# Side-by-Side Guidelines

Version 7.9 - December 4, 2019

# Part 1: Side-by-Side Rating Guidelines

- 1.0 Introduction
- 1.1 What is SxS Rating?
- 2.0 How to Rate Standard Needs Met SxS Tasks
- 2.1 Rating Philosophy
- 2.2 Summary of Rating Steps for Standard Needs Met SxS Tasks
- 2.3 Thinking about Result Sets
- 2.4 SxS Rating Scale
- 2.5 SxS Rating Do's and Don'ts
- 2.6 SxS Rating Comments

# Part 2: Side-by-Side Rating Examples

- 3.0 Standard Needs Met SxS Rating Examples
- 4.0 Standard Needs Met SxS Rating Scenarios

# Part 3: Using the Evaluation Platform for SxS Rating

5.0 The Needs Met SxS Rating Interface

# Part 1: Side-by-Side Rating Guidelines

### 1.0 Introduction

Welcome to Side-by-Side (SxS) rating! These guidelines discuss how to assign SxS ratings in Needs Met rating tasks that have a Side-by-Side rating scale. For more information about Needs Met rating, please refer to the <a href="General Guidelines">General Guidelines</a>.

### 1.1 What is SxS Rating?

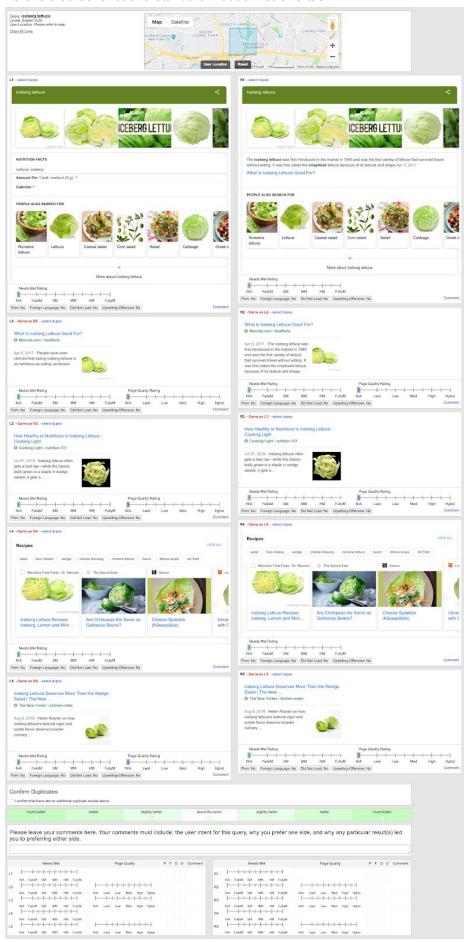
Side-by-Side rating asks you to compare two sets of things—in most Needs Met SxS tasks that means two sets of results for the same query—and then tell us which side is better using the scale: **much better**, **better**, **slightly better**, or **about the same**. Please note the following:

- Most standard Needs Met SxS tasks have 5 result blocks on each side, but some tasks may vary.
- Most result blocks in standard Needs Met SxS tasks have a result block link, but others may not. See Section 12.8 in the <u>General Guidelines</u> for more details on different types of result blocks, and Section 13.1 for guidance on which part of the result to rate.
- Many standard Needs Met SxS tasks have different types of result blocks. For example, you may see web, video, and image result blocks in the same Needs Met SxS task.
- Sometimes, special projects will ask you to compare other things, such as image results, video results, map
  results, product results, snippets, translations of text, etc. Special projects will usually have project-specific
  instructions that explain how to rate the results, and some may use different scales and buttons. For these
  projects, pay very close attention to the project-specific instructions and follow them.

For most SxS projects, you will complete each task by indicating which side you like better (and how much better), then explaining why. SxS rating allows you to tell us what you like and what you think a user would like to see, and your job is to share a well thought-out opinion.

Don't be shy—tell us what you think! If one side is truly better than the other, use the full SxS scale to indicate your preference. On the other hand, if the difference between the two sides is very small or would make no difference to users, **about the same** is probably the best rating.

#### Here is a screenshot of a standard Needs Met SxS task:



#### 2.0 How to Rate Standard Needs Met SxS Tasks

These guidelines apply to standard Needs Met SxS tasks. Remember that for other kinds of SxS tasks (images, videos, map results, products, etc.), you should read and follow the project-specific instructions.

### 2.1 Rating Philosophy

First, here are the most important points to remember about SxS rating:

- Users should find what they are looking for. Respect their intent.
- Users should get the best results in the most helpful order. Order matters, especially on mobile devices with smaller screens.
- Users should have many options and choices. Not everyone likes the same websites or types of content.
- Results should feel right for the user location and language. Use your knowledge and rate from a local
  perspective.

### 2.2 Summary of Rating Steps for Standard Needs Met SxS Tasks

Here is a rough outline of the steps you should follow when rating standard Needs Met SxS tasks.

- 1. **Understand the query.** Think about likely user intents and query interpretations. What are users looking for when issuing the query?
- 2. **Assign Needs Met and Page Quality ratings to all results on each side.** Note that some special content result blocks may not ask for a Page Quality rating. Also, assign flags and mark duplicate results as needed.
- 3. **Decide which set of results is better for users in the rating language, locale, and user location.** This may be difficult when there are lots of factors to consider, so use your best judgment.
- 4. **Choose an overall SxS rating based on the results.** After deciding which set of results is better, choose a SxS rating. The rating depends on how much better one side is for users issuing the guery.
- 5. **Write a comment to explain your rating.** Please explain your interpretation of the likely user intent and the reasoning behind your rating. In your comment, include notes on the specific results that you think would be most or least helpful for users. Note that comments are not required when you select **about the same**, but are encouraged because they help us understand your reasoning.

Your rating quality for these tasks is evaluated based on your Needs Met ratings, Page Quality ratings, overall SxS rating, and the reasoning behind your ratings as explained in your comments.

# 2.3 Thinking about Result Sets

When assigning SxS ratings, the most important consideration is how helpful each result set is as a whole. Because you are looking at a set of results, additional factors should be considered such as order and diversity of the results.

The considerations below are not in any particular order. For some queries, one consideration may be more important than another.

Result Set Consideration	Description	
How helpful is the set of results?	The <b>set</b> of results must be considered as a whole—you should not just consider results individually.	
	The helpfulness of an individual result may depend on other results in the set. True duplicate results may be useless; near duplicates may or may not add value depending on the query.	
Order of results	The order of the results is extremely important, especially for mobile users with smaller devices. Most users look at the top results, so the best and most helpful results should appear at the top of the page. Results in lower positions get less attention from users, even if they are helpful. Useless results may discourage users from looking further down the page. Therefore, results should appear in order of helpfulness to users.	
	Think about query interpretation and user intent. Really great results for the dominant interpretation or most likely user intent should be at the top. Mediocre or average results for a minor interpretation may not be worth showing at all.	
Diversity of results	The results in the set should work together to address different aspects of query interpretation/user intent and provide a helpful collection of resources for users. We will call this rating consideration "diversity". Here are a few types of diversity to think about:	
	1. <b>Diversity in query interpretation/user intent</b> : If there is more than one reasonable user intent or query interpretation, the result set should reflect this. When the query is ambiguous or vague, a result set with many interpretations or user intents may be helpful. However, if there is only one interpretation or intent, the entire result set should address that one interpretation.	
	2. Diversity of information sources/websites: Sometimes, many different kinds of websites are helpful for a query. For travel queries, trip planning websites, reviews, and hotel and flight aggregators may all be helpful. Even when users are looking for specific information or a particular type of content, such as song lyrics, the result set should include several different sources. This gives users choices and provides confirmation that the information is correct.	
	3. <b>Diversity of result types</b> : Some queries really have only one type of helpful result. If the query asks for videos, then videos are the appropriate type of result. Some queries are well served by different types of results. For example, news, videos, blogs, and images may all be appropriate for a celebrity query. Also, some queries may benefit from having helpful special content result blocks, which provide immediate, on-page content. For example, special content result blocks may be helpful when users are looking for a specific fact or piece of information.	
	<ol> <li>Diversity of people and cultures represented: For broad queries, consider the diversity of individuals and demographic backgrounds represented in the result set.</li> </ol>	
	For example, for a query like [weddings], users should be able to find results about all kinds of weddings: held in different settings, from different traditions and styles, large and small in size, etc. However, if the query is more specific, the result set should reflect diversity within the scope of the query. For example, [argentinian writers] should only return results about writers from Argentina, but would benefit from a diversity of authors by time period, genre, subject matter, gender, race, etc.	
	arversity of authors by time period, genie, subject matter, genuer, race, etc.	

Result Set Consideration	Description	
Number of results	Some result sets have many results; others have few results. Please keep in mind:	
	More results is not necessarily better. Think about the helpfulness of the whole set.	
	Truly useless results add little to nothing and may just be distracting or discouraging.	
	Unsolicited pornography and malicious downloads are a poor user experience; none should appear in a result set.	
Content in the Result Block	Some results show helpful content and/or media directly in the result block, saving users from clicking to load another page. You should prefer result sets with helpful content in result blocks that immediately satisfies mobile users.	
	However, remember to think about query interpretation and user intent! Some types of content shown directly in result blocks might not be helpful at all. For example, a result block showing a map and nearby stores for the broad query [science] would be distracting and useless for nearly all users.	
User Location and Language	Imagine that you are a mobile user in the specified user location, and think about what that user would want to see—the results need to feel right for a user in that location. For SxS ratings, think about which result set better addresses users in that location. Refer to the <a href="General Guidelines">General Guidelines</a> for more information on user location and language.	
Page Quality	Users need results they can trust, therefore you should prefer high quality result sets to low quality result sets. Think about the trustworthiness of each result for mobile users, and evaluate which result set is more trustworthy as a whole.	
Freshness	Some queries (for example, queries for current or recurring events) demand very recent or "fresh" information. Think about mobile users and which result set better addresses their immediate needs. Refer to the <u>General Guidelines</u> for more information on freshness.	

### 2.4 SxS Rating Scale

For standard Needs Met SxS rating tasks, your rating choices will be **much better**, **better**, **slightly better**, or **about the same** for the two sets you are comparing.

The strength of the rating depends on the difference in helpfulness between the two sides—the bigger the difference for the user, the stronger the rating.

For example, the rating **much better** should be used when the user would have a much easier time accomplishing a goal or finding information. Higher placement of the best results, much better ordering of the results, or really helpful unique results could be reasons for a **much better** rating. Presence of an unsolicited upsetting or offensive result on one side could also justify a **much better** rating for the other side. If users would prefer one side but the difference is not as large, **slightly better** or **better** may be more appropriate.

However, sometimes even large differences between result sets make no difference to the user, making **about the same** appropriate. For example, both sides could have completely different, yet equally helpful or unhelpful sets of results.

# 2.5 SxS Rating Do's and Don'ts

Here is a table of Do's and Don'ts that reflect common rating scenarios.

SxS Rating Do's	SxS Rating Don'ts
Do think about how helpful the result set would be for mobile users.	Don't ignore how easy it is to see or find information users are looking for, or how the experience would be for mobile users.
Do compare the sets of results as a whole.	Don't just compare the first result on the left to the first result on the right, and then the second result on the left to the second result on the right, etc.
<b>Do</b> look at all results, including those that appear on both sides.	Don't ignore results that are the same on both sides.  Don't rate based only on results that are different.
Do think about the differences in the result sets from a user perspective.	Don't use much better or better for changes that make little or no difference to users.
<b>Do</b> use the full rating scale. Good raters know when to use <b>about the same</b> and when to use <b>much better</b> .	<b>Don't</b> be afraid to use <b>much better</b> or <b>better</b> when one side is truly better than the other.
Do think about different query intents. A query can have more than one interpretation.	Don't assume all users are looking for the same thing.
Do think about user intent for result sets with special content shown directly in result blocks.	Don't automatically give a higher rating to the side with more special content in result blocks. For some queries, special content or media might be exactly what the user needs. For other queries, special content might be completely unhelpful.
Do give a higher rating to the side that is truly more helpful, even if it has fewer results.	Don't automatically give a higher rating to the side with more results. More results are not necessarily better.  Don't assume that "something is better than nothing".  This is very important for extra features or for comparing a side with no results to a side with results.
<b>Do</b> think about diversity. For some queries, diversity may be very important. For other queries, diversity may not be important at all.	Don't automatically give a higher rating to the side with more diversity.
Do use common sense and good judgment.	<b>Don't</b> blindly follow instructions without understanding the task.

### 2.6 SxS Rating Comments

Comments on your overall SxS rating are a very important part of SxS tasks. When your SxS ratings are reviewed, we look at these SxS task comments to understand your rating and review the quality of your work.

A comment explaining your rating is required when you select **slightly better**, **better**, and **much better** ratings. Rating comments are not required when you select **about the same**, but are encouraged because they help us understand your reasoning.

Your comment must include the following information:

- 1. Describe the user intent for the given query.
- 2. Describe why you prefer one side over the other, or (optionally) why you think both sides are about the same.
- 3. Describe any particular result(s) that led you to prefer either side.

**All comments must be written in English**. If you want to write full sentences, please do so. However, you can use abbreviations for results or objects in a list, as long as you adopt the same abbreviations found in the General Guidelines.

Please note the following:

- L1 refers to the first result on the left side; R3 refers to the third result on the right side.
- If you use shorthand notation, you should use uppercase "L" and "R", not lowercase "l" and "r".
- Do not use Internet slang types of abbreviations, such as IMO (an abbreviation for "in my opinion").
- The SxS rating interface allows you to write comments (take notes) about individual results as you rate, which may be helpful to you when assigning a SxS rating. You are not required to write comments about individual results as these comments are primarily for your own use.

The following tables display examples of good and bad rating comments. Please note that some of these comments are from tasks rated **about the same**.

#### **Good Comment Examples**

The query has a dominant interpretation (the government spy agency). The user probably wants to go to the homepage. The query also has some other interpretations (a cooking school, an art school), but they are less likely. The L side has minor interpretations at the top followed by FailsM results. R1, R2, R3 give helpful coverage of the dominant interpretation, including the homepage, and a nice mix of other interpretations at R4, R5.

User intent is to go the official homepage. L side has better results. R side has 5 foreign sites that are not helpful for users in the U.S. In addition, the R side has 3 minor dealer sites. By contrast, L side starts with the FullyM (missing from the R side) and has several other helpful results (L2, L4, L6).

User wants to purchase a treadmill. Both sides are bad. L1 and R1 are the same and just barely helpful. L2 is a subpage of L1 that adds no value (and L1 wasn't helpful anyway). L4/R2 and L7/R3 no longer load. L3/R4 and L6/R5 are short-lived classified ads. R5 is keyword stuffing without real content.

All of these comments are helpful. They provide specific information that helps us understand why the rating was chosen.

The user wants to go to the official page of the medicine or find information about it. R side order is much better because the most helpful result is at the top, as well as a Mayo Clinic page about the medicine. L side has useless results at the top: L1/R4, L2/R5.

The left side is much better because R1 is an inappropriate porn result. This children's TV-show query shouldn't have any porn.

The user may want to purchase this skin care product or read reviews about it. R side is mostly reviews and decent shopping sites (R1, R2, R4, R6). L2, L3, L4 are low quality commercial sites.

The images on both sides are unrelated to this Mexican holiday and thus are equally useless. I don't have a preference for either side.

L has more malicious pages (L3, L4, L5) but also more helpful results – L1, L2 are really helpful information results. R side is bad- R1, R2, R3 are all not related to the query.

The user probably wants to make a reservation at a hotel in Dallas. R has the top travel websites (Tripadvisor, Travelocity, Orbitz, etc) and L has low quality travel sites and some malicious pages. I tried to make a reservation on L1 and L2, but wasn't able to.

Bad Comment Examples		
Hard to read—too many symbols	L1=R5, R5=L2, L4=R4, R1 >>L3, L5 <r4, r2="FailsM&lt;/td"></r4,>	
	I prefer the left side.	
Non-specific and	Better order on L.	
unhelpful. These comments do not	They are about the same.	
explain the rating or help us understand	Left side is more helpful	
which results are helpful.	Right side is better.	
	The left side is better than the right side because the results on the left side are more satisfying for the query. The results on the right side are not as satisfying as the results on the left side.	
Long and unhelpful. This comment does not explain the rating or help us understand which results are helpful.	On the right side, a few of the results have the complete biography of Barack Obama, which I like because I am a Barack Obama fan and I think any result about Barack Obama should tell the story of his life. I also liked L2 because of the picture of Barack Obama and his family. They look really nice and I love the dress Michelle is wearing. L10 is really helpful because it explains exactly what kind of dog Barack Obama is looking for, even though the query doesn't ask for it but I think it's important to consider anyway. And I didn't like L5 because one link on the left wasn't working when I tried it but I think all the others are working so maybe it isn't so bad but I still didn't like the site.	

# Part 2: Side-by-Side Rating Examples

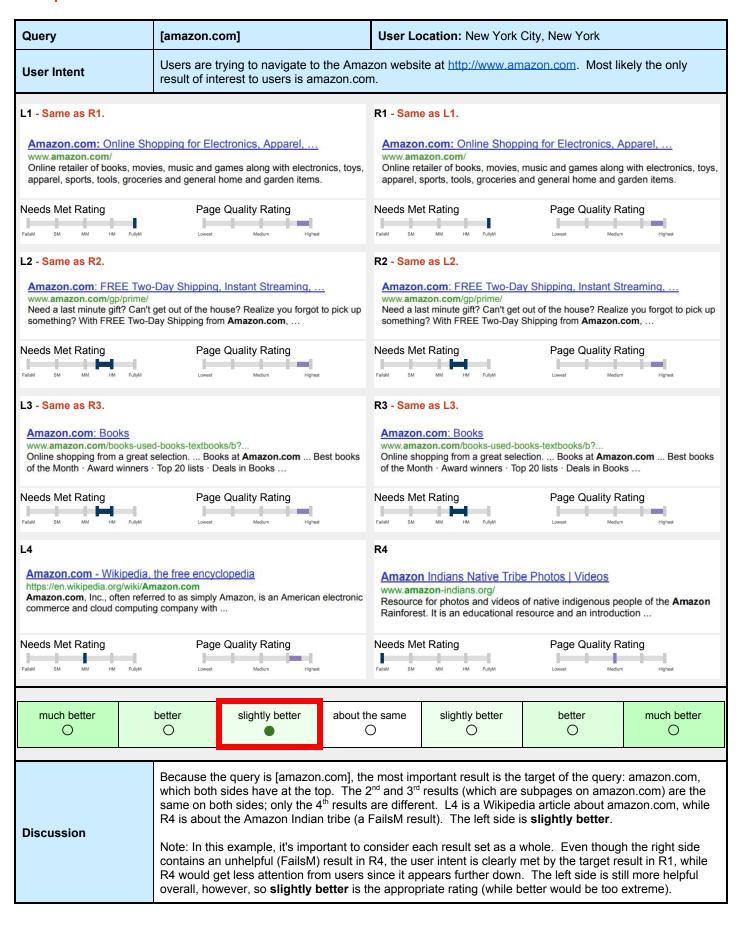
This section has some standard Needs Met SxS rating examples and rating scenarios.

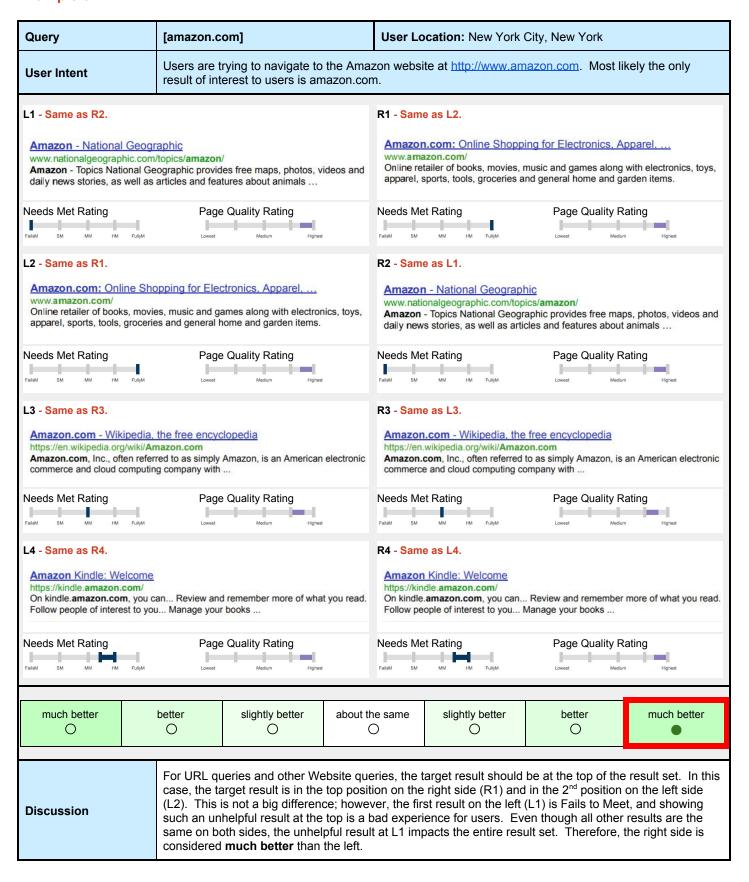
Each example will include a User Location in the form of a city and state. In cases where the location doesn't matter, it will show as New York City, New York. Note that in actual rating tasks, you will need to infer the User Location based on the map that is provided, as discussed in Section 28.0 of the General Guidelines.

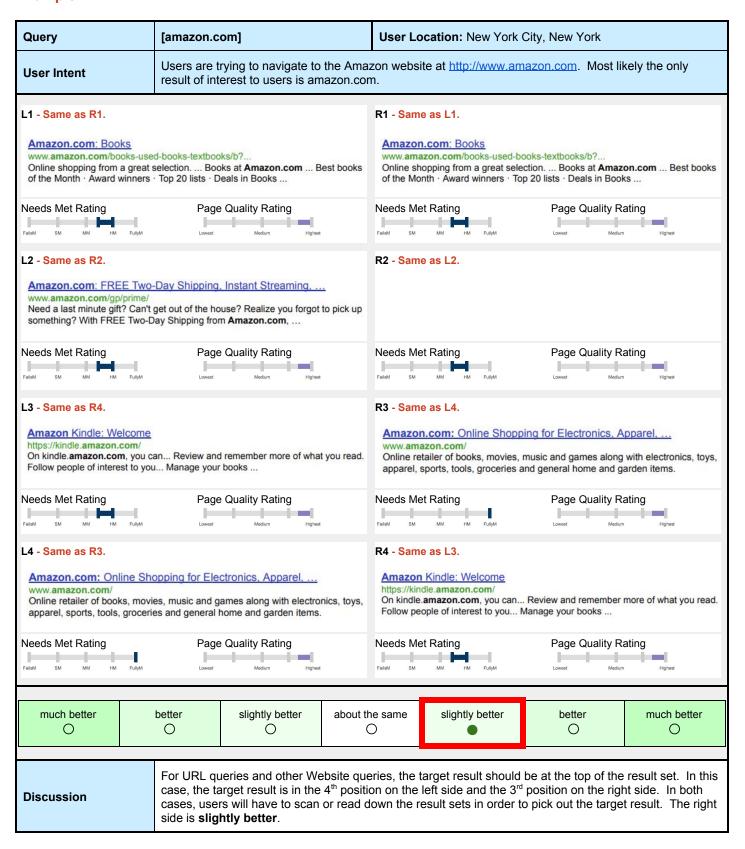
### 3.0 Standard Needs Met SxS Rating Examples

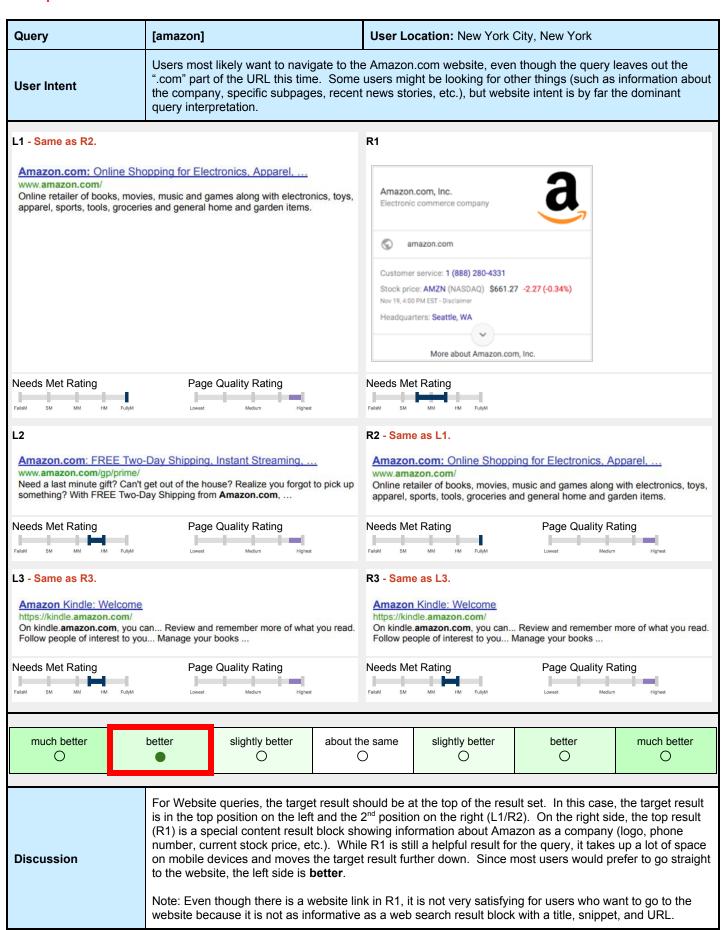
Please note that the result blocks in the following examples are not clickable, and we are not asking you to try to visit the landing pages. Also keep in mind that websites change rapidly, so we use images or "snapshots" of result blocks and landing pages throughout. Please focus on the concepts illustrated by each example rather than particular website URLs or images.

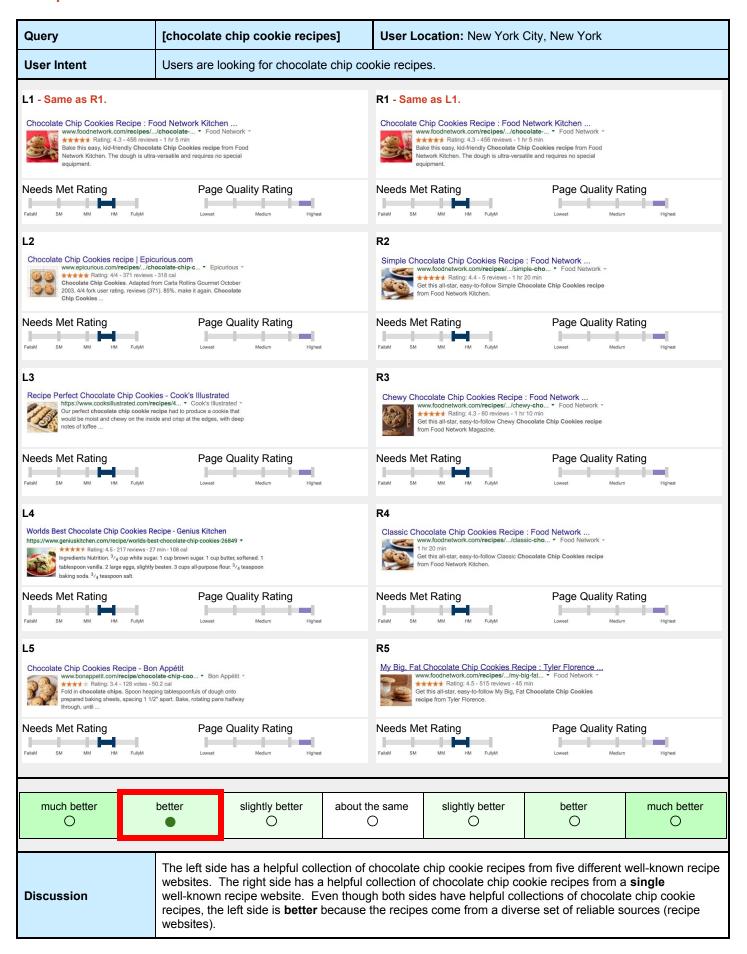
Query	[amazon.com]	User Location: New York City, New York	
User Intent	Users are trying to navigate to the Amazon website at <a href="http://www.amazon.com">http://www.amazon.com</a> . Most likely the only result of interest to users is amazon.com.		
L1 - Same as R4.		R1 - Same as L4.	
Amazon.com - Wikipedia. https://en.wikipedia.org/wiki/Arr Amazon.com, Inc., often refer commerce and cloud computin	nazon_(company) red to as simply Amazon, is an American electronic	Amazon.com: Online Shopping for Electronics, Apparel, www.amazon.com/ Online retailer of books, movies, music and games along with electronics, toys, apparel, sports, tools, groceries and general home and garden items.	
Needs Met Rating	Page Quality Rating  Lowest Medium Highest	Needs Met Rating Page Quality Rating  FallaM SM MM HM FullyM Lovest Medium Highest	
L2 - Same as R2.		R2 - Same as L2.	
www.amazon.com/gp/prime/ Need a last minute gift? Can't	-Day Shipping, Instant Streaming, get out of the house? Realize you forgot to pick up bay Shipping from Amazon.com,	Amazon.com: FREE Two-Day Shipping, Instant Streaming, www.amazon.com/gp/prime/ Need a last minute gift? Can't get out of the house? Realize you forgot to pick up something? With FREE Two-Day Shipping from Amazon.com,	
Needs Met Rating	Page Quality Rating	Needs Met Rating Page Quality Rating	
FailsM SM MM HM FullyM	Lowest Medium Highest	FallsM SM MM HM FullyM Lowest Medium Highest	
L3 - Same as R3.  Amazon to open second brick and mortar store - USA TODAY http://www.usatoday.com > Tech Mar 7, 2016 - Amazon plans to open a second brick and mortar bookstore, in San Diego, this summer. The online sales giant opened its		R3 - Same as L3.  Amazon to open second brick and mortar store - USA TODAY http://www.usatoday.com > Tech Mar 7, 2016 - Amazon plans to open a second brick and mortar bookstore, in San Diego, this summer. The online sales giant opened its	
Needs Met Rating	Page Quality Rating  Lowest Medium Highest	Needs Met Rating Page Quality Rating  FallsM SM MM HM FullyM Lowest Medium Highest	
L4 - Same as R1.		R4 - Same as L1	
Amazon.com: Online Shopping for Electronics, Apparel, www.amazon.com/ Online retailer of books, movies, music and games along with electronics, toys, apparel, sports, tools, groceries and general home and garden items.		Amazon.com - Wikipedia, the free encyclopedia https://en.wikipedia.org/wiki/Amazon_(company) Amazon.com, Inc., often referred to as simply Amazon, is an American electronic commerce and cloud computing company with	
Needs Met Rating Page Quality Rating  FallaM SM MM HM FullyM Lowest Medium Highest		Needs Met Rating Page Quality Rating  FallsM SM MM HM FullyM Lowest Medium Highest	
much better		ne same slightly better better much better	
Discussion	case, the target result is in the top positi	eries, the target result should be at the top of the result set. In this ion on the right (R1), but in the 4 <sup>th</sup> position on the left (L4). Users amazon.com on the right side, so the right side is <b>much better</b> .	



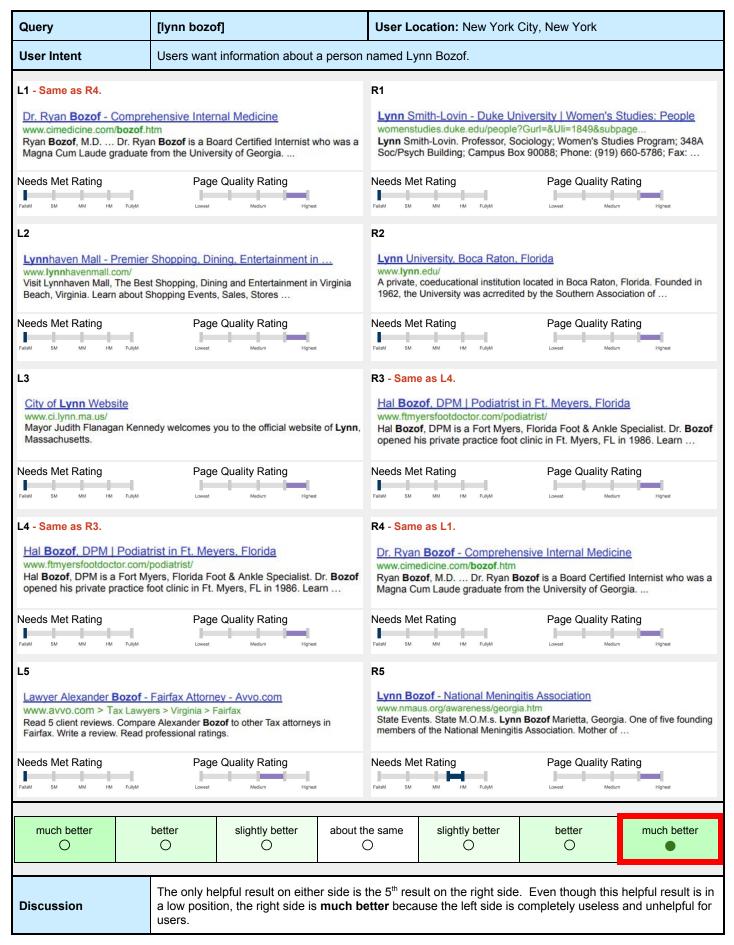


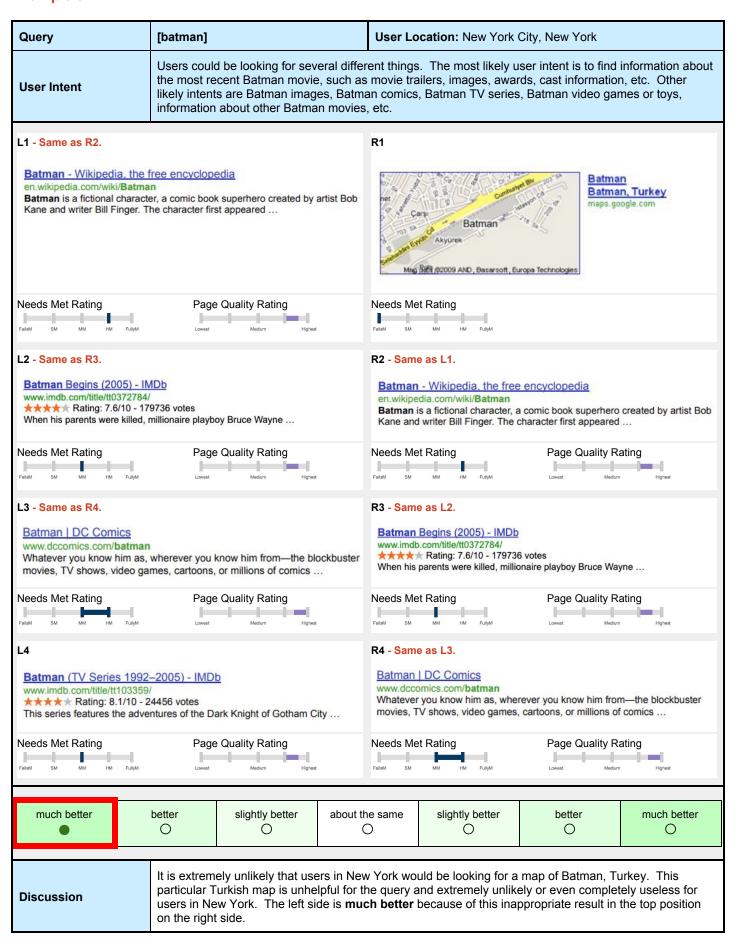


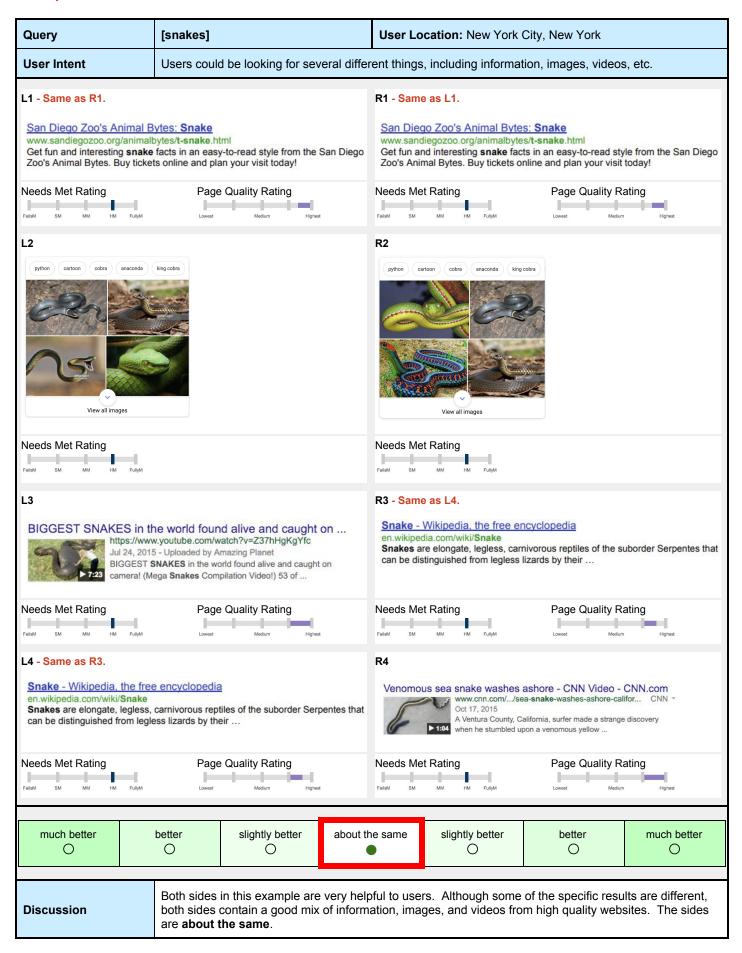


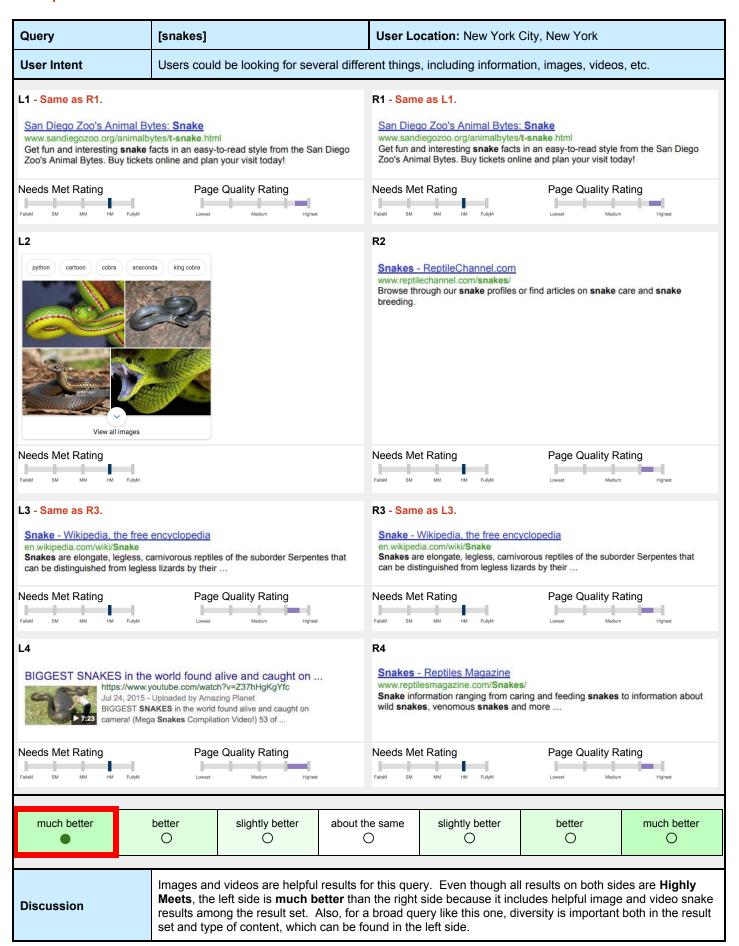


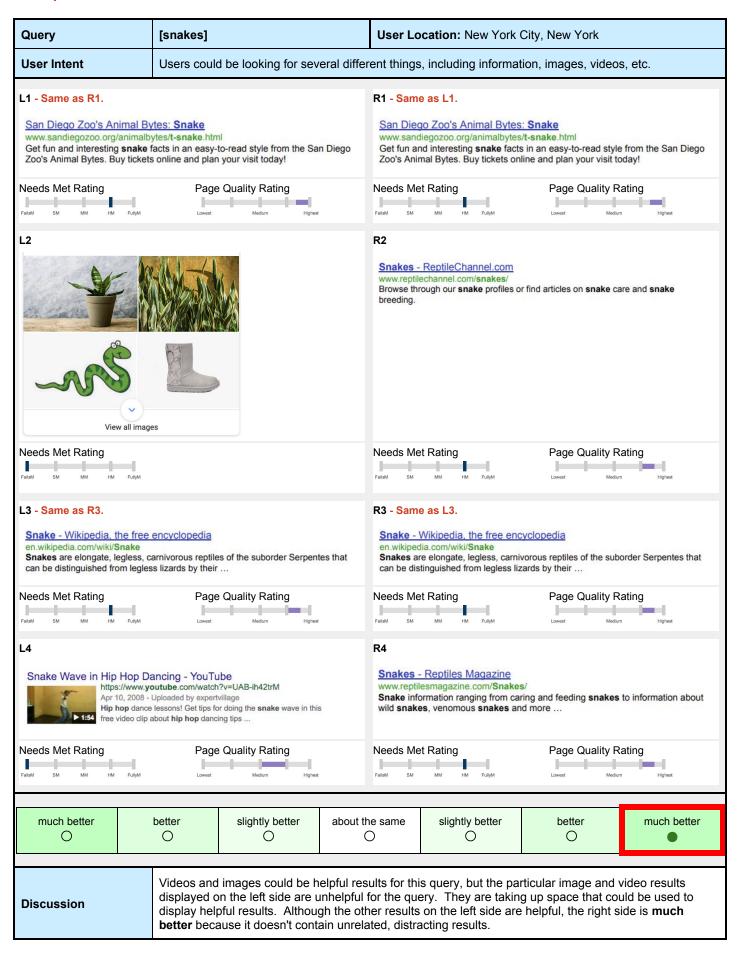
Query	[karate 7304 lakewood dr w, lakewood, wa]	User Location: Lakewood, Washington
User Intent	Users want information about the karate organization at this address in Washington state or to go to the organization's homepage.	
	artial Arts - Yelp ness & Instruction > Martial Arts	Champion Karate American Kenpo System   Lake Martial Arts   Www.championkaratefl.com   Champion Karate is a family oriented martial arts school in Lake Mary that teaches adults and children of all abilities in specialized programs.  Needs Met Rating   Page Quality Rating   Faladi   BM   MM   IM   FalyM   Page   Page   Page   Page   Rarate classes in Denver, Colorado at Denwakan are friendly but formal. Come in for a free class and see for yourself!  Needs Met Rating   Page Quality Rating   Faladi   BM   MM   IM   FalyM   Page   Page   Page   Www.jpan/smaratialarts.com/   Welcome to J.G. Parks US Taekwondo Academy - the best provider of martial arts classes in the Lakewood, Long Beach and Cerritos area  Needs Met Rating   Page Quality Rating   Faladi   BM   MM   IM   FalyM   Indeed.com   Www.indeed.com/l-Lakewood, WA   Indeed.com   Www.indeed.com/l-Lakewood, WA
		Korean Karate Center - Lakewood, Ohio www.koreankaratecenter.com/ At the Korean Karate Center, whether Sa Bom Nim Marlene Kachevas is teaching men, women, or children, the focus is on the total development
		Needs Met Rating Page Quality Rating  FalsM SM MM HM FullyM Lowest Medium Highest
much better	better slightly better	about the same Slightly better better much better
Discussion	right side has five results, but t	but it contains information that would be very helpful to most users. The ey are all unrelated to the query and completely unhelpful to users. Ever a result, it is <b>much better</b> than the right side.

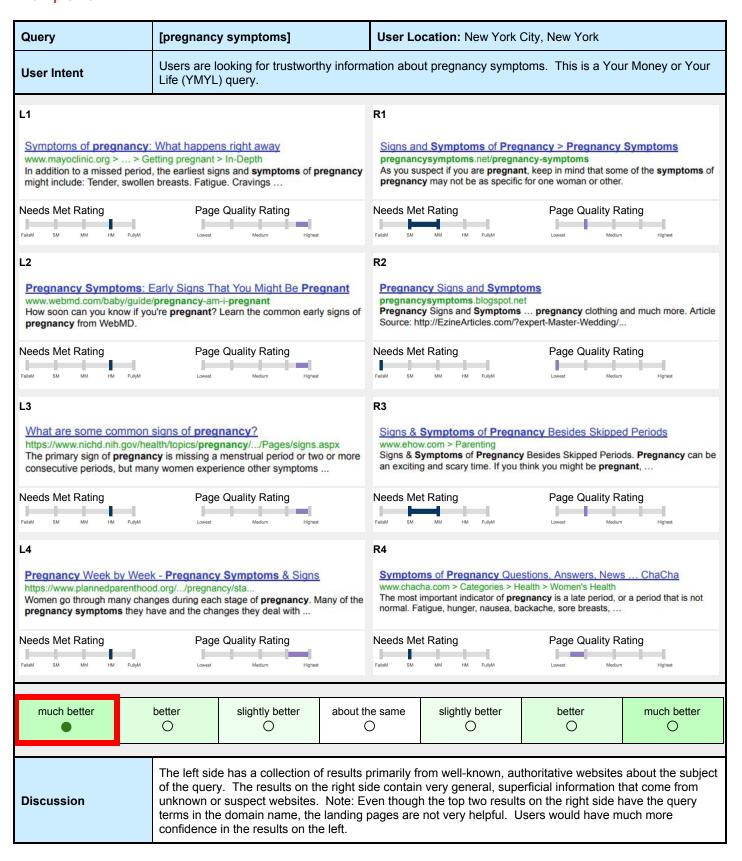


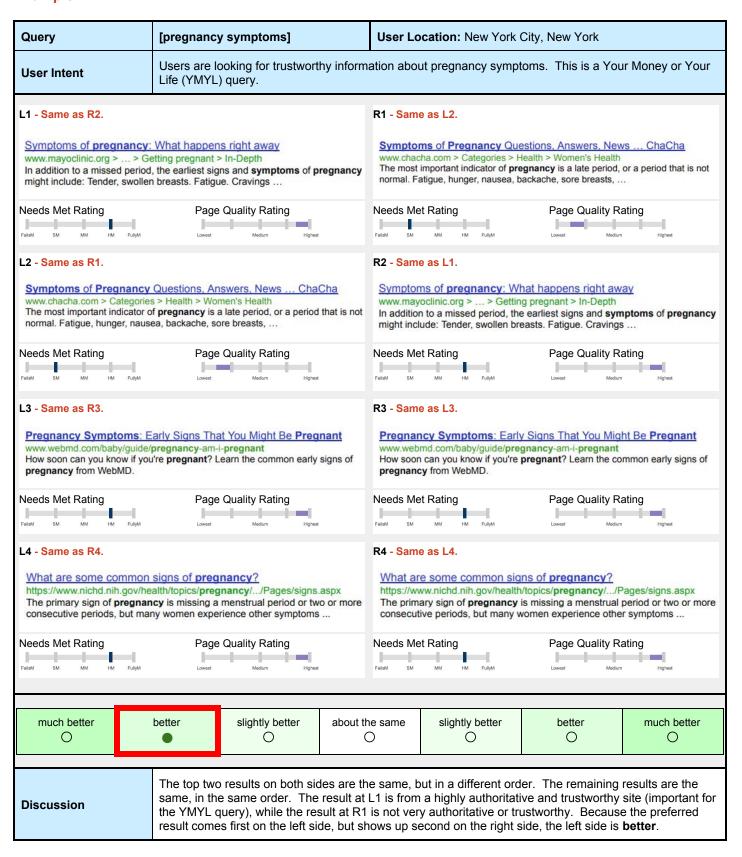


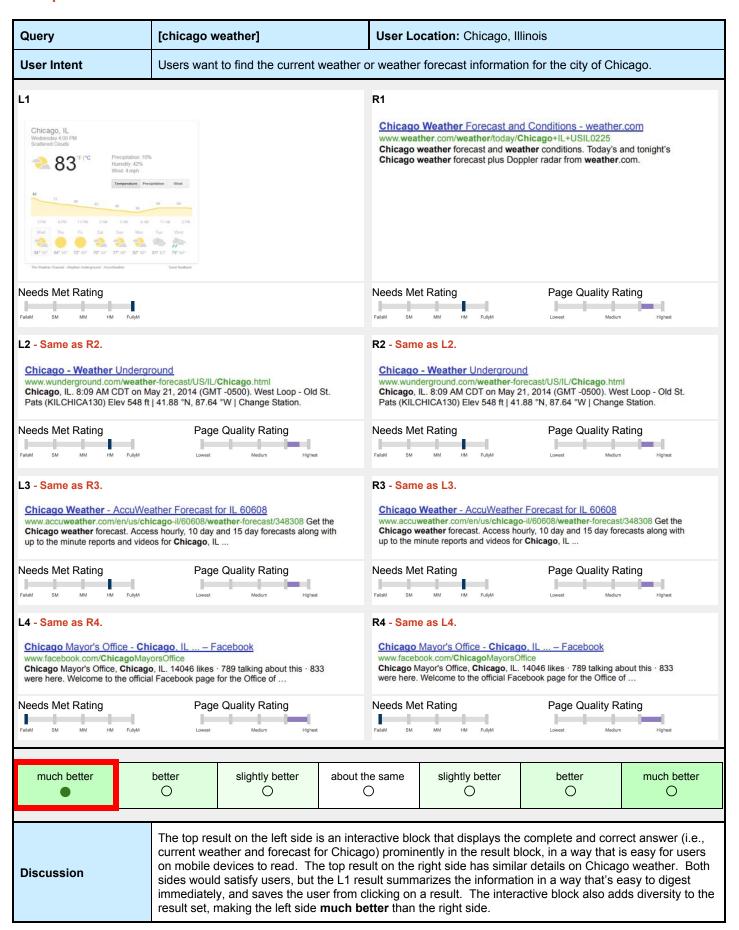


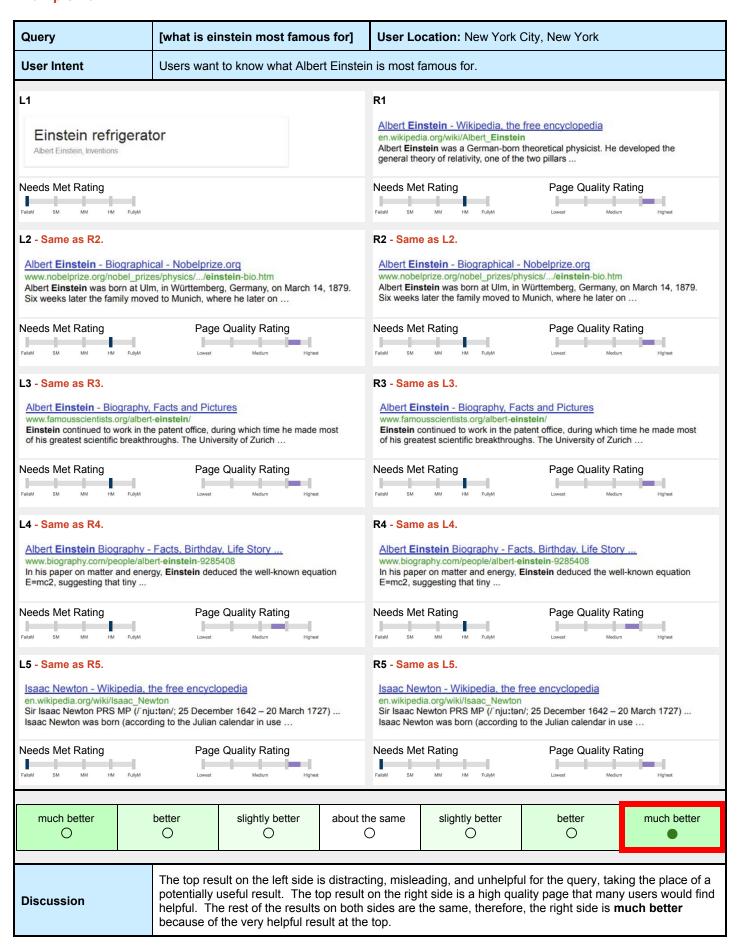


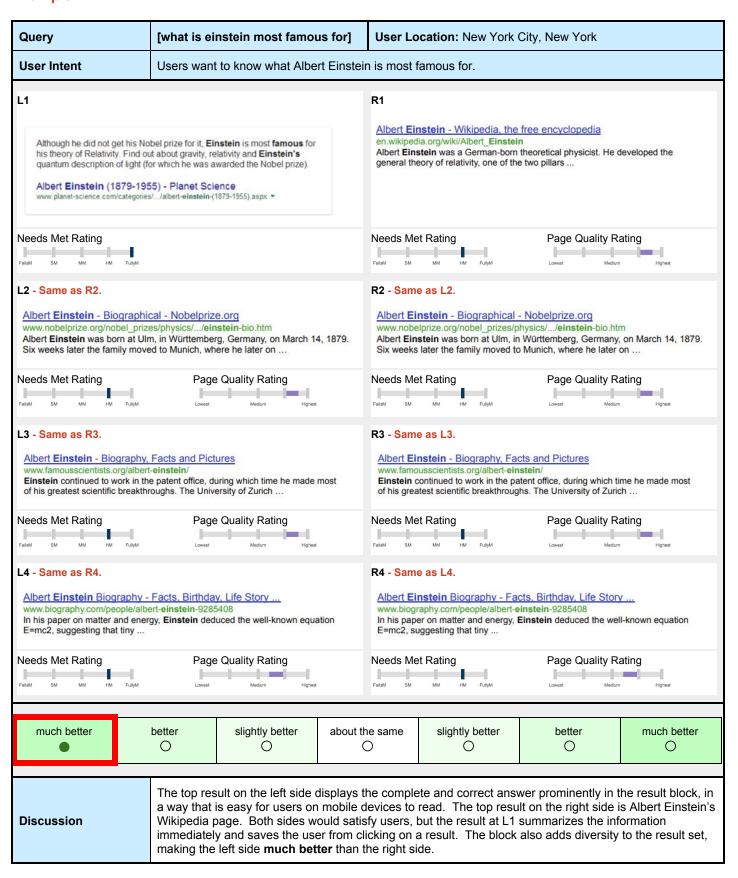


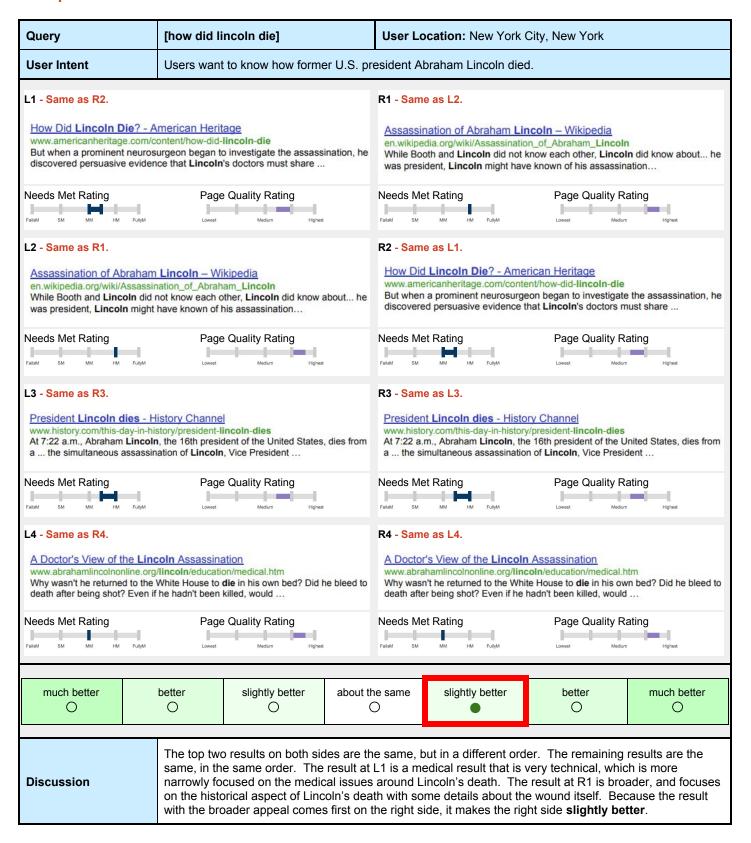


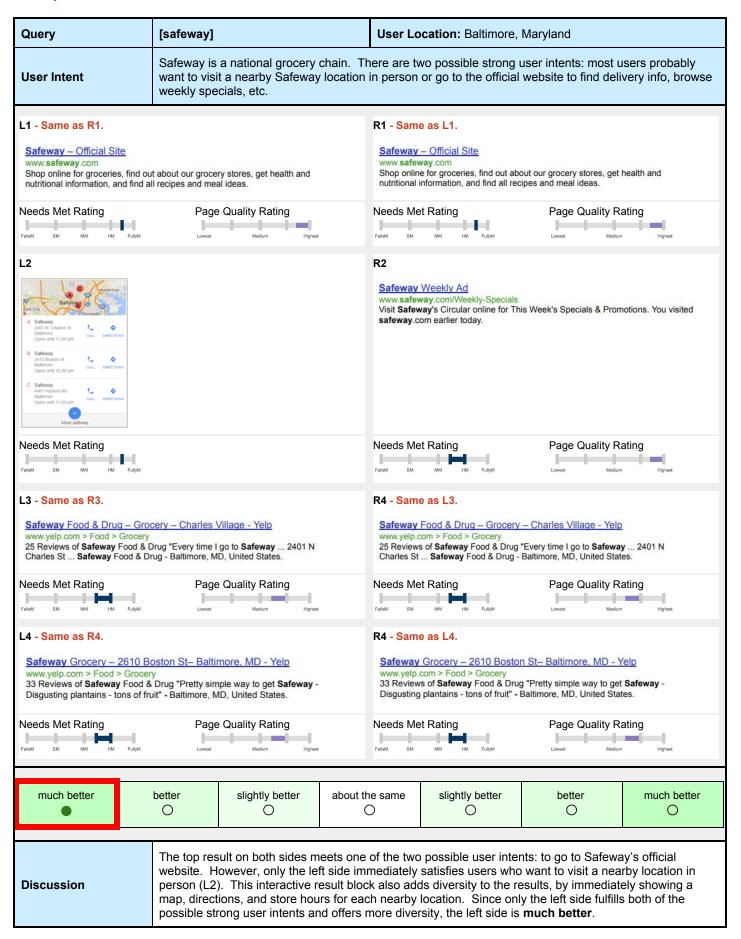


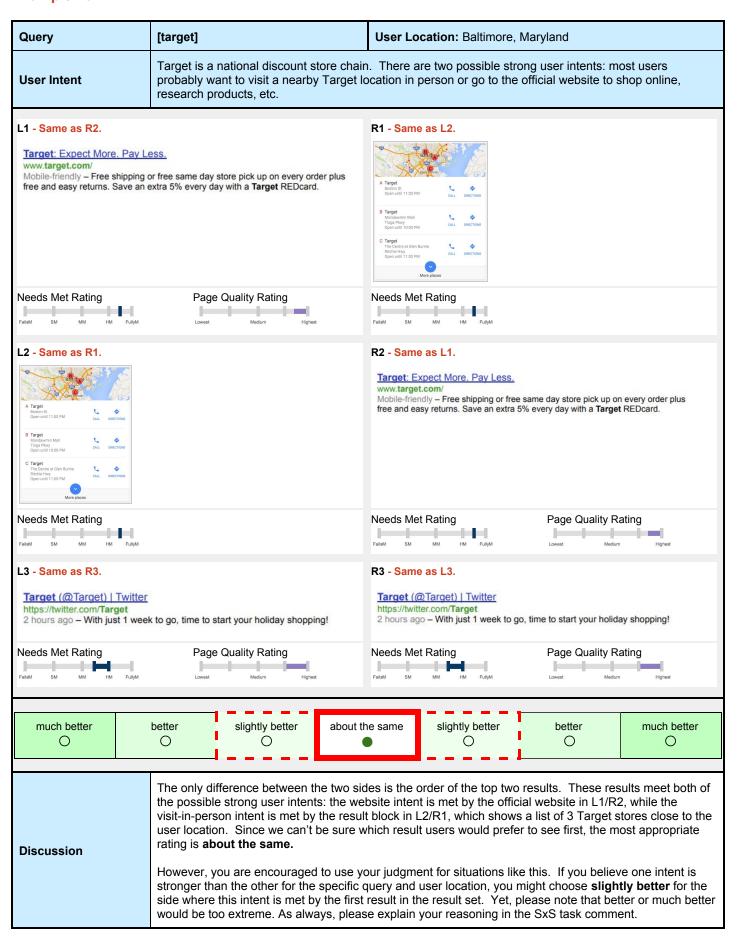


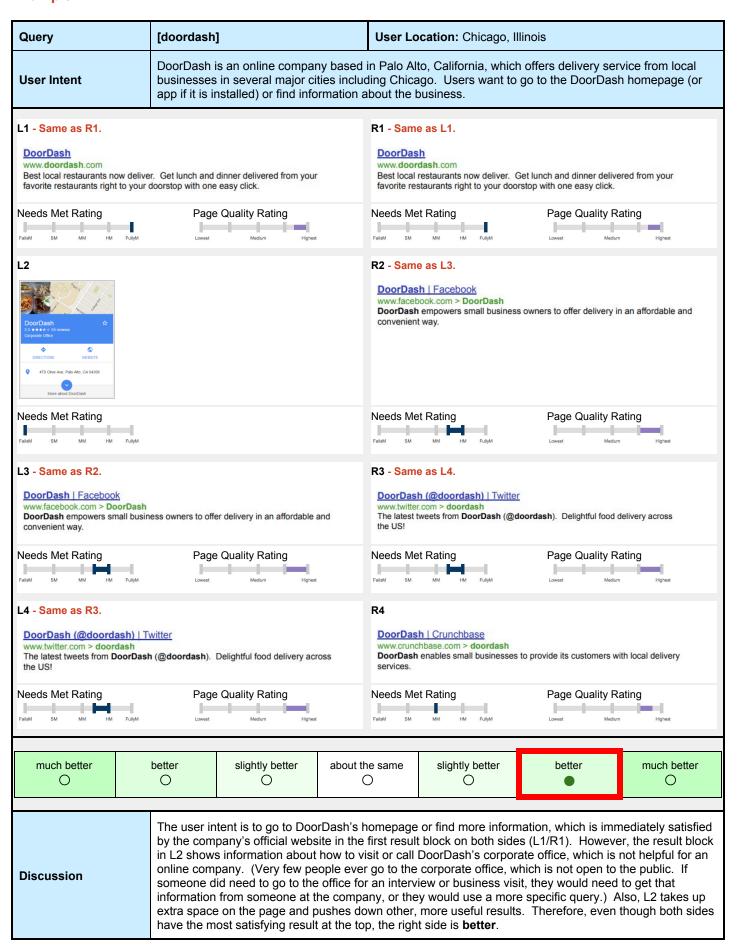


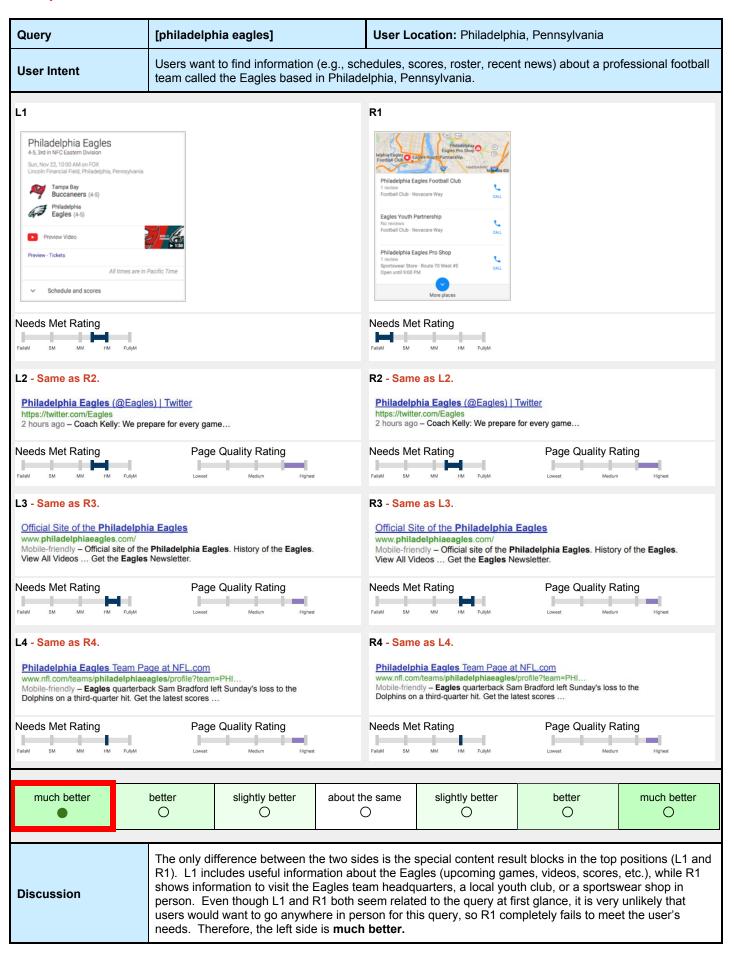


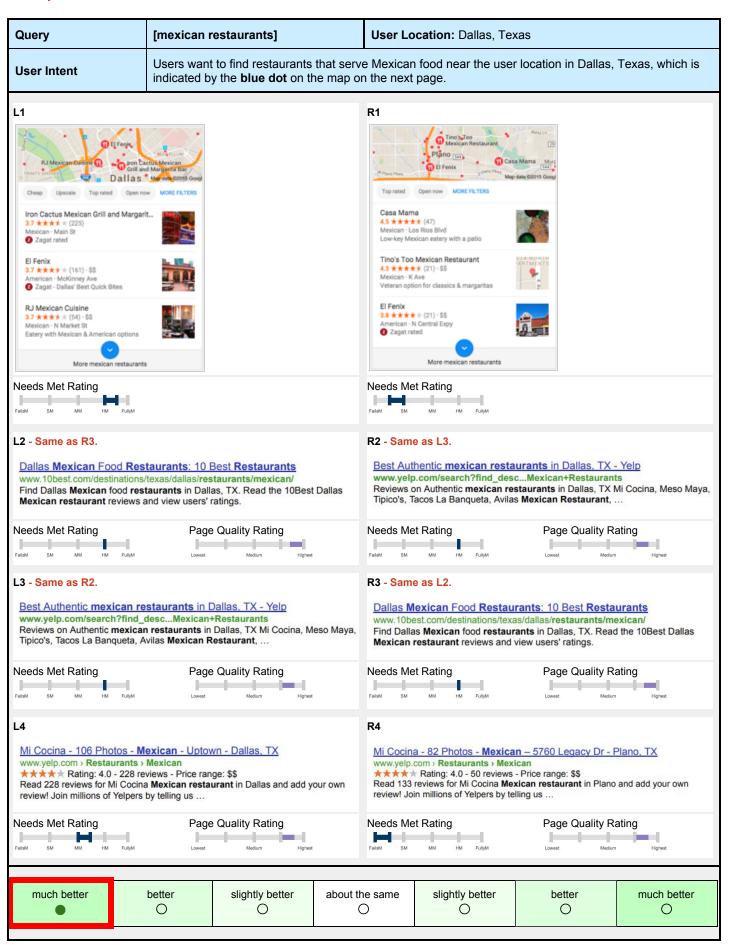




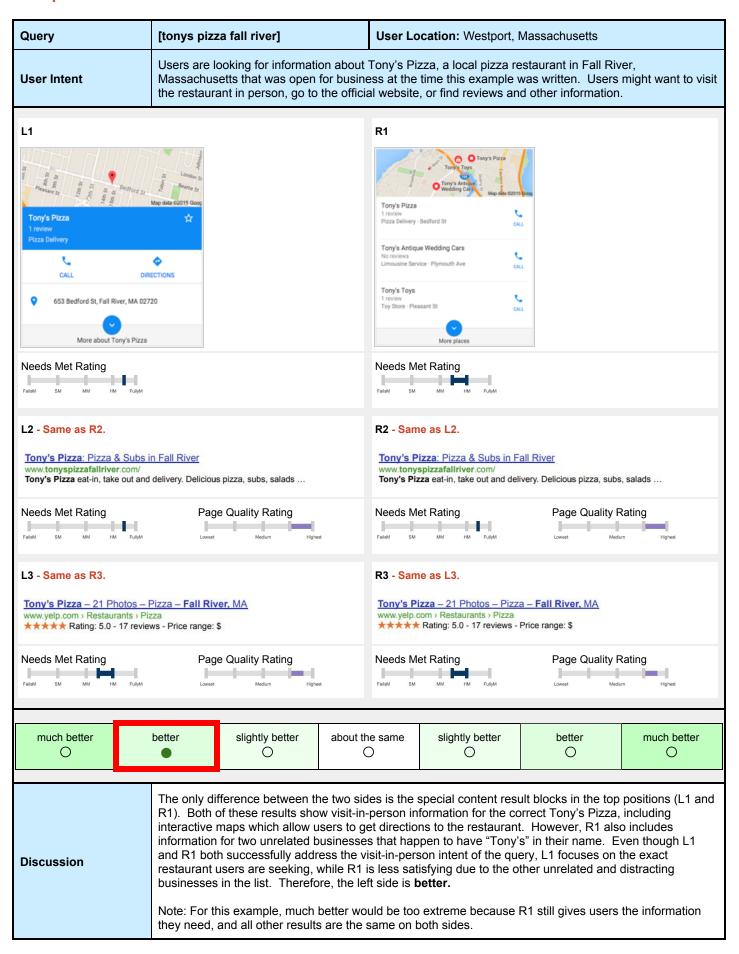


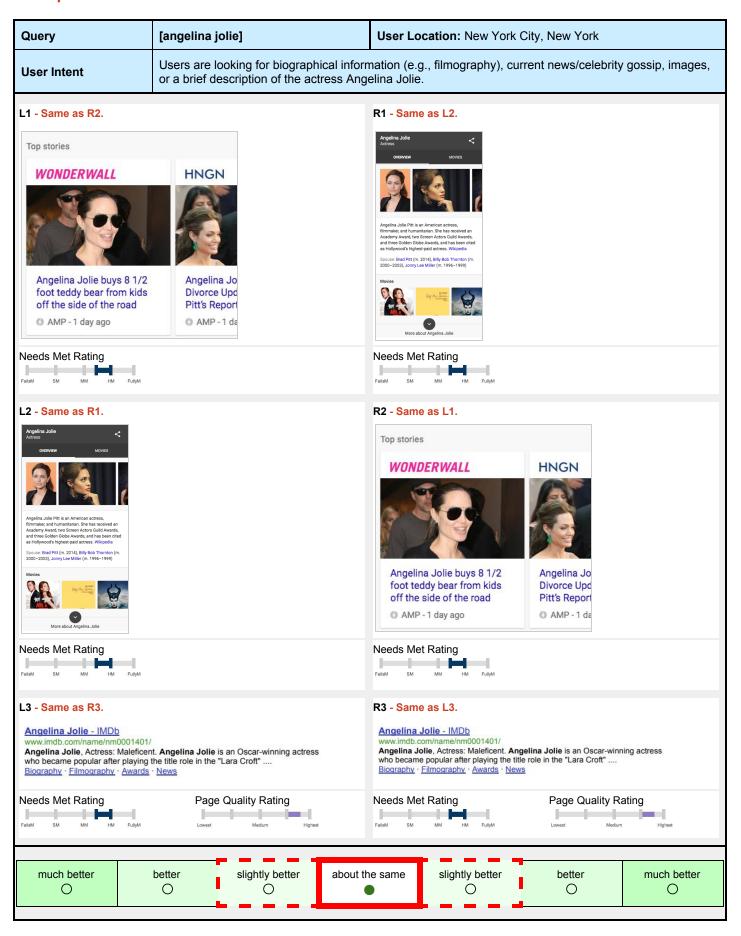






Query	[mexican restaurants]	User Location: Dallas, Texas
User Intent	Users want to find restaurants that serve Mexican food near the user location in Dallas, Texas, which is indicated by the <b>blue dot</b> on the map below.	
Discussion	Users searching for Mexican restaurants in Dallas would like to see a list of choices nearby. There are many high-quality and diverse options in the city, therefore most users would not be willing to drive a long distance.  The top result on both sides shows interactive blocks with a list of options. However, the restaurants in R1 are in the suburb of Plano. As you can see in the map, Plano is located outside of Dallas, approximately 20 miles north of the user location (blue dot).  Similarly, L4 and R4 are Yelp pages for two different locations of a popular Mexican restaurant (Mi Cocina), but R4 is in Plano. Since users would prefer Mexican restaurants nearby, the left side is much better for showing results in the user location.	Carrollton  Richaldson  Branch  Carland  University Park  Park  Balch Spring  Plano  Murphy  Richaldson  Richaldson  Richaldson  Branch  Garland  S35  Mesquest  1 Prairie  S56  Balch Spring





For very broad Know queries like this one, different users may be looking for different things: biographical information, news, images, etc. For this reason, different users may prefer different types of results at the top—no exact ordering would be perfect for all users.

#### **Discussion**

All of the results on both sides are highly satisfying. The only difference between the two sides is in the order of the interactive SCRBs in the top two positions. The SCRB in L1/R2 shows a selection of recent news stories focused on Angelina Jolie. The SCRB in L2/R1 shows a variety of helpful content about her directly on the page with an option to see more. While no single particular result will satisfy all users, most users could easily find a satisfying result for their needs in either set of results. Therefore, since either side would offer a highly satisfying user experience and the overall differences in user experience are small, the most appropriate rating is **about the same**.

However, there is some room for judgement in situations like this. If the recent news stories shown are particularly interesting, you might choose **slightly better** for the left side where the news appears at the top. Alternatively, if the recent news seems less interesting, you might decide the right side is **slightly better**. Yet, please note that better or much better for either side would be too extreme for this particular set of results.

### 4.0 Standard Needs Met SxS Rating Scenarios

Here are some examples of standard Needs Met SxS rating task scenarios including general advice you should follow. Please assume that these tasks have 5 results on each side.

We are aware that these scenarios do not include all the information you would have in real rating tasks, but we hope they provide some additional guidance. We understand that raters will have different opinions and that you may disagree with our suggested ratings.

Please note that this is just a small sample of the types of rating scenarios you may encounter, and that we are providing general advice, not strict rules.

Scenario 1		
Query	[nintendo.com]	User Location: New York City, New York
User Intent	This is a <b>Website</b> query. The target of the query is www.nintendo.com.	
Description	<ul> <li>One side has the www.nintendo.com result in the 5<sup>th</sup> position, with nintendo.com subpages and news/information pages about Nintendo in the other positions.</li> <li>The other side has nintendo.com subpages and news/information pages about Nintendo, but does not have the target of the query.</li> </ul>	
Rating	The side with the target of the query is <b>much better</b> than the other side, even though the result is not at the top of the result set.	

Scenario 2		
Query	[nintendo.com]	User Location: New York City, New York
User Intent	This is a Website query. The target of the query is	www.nintendo.com.
Description	<ul> <li>The target result www.nintendo.com is at the top on both sides.</li> <li>The second result on one side is a porn result.</li> <li>The second result on the other side is a Wikipedia article on leisure activities in general, which is completely unhelpful for the query intent.</li> <li>The other results are the same on both sides.</li> </ul>	
Rating	The side without the porn result is <b>much better</b> . Even though the target result is at the top on both sides, displaying unsolicited porn leads to such a poor user experience that the <b>much better</b> rating is appropriate.  Note: The same reasoning should apply to other types of results that cause very poor user experiences, such as malicious download pages, results with unsolicited upsetting or offensive content, etc. Please use your judgment and represent users in your locale.	

Scenario 3	Scenario 3		
Query	[how long does a teenager have to pay the fine for a speeding ticket in carrollton, texas]	User Location: New York City, New York	
User Intent	The user wants to know the answer to this question.		
Description	None of the results on either side answers the question.		
Rating	The sides are equally bad or unhelpful. The rating should be <b>about the same</b> .		

Scenario 4		
Query	[britney spears]	User Location: New York City, New York
User Intent	Users are probably looking for Britney Spears' official homepage, gossip, images, videos, recent news, etc.	
Description	Both sides have excellent collections of results for the query, including her homepage in the top position, important recent articles, great information pages, images, etc.	
Rating	The sides are equally helpful. The rating should be <b>about the same</b> .	

Scenario 5	Scenario 5		
Query	[britney spears]	User Location: New York City, New York	
User Intent	Users are probably looking for Britney Spears' offic	ial homepage, images, videos, recent news, etc.	
Description	<ul> <li>The left side has her official homepage in the top position.</li> <li>The right side has a special content result block showing recent news stories in the top position, followed by her official homepage further down in the results.</li> <li>The other results on both sides are equally helpful and diverse, including more news stories, images, videos, music, and excellent information pages.</li> </ul>		
Rating	The best rating depends on the helpfulness of recent news about Britney Spears, which changes on a daily basis.  If the news stories indicate that she recently made breaking news or released a new album at the top of the charts, you might rate the right side as <b>better</b> since showing news stories directly in the results would be very useful for many users, or even <b>much better</b> if the news is truly stunning and worthy of front-page headlines. If the news stories are mostly random gossip or trivial articles, you might prefer the left side with her official homepage at the top since the news is not especially fresh or helpful.  Please explain your reasoning in your SxS task comment in situations like this.		

Scenario 6		
Query	[map of inca]	User Location: New York City, New York
User Intent	The user wants to find a map of the Inca Empire.	
Description	<ul> <li>One side has three results that display maps of the Inca Empire and are easy to navigate on a mobile phone. The remaining two results on the side are not very helpful.</li> <li>The other side has no maps of the Inca Empire, but one result displays a map of Inca Street in Denver, Colorado, a highly unlikely interpretation.</li> </ul>	
Rating	The side which has three map results of the Inca Empire should be rated <b>much better</b> .	

Scenario 7		
Query	[apple]	User Location: New York City, New York
User Intent	<ul> <li>The dominant query interpretation is Apple Inc., maker of computers, iPhones, iPods, etc.</li> <li>Many users want to go to the Apple company homepage www.apple.com.</li> <li>Some users might want information about Apple products, such as the computers, iPhones, iPods, etc.</li> <li>There are minor interpretations for this query as well, such as the apple fruit or the singer Fiona Apple.</li> </ul>	
Description	<ul> <li>Both sides have apple.com at the top.</li> <li>Both sides have a helpful result for the iPhone and iPod.</li> <li>One side has Fiona Apple's official homepage at the bottom of the result set.</li> </ul>	
Rating	Most users are looking for the Apple website or information on Apple products. It's important that the result sets do a good job with this interpretation. A few people might be interested in results about Fiona Apple, but it's unlikely.  Diversity of query interpretation isn't as important for this query. The correct rating is <b>about the same</b> or perhaps <b>slightly better</b> for the side that contains the result about Fiona Apple.	

Scenario 8	Scenario 8		
Query	[ADA]	User Location: New York City, New York	
User Intent	<ul> <li>There are many common interpretations of this query. The four most common are American Dental Association, American Diabetes Association, Americans with Disabilities Act and American Dietetic Association. There are many other less likely possibilities as well, such as Auto Design Association, American Dance Awards, Alternative Distribution Alliance, etc.</li> <li>Users are either looking for information about one of these interpretations or are trying to navigate to the homepage of one of these interpretations.</li> </ul>		
Description	<ul> <li>One side has a diverse mix of very minor interpretations.</li> <li>The other side only has results for the American Dental Association, including the homepage of the organization at the top and informational results below.</li> </ul>		
Rating	Neither side is helpful. Because there are so many possible meanings for this query, the result set should be diverse in query interpretation. But, it's also important for the result set to include the most likely interpretations.		
	This kind of task is hard, and the exact rating would depend on the actual results. But here is the question to ask: which side will be helpful for more users? If you think more users would be more likely to find what they are looking for on the side with the common interpretation, use <b>slightly better</b> or <b>better</b> . If you think both sides are equally bad, choose <b>about the same</b> . Our rating choice here is <b>slightly better</b> for the side with the American Dental Association interpretation.  Because both sides are bad, the <b>much better</b> rating is too extreme. For some or many users, both sides would be equally unhelpful.		
	When the task is difficult to rate, make sure you explain your choice.		

Scenario 9		
Query	[wizard of oz]	User Location: New York City, New York
User Intent	The user is most likely looking for information about the popular movie, but minor interpretations include the Andrew Lloyd Webber musical or the novel.	
Description	<ul> <li>The left side has helpful results about the movie, which is the dominant interpretation for this query.</li> <li>The right side has results from several local, very small theater groups performing the Wizard of Oz musical. Some of the results are stale, including the top two results, because the performance dates have passed.</li> </ul>	
Rating	In this example, both sides have only addressed one interpretation. The left side is the dominant interpretation and has very helpful results. The right side addresses a minor interpretation with unhelpful results at the top. Therefore, the left side is <b>much better</b> than the right side.	

Scenario 10		
Query	[red wine vs. white wine making]	User Location: New York City, New York
User Intent	The user wants information on this topic.	
Description	<ul> <li>The first result on both sides is a Wikipedia article on wine in general, which is not helpful for the query intent.</li> <li>The second result on the left side doesn't load, while the second result on the right side is an article that explains the differences between making red and white wines.</li> <li>The rest of the results on both sides are the same. Of those results, only the 5th result on both sides is helpful.</li> </ul>	
Rating	The right side should be rated <b>better</b> or <b>much better</b> than the left side because it has one of the two helpful results in a high position.	

Scenario 11		
Query	[plane crash john f kennedy jr]	User Location: New York City, New York
User Intent	John F. Kennedy, Jr. died in a plane crash in 1999. The user wants information on this topic.	
Description	<ul> <li>The left side has 5 informative newspaper articles on the subject of the query. All of the articles were published in 1999 and 2000.</li> <li>The right side has 5 informative results about other members of the Kennedy family. All of the articles were published within the last month.</li> </ul>	
Rating	The left side should be rated <b>much better</b> than the right side because it has articles related to the topic of the query. Because the plane crash occurred in 1999, it is logical that most helpful results would have been published around that time. Very recent articles that are unrelated to the topic of the query are not helpful.	

Scenario 12		
Query	[olympic figure skating results]	User Location: New York City, New York
User Intent	The user wants information on this topic.	
Description	<ul> <li>The left side has Olympic figure skating results from 2006 and 2010 only. There are no results from the most recent Winter Olympics.</li> <li>The right side has Olympic figure skating results from the most recent Winter Olympics (which took place in 2016 at the time this scenario was written).</li> </ul>	
Rating	Unless the user specifies otherwise, we will assume that he or she is interested in the most recent results. The right side is <b>much better</b> than the left because it displays results from the most recent Winter Olympics.	

Scenario 13		
Query	[olympic figure skating results 2002]	User Location: New York City, New York
User Intent	The user wants information on this topic.	
Description	<ul> <li>The left side has Olympic figure skating results for 2002.</li> <li>The right side has Olympic figure skating results from the most recent Winter Olympics (which took place in 2016 at the time this scenario was written).</li> </ul>	
Rating	The user is specifically interested in Olympic figure skating results for 2002. The left side is <b>much better</b> than the right side.	

Scenario 14		
Query	[toys]	User Location: New York City, New York
User Intent	The user wants to find children's toys to buy.	
Description	<ul> <li>All results on the left are related to children's toys. There are three shopping-related results and two news results with articles about toy safety.</li> <li>Four of the results on the right are the same as or similar to the results on the left side.</li> <li>The right side has one result related to adult sex toys.</li> </ul>	
Rating	Porn-related results are unacceptable for this query which clearly has non-porn intent. The left side should be rated <b>much better</b> than the right side.	

Scenario 15	
Description	<ul> <li>One side has no search results and displays the message "This side did not generate any results".</li> <li>The other side displays search results, but none are related to the subject of the query.</li> </ul>
Rating	In this scenario, we have to ask ourselves whether it is better to display no search results or to display unhelpful search results.  When both sides are so obviously unhelpful, usually the best rating is <b>about the same</b> . In some situations, you might find results that are borderline <b>SM/FailsM</b> and could possibly be of use to some users. In this case, the side with results could be <b>slightly better</b> . On the other hand, you may encounter a situation where the results are so clearly Fails to Meet that they would just frustrate users or waste their time. In that case, the side with no results could also be <b>slightly better</b> . In any of these situations, both sides are still nearly useless and you should not use strong ratings. Please explain your reasoning in your SxS task comment in situations like this.

Scenario 16	Scenario 16		
Query	[google adsense]	User Location: New York City, New York	
User Intent	The user wants to go to the AdSense website, learn more about the program, or sign in to their account.		
Description	<ul> <li>The top result on both sides is the homepage, which is rated Fully Meets.</li> <li>The second result on the left side is a dupe of the top (Fully Meets) result.</li> <li>The bottom three results on each side are the same.</li> </ul>		
Rating	The Fully Meets result in the second position on the left has no added value, because it is a dupe of the top result. Because the right side has no dupes, and the bottom results are the same, the right side has an additional unique result. If R2 has a Highly Meets rating, then the right side is <b>much better</b> . If R2 has a Fails to Meet rating, then the two sides are <b>about the same</b> .  In this scenario, the final SxS rating depends on the Needs Met rating of the additional unique result since the other results are the same. Therefore, a rating in the range from <b>about the same</b> to <b>much better</b> could be justified. Please explain your reasoning in your SxS task comment in situations like this.		

Scenario 17		
Query	[cinemark san mateo]	User Location: Redwood City, California
User Intent	<ul> <li>The user is looking for information about a Cinemark movie theater in San Mateo, California, which is close to the user location.</li> <li>Most users want to visit the theater in person, buy tickets, and/or look up information in advance such as showtimes, prices, movie information, etc.</li> </ul>	
Description	<ul> <li>Both sides show helpful results about the correct movie theater, including special content result blocks in the top positions for the specific Cinemark location the user wants.</li> <li>The special content result block on the left shows movie showtimes for the theater.</li> <li>The special content result block on the right shows visit-in-person information for the theater, including an interactive map to get directions.</li> </ul>	
Rating	Both sides could potentially provide exactly what users are looking for when issuing this query. Some users might be immediately ready to visit the theater in person, while other users might need to see showtimes, purchase tickets, or find other information first.  This kind of task is hard, because it's difficult to know which situation applies to most users, and both sides contain excellent results for the query. You might choose <b>about the same</b> if you think showtimes and visit-in-person information are equally useful, or you might decide one side is <b>slightly better</b> if you think seeing one type of content directly in result blocks is more immediately helpful than the other. However, please note that in this type of scenario, better or much better would be too extreme.  Please use your best judgment in choosing a rating, and explain your reasoning in your SxS task comment.	

Scenario 18		
Query	[hotels in pacifica]	User Location: San Francisco, California
User Intent	The user is looking for information about hotels located in Pacifica, California, in order to research, call, visit, or make a reservation online.	
Description	<ul> <li>The top result on the left side is an interactive result block showing information about several hotels in Pacifica (including reviews, contact info, directions, etc. for each hotel).</li> <li>The top result on the right side is an interactive result block showing information for a single hotel in Pacifica (including similar information such as reviews, contact info, etc. for this specific hotel).</li> <li>The remaining results are the same on both sides.</li> </ul>	
Rating	Users are searching broadly for hotels in the area, rather than one specific hotel. Even if the single hotel shown on the right is a really excellent, top-rated hotel, it's important to give users many options and choices for this query. Therefore, the left side should be rated <b>much better</b> because the top/most prominent result offers multiple hotel options and choices to meet users' needs.	

# Part 3: Using the Evaluation Platform for SxS Rating

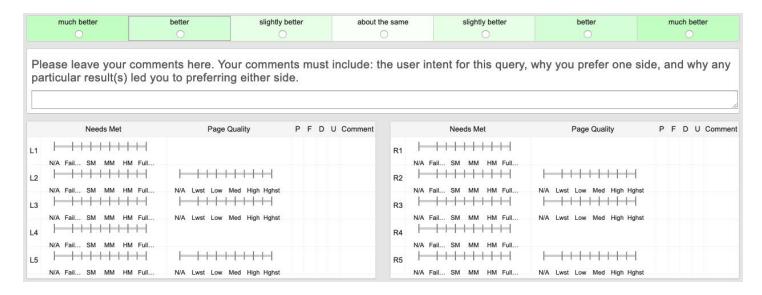
This section of the Side-by-Side Rating Guidelines is about aspects of the Needs Met template that pertain to SxS rating. Please see Part 4 of the <u>General Guidelines</u> for general information about using the Evaluation Platform.

### 5.0 The Needs Met SxS Rating Interface

The Needs Met SxS rating interface includes the following at the bottom of the task page:

- Side-by-Side rating scale
- Summary tables that show your Needs Met ratings, Page Quality ratings, flags, and comments for individual results

Here is a screenshot of this portion of the task page.



See <u>Section 1.1</u> for a screenshot of the full standard SxS Needs Met task page.