

Information Gain in Google's Top-Ranking Results: A Measurement of 150 Pages

Eric Lancheres · On-Page.ai Research | June 12, 2026 | First edition of a quarterly index

ABSTRACT

We measured how much new information top-ranking pages on Google contribute beyond the pages they rank alongside. Using the Information Gain Score — a 0–100 measure of how much a page adds beyond the live ranking cohort for its keyword, compared by meaning rather than exact wording — we scored 150 pages holding top-3 organic positions across 50 keywords in ten verticals. The median page scored 52/100. 21% of pages graded highly original (70+); 24% graded mostly shared (below 40), adding little beyond what already ranks alongside them. An exploratory 84-page extension at positions 4, 7, and 10 found mostly-shared content substantially more common below the top 3 (37–40% of pages against 24%). Within the top 3, position was unrelated to the score; the spread inside a single SERP, however, averaged 32 points, and in 90% of SERPs at least one common topic question — from a question set generated per keyword by the analysis — went unanswered by every scanned top-3 page. Scores were weakly related to content length and most strongly associated with original quantitative evidence: pages carrying 15 or more unique data points averaged 62/100 against 40/100 for pages carrying at most one. Vertical medians ranged from 42 (health) to 62 (legal). Originality is evidently not what separates positions within the top 3; the distributions instead suggest substantial unclaimed headroom for pages that contribute original information.

KEY FINDINGS

- Top-ranking pages are not as original as most SEOs would assume. The median top-3 page added only a moderate amount of information beyond the other pages ranking for the same query.
- Being #1 did not reliably mean a page was more original than #2 or #3; however, top-3 pages as a group were more original than positions 4 through 10, which were more likely to repeat what already ranked (an exploratory extension).
- Length is not originality: the longest pages scored only modestly higher, and the middle third broke the pattern.

1. Background

“Information gain” describes the amount of new information a document adds relative to documents a reader has already seen. The concept entered search-industry discussion through a Google patent, “Contextual Estimation of Link Information Gain” [1], which describes scoring documents by how much they add beyond previously-presented results. A patent describes a capability, not a confirmed ranking system; this study makes no claim about how Google weights such signals. Google's public content guidance nonetheless emphasizes original information, research, and analysis beyond restating existing sources [2, 3], and the concept has since been taken up in industry analyses [4, 5].

Independent of ranking mechanics, the question the metric answers is of practical interest: when a page competes with nine other results for the same query, how much of it says something the others do not? To our knowledge, no published measurement has quantified this across live top-ranking results. This first edition establishes the baseline for a planned quarterly index.

2. Method

Sample. 50 informational and commercial keywords were drawn from ten verticals (marketing/SEO, B2B SaaS, health, personal finance, ecommerce, legal, travel, food & recipes, technology/dev, home improvement; five keywords per vertical). For each keyword we identified the top three organic results at scan time (June 12, 2026; US, English), yielding 150 unique page observations, 15 per vertical, 50 per SERP position. An exploratory extension drew positions 4, 7, and 10 from an independent SERP data source [6] for the same keywords on the same day; after removing URLs already present in the top-3 sample (the two sources count positions differently around SERP features) and pages that did not yield scoreable text — including three forum threads holding page-one positions — 84 lower-position pages were scored (20, 38, and 26 at positions 4, 7, and 10).

Measurement. Each page was scored with the Information Gain Score: a 0–100 measure of how much of the page's main content adds beyond the pooled content of the pages currently ranking for the same keyword. The comparison is semantic — paraphrased restatements of competitor content count as overlap, not originality. The score grades as: highly original (70–100), moderately original (40–69), mostly shared (0–39). We additionally counted unique numeric data points (figures present on the scored page that appear nowhere else in its cohort), bucketed pages into terciles by length of extracted main content, and labeled keyword intent by template (comparison, “best,” and cost queries as commercial-investigational; the remainder informational).

Scope. All scans were executed against the live web on a single day. No page content was modified, no rankings were tracked over time, and no per-site results are published — findings below are aggregates.

Notes. Keywords were hand-constructed for breadth across the ten verticals and balanced between informational and commercial query templates, not randomly sampled — a limitation noted in §5. Scoring runs on each page's extracted main content; navigation, boilerplate, and template chrome are not scored. Unique-data-point counts can include trivial figures — the count is reported as a correlate of the score, not a target: padding a page with junk numbers adds numbers, not information. The unanswered-

question analysis checks cohort pages against a set of topic questions generated per keyword; the set samples plausible reader questions and is not an exhaustive inventory of what a SERP could answer.

3. Results

3.1 Score distribution

Scores ranged from 0 to 95. The median top-3 page scored 52/100 – roughly half of a typical top-ranking page's content has a close semantic equivalent elsewhere in its own cohort. The distribution is broad: one page in ten scored below 26, and one in ten scored 79 or above. Eleven pages scored in single digits; three exceeded 90.

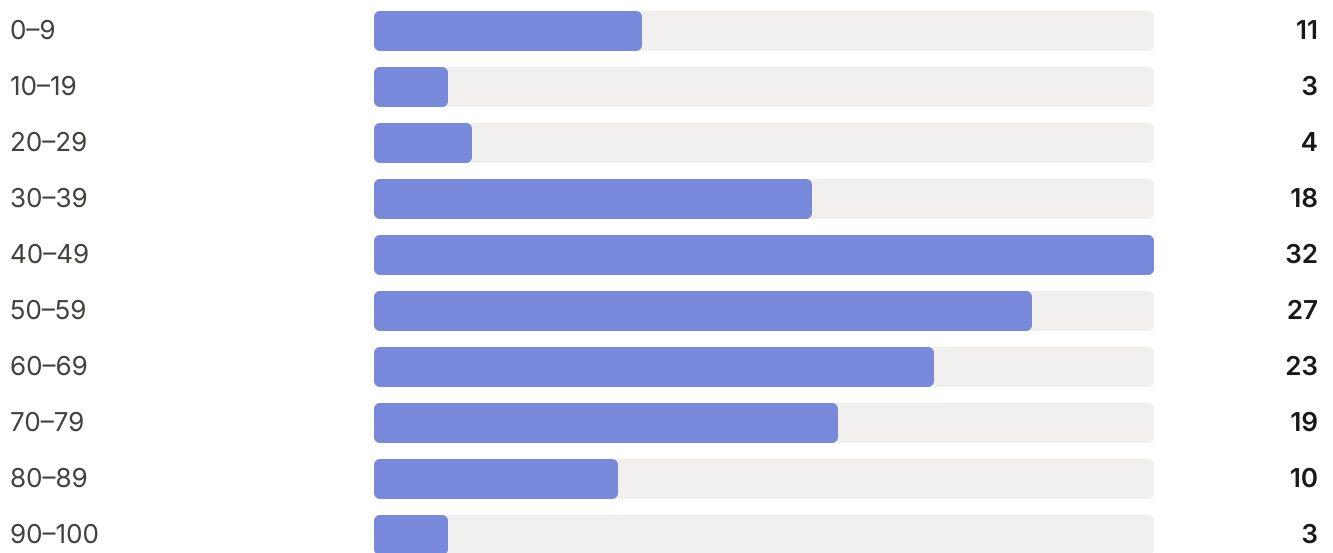


Figure 1. Distribution of Information Gain Scores across 150 top-3 ranking pages, by score decile (scanned June 12, 2026).

Statistic	Information Gain Score
Minimum	0
10th percentile	25.8
25th percentile	41.3
Median	52
75th percentile	67.8
90th percentile	79
Maximum	95

Table 1. Percentiles of the Information Gain Score (n = 150).

3.2 Grade distribution

Roughly one page in four graded mostly shared — its content adds little beyond what is already present in the rest of its cohort. One in five graded highly original.

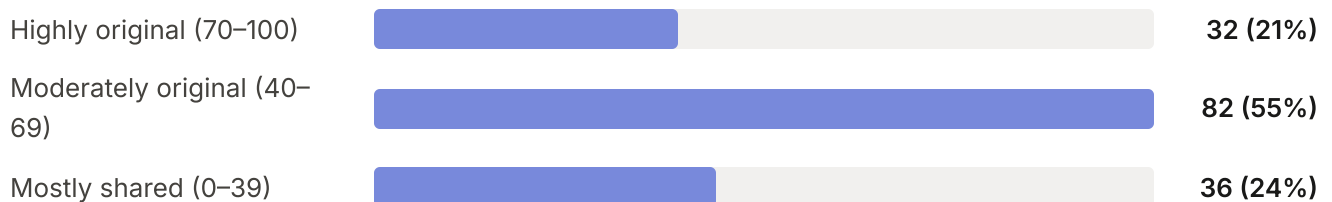


Figure 2. Grade distribution of 150 top-3 ranking pages. Bars show page counts; percentages of sample in parentheses.

3.3 Position within the top 3

Scores showed no consistent relationship with position. Medians for positions 1, 2, and 3 were 52, 51.5, and 52 respectively. The page holding the #1 position is, on average, no more original than the results immediately below it — whatever sustains a #1 ranking, it is not distinguishable in these data by information gain.

Position	n	Mean	Median
Position 1	50	50.2	52
Position 2	50	54.9	51.5
Position 3	50	49.2	52

Table 2. Information Gain Score by SERP position (n = 50 pages per position).

3.4 Below the top 3 (exploratory)

The lower-position extension shows a modest decline in average score (means 51.4 for positions 1–3 against 44.5–49.3 at positions 4, 7, and 10) and a markedly different grade mix: mostly-shared pages are substantially more common below the top 3 – 37–40% of pages at positions 4, 7, and 10, against 24% in the top 3. The share of highly original pages is similar across positions. Read together with §3.3: originality does not separate positions *within* the top 3, but pages lower on page one are considerably more likely to be largely redundant with their cohort. Pooled, pages below the top 3 average 47.1 against 51.4 for the top 3 – a 4–5 point gap – with position 10 the least original bucket in the sample (mean 44.5). In a ranking system shaped by many factors, a consistent univariate gap of this size is small but notable. Subsamples are small and drawn from a different SERP source; we treat this as exploratory.

Average Information Gain Score by position (axis 40–55)

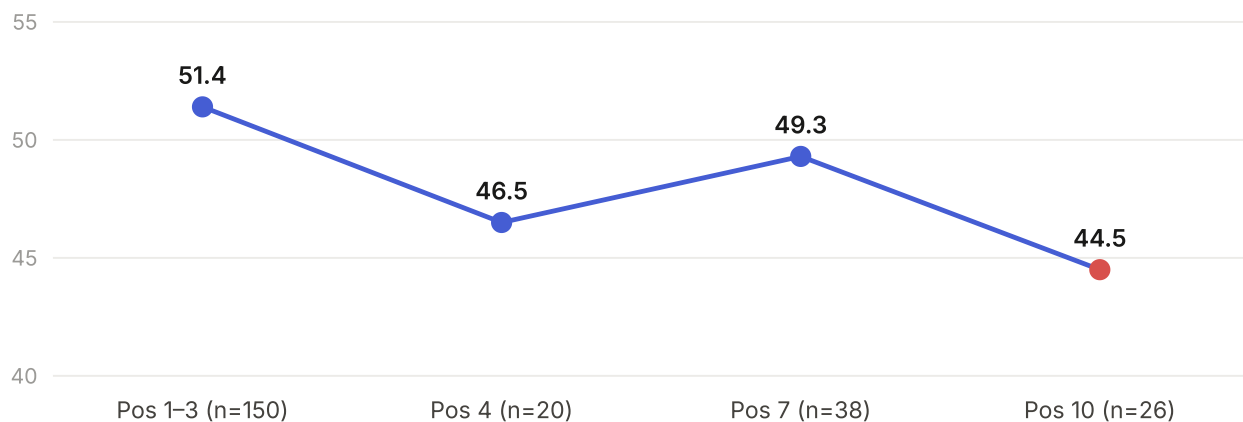


Figure 3. Average Information Gain Score by SERP position. The y-axis is truncated to 40–55 to make the 4–5 point gap visible; position 10 (red) is the least original bucket. Positions 4/7/10 are an exploratory extension drawn from an independent SERP source.

Bucket	n	Mean	Median	Mostly shared
Positions 1-3	150	51.4	52	24%
Position 4	20	46.5	47	40%
Position 7	38	49.3	50.5	37%
Position 10	26	44.5	48	38%

Table 3. Information Gain Score and mostly-shared share by position bucket. Positions 4/7/10 are an exploratory extension (n = 84).

Pages that mostly repeat what already ranks (% of pages scoring below 40)

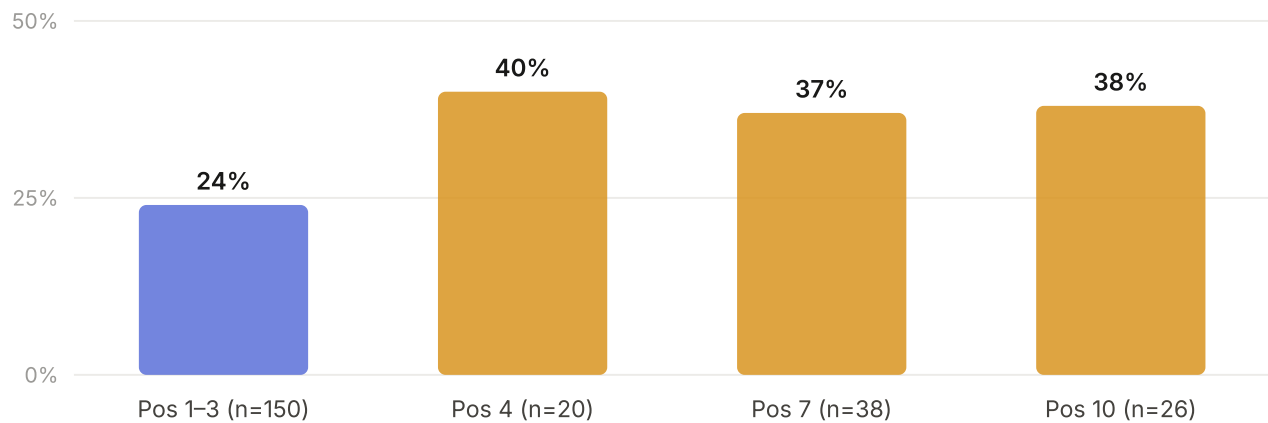


Figure 4. Share of pages that mostly repeat what already ranks (score below 40), by position — nearly twice as common below the top 3 (amber) as within it (blue).

3.5 Variation by vertical

Vertical medians spanned twenty points, from 42 (health) to 62 (legal). Legal content was the most consistently original vertical — seven of its fifteen pages graded highly original and only one graded mostly shared. Health, ecommerce, and B2B SaaS occupied the bottom of the range. Within-vertical spread remained large relative to between-vertical differences: several verticals contained both zero-scoring pages and pages in the 80s or 90s. Vertical samples are small (n = 15 each); these contrasts should be read as indicative.

Median Information Gain Score by vertical (n = 15 per vertical)

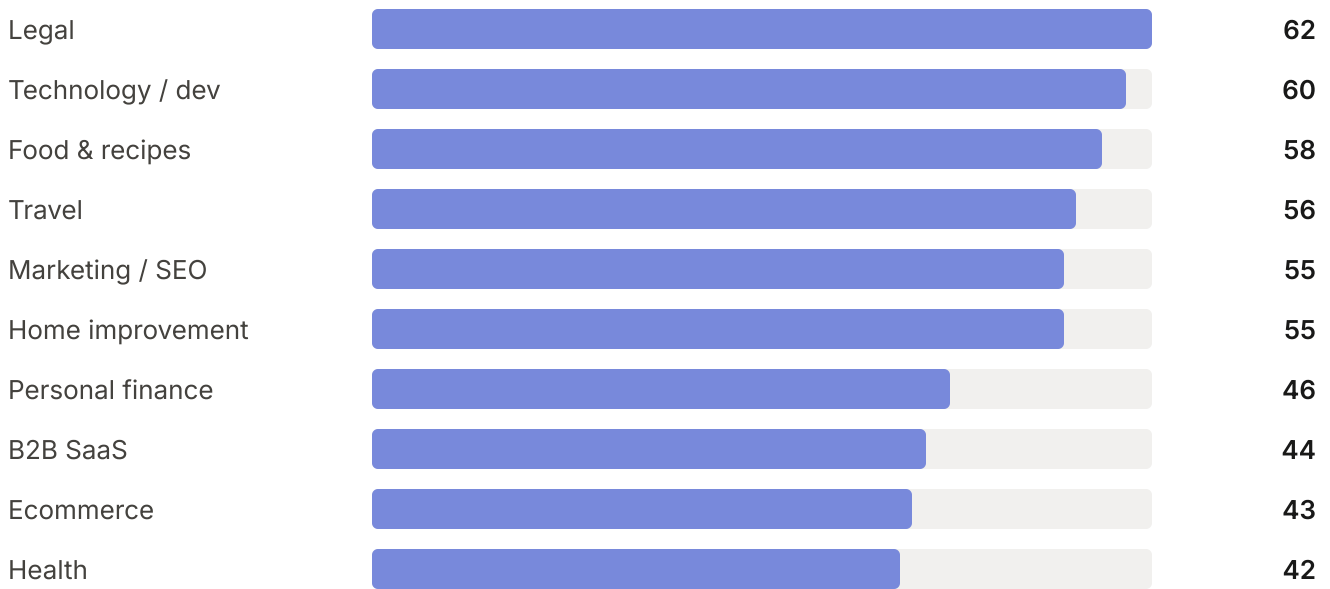


Figure 5. Median Information Gain Score by vertical (n = 15 pages per vertical). Means: legal 64.1, home improvement 57.5, technology/dev 56.2, marketing/SEO 53.3, travel 51.7, B2B SaaS 50.6, food & recipes 49.3, personal finance 47.7, health 44.4, ecommerce 39.5.

3.6 Unique data points

The median top-3 page carried 4 numeric data points found nowhere else in its cohort, against a cohort-wide average of approximately 8. The distribution is heavily right-skewed (the richest page carried 158 unique figures), and original quantitative evidence concentrated in a minority of pages: 30% of the sample carried at most one unique figure.

Unique data points were the strongest page-level correlate of the score in this sample. Mean scores rose monotonically across data-point bins, from 40.2 for pages with at most one unique figure to 62.1 for pages with fifteen or more (Pearson $r = 0.17$: the linear correlation is weak — both distributions are heavily skewed — but the bucketed difference is practically large). Part of this association is structural — original figures are original content — but the gap between bins is large enough to be practically meaningful: pages rich in original data are, on average, a full grade band above pages without it.

Unique data points	n	Mean score	Median score
0–1 unique data points	45	40.2	47
2–5 unique data points	45	54.4	52
6–14 unique data points	38	55.1	55.5
15+ unique data points	22	62.1	60

Table 4. Information Gain Score by unique-data-point count (n = 150).



Figure 6. Mean Information Gain Score by unique-data-point count. The gradient is monotonic: pages rich in original figures average a full grade band above pages without them.

3.7 Content length and keyword intent

Length explains little. Pages in the longest third of the sample scored somewhat higher than the shortest third (median 57.5 against 50.5), but the middle tercile broke the pattern (median 49) — adding length is evidently not, by itself, adding information.

Length tercile	n	Mean score	Median score
Shortest third	50	48	50.5
Middle third	50	50.4	49
Longest third	50	55.9	57.5

Table 5. Information Gain Score by main-content length tercile (n = 50 per tercile).

Keyword intent showed a modest gap: pages ranking for informational queries scored a median of 54 against 48 for commercial-investigational queries (means 53.8 and 46.8; n = 99 and 51). Pages competing for commercial terms restate their cohorts more — consistent with the heavy templating of comparison and “best” content.

3.8 The gap inside the SERP

Aggregates understate how uneven individual SERPs are. Within a single keyword's top 3, the spread between the most and least original page averaged 31.6 points (median 25): 64% of SERPs carried a gap of 20 points or more, and 40% a gap of 30 or more. Nearly every SERP, in other words, has a clear originality leader and a clear laggard.

Unanswered questions are similarly widespread. For each keyword, the analysis generates a set of common topic questions and checks them against every page in the cohort — a sample of plausible reader questions, not an exhaustive inventory. In 90% of SERPs, at least one such question went unanswered by every scanned top-3 page (75% of individual pages had at least one in view). And pages that carried more unique data points than their own cohort's average scored a median of 60 against 48 for pages that did not — out-evidencing the cohort co-occurs with roughly a grade-band advantage.

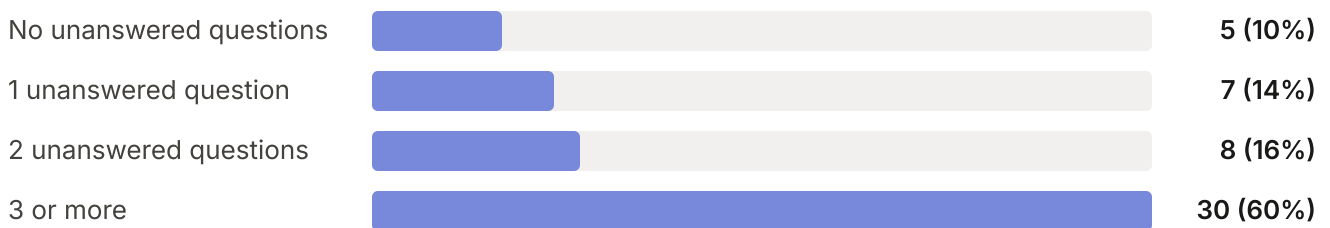


Figure 7. SERPs by number of distinct topic questions — from the analysis's per-keyword question set — left unanswered by every scanned top-3 page (n = 50 SERPs). 90% have at least one; 60% have three or more.

4. Discussion

Three observations stand out. First, originality is evidently not a prerequisite for ranking: a quarter of top-3 pages grade mostly shared, position within the top 3 is unrelated to the score, and pages scoring zero hold top-3 positions in four verticals. Whatever

combination of authority, relevance, and history sustains those rankings, it does not currently require contributing new information.

Second, the most actionable correlate of information gain in these data is original quantitative evidence. The unique-data-point gradient (Table 3) is monotonic and large, and unlike length it does not saturate: pages carrying original figures their cohort lacks are, on average, a grade band above pages that restate existing numbers.

Third, the headroom is large and unevenly distributed. If search systems or AI assistants increase the weight placed on a page's marginal contribution — as the patent literature contemplates and as synthesis-based answer engines structurally favor — the median incumbent in verticals like health, ecommerce, and B2B SaaS contributes little that is not already present in its cohort. A page that contributes original claims, figures, or answers competes in a field where the median top-3 incumbent carries four unique data points.

Fourth, the exploratory lower-position data suggests the redundancy burden grows down the page: mostly-shared content is half again as common at positions 4–10 as in the top 3. Combined with the within-SERP spread (§3.8), the practical picture is of SERPs with one or two pages carrying most of the original information and the remainder restating it — an originality lead that is, in most SERPs, held thinly.

We do not claim a causal relationship between the score and rankings or citations. The study measures the state of the SERP, not the mechanism behind it.

5. Limitations

This first edition has deliberate constraints. The sample, while doubled from the initial design (150 pages, 50 keywords), remains small per vertical ($n = 15$), and the lower-position extension is smaller still ($n = 84$, from a different SERP source, with position-4 coverage reduced by source overlap) — its contrasts are directional, not definitive. All scans were taken on a single day in one locale (US, English); SERPs change. Keyword intent was labeled by query template, not by SERP inspection. The ten verticals were

chosen for breadth, not representativeness. The Information Gain Score is our own metric — it measures semantic contribution relative to a cohort, not any search engine's internal signal. The planned full edition extends the sample to 300 keywords and positions 1–10.

6. Implications for practitioners

The findings translate into a straightforward editorial check for any page competing on an established keyword:

- Compare the page against the pages currently ranking for its keyword, not against an abstract quality bar — originality is relative to the live cohort.
- Treat consensus content as the cost of entry: cover what every ranking page covers briefly, and spend the page's depth on what none of them have.
- Add original quantitative evidence — measurements, tests, first-party numbers. It was the strongest single correlate of originality in this sample.
- Ask what common reader questions the ranking pages leave unanswered; measured against a generated per-keyword question set, nine out of ten SERPs in this sample had at least one.

References

1. “Contextual Estimation of Link Information Gain,” U.S. Patent 11,354,342 B2, Google LLC. patents.google.com/patent/US11354342B2 (granted continuations include US 11,720,613 B2 and US 12,013,887 B2).
2. “Creating helpful, reliable, people-first content,” Google Search Central documentation. developers.google.com/search/docs/fundamentals/creating-helpful-content
3. “Search Quality Evaluator Guidelines,” Google. guidelines.raterhub.com (PDF)
4. Information gain, Backlinko. backlinko.com/information-gain
5. Information gain in SEO, Searchbloom. searchbloom.com/blog/information-gain-seo
6. Serper, Google Search API (SERP data source for the lower-position extension). serper.dev
7. On-Page.ai Information Gain Score: scoring output reference, [report schema documentation](#).

DATA AND METRIC AVAILABILITY

Aggregate statistics are reported above; per-URL results are not published. The metric used in this study is publicly available: any page can be scored against its live ranking cohort with the [free checker](#) or programmatically via the [On-Page.ai API](#).

Cite this study: Lancheres, E. (2026). *Information Gain in Google's Top-Ranking Results: A Measurement of 150 Pages*. On-Page.ai Research. <https://api.on-page.ai/research/information-gain-study>