(<actual range>, <expected range>)`

	A	B	C	D	E
1					
2		Task 1	Task 2	Sum	
3	Expected	8	8	16	
4	Observed	12	4	16	
5					
6	Chi-square	0.0455			
7					

Excel example:



## Tip 4: Compare Means



# Comparing Means

---

- T-test:
  - Independent samples (between subjects)
  - Paired samples (within subjects)

Excel example:

Apollo websites,  
task times

Excel example:

Haptic mouse  
study

# T-tests in Excel

`=TTEST(<array1>,<array2>,x,y)`

Independent Samples:

	A	B	C	D	E	F	G
1	Task Times (mins)	NASA	Wikipedia			NASA	Wikipedia
2		1.0	1.2		Means:	2.32	1.83
3		2.8	1.2				
4		0.7	2.2		t-test:	0.018133	
5		1.9	1.5				
6		2.8	1.5				
7		5.1	1.0				
8		1.9	1.9				
9		6.7	2.2				

Paired Samples:

	A	B	C	D	E
1	Accuracy:	haptic	non-haptic	t-test:	
2	P1	67%	33%	0.03	
3	P2	50%	17%		
4	P3	83%	83%		
5	P4	100%	67%		
6	P5	67%	0%		
7	Means:	73%	40%		
8					
9					

$x = 2$  (for two-tailed test) in almost all cases

$y = 2$  (independent samples)

$y = 1$  (paired samples)

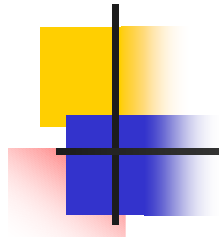
# Comparing Multiple Means

- Analysis of Variance (ANOVA)

	A	B	C	D	E	F	G
1	Anova: Single Factor						
2							
3	SUMMARY						
4	<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>		
5	Nav 1	214	162.4	75.9%	0.04		
6	Nav 2	213	157.15	73.8%	0.06		
7	Nav 3	232	162.75	70.2%	0.05		
8	Nav 4	231	172.25	74.6%	0.06		
9							
10							
11	ANOVA						
12	<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
13	Between Groups	0.41	3	0.14	2.56	<b>0.05</b>	2.61
14	Within Groups	47.33	886	0.05			
15							
16	Total	47.74	889				
17							

“Tools” > “Data Analysis” > “Anova: Single Factor”

Excel example: Study comparing 4 navigation approaches for a website



## Tip 5: Consider Using Expectation Measures



# Expectation Measure

---

- Based on “expectancy disconfirmation” theory from Market Research
- Basic approach:
  - Ask the users how easy/difficult they *think* each task is going to be, before doing them.
  - Ask them how easy/difficult it *actually was* immediately after each task.
  - Same 7-point scales in both cases:
    - Very Difficult to Very Easy

# Expectation Measure: Before/After

Before:

**Task 1. Find out how many people report to Jeremy George's manager. (Include Jeremy in this number.)**

How Difficult or Easy do you expect this task to be?

Very Difficult      Very Easy

**Task 2. You need to return a notebook to someone but only know that he works on the 4th floor of 245 Summer St (building code V, floor 4), and that his manager's name is Tom. Find this man and report his corpID.**

How Difficult or Easy do you expect this task to be?

Very Difficult      Very Easy

**Task 3. You remember talking with someone named John and want to contact him, but don't remember his last name. You only know that his last name starts with S. You also remember that he works at 500 Salem St. in Smithfield (Building code OS) and works in FISC. Find this man and report his corpID.**

How Difficult or Easy do you expect this task to be?

Very Difficult      Very Easy

**Task 4. You know a woman who works in FISC whose first name is Edyta and you need to call her. What are the last four digits of her phone number?**

How Difficult or Easy do you expect this task to be?

Very Difficult      Very Easy

After:

The screenshot shows a web browser window titled "Tasks1 - Microsoft Internet Explorer". The page content is as follows:

**Task 1 of 7:** Find out **how many people** report to Jeremy George's manager. (Include Jeremy in this number.)

Enter answer:

Overall, this task was:  
Very Difficult      Very Easy

NEXT Task >>



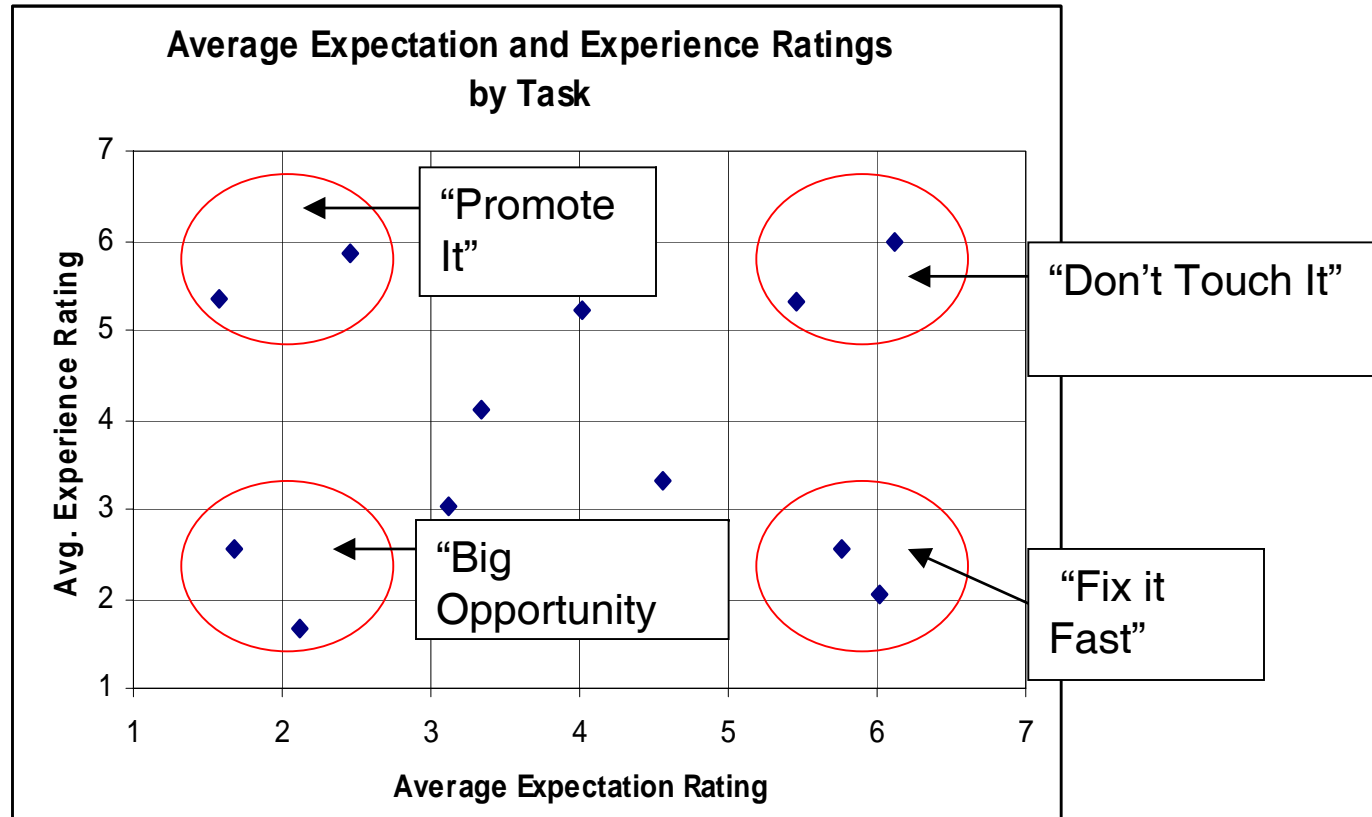
# Expectation Measure

---

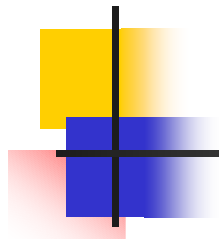
- For each task, calculate two averages:
  - Average expectation score
  - Average experience score
- Visualize the data as a scatterplot with two axes:
  - Expectation rating
  - Experience rating
  - The four quadrants are interesting

# Expectation Measure

1=Difficult  
...  
7=Easy



1=Difficult; 7=Easy



## Tip 6: Use the System Usability Scale (SUS)



# SUS

---

- Developed at Digital Equipment Corp.
- Consists of ten items.
- Adapted by replacing “system” with “website”.
- Each item is a statement (positive or negative) and a rating on a five-point scale of “Strongly Disagree” to “Strongly Agree” .

For details see

<http://www.usability.serco.com/trump/documents/Suschapt.doc>



	Strongly Disagree				Strongly Agree
1. I think I would like to use this website frequently.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. I found the website unnecessarily complex.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. I thought the website was easy to use.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. I think I would need Tech Support to be able to use this website.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. I found the various functions in this website were well integrated.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. I thought there was too much inconsistency in this website.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. I would imagine that most people would learn to use this website very quickly.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. I found the website very cumbersome to use.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9. I felt very confident using the website.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10. I need to learn a lot about this website before I could effectively use it.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



# SUS Scoring

---

- SUS yields a single number representing a composite measure of the overall usability of the system being studied. Note that scores for individual items are not meaningful on their own.
- To calculate the SUS score:
  - Each item's score contribution will range from 0 to 4.
  - For items 1,3,5,7,and 9 the score contribution is the scale position minus 1.
  - For items 2,4,6,8 and 10, the contribution is 5 minus the scale position.
  - Multiply the sum of the scores by 2.5 to obtain the overall SUS score.
- SUS scores have a range of 0 to 100.

[http://www.measuringux.com/SUS\\_Calculation.xls](http://www.measuringux.com/SUS_Calculation.xls)

# SUS Scoring Example

	Strongly disagree					Strongly agree	
1. I think that I would like to use this system frequently	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>		4
	1	2	3	4	5		
2. I found the system unnecessarily complex	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		1
	1	2	3	4	5		
3. I thought the system was easy to use	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		1
	1	2	3	4	5		
4. I think that I would need the support of a technical person to be able to use this system	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		4
	1	2	3	4	5		
5. I found the various functions in this system were well integrated	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		1
	1	2	3	4	5		
6. I thought there was too much inconsistency in this system	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2
	1	2	3	4	5		
7. I would imagine that most people would learn to use this system very quickly	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		1
	1	2	3	4	5		
8. I found the system very cumbersome to use	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		1
	1	2	3	4	5		
9. I felt very confident using the system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>		4
	1	2	3	4	5		
10. I needed to learn a lot of things before I could get going with this system	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		3
	1	2	3	4	5		

Total= 22      SUS Score= 22 \* 2.5 = 55

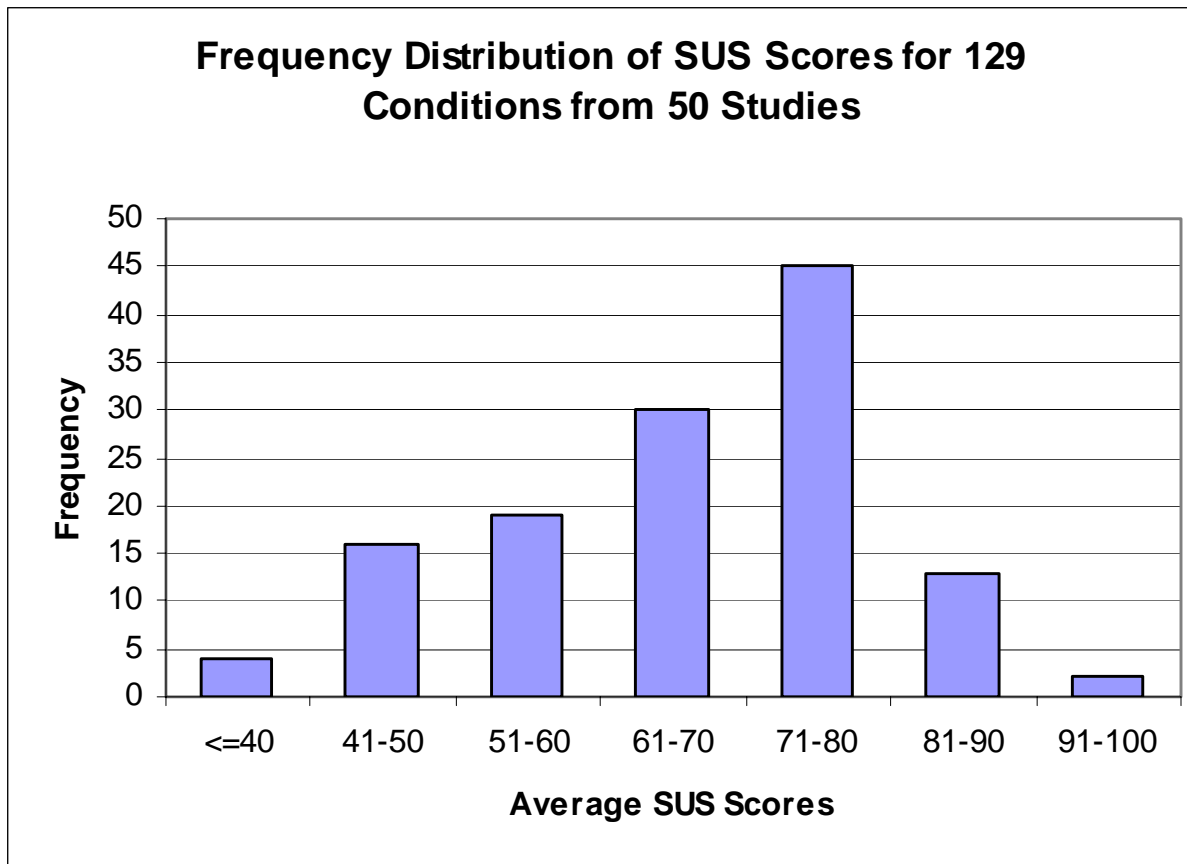


## SUS Usage

---

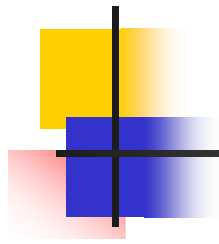
- “SUS has been made freely available for use in usability assessment, and has been used for a variety of research projects and industrial evaluations; the only prerequisite for its use is that any published report should acknowledge the source of the measure.”

# SUS Data from 50 Studies



Percentiles:	
10th	47.4
25th	56.7
50th	68.9
75th	76.7
90th	81.2
<b>Mean</b>	<b>66.4</b>

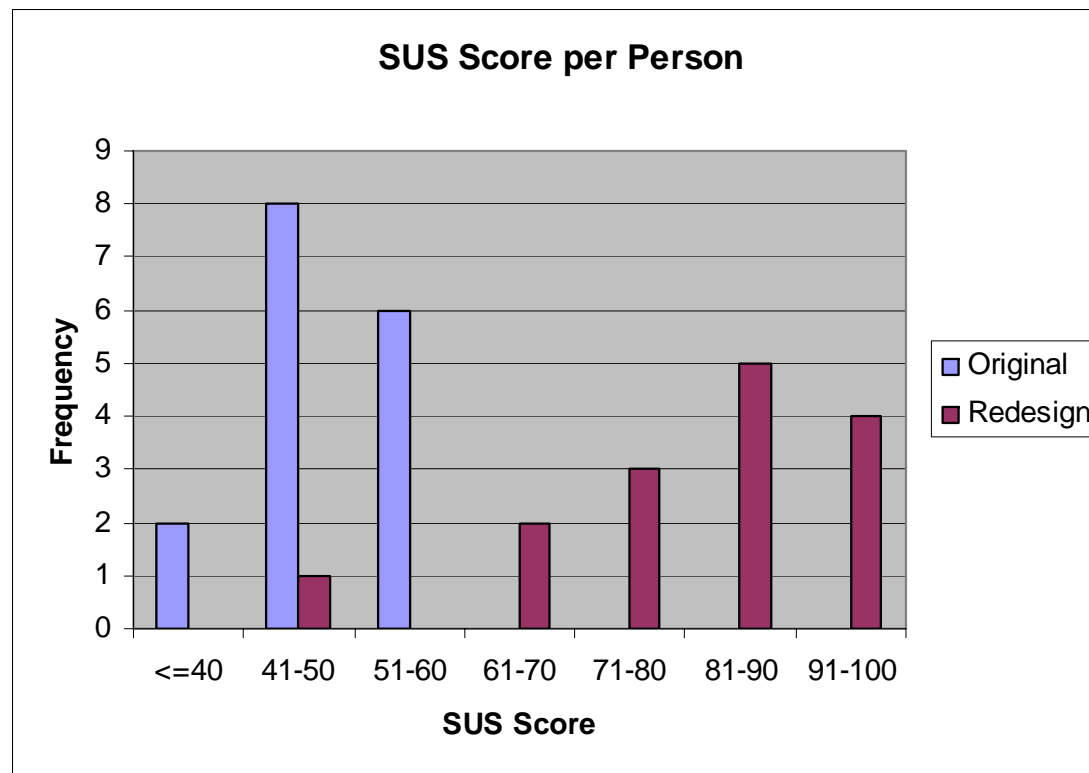
<http://www.measuringux.com/SUS-scores.xls>



## Tip 7: Show Frequency Distributions

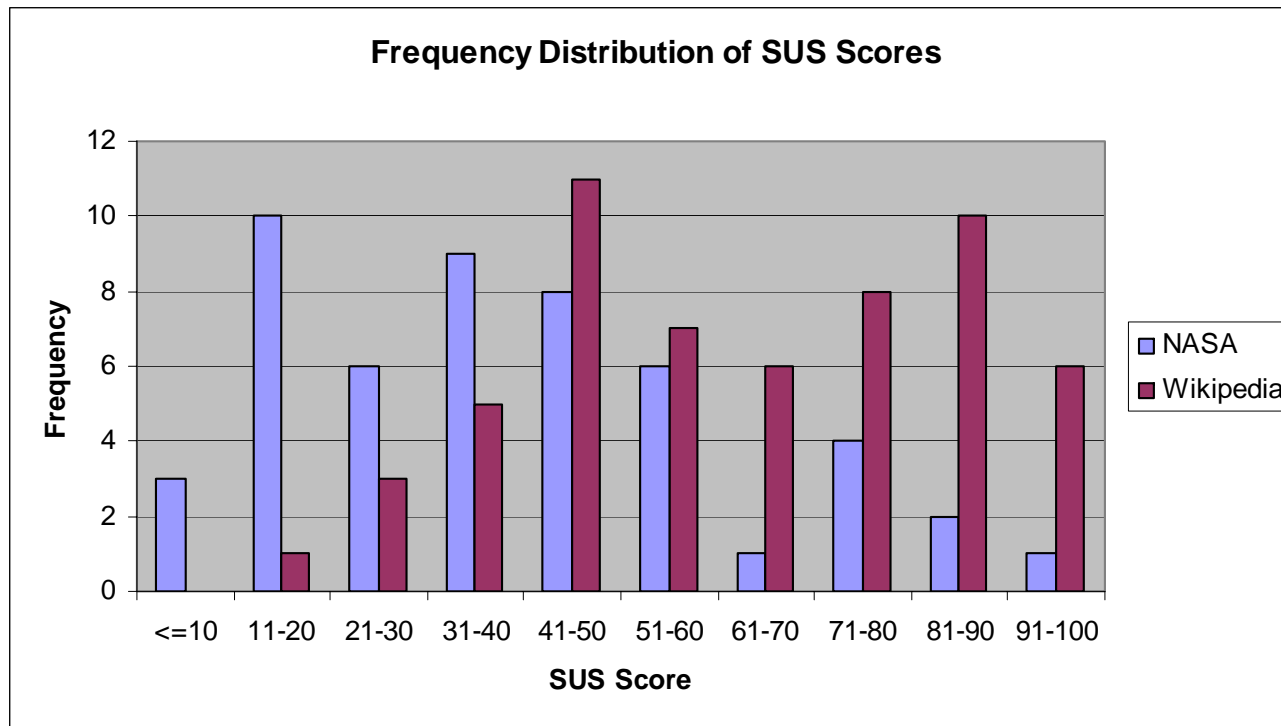
# Frequency Distributions

Frequency distributions are a useful way of visualizing variability if you have a reasonably large number of data points.



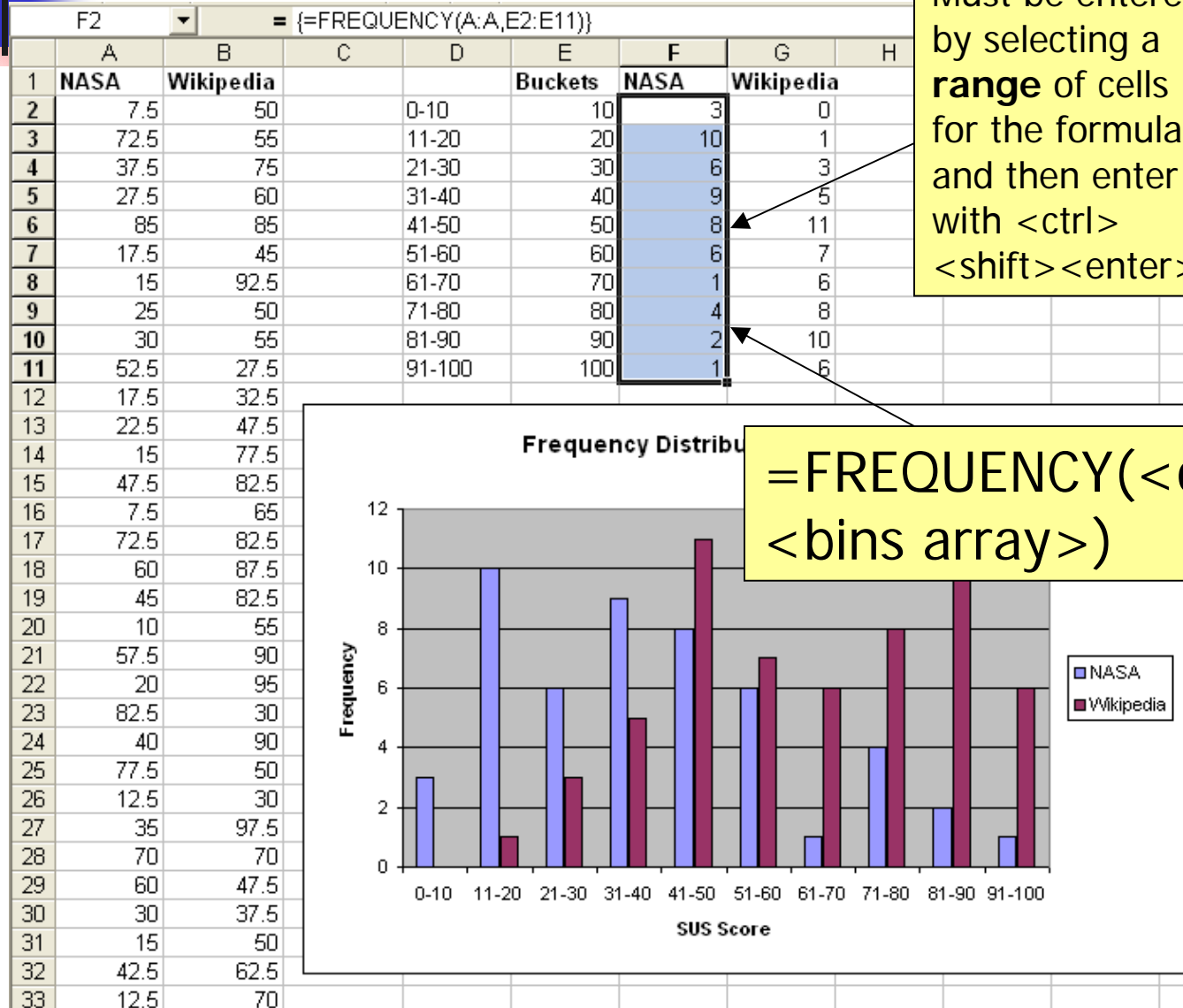
Comparison of SUS scores cross two tests in an iterative process:  
(LeDoux, Connor, & Tullis, 2005)

# Frequency Distributions



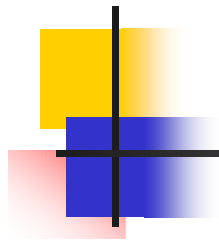
[Excel sample](#)

# How-to



Must be entered by selecting a **range** of cells for the formula and then enter with <ctrl> <shift> <enter>.

=FREQUENCY(<data array>, <bins array>)



## Tip 8: Combine Different Metrics



# Combined Metrics

---

- Often it's useful to combine different metrics to get an overall usability measure.
- Challenge is combining metrics that have different scales, e.g.
  - Task completion: % correct
  - Task time: Seconds
  - Subjective rating: SUS score
- Two common techniques:
  - Combine using percentages
  - Combine using z-scores



# Combine Based on Percentages

---

- Basic idea is to convert each of the metrics to a percentage and then average those together.
- For each metric to be transformed, you want:
  - 0% to represent the *worst* possible score
  - 100% to represent the *best* possible score
- Some metrics already are a percentage:
  - SUS scores
  - % correct tasks



# Sample Data

Original data:

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
<b>Averages</b>	<b>49.7</b>	<b>8.9</b>	<b>2.8</b>



# Sample Data

Original data with percentage transformations added:

Participant #	Time per Task (sec)	Tasks Completed (of 15)	Rating (0-4)	Time %	Tasks %	Rating %	Average
1	65	7	2.4	38%	47%	60%	48%
2	50	9	2.6	50%	60%	65%	58%
3	34	13	3.1	74%	87%	78%	79%
4	70	6	1.7	36%	40%	43%	39%
5	28	11	3.2	89%	73%	80%	81%
6	52	9	3.3	48%	60%	83%	64%
7	58	8	2.5	43%	53%	63%	53%
8	60	7	1.4	42%	47%	35%	41%
9	25	9	3.8	100%	60%	95%	85%
10	55	10	3.6	45%	67%	90%	67%
<b>Averages</b>	<b>49.7</b>	<b>8.9</b>	<b>2.8</b>	<b>57%</b>	<b>59%</b>	<b>69%</b>	<b>62%</b>

[Excel spreadsheet](#)



# Combine Using Z-scores

---

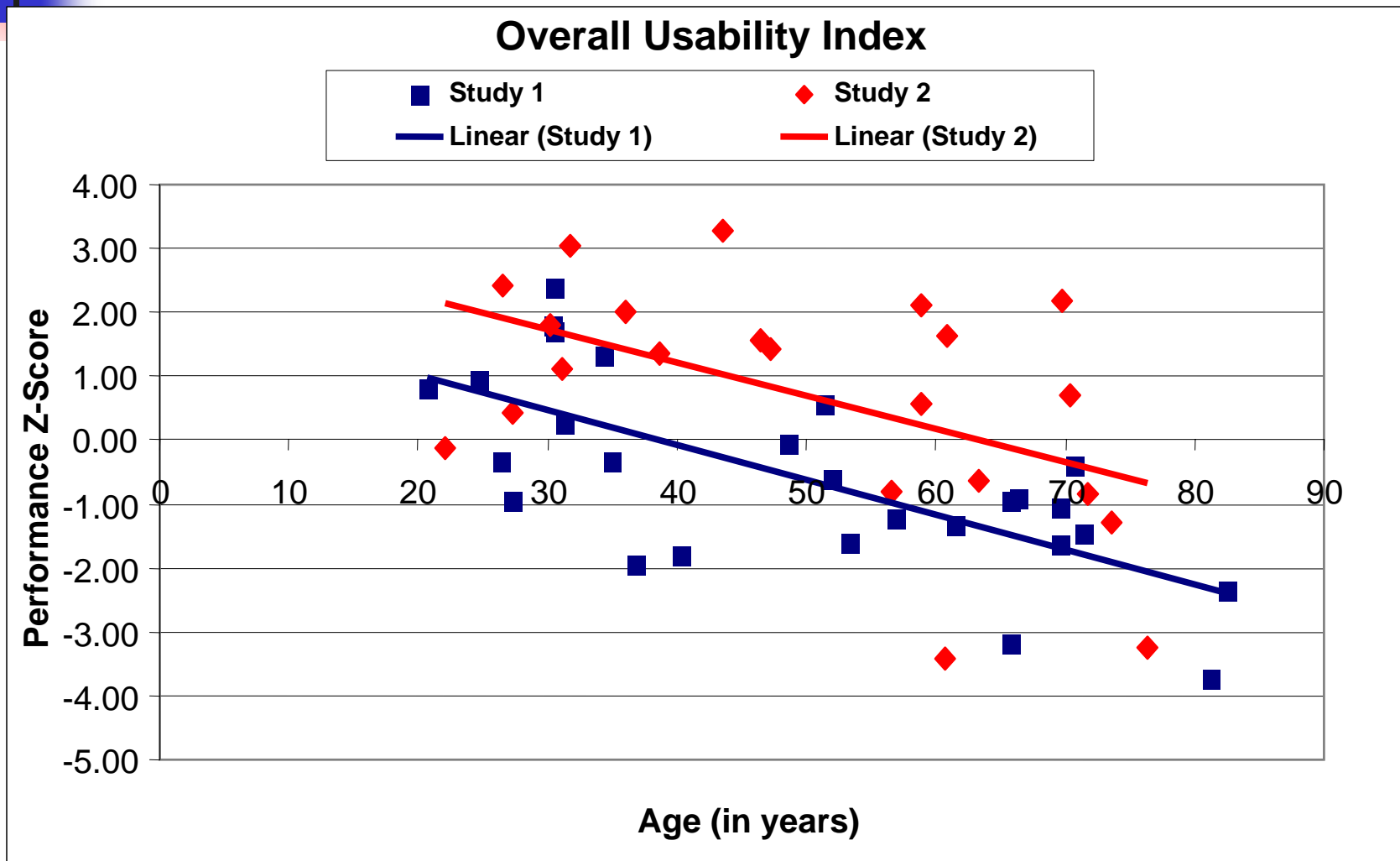
- Another method sometimes used is z-score transformation:
  - Convert each participant's score for each metric to a z-score.
    - Z-scores are based on the normal distribution.
    - They have a mean of 0 and standard deviation of 1.
    - Use the "standardize" function in Excel.
  - Average the z-scores for each person to get an overall z-score.
    - Make sure all scales go the same direction.
  - Must decide whether each score is going to be given equal weight.

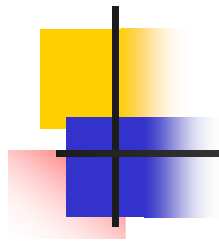
# Z-score Transformation Example

=standardize(B2,\$B\$12,\$B\$13)

	A	B	C	D	E	F	G	H	I
1	Participant #	Time per Task	Task Completion	SUS Score	z-time	z-completion	z-SUS	z-time * -1	Average z
2	1	65	52%	65	0.98	-0.89	-0.33	-0.98	-0.74
3	2	50	60%	68	0.02	-0.18	-0.14	-0.02	-0.11
4	3	34	78%	74	-1.01	1.43	0.25	1.01	0.90
5	4	70	42%	48	1.30	-1.78	-1.44	-1.30	-1.51
6	5	28	71%	82	-1.39	0.80	0.77	1.39	0.99
7	6	52	68%	77	0.15	0.54	0.45	-0.15	0.28
8	7	58	64%	70	0.53	0.18	-0.01	-0.53	-0.12
9	8	60	49%	42	0.66	-1.16	-1.83	-0.66	-1.22
10	9	25	70%	90	-1.59	0.71	1.29	1.59	1.20
11	10	55	66%	85	0.34	0.36	0.97	-0.34	0.33
12	<b>Mean</b>	<b>49.7</b>	<b>62.0%</b>	<b>70.1</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	
13	<b>Std Deviation</b>	15.57	0.11	15.39	1.00	1.00	1.00	1.00	

# A Real Example: Older Users

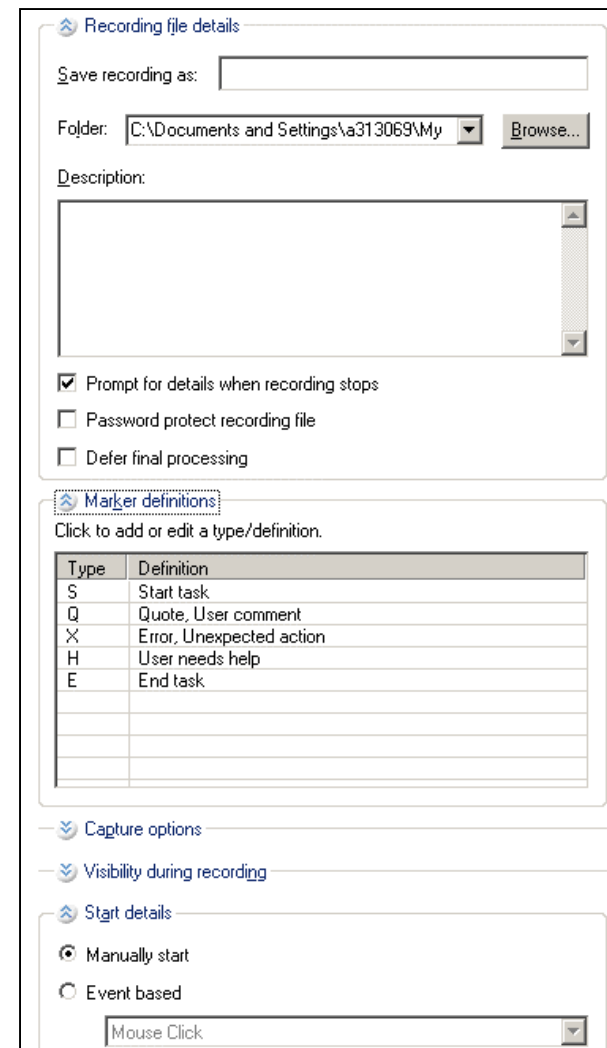
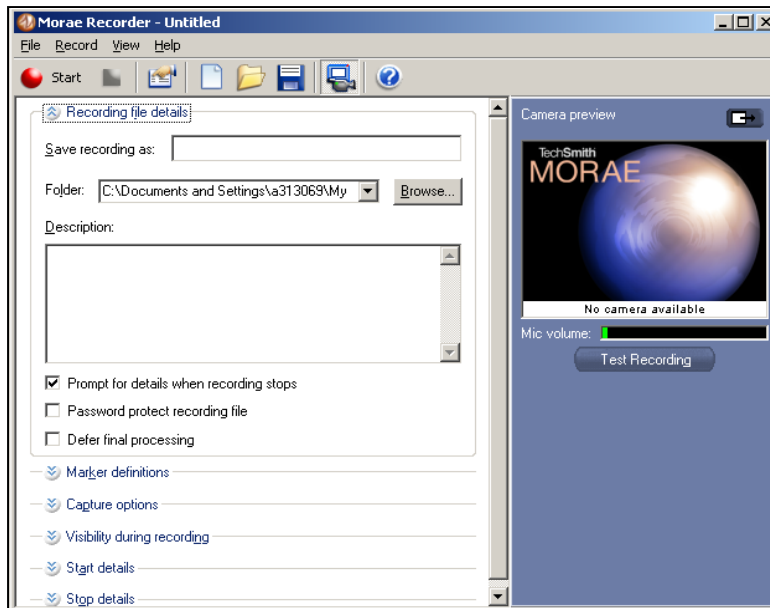




## Tip 9: Use Appropriate Tools

# Morae and Camtasia

- Excellent tools for capturing video, logging issues, and creating highlight tapes ([www.techsmith.com](http://www.techsmith.com)).



# Data Logger

- Free Excel program used to collect and analyze usability test data.
- Records task success, time-on-task, survey questions, and automatically generates charts.

([www.userfocus.co.uk/resources/datalogger.html](http://www.userfocus.co.uk/resources/datalogger.html))

**Usability Datalogger**

Admin  
Instructions: Enter project data below in WHITE CELLS

Project

Participant	Name (first last)	Date	Time (am/pm)	M/F	Age	Additional demographics...	Task Order for each u
1							1 2 3 4 5
2							1 2 3 4 5
3							1 2 3 4 5
4							1 2 3 4 5
5							1 2 3 4 5
6							1 2 3 4 5
7							1 2 3 4 5
8							1 2 3 4 5
9							1 2 3 4 5
10							1 2 3 4 5
11							1 2 3 4 5
12							1 2 3 4 5

Use the options below to configure the tool to meet your needs.

Effectiveness	Menu label	Description	Validation range
1	-	Menu default value	1 2 3 4 5
2	Easy	1st or 2nd try - no problem	1 1 1 1 1
3	Med	1st or 2nd try - observed difficulty	1 1 1 1 1
4	Hard	2nd or 3rd try - expressed difficulty	1 1 1 1 1
5	Assist	Succeeded with assistance	1 1 1 1 1
6	Fail	Failed or gave up	1 1 1 1 1

\*The above labels appear in the pulldown menus on each participant sheet

Tool name:

M/F range:

**Usability Datalogger**

Tasks & Questions  
Instructions: Enter tasks and questions below. Long version is optional. Short version is needed to po

#	Task Description (long version)	(short version)
1	<>	<>
2	<>	<>
3	<>	<>
4	<>	<>

**Usability Datalogger**

TIMER

**START** 0:00:00 Start  
0:00:00 End  
0:00:00 Segment  
0:00:00 Total Elapsed (h:mm:ss)  
0 Total Elapsed (s)

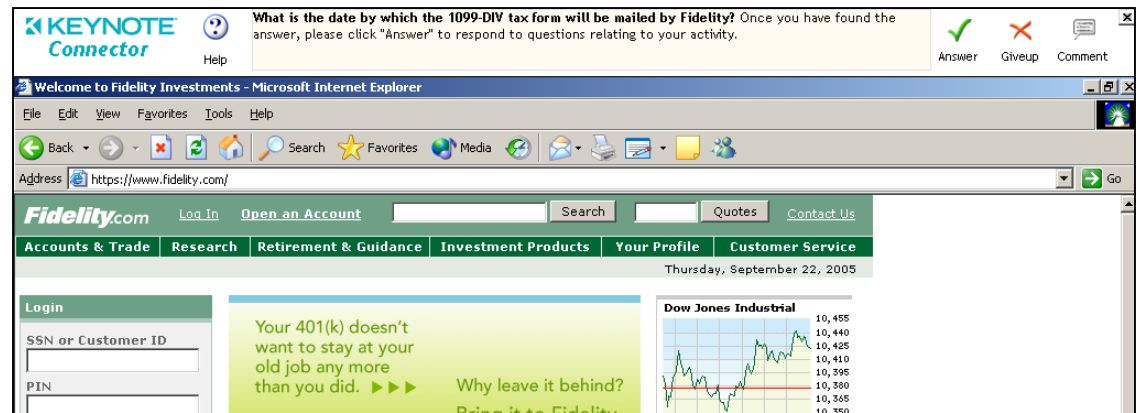
**RESET**

Task	Short description	Score	Time (s)	Observations
1	<>	-		
2	<>	-		
3	<>	-		

# Keynote (formerly Vividence)

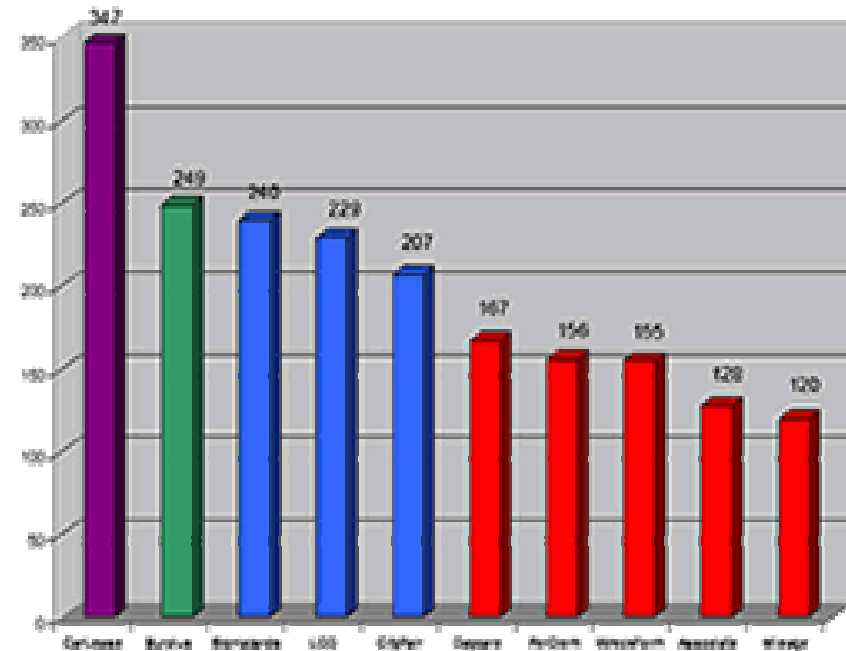
- Data collection and analysis tools designed to run automated usability studies.
- Metrics include task success, time-on-task, clicks, and click-stream.

([www.keynote.com](http://www.keynote.com))



# UTE (Usability Testing Environment)

- Bob Bailey at Mind Design Systems ([www.mindd.com](http://www.mindd.com))
- Comprehensive tool for capturing and analyzing usability data
- Automatically calculates success rates, time-on-task, and many other metrics



# Ergo Browser

- Modified Internet browser designed to collect usability data including time-on-task, clicks, pages, and scroll usage. Reasonably priced and easy to use. ([www.ergolabs.com](http://www.ergolabs.com)).



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	test02	11/11/1999 10:06														
2	TIME	METHOD	URL	TASK NUM BER	T-o-T	TOTAL PAGES	Back	Forward	Stop	Reload	Print	Go	VERT SCR Clicks	HORIZ SCR Clicks	Total Clicks	KEYS: KEY RE
3																
4																
5		BEGIN TASK		1												
6	22.85	DOCUMENT	http://www.yahoo.com/										3	0	2	13 <Shift>;f
7	35.04	DOCUMENT	http://search.yahoo.com/bin/search?p=Madagascar										0	0	1	0
8	56.02	DOCUMENT	http://dir.yahoo.com/Recreation/Travel/By_Region/Countries/Madagascar/										2	0	1	0

# The Observer from Noldus

- Sophisticated software for usability data collection, analysis and presentation.
- Ability to integrate multiple video feeds, analyze eye-tracking data, and other physiological measurements.

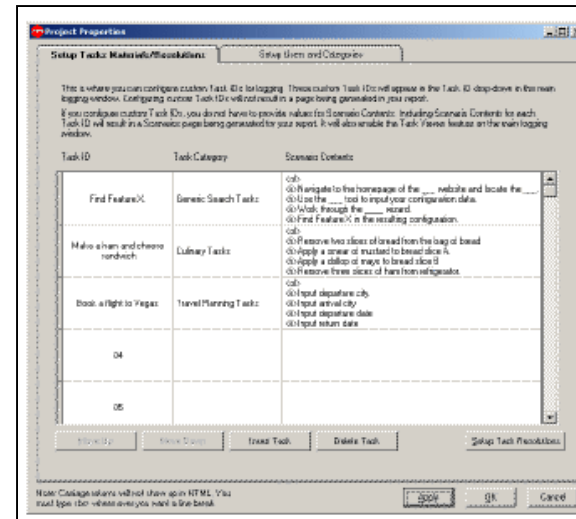
([www.noldus.com](http://www.noldus.com))



# Ovo Studios

- Software for collecting and analyzing usability data. Allows you to log issue, collects time-on-task, create video highlights, and helps in report generation.

([www.ovostudios.com](http://www.ovostudios.com))



Source 3: Unused  
Source 4: Unused

User ID "Jane" started Task ID "Book a flight to Vegas." at this point.

(left-click to zoom)

Update Observation Delete Observation

Observation Viewer

Timeline Shows: Start Capture at 10:41:04 AM on 4/29/2005

- Grid View:** The Ovo Logger Grid View provides a textual interface for searching, filtering, viewing, and editing user data. Simply select any field in the grid to directly edit its contents. You can configure a multivariate filter using simple boolean logic (e.g. contains, does not contain, starts with, ends with, etc.) to quickly find the data you are interested in.

ser IDs	All Task IDs	Time On Task	All Category One Values	All Category Two Values
	Find Feature X.	-1	All Category One Values	Assistance
	Find Feature X.	325.188	Custom	Feedback
	Make a ham and	254.36	Clearly Marked Exits	Terminology
	Book a flight to	226.172	Feedback	Standards & Guidelines
	Find Feature X.	64.531	Good Error Messages	Assistance
	Find Feature X.	124.531	Quick Log	Assistance
	Find Feature X.	184.531	Clearly Marked Exits	Assistance
	Find Feature X.	244.531	Clearly Marked Exits	Assistance

# "SUM" Calculator

- Another free tool developed by Jeff Sauro to calculate a single usability score "SUM" ([www.measuringusability.com](http://www.measuringusability.com)).

Copyright © 2005 Measuring Usability LLC <http://www.measuringusability.com>

### Summary

	SUM	Completion	Sat	Time	Errors
High	81.6%	91.1%	64.7%	89.8%	73.6%
<b>Point Estimate</b>	<b>67.1%</b>	<b>77.4%</b>	<b>50.0%</b>	<b>79.7%</b>	<b>61.1%</b>
Low	55.5%	63.7%	35.3%	69.7%	48.6%

### Task Raw Data

User	Satisfaction	Completion	Errors	Times	Standardized Avg	% Avg
1	4.33	1	1	88	0.42	66%
2	4.00	1	1	83	0.35	64%
3	3.50	1	0	182	-0.05	48%
4	4.00	1	1	50	0.74	77%
5	2.67	1	2	101	-0.59	28%
6	4.00	1	0	95	0.61	73%
7	4.00	1	0	98	0.58	72%
8	3.67	1	1	145	-0.18	43%
9	5.00	0	1	113	-0.15	44%
10	4.00	1	1	86	0.33	63%
11	4.00	1	0	94	0.62	73%
12	2.50	1	2	131	-0.84	20%
13	4.00	1	1	57	0.64	74%
14	4.00	1	0	101	0.56	71%
15	5.00	1	0	88	0.99	84%
16	4.33	0	0	81	0.24	60%
17	4.33	1	1	151	0.00	50%
18	2.33	1	2	73	-0.45	33%
19	5.00	0	1	64	0.28	61%
20	4.33	1	1	100	0.32	63%
21	5.00	0	0	59	0.70	76%

### Criteria

Confidence Level: 90%

Time Percentile: 95%

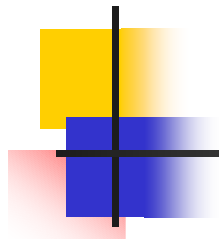
Sat Spec: 4

Time Spec:   
 Derived: 124

Error Opportunities: 2

For Task Time Use: Log Data

For Sample Estimate Use: Point Estimate



## Tip 10: Present Data Appropriately

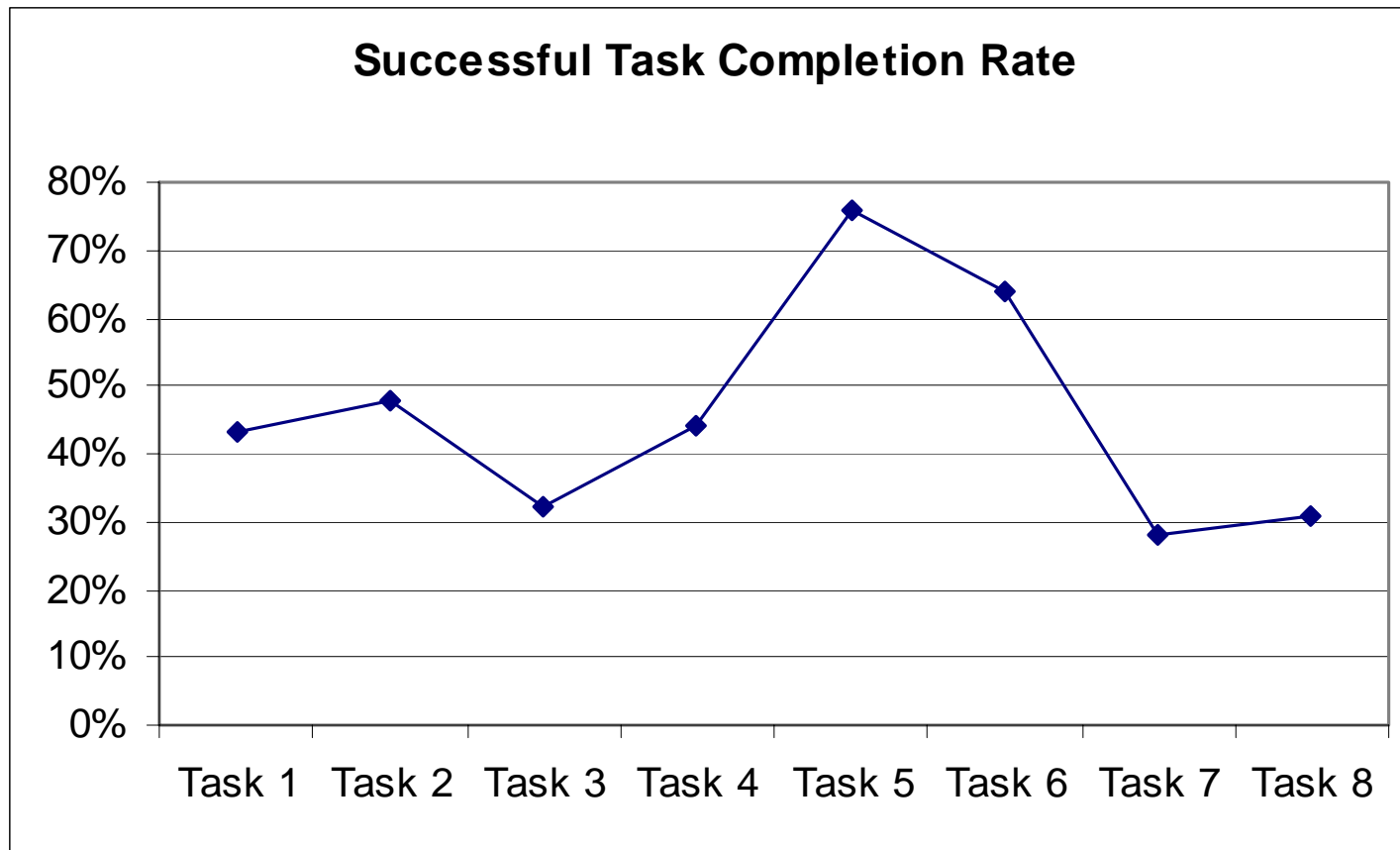


# Top 10 Mistakes in Data Graphs

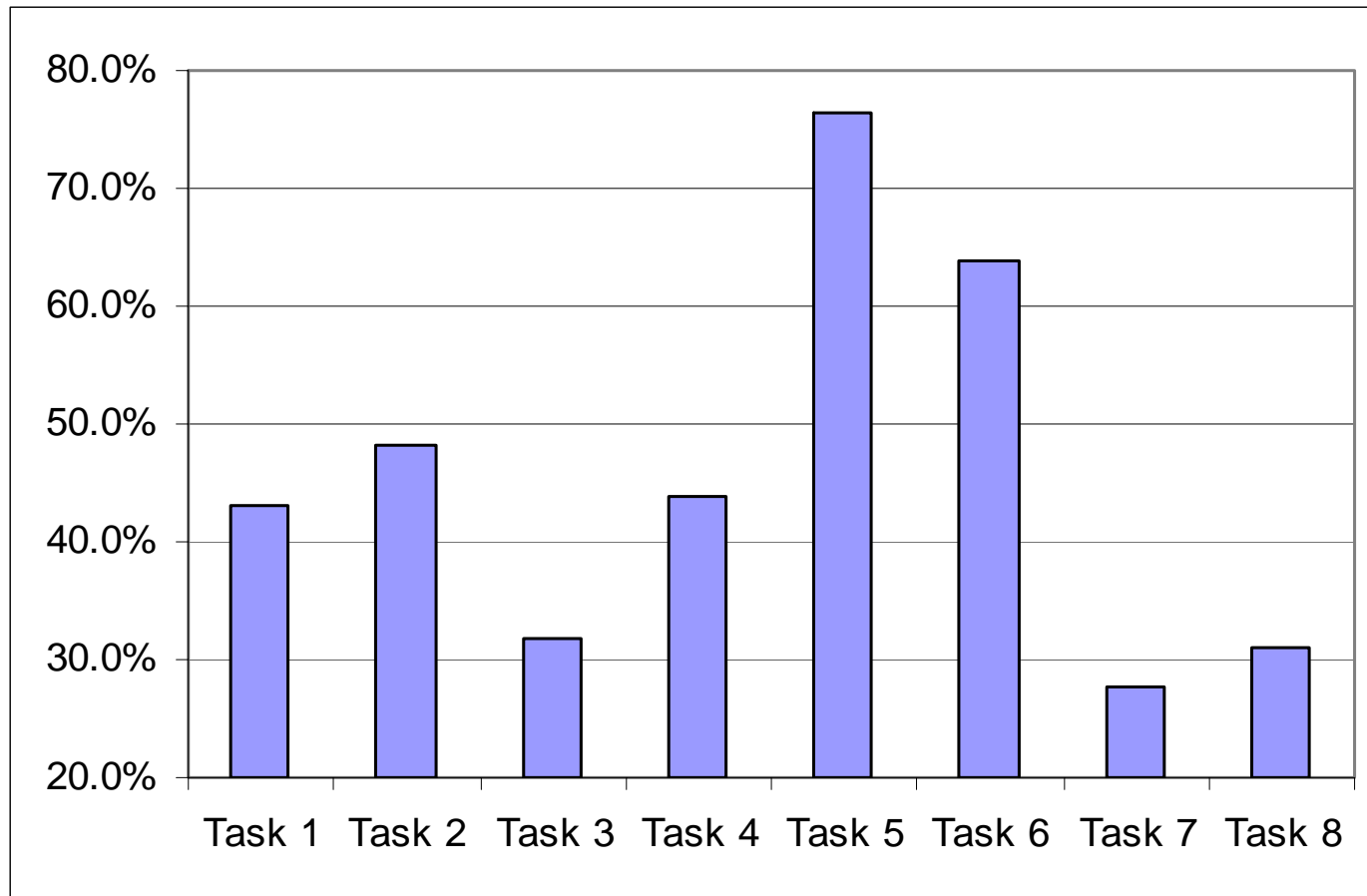
---

1. Not labeling axes or units.
2. Implying more precision in your data than it deserves.
3. Not showing confidence intervals when you can.
4. Not starting a bar graph at the lowest possible value for the y-axis (usually 0).
5. Using a line graph when it should be a bar graph.
6. Using 3-D when it doesn't add any value.
7. Trying to include too much.
8. Poor labeling of pie charts.
9. Using color coding as the only way to convey information.
10. Not knowing when to use stacked bar graphs.

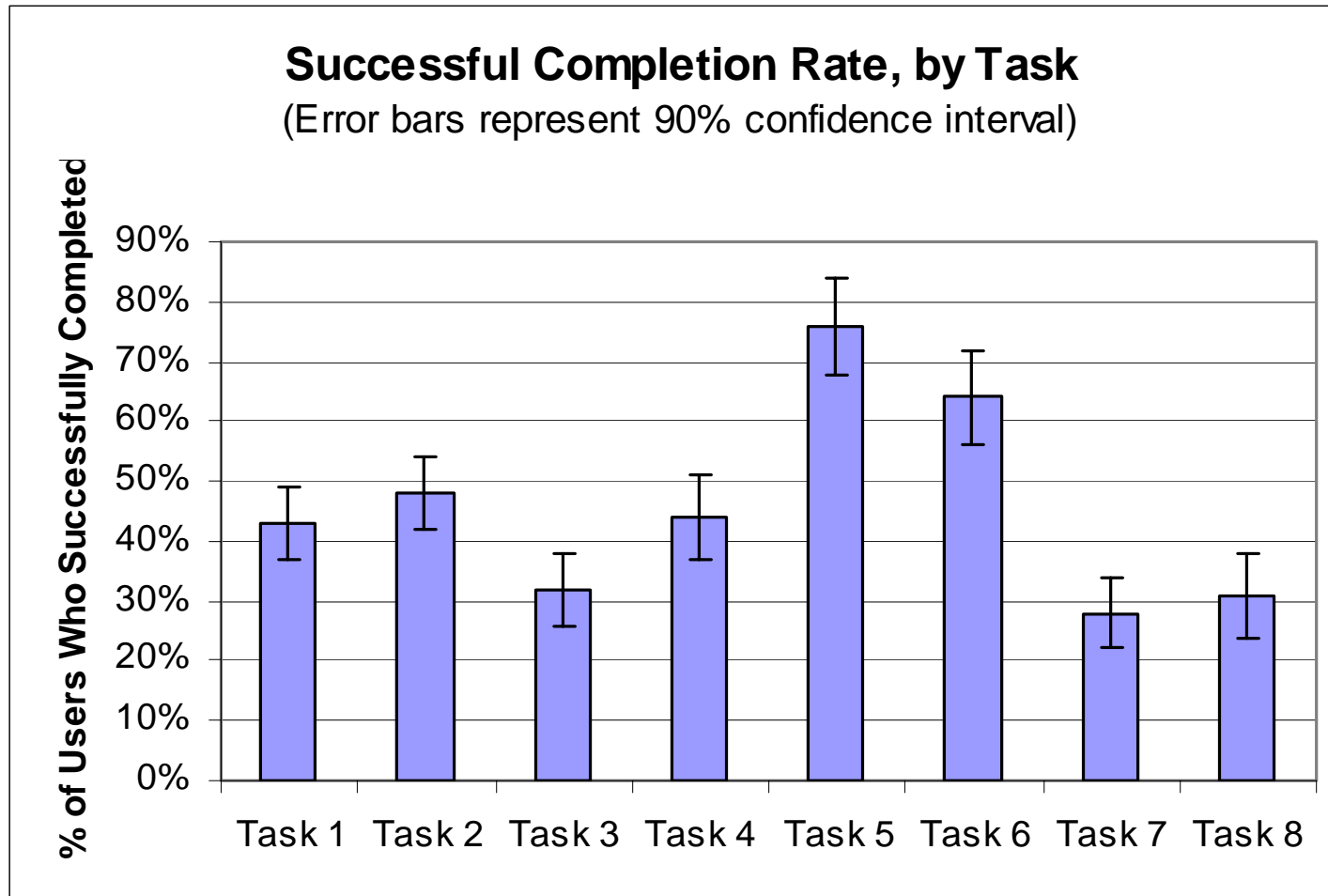
# What's Wrong With This?



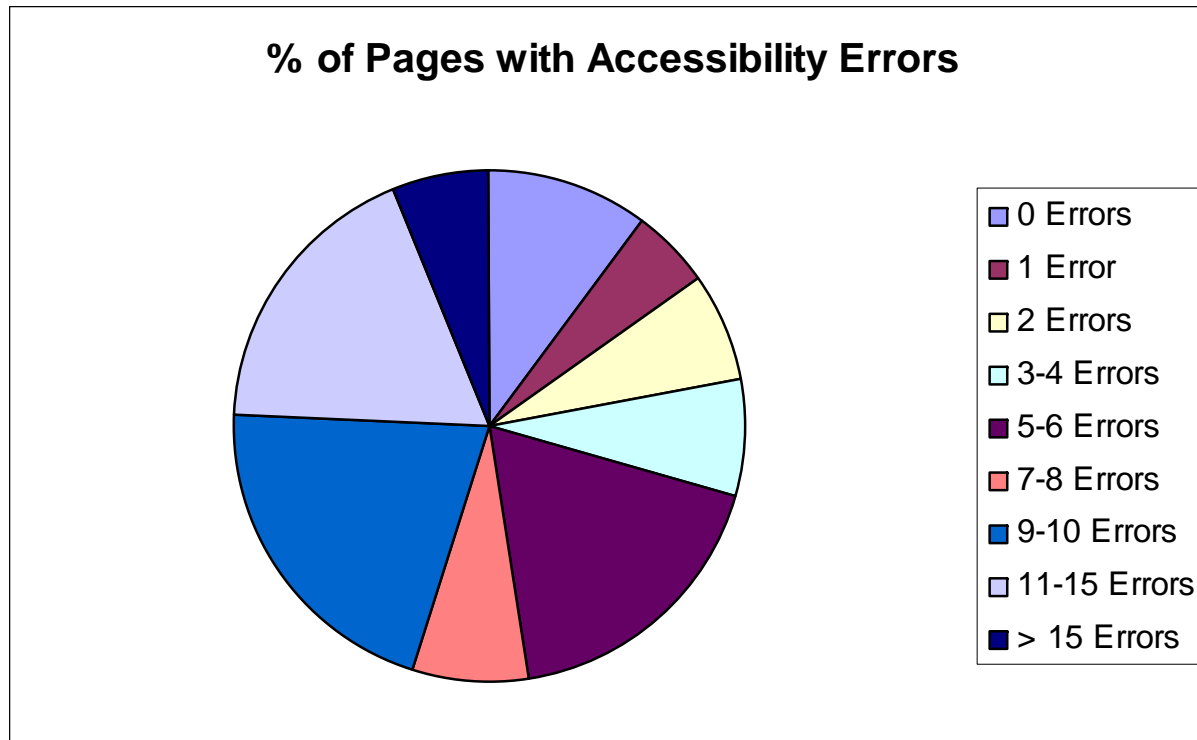
# What's Wrong With This?



# Better Version

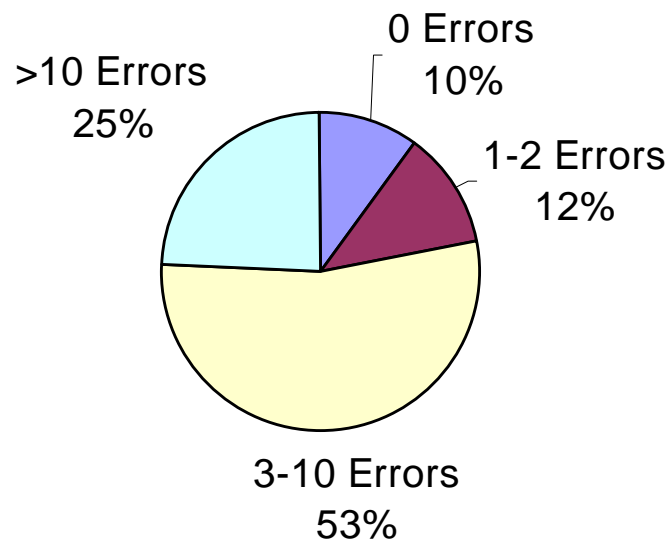


# What's Wrong With This?

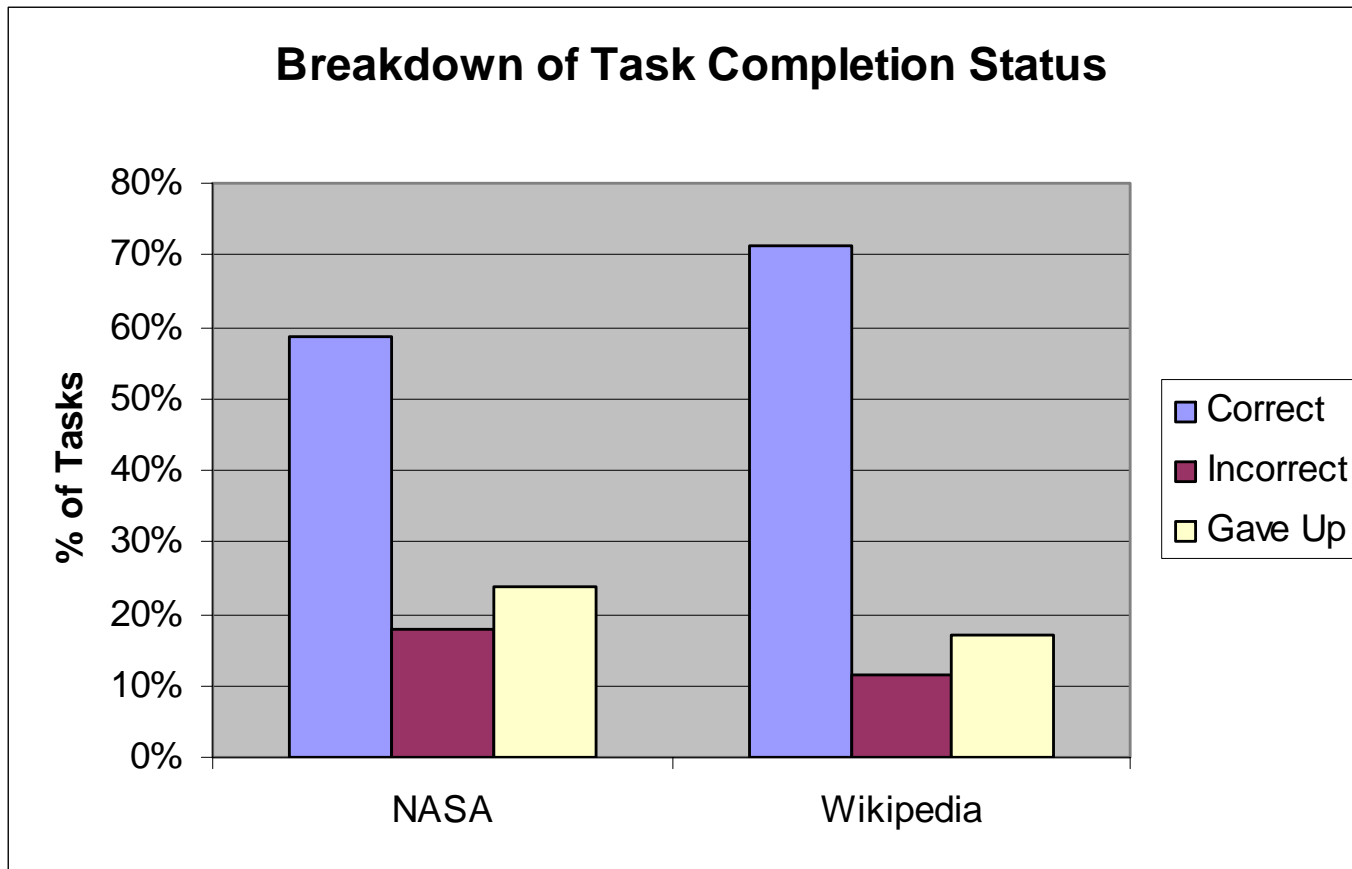


# Better Version

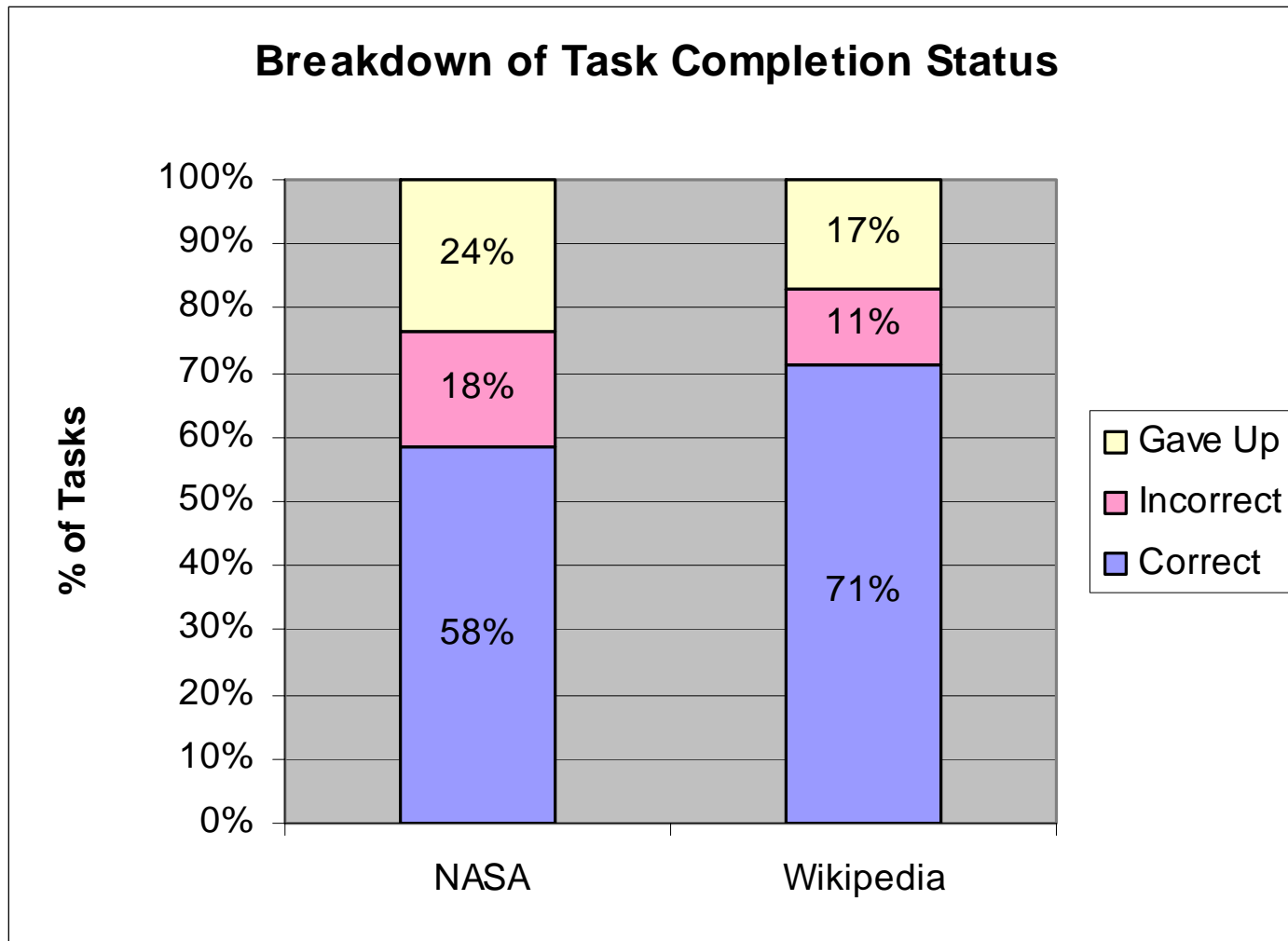
**% of Pages with Accessibility Errors**



# How Could This Be Better?



# Better Version





# Thank You!

For more information: [www.measuringux.com](http://www.measuringux.com)

