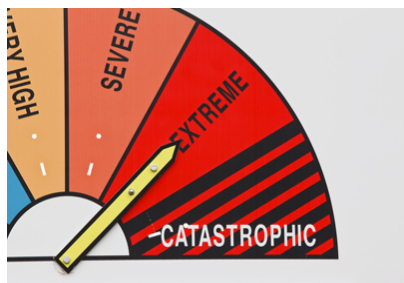


Rating The Severity Of Usability Problems

Jeff Sauro • July 30, 2013



If only one out of 1000 users encounters a problem with a website, then it's a minor problem.

157

Tweet

If that sentence bothered you, it should.

It could be that that single problem resulted in one visitor's financial information inadvertently being posted to the website for the world to see.

Or it could be a slight hesitation with a label on an obscure part of a website.

It's part of the responsibility of user experience professionals to help developers make decisions about what to fix.

Accounting for problem frequency and severity are two critical ingredients when communicating the importance of usability problems. They are also two of the inputs needed for [a Failure Modes Effects Analysis \(FMEA\)](#), a more structured prioritization process.

Problem Frequency

Measuring the frequency of a problem is generally straightforward. Take the number of users that encounter a problem divided by the total number of users. For example, if 1 out of 5 users encounter a problem, the problem frequency is .20, or 20%. The problem frequency can then be presented in a [user-by-problem matrix](#). It can also be used to estimate the [sample size needed](#) to discover a certain percent of the problems.

Problem Severity

Rating the severity of a problem is less objective than finding the problem frequency. There are a number of ways to assign severity ratings. I've selected a few of the more popular approaches described in the literature, and I'll contrast those with the method we use at Measuring Usability.

While there are differences in approaches, in general each method proposes a similar structure: a set of ordered categories reflecting the impact the problem has on the user, from minor to major.

Jakob Nielsen

[Jakob Nielsen](#) proposed the following four-step scale a few decades ago:

- 0 = I don't agree that this is a usability problem at all
- 1 = Cosmetic problem only: need not be fixed unless extra time is available on project
- 2 = Minor usability problem: fixing this should be given low priority
- 3 = Major usability problem: important to fix, so should be given high priority
- 4 = Usability catastrophe: imperative to fix this before product can be released

Jeff Rubin

In Jeff's [influential 1994 book](#), he outlined the following scale for problem severity:

- 4: Unusable:** The user is not able to or will not want to use a particular part of the product because of the way that the product has been designed and implemented.
- 3: Severe:** The user will probably use or attempt to use the product here, but will be severely limited in his or her ability to do so.
- 2: Moderate:** The user will be able to use the product in most cases, but will have to

NEWSLETTER SIGN UP

Receive bi-weekly updates.

[3964 Subscribers]

Join

CONNECT WITH US

Follow Us On Twitter 13914 Followers

Subscribe to RSS **1750 readers**
BY FEEDBURNER

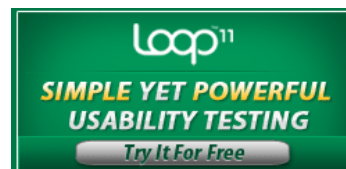
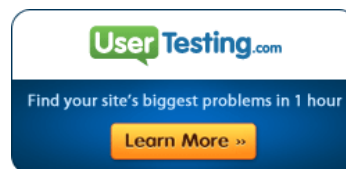
Newsletter Sign Up [3964 Subscribers]

UX BOOTCAMP

Denver CO, Aug 20-22nd 2014

3 Days of Hands-On Training on User Experience Methods, Metrics and Analysis. [Learn More](#)

OUR SUPPORTERS



Most Recent Most Popular

The Most Fundamental Concept in Usability 18 hours ago

5 Tips for Conducting a Mobile Usability Test 8 days ago

undertake some moderate effort in getting around the problem.

1: Irritant: The problem occurs only intermittently, can be circumvented easily, or is dependent on a standard that is outside the product's boundaries. Could also be a cosmetic problem.

Dumas and Redish

Joe Dumas and Ginny Redish, in their seminal book, [A Practical Guide to Usability Testing](#), offer a similar categorization as Rubin and Nielsen but add a global versus local dimension to the problems. The idea is that if a problem affects the global navigation of a website, it becomes more critical than a local problem only affecting, say, one page.

- Level 1:** Prevents Task Completion
- Level 2:** Creates significant delay and frustration
- Level 3:** Problems have a minor effect on usability
- Level 4:** Subtle and possible enhancements/suggestions

Chauncey Wilson

[Chauncey Wilson suggests](#) that usability severity scales should match the severity rating of bug-tracking systems in a company. He offers a five-point scale with the following levels. Earlier, he's used a similar [four-point variant\[pdf\]](#).

- Level 1:** Catastrophic error causing irrevocable loss of data or damage to the hardware or software. The problem could result in large-scale failures that prevent many people from doing their work. Performance is so bad that the system cannot accomplish business goals.
- Level 2:** Severe problem, causing possible loss of data. User has no workaround to the problem. Performance is so poor that the system is universally regarded as 'pitiful'.
- Level 3:** Moderate problem causing no permanent loss of data, but wasted time. There is a workaround to the problem. Internal inconsistencies result in increased learning or error rates. An important function or feature does not work as expected.
- Level 4:** Minor but irritating problem. Generally, it causes loss of data, but the problem slows users down slightly. There are minimal violations of guidelines that affect appearance or perception, and mistakes that are recoverable.
- Level 5:** Minimal error. The problem is rare and causes no data loss or major loss of time. Minor cosmetic or consistency issue.

The Wilson and Dumas & Redish scales have the more severe problem with lower numbers. That is because in the early days of computing, severe bugs were called "level 1 bugs" and those had to be fixed before product release (Dumas, Personal Communication 2013). In this scale, the problems are defined in terms of data loss rather than their impact on users' performance or emotional state.

Molich & Jeffries

Rolf Molich is famous for his series of [comparative usability evaluations \(CUE\)](#). He's also famous for reviewing and writing (often critically) about the quality of usability reports. He and Robin Jeffries offered a three-point scale.

1. **Minor:** delays user briefly.
2. **Serious:** delays user significantly but eventually allows them to complete the task.
3. **Catastrophic:** prevents user from completing their task.

This three-point approach is simpler than others but tends to rely heavily on how the problem impacts time on task.

Our Approach

Originally we started with a 7-point rating scale where evaluators assigned the problem severity a value from cosmetic (1) to catastrophic (7) but we found it was difficult to distinguish easily between levels 2 and 6. We reduced this to a four-point scale similar to Rubin, Nielsen and Dumas/Redish above and treated them more as categories than a continuum.

While there was much less ambiguity with four points, we still found a murky distinction between the two middle levels in both assigning the severity and reporting the levels of problems to clients.

So we reduced our severity scale to just three levels, along with one for insights, user

10 Surprisingly Common But Painful User Experiences 15 days ago

7 Reasons Usability Problems Don't Get Fixed 22 days ago

JEFF'S BOOKS



Quantifying the User Experience: Practical Statistics for User Research

The most comprehensive statistical resource for UX Professionals

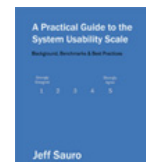
[Buy on Amazon](#)



Excel & R Companion to Quantifying the User Experience

Detailed Steps to Solve over 100 Examples and Exercises in the Excel Calculator and R

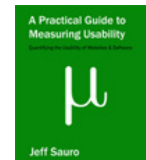
[Buy on Amazon](#) | [Download](#)



A Practical Guide to the System Usability Scale

Background, Benchmarks & Best Practices for the most popular usability questionnaire

[Buy on Amazon](#) | [Download](#)



A Practical Guide to Measuring Usability

72 Answers to the Most Common Questions about Quantifying the Usability of Websites and Software

[Buy on Amazon](#) | [Download](#)

suggestions or positive attributes.

1. **Minor:** Causes some hesitation or slight irritation.
2. **Moderate:** Causes occasional task failure for some users; causes delays and moderate irritation.
3. **Critical:** Leads to task failure. Causes user extreme irritation.

Insight/Suggestion/Positive: Users mention an idea or observation that does or could enhance the overall experience.

Summary

I've put abbreviated versions of these scales below in the table to show the similarities in some of the terms and levels. I've also aligned the scales so higher numbers indicate more severe problems.

Level	Nielsen	Rubin	Dumas	Wilson	Molich & Jeffries	Sauro
0	Not a Problem					Insight/ Suggestion/ Positive
1	Cosmetic	Irritant	Subtle & possible enhancements/ suggestions	Minor cosmetic or consistency issue	Minor (delays user briefly)	Minor : Some hesitation or slight irritation
2	Minor	Moderate	Problems have a minor effect on usability	Minor but irritating problem		
3	Major	Severe	Creates significant delay and frustration	Moderate problem	Serious (delays user significantly but eventually)	Moderate: Causes occasional task failure for some users; causes delays and moderate irritation
4		Unusable	Prevents Task Completion	Severe problem		Critical: Leads to task failure. Causes user extreme irritation.
5	Catastrophe			Catastrophic error	Catastrophic (prevents user from completing their task)	

Some lessons from these problem severity levels:

1. **Don't obsess over finding the right number of categories or labels:** Three categories is probably sufficient, but merging scales with bug tracking levels or having more levels to generate more internal buy-in are both legitimate reasons to have more points. Once you pick a system, try and stick with it to allow comparison over time.
2. **There will still be inter-rater disagreement and judgment calls:** These are rough guides, not precise instruments. Different evaluators [will disagree](#), despite the clarity of the severity levels. One of the best approaches is to have multiple evaluators rate the severity independently, calculate the agreement, and then average the ratings.
3. **The numbers assigned to each level are somewhat arbitrary:** Don't obsess too much over whether higher severity problems should have higher numbers or lower ones. I prefer the latter, but it's the order that has meaning. While the intervals between severities of 1, 2 and 3 are likely different, the ranks can be used for additional analysis when comparing different evaluators or problems severity and frequency.

4. **Don't forget the positives:** [Dumas, Molich & Jeffries](#) wrote a persuasive article talking about the need for pointing out positive findings. While a usability test is usually meant to uncover problems, understanding the positives encourages the developers and doesn't make you or your team come across as the constant harbingers of bad news.
5. **Treat frequency separately from severity:** We report the frequency of an issue along with its severity. When possible, we have a separate analyst rate the severity of a problem without knowing its frequency—a topic for a future blog.

Thanks to Joe Dumas for commenting on an earlier draft of this article.

157

Tweet

About Jeff Sauro



Jeff Sauro is the founding principal of Measuring Usability LLC, a company providing statistics and usability consulting to [Fortune 1000 companies](#). He is the author of over [20 journal articles](#) and 4 books on statistics and the user-experience.

[More about Jeff...](#)

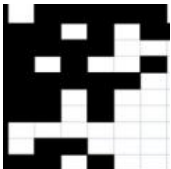
Learn More

[UX Bootcamp: Aug 20th-22nd in Denver, CO](#)

[Best Practices for Remote Usability Testing](#)

[The Science of Great Site Navigation: Online Card Sorting + Tree Testing Live Webinar](#)

You Might Also Be Interested In:



[Report Usability Issues in a User by Problem Matrix](#)



[Do severe problems affect more users than trivial ones?](#)



[Prioritizing Problems in the User Experience: The FMEA](#)

Related Topics

[Usability Problems](#)

Posted Comments

There are 3 Comments

August 28, 2013 | [jordisan](#) wrote:

Why not:

frequency: 0 to 3

severity: 0 to 3

overall priority: frequency x severity