A large, stylized magnifying glass icon is positioned behind the title text. The handle of the magnifying glass extends from the bottom right towards the bottom left, pointing towards the title area. The lens is a light blue circle, and the handle is a darker blue.

Search Quality Rater Guidelines: An Overview

Download the full [Search Quality Rater Program Guidelines](#)

December 2022



Introduction

Section 01 How Search Works

Section 02 Improving Search

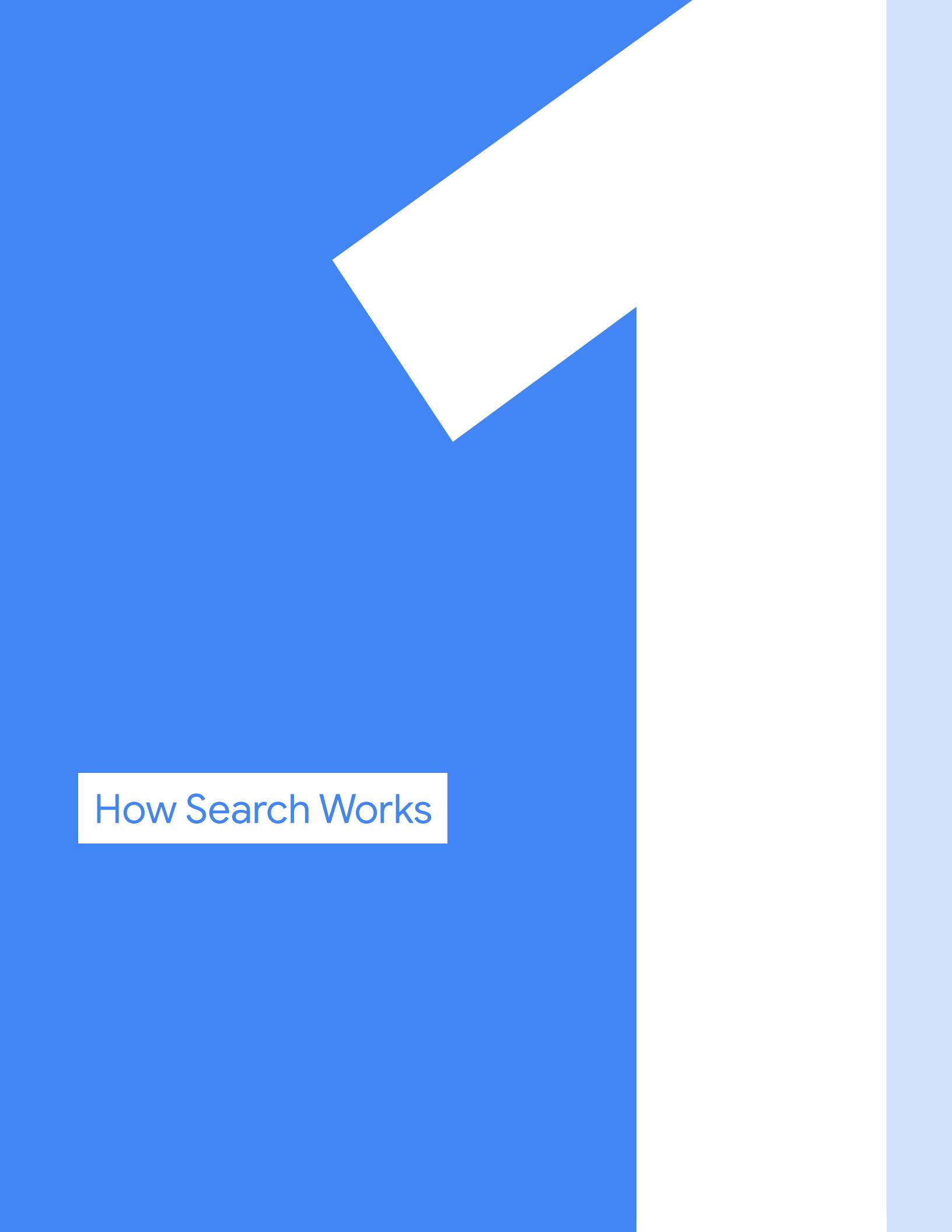
Section 03 The Search Quality Rating Process

Section 04 Summary

Search engines exist to help people find helpful, relevant, and reliable information. To do that, search engines must provide a diverse set of high quality search results, presented in the most helpful way.

At Google, we like to say that Search is not a solved problem: We're constantly making improvements to make Search work better for our users. We put all proposed improvements to our Search product through a rigorous evaluation process. This process includes soliciting feedback from "Search Quality Raters", who help us measure how people are likely to experience our results.

In this document, we briefly explain how Search works and how we improve Search through these search quality evaluations.



How Search Works

For many people, Google Search is a place they go when they want to find information about a question, whether it's to learn more about an issue or fact check a friend quoting a stat about last weekend's football game.

People use Search for billions of queries every day, and one of the reasons they continue to come to Google is they know that they can often find relevant, reliable information that they can trust. To help us do this, we rely on three key elements that inform our approach to information quality:



High-quality automated ranking: We design our automated ranking systems to identify information that people are likely to find useful and reliable.



Helpful Search features: We have developed a number of Search features that not only help our users make sense of all the information they're seeing online, but that also provide direct access to information from authoritative sources — including intergovernmental organizations or government entities.



Content policies: We have policies for what can appear in Search to make sure that we're showing high quality and helpful content.

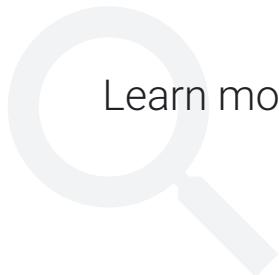
In less than half a second, our systems sort through hundreds of billions of web pages to try and find the most relevant and reliable results available. Our ranking systems are therefore developed with the goals of returning relevant and reliable results. To ensure relevance of search results, we have a variety of systems that aim to match the words and concepts in users' queries with related information in our index. This ranges from systems that understand things like misspellings or synonyms to more advanced AI-based systems that can understand more complex, natural-language queries.

However, when it comes to reliability – that is, finding high-quality, trustworthy information – even with our advanced information understanding capabilities, search engines like Google do not understand content the way humans do. Our systems often can't tell from the words or images alone if something is exaggerated, incorrect, low quality, or otherwise unhelpful.



Search engines largely understand the quality of content through what are commonly called “signals.” You can think of these as clues about the characteristics of a page that align with what humans might interpret as high quality or reliable. For example, the number of quality pages that link to a particular page is a signal that a page may be a trusted source of information on a topic.

Learn more about How Search Works [here](#)





Improving Search

We're constantly making improvements to Search — more than 5,150 in 2021 alone. These changes can be big launches or small tune-ups, but they're all designed to make Search work better for our users and to make sure users can find relevant, high quality information when they need it.

We put all possible changes to Search through a rigorous evaluation process to analyze metrics and decide whether to implement a proposed change. Data from these evaluations and experiments go through a careful review by experienced data scientists, product managers, and engineers who then determine if the change is approved to launch. Launches also go through additional evaluations, such as privacy and legal review.

1

We have an idea of how to improve Search

2

We develop the idea into a possible change

3

We evaluate to see if the idea returns more useful search results through...

- User Research
- Live Experiments
- Search Quality Tests

4

If successful, we launch the change to Search

How do we evaluate search?

We evaluate Search in multiple ways, including Search Quality Tests, User Research and Live Experiments:

Search Quality Tests

To evaluate whether our systems do, in fact, provide information that people searching find relevant and reliable as intended, we solicit feedback on proposed improvements from Search Quality Raters who use our guidelines to review the changes. These Search Quality Raters are from different locales around the globe who collectively perform millions of sample searches and rate the quality of the results according to the signals we previously established.

Our [Search Quality Rating guidelines](#) outline what our Raters look for. In Section 3, we explain the Search Quality Rating Process in more depth.



User Research

We have a user research team whose job it is to talk to people all around the world to understand how Search can be more useful and how people in different communities access information online.

We also invite people to give us feedback on different iterations of our projects.



Live Experiments

We conduct live traffic experiments to see how real people interact with the proposed change, before launching it to everyone. We enable the feature in question to just a small percentage of people, usually starting at 0.1% or smaller. We then compare the experiment group to a control group that did not have the feature enabled.

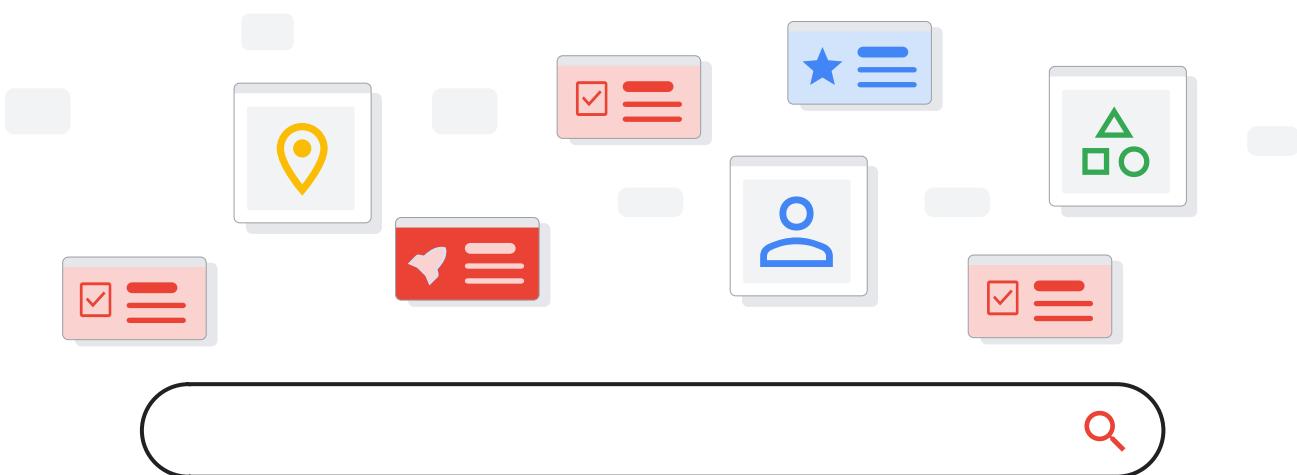
We look at a very long list of metrics, such as what people click on, how many queries were done, whether queries were abandoned, how long it took for people to click on a result, and so on. We use these results to measure whether engagement with the new feature is positive, to ensure that the changes that we make are increasing the relevance and usefulness of our results for everyone.



The Search Quality Rating Process

Our Search Quality Rating Process measures the quality of Search results on an ongoing basis. We work with ~16,000 external Search Quality Raters who provide ratings based on our guidelines and represent real users and their likely information needs, using their best judgment to represent their locale.

To assess the quality of search results, Raters work from a common set of guidelines – and are given specific tasks to complete.



How a Rater Task works

We first generate a sample of searches (say, a few hundred) to analyze a particular kind of search or potential ranking change. A group of Raters will be assigned this set of searches, and are asked to perform certain tasks on them. For example, one type of task is a side-by-side experiment, in which the Rater is shown two different sets of Search results: one with the proposed change already implemented and one without. We ask them which results they prefer and why. Raters also give valuable feedback on other systems, such as spelling. To assess proposed improvements to the spelling system, we ask Raters if the original query is misspelled, and whether the correction generated from the improved spelling system is accurate for individual queries.

Raters also review every page listed in the results set and evaluate that page against rating scales set out in our rater guidelines. **The Search Quality Rating** consists of two parts:



Page Quality

How well the page achieves its purpose

- Determine the purpose
- Assess if page is harmful
- Determine the rating



Needs Met

How useful a result is for a given search

- Determine the user intent
- Determine the rating

It is important to note that no single rating – or single rater – directly impacts how a given page or site ranks in Search. With trillions of pages that are constantly changing, there is no way that humans could evaluate every page on a recurring basis. Simply using Search Quality Ratings for ranking purposes would not be feasible, as humans could never individually rate each page on the open web. Indeed, this is why search engines were developed.

In addition to being an impossible task, using only Search Quality Ratings to determine ranking would not provide enough signals to determine how Search ranking should work. There are so many dimensions to quality and relevance that are critical, such as signals that:

- tell us whether something might be spammy;
- flag that a site may be unsafe or unsecure; or
- indicate that a page may be outdated.

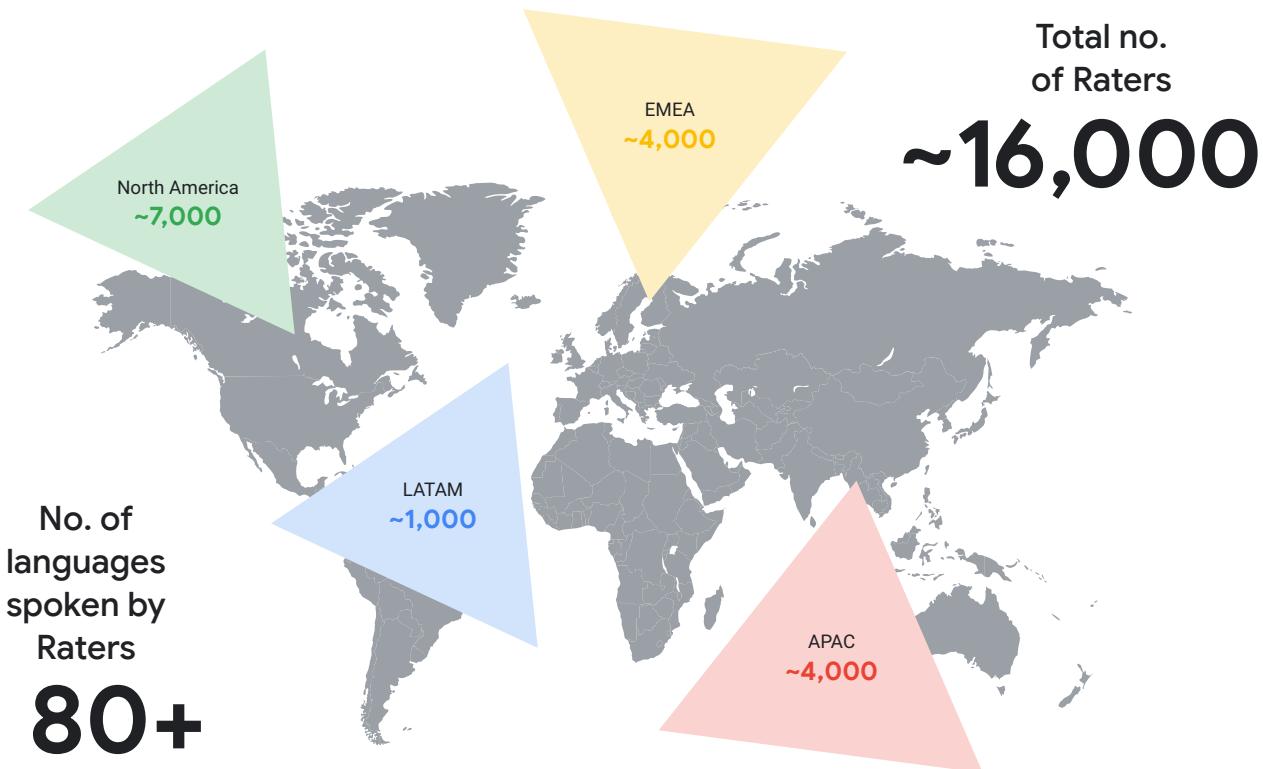
Thus, no single source of information – like a Search Quality Rating – would ever capture every dimension that's important for a task as complex as ranking.

We use aggregated ratings to measure how well our systems are working to deliver helpful content around the world. In addition, ratings are used to improve our systems by giving them positive and negative examples of search results.

The Raters

Raters represent Search users in their locale. They must be very familiar with the task language and location in order to represent the experience of people in their locale, and be familiar and comfortable with using a search engine.

As such, the guidelines remind Raters that users may have needs that differ from their own, as they come from many different backgrounds – with a great diversity in terms of ages, genders, races, religions, and political affiliations.



The Raters are employed by external vendors, and these numbers may vary based on operational needs. Because of the importance of scaling the Search Quality Rating process across regions, diversity of locale remains a priority.

Ratings should not be based on the Raters' personal opinions, preferences, religious beliefs, or political views. Instead, Raters are instructed to always use their best judgment and represent the cultural standards of their rating locale. Raters are also required to pass a test on the guidelines to ensure our own quality standards are met.

Page Quality rating

The goal of Page Quality (PQ) rating is to evaluate how well the page achieves its purpose.

There are three steps to the PQ rating process:



Page Quality

How well the page achieves its purpose

- Determine the purpose
- Assess if page is harmful
- Determine the rating



Needs Met

How useful a result is for a given search

- Determine the user intent
- Determine the rating

Step 1: Determining the purpose of the page



Page Quality

How well the page achieves its purpose

- Determine the purpose
- Assess if page is harmful
- Determine the rating

In order to assign a rating, Raters must first understand the purpose of the page. For example, the purpose of a news website homepage is to inform users about recent or important events.

This enables the Rater to better understand what considerations are important when evaluating the quality of that particular page in Step 3. Because different types of websites and webpages can have very different purposes, our expectations and standards for different types of pages are also different.



Here are a few examples where it is easy to understand the purpose of the page:

Type of page

Purpose of the page



News website
homepage

To inform users about recent or important events.



Shopping page

To sell or give information about the product.



Video page

To share a video.



Currency
converter page

To calculate equivalent amounts in different currencies.

Websites and pages should be created to help people. If that is not the case, a rating of Lowest may be warranted. As long as the page is created to help people, we will not consider any particular page purpose or type to be higher quality than another. For example, encyclopedia pages are not necessarily higher quality than humor pages.

There are highest quality and lowest quality webpages of all different types and purposes: shopping pages, news pages, forum pages, video pages, pages with error messages, PDFs, images, gossip pages, humor pages, homepages, and all other types of pages. The type of page does not determine the PQ rating – you have to understand the purpose of the page to determine the rating.



Step 2: Assess if the purpose of the page is harmful, or the page has the potential to cause harm



Page Quality

How well the page achieves its purpose

- Determine the purpose
- Assess if page is harmful
- Determine the rating

If the website or page has a harmful purpose or is designed to deceive people about its true purpose, it will immediately be rated the Lowest quality on the PQ rating scale. This includes websites or pages that are harmful to people or society, untrustworthy, or spammy as specified in the guidelines.

There is lots of content on the Internet that some would find controversial, one-sided, off-putting, or distasteful, yet would not be considered harmful as specified in the guidelines. Raters are instructed to follow the standards outlined in Section 4.0 of the guidelines which defines what is considered harmful.

Step 3: Determining the rating



Page Quality

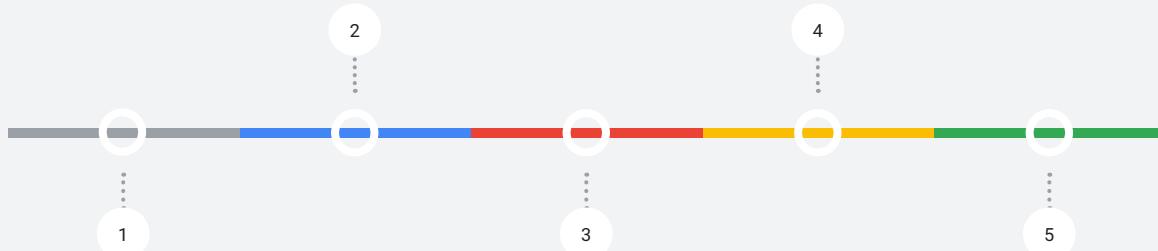
How well the page achieves its purpose

- Determine the purpose
- Assess if page is harmful
- Determine the rating

The PQ rating is based on how well the page achieves its purpose on a scale from lowest to highest quality.

Low: Low quality pages may have been intended to serve a beneficial purpose. However, they do not achieve their purpose well because they are lacking in an important dimension.

High: A High quality page serves a beneficial purpose and achieves its purpose well.



Lowest: Lowest quality pages are untrustworthy, deceptive, harmful to people or society, or have other highly undesirable characteristics. (Examples of pages that may be harmful are included [in Section 4.0](#) of the guidelines).

Medium: The page has a beneficial purpose and achieves its purpose; however it does not merit a High quality rating, nor is there anything to indicate that a Low quality rating is appropriate.
OR The page or website has strong High quality rating characteristics, but also has mild Low quality characteristics. The strong High quality aspects make it difficult to rate the page Low.

Highest: A Highest quality page serves a beneficial purpose and achieves its purpose very well.



In determining page quality, Raters must consider EEAT:

- The first-hand **e**xperience of the creator.
- The **e**xpertise of the creator.
- The **a**uthoritativeness of the creator, the main content itself and the website.
- And **trust**: the extent to which the page is accurate, honest, safe, and reliable.

Raters determine a page quality rating by:

- Reviewing the Main Content (MC) quality: As a general rule, MC is high quality if it takes a significant amount of effort, originality, and talent or skill to create the content.
- Reviewing the information available about the website and its creator: It should be clear who is responsible for the website and who created the content on the page, even if through an alias or username.
- Researching the reputation of the website and creator: A website's reputation is based on the experience of real users, as well as the opinion of people who are experts in the topic of the website.



Pages on the World Wide Web are about a vast variety of topics. Some topics require different standards for quality than others. For example, some topics could significantly impact the health, financial stability, or safety of people, or the welfare or well-being of society. We call such topics “Your Money or Your Life”, or YMYL. Raters apply very high PQ standards for pages on YMYL topics because low quality pages could potentially negatively impact the health, financial stability, or safety of people, or the welfare or well-being of society. Examples of YMYL topics can be found in Section 2.3 of the guidelines.

Similarly, other websites or pages that are harmful to people or society, untrustworthy, or spammy, as specified in these guidelines, should receive the Lowest rating. Examples of harmful pages are detailed in Section 4.0 of the guidelines.

The Search Quality Rating Process



Example

Webpage/ Type of Content	Quality Characteristics	PQ Rating and Explanation
Page about seasonal flu	<p>Very high level of E-E-A-T for the purpose of the page</p> <p>Very positive website reputation for the topic of the page</p>	<p>Lowest Medium Highest</p> <p>Many patient hospitalizations and deaths occur due to the flu each year. This topic could significantly impact a person's health. This is a YMYL topic.</p> <p>This is an influenza reference page on a trustworthy and authoritative medical website.</p> <p>This website has a reputation of being one of the best web resources for medical information of this type.</p>

Needs Met rating

Needs Met rating tasks focus on user needs and how useful the result is for people who are using Google Search.

How useful a search result is concerns the intent of the user, as interpreted from the query, and how fully the result satisfies this intent. There are therefore two steps to the Needs Met rating:



Page Quality

How well the page achieves its purpose

- Determine the purpose
- Assess if page is harmful
- Determine the rating



Needs Met

How useful a result is for a given search

- Determine the user intent
- Determine the rating

Step 1: Determining the user intent from the query



Needs Met

How useful a result is for a given search

- Determine the user intent
- Determine the rating

A query is what a user types or speaks into our Search system. We use the content of their query and, if relevant, the user location to determine the intent of the user.

For example, if a user searches for “coffee shops” and they are based in London, we may determine their intent is to find coffee shops in London. However, many queries also have more than one meaning. For example, if a user searches for “mercury,” their intent may be to learn more about the planet mercury or the chemical element.

We assume users are looking for current information about a topic, and Raters are instructed to think about the query’s current meaning as they are rating.

Step 2: Determining the rating



Needs Met

How useful a result is for a given search

- Determine the user intent
- Determine the rating

Needs Met rating

Fails to Meet (FailsM)

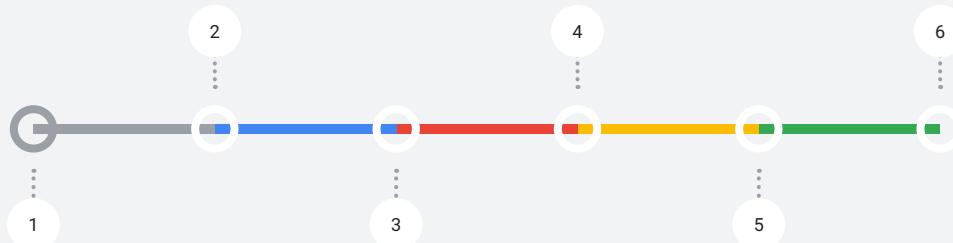
Completely fails to meet the needs of the users. All or almost all users would wish to see additional results.

Moderately Meets (MM)

Helpful for many users OR very helpful for some users. Some or many users may wish to see additional results.

Fully Meets (FullyM)

A special rating category, which only applies to certain queries and results. All or almost all users would be immediately and fully satisfied by the result and would not need to view other results to satisfy their need.



N/A

Slightly Meets (SM)

Helpful for fewer users. There is a connection between the query and the result, but not a strong or satisfying connection. Many or most users would wish to see additional results.

Highly Meets (HM)

Very helpful for many or most users. Some users may wish to see additional results.

The Search Quality Rating Process

The NM rating is based on how well a search result meets the intent of the user.

In determining the rating, the Rater considers the extent to which the result:

- “fits” the query
- is comprehensive
- is up-to-date
- comes from an authoritative source
- satisfies the user, or whether the user would want to see additional results

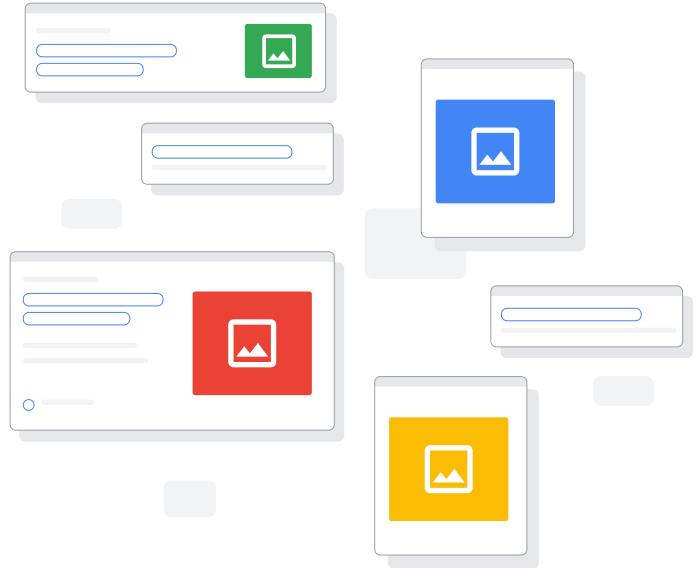
Example

Query & User Intent	Result	Rating	Explanation
<p>Query: [batman]</p> <p>User Location: Anaheim, California</p> <p>User Intent: Find information about the fictional superhero that appears in American comic books, movies, and television shows.</p>		<p>FailsM SM MM HM FullyM</p> <p>A horizontal scale with five points: FailsM, SM, MM, HM, and FullyM. The first point is marked with a red dot.</p>	<p>It is extremely unlikely that this query is looking for information on a city in Turkey called Batman, given that the user is located in the United States. No or almost no users would be satisfied with this result.</p>

Summary

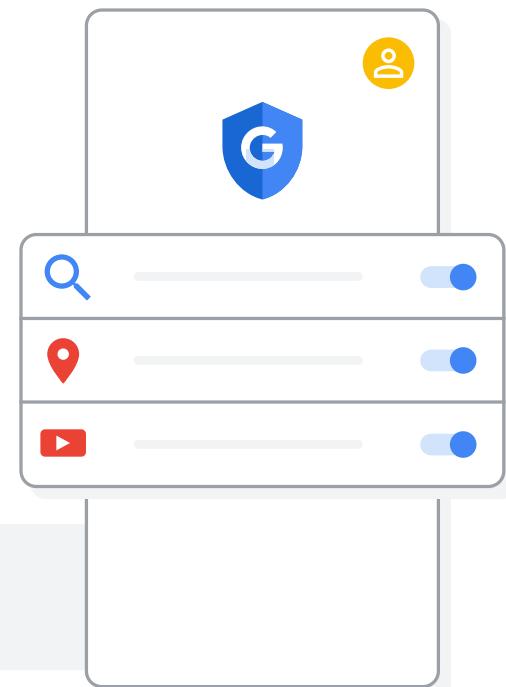
Access to information is at the core of Google's mission, and every day we work to make information from the web available to everyone. We fundamentally design our ranking systems to identify information that people are likely to find useful and reliable.

Because the web and people's information needs keep changing, we make a lot of improvements to our Search algorithms to keep up. The improvements we make go through an evaluation process designed so that people around the world continue to find Google useful for whatever they're looking for.



A key part of our evaluation process is getting feedback from everyday users about whether our ranking systems and proposed improvements are working well. We have ~16,000 Search Quality Raters from different regions across the globe, who, following instructions anyone can read in our Search Quality Rater guidelines, evaluate results for sample queries and assess how well the pages listed appear to demonstrate characteristics of reliability and usefulness.

Search Quality Raters are required to conduct a Page Quality Rating task which assesses the reliability of our Search results. For this task, Raters consider the quality of the webpage using criteria that we call E-E-A-T: Experience, Expertise, Authoritativeness and Trust. We have very high Page Quality rating standards for pages on “Your Money or Your Life” (YMLY) topics because low quality pages could potentially negatively impact the health, financial stability, or safety of people, or the welfare or well-being of society.



Pages which are considered harmful to users, because, for example, they mislead the user or contain dangerous content, receive the lowest page quality rating.

The Search Quality Rater also conducts a Needs Met Rating Task, which assesses the usefulness of the results to the user by looking at whether it satisfies the intent of the user as interpreted from the search query and user location. Raters are required to consider the extent to which the search results are comprehensive, current or authoritative. For example, a result that is rated as “fully meets” means that a user would be immediately and fully satisfied by the result and would not need to view other results to satisfy their need.



Search Quality Ratings alone will not directly affect how a particular webpage, website, or result appears in Google Search, nor will they cause specific webpages, websites, or results to move up or down on the search results page. Instead, the ratings will be taken in aggregate and used to measure how well search engine algorithms are performing for a broad range of searches.

You can find additional information on Search Quality Rating at the following resources:



[Search Quality Rater guidelines](#)

[How Search Works](#)

[How Google delivers reliable information in Search](#)

[How insights from people around the world make Google Search better](#)



[How real people make Google Search better](#)

[Could you walk in Google's shoes? Making tough calls with Search listings](#)

