Characterizing Perceived Readability in Data Visualization: Design and Reader Factors

Anne-Flore Cabouat*
Université Paris-Saclay, Inria, LISN, Télécom Paris, Institut Polytechnique de Paris, i3, CNRS

Kuno Kurzhals†
University of Stuttgart

Laura Matzen‡
Sandia National Laboratories

Samuel Huron§
Télécom Paris, Institut Polytechnique de Paris, i3, CNRS

Tobias Isenberg*
Université Paris-Saclay, Inria, LISN, CNRS

Petra Isenberg*
Université Paris-Saclay, Inria, LISN, CNRS

ABSTRACT

We report results from an online study designed to characterize perceived readability across a controlled set of static visual data representations.

Index Terms: Readability.

1 STUDY DESIGN

We design a within-participant study to address the following exploratory research questions:

- **RQ1:** How do visualization design characteristics (encoding family, idiom commonness, and visual clarity) influence different perceived readability facets ◆◆◆◆?
- **RQ2:** How do visualization reading short test scores relate to perceived readability?
- **RQ3:** How do subjective familiarity and confidence relate to perceived readability, and how do these effects compare to those of reading test scores?

In addition to these exploratory analyses, we include two preregistered (osf.io/7at2d) within-participant confirmatory analyses to validate the effectiveness of our stimulus manipulations and establish a baseline for perceived readability across our set of 16 stimuli visualizations:

- **H1:** Within each encoding family, common visualization idioms yield higher ◆ **UNDERSTAND** ratings than uncommon idioms.
- **H2:** Within each visualization idiom, uncluttered layouts yield higher ◆ **LAYOUT** ratings than cluttered layouts.

2 VISUALIZATION STIMULI AND TASKS

2.1 Visual design

Our set of static visual stimuli includes visualizations from four encoding families: A-area, B-position, C-color magnitude, and D-node-link. Each family includes two levels of idiom: a-common and b-uncommon. For example, the A-area family includes geographical maps (a) and treemaps (b). Each idiom is presented in two levels of layout clarity: 1-uncluttered and 2-cluttered. Layout variants within a single idiom differ in the amount of visual elements and density of information in the visualization. The resulting stimulus space consists of 16 unique combinations of encoding family, idiom familiarity, and layout clarity presented in Figure 1. Each stimulus corresponds to a specific visualization instance that all participants evaluated.

2.2 Reading tasks

To ensure that participants engaged with each visualization before reporting perceived readability, we designed two tasks per stimulus: a local task and an aggregate task. We tailored each task to the visual encoding of each family and to elicit different aspects of the reading experience. Local tasks required retrieval or comparison of individual data elements, focusing on information extraction. Aggregate tasks required comparison or estimation of grouped values or overall patterns, engaging higher-level interpretation of the data structure. Including both task types ensured that perceived readability ratings reflected multiple levels of engagement with the visualization. Table 1 summarizes the study's reading tasks. Within each idiom, we aimed to balance the difficulty of tasks between cluttered and uncluttered variants, and, as far as possible, across idioms within the same encoding family. Task performance is not analyzed in the present work.

3 STUDY DESIGN

3.1 Measures

We measure perceived readability using the PREVis questionnaire [1], a validated instrument that captures multiple aspects of the reading experience. The four PREVis scales include perceived understandability (◆ **UNDERSTAND**, 3 items, $\alpha = .95$ and $\omega = .95$), layout clarity (◆ **LAYOUT**, 3 items, $\alpha = .93$ and $\omega = .93$), information retrieval (◆ **DATAREAD**, 3 items, $\alpha = .94$ and $\omega = .94$), and feature visibility (◆ **DATAFEAT**, 2 items, $\alpha = .91$). We present a screenshot of the rating interface in Figure 2.

We use the MiniVLAT [2], a short test of heterogeneous visualization reading skills, as a composite proxy of general visualization reading ability. The test consists of 12 reading questions with multiple-choice options, and participants have 25 seconds for each question. Internal consistency is moderate: $\alpha = .34$ and $\omega = .65$.

We ask two questions to assess subjective familiarity with visualization types (see screenshots in Figure 3):

- Confidence: "How confident do you feel in your ability to correctly read this type of visual representation?" (1=Extremely unsure, 7=Extremely confident)
- Frequency: "How frequently have you read this type of visual representation in recent years (before this study)?" (1=Almost never, 7=Extremely often)

*e-mail: given-name.family_name@inria.fr
†email:given_name.family_names@visus.uni-stuttgart.de
‡email:given_name.family_names@sandia.gov
§e-mail: given_name.family_name@telecom-paris.fr

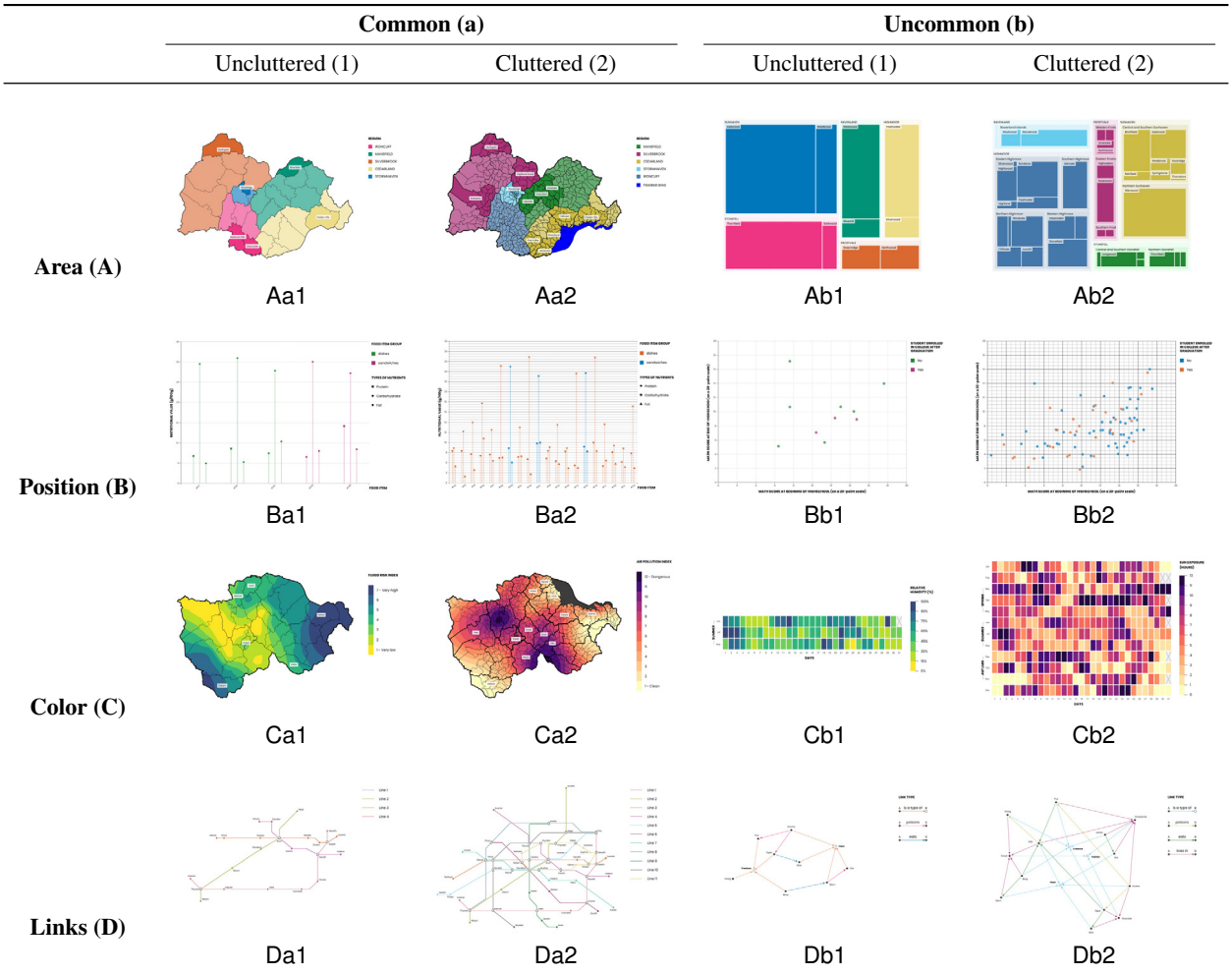


Figure 1: All 16 stimuli organized by encoding family (rows), idiom type (familiar vs. uncommon), and clutter level (clear vs. cluttered).

Answer type (input modality)	*Task type*	*Visual encoding to read*	*Visualization type*	*Target level*	*Closest Mini-VLAT item*
Continuous numerical (Open-ended field)	Retrieve value + Compare	Area	Geographical map	Entity	Item 9
		Area	Treemap	Entity	Item 9
		Position	Lollipop plot	Entity	Item 3
		Position	Scatterplot	Entity	Item 3
Ordinal numerical (Multiple-choice with ordered bins)	Retrieve value	Color magnitude	Density map	Entity	Item 3 / 4
		Color magnitude	Calendar heatmap	Entity	Item 3 / 4
Categorical (Multiple-choice yes/no or binary choice)	Derive value + Compare	Area	Geographical map	Dataset	Item 5
		Area	Treemap	Dataset	Item 5
		Position	Lollipop plot	Dataset	Item 6
		Position	Scatterplot	Dataset	Item 6
		Color magnitude	Density map	Dataset	Item 4
		Color magnitude	Calendar heatmap	Dataset	Item 4
		Link	Node-link diagram	Dataset	N/A
		Link	Bipartite network	Dataset	N/A
Categorical (Multiple-choice yes/no)	Accessibility	Link	Node-link diagram	Entity	N/A
		Link	Node-link diagram	Entity	N/A
		Link	Bipartite network	Entity	N/A
		Link	Bipartite network	Entity	N/A

Table 1: Overview of the reading tasks used in the study. Each participant completed all tasks.

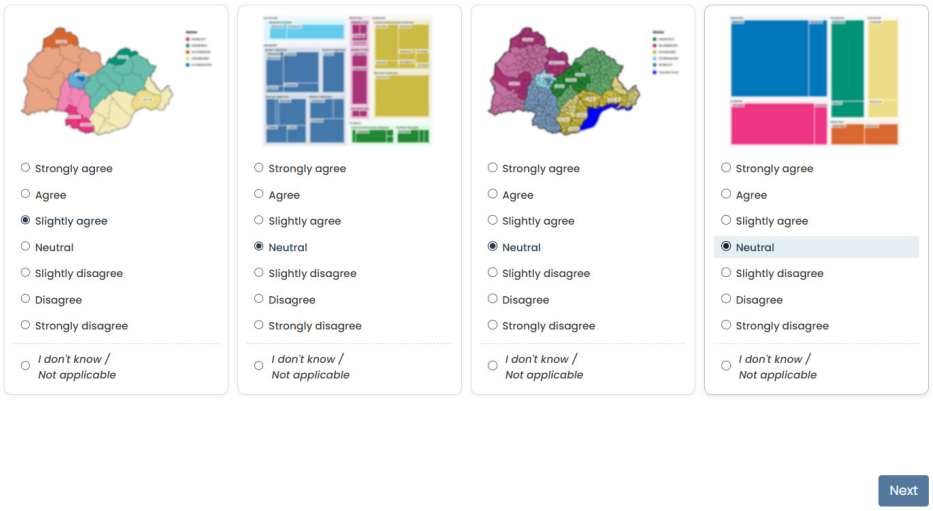
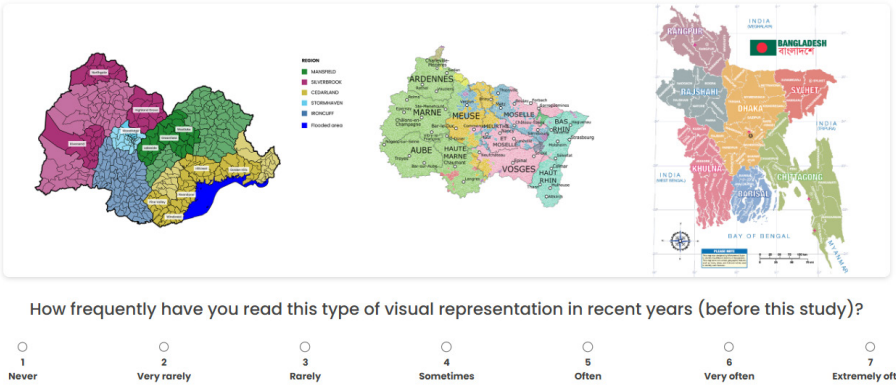
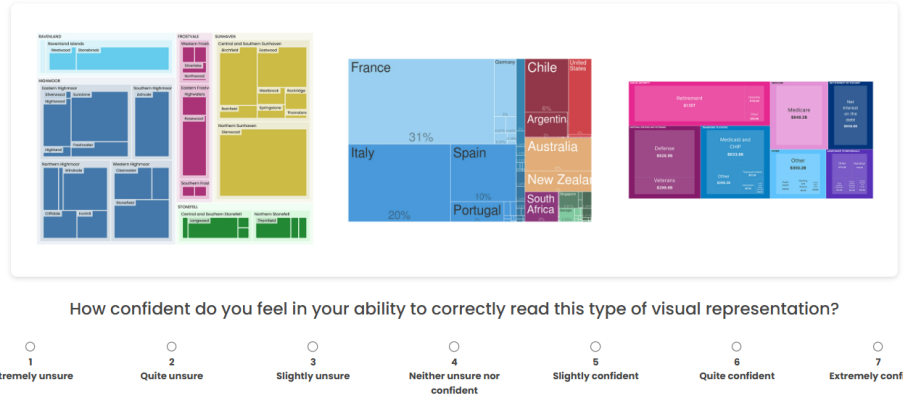


Figure 2: Screenshot of the survey UI during PREVis rating at the end of each visual encoding session: all 4 visualization stimuli are presented, following the order in which the participant saw them, and rated concurrently on each item of each scale. Scale order was randomized, as well as the order of items within each scale. The stimulus images were deliberately blurred, with the intent to elicit ratings from recall of the reading experience rather than on-the-spot comparison.



(a)



(b)

Figure 3: Screenshots of the survey UI during idiom rating at the end of a session: for each idiom level (a-common and b-uncommon), we presented the cluttered stimulus along with 2 other examples found in-the-wild and collected Frequency (a) and Confidence (b) ratings.

Age range	N
18–24	2
25–34	28
35–44	13
45–54	9
55–64	6
65+	2

Nationality	N
United Kingdom	21
United States	11
Italy	4
Nigeria	4
Hungary	3
Korea	3
South Africa	3
China	2
Other	11

Table 2: Participant demographics.

3.2 Participants

We recruited participants through Prolific in a four-session longitudinal study, using allowlists to invite participants from one session to the next. Eligibility criteria, quality-control procedures, and exclusion criteria were preregistered (osf.io/7at2d). After exclusions, the final sample comprised 60 participants, including 30 men, 29 women, and 1 non-binary participant. We report the distribution of agre ranges and nationalities in Table 2.

3.3 Procedure

We conducted an online study using a fully within-subject design and divided into four sessions (each estimated to last 15-20 minutes based on pilot tests) to minimize participant fatigue. Sessions were spaced apart by a minimum of 1 hour and a maximum of 4 days. Each session focused one visualization encoding family (A, B, C, and D). The order of the sessions was fully counterbalanced using a Latin-square design.

Within each session, participants view four stimuli representing combinations of idiom commonness (a-common vs b-uncommon) and visual clarity (1-uncluttered vs 2-cluttered). For each stimulus, participants completed two tasks and then reported their perceived readability. Within each session, both stimulus order and task order were randomized using the randomization function available on our survey platform, Limesurvey. At the end of each session, participants also reported their subjective familiarity for each idiom presented. Finally, during the first session, they completed the Mini-VLAT [2] visualization reading ability test.

4 Confirmatory Analyses Results

We report here the results from pre-registered (osf.io/7at2d) confirmatory analyses testing the effectiveness of our stimulus manipulations on perceived readability measures. Exploratory analyses addressing the research questions outlined in section 1 will be part of a future publication.

4.1 H1: Idiom commonness and perceived understandability

For each participant, we averaged PREVis ◆ **Understand** scores across the two layout conditions (1-uncluttered and 2-cluttered) within each idiom level and encoding family, yielding one score per idiom level (a-common, b-uncommon) per family (A, B, C, and D). Within-participant comparisons of the resulting average PREVis ◆ **Understand** scores show an overall tendency for b-uncommon idioms to be rated as less easy to understand than a-common across encoding families, as shown in Figure 4. However, the magnitude of this effect is modest, remaining below one point on the 7-point scale across all families. The effect is most clear for B-position, and to a lesser extent for A-area. In contrast, no reliable difference is observed for C-color magnitude and D-node-link, where the confidence interval overlaps zero; what's more, the point estimate for D is slightly positive.

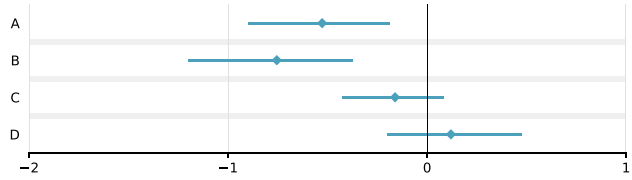


Figure 4: Within-participant ◆ **Understand** rating differences (i.e., paired-sample t-test) between the b-uncommon and the a-common idioms conditions, for each encoding family, with 95% confidence interval.

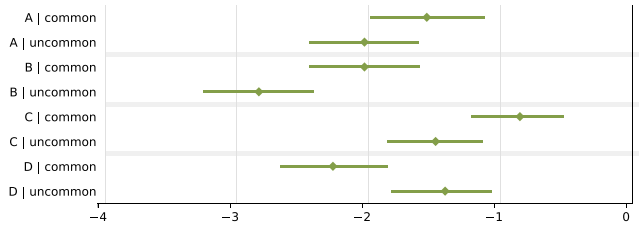


Figure 5: Within-participant ◆ **Layout** rating differences (i.e., paired-sample t-test) between the 2-cluttered and 1-uncluttered layout conditions, for each encoding family × idiom level combination, with 95% confidence interval.

Taken together, these results only provide partial support for H1, indicating that idiom commonness is associated with perceived understandability, but that this relationship varies across encoding families and remains limited in magnitude. We report the values in Table 3.

Table 3: uncommon - common

Channel	Estimate	Lower CI	Upper CI
A	-0.528	-0.908	-0.192
B	-0.756	-1.153	-0.369
C	-0.161	-0.408	0.072
D	0.119	-0.197	0.439

Table 4: Within-participant ◆ **Layout** rating differences between clear and cluttered layout within each idiom.

Channel	Idiom	Estimate	Lower CI	Upper CI
A-area	a-common	-1.561	-1.978	-1.128
	b-uncommon	-2.033	-2.439	-1.628
B-position	a-common	-2.033	-2.439	-0.622
	b-uncommon	-2.833	-3.245	-2.422
C-color magnitude	a-common	-0.856	-1.217	-0.533
	b-uncommon	-1.494	-1.85	-1.144
D-node-link	a-common	-2.272	-2.661	-1.861
	b-uncommon	-1.422	-1.822	-1.078

4.2 H2: Within-idiom visual clutter and perceived layout clarity

Within-participant comparisons of PREVis ◆ **Layout** scores reveal a clear and consistent effect of visual clarity on perceived layout readability for each visual encoding family × idiom commonness level, as shown in Figure 5. Across all encoding families and idiom levels, cluttered visualizations are rated as less readable than their uncluttered counterparts. The magnitude of these differences is larger than that observed for idiom commonness in H1, ranging

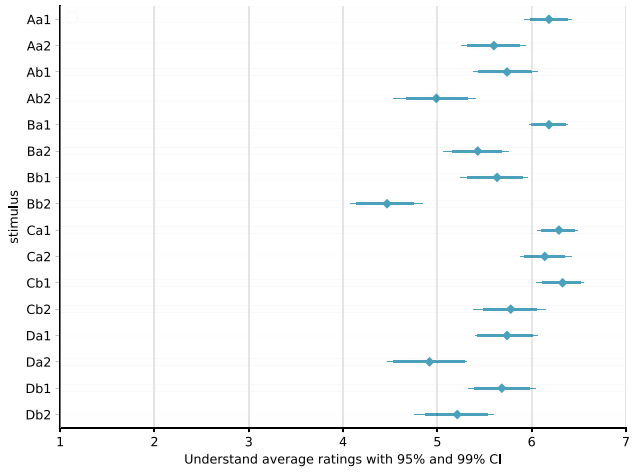


Figure 6: Point-estimate with 99 and 95% confidence intervals for all 16 stimuli on the Understand readability scale.

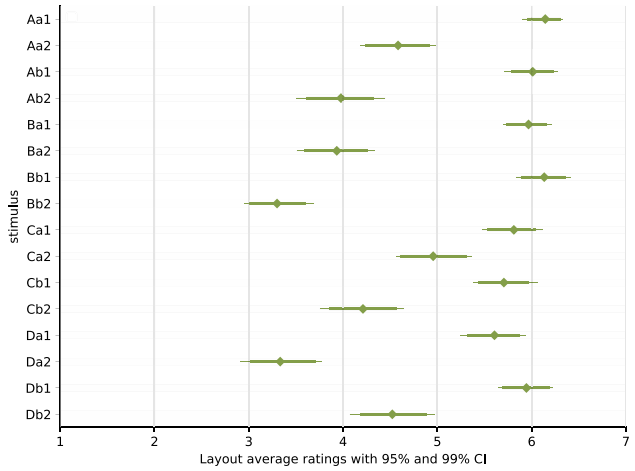


Figure 7: Point-estimate with 99 and 95% confidence intervals for all 16 stimuli on the Layout readability scale.

from 1 to 3 points on the scale. These results provide strong support for H2, showing that adding visual clutter heavily reduces perceived layout clarity across all tested visualization conditions. We report the detailed values in Table 4.

5 Perceived Readability Ratings Across All Stimuli

To further characterize the readability of each of our 16 stimuli, we present below the point-estimate and boostrapped 95% Ci for all 16 stimuli on each of the four PREVis ◆◆◆◆ scales: Understand (Figure 9).

References

[1] A.-F. Cabouat, T. He, P. Isenberg, and T. Isenberg. Previs: Perceived readability evaluation for visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 31(1):1083, Jan. 2025. doi: 10.1109/TVCG.2024.3456318 1

[2] S. Pandey and A. Ottley. Mini-vlat: A short and effective measure of visualization literacy. *Computer Graphics Forum*, 42(3):1–11, 2023. doi: 10.1111/cgf.14809 1, 4

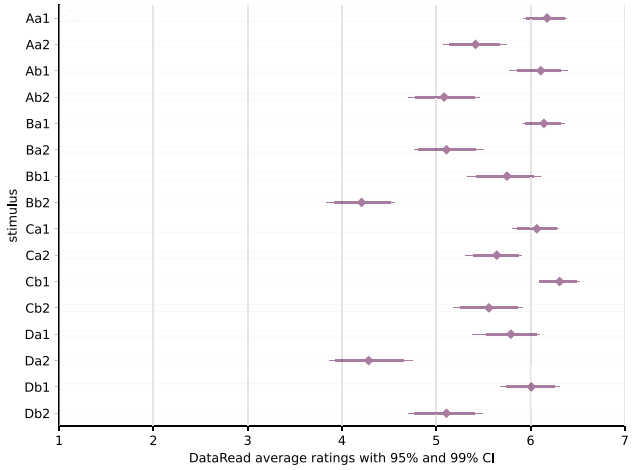


Figure 8: Point-estimate with 99 and 95% confidence intervals for all 16 stimuli on the Data Reading readability scale.

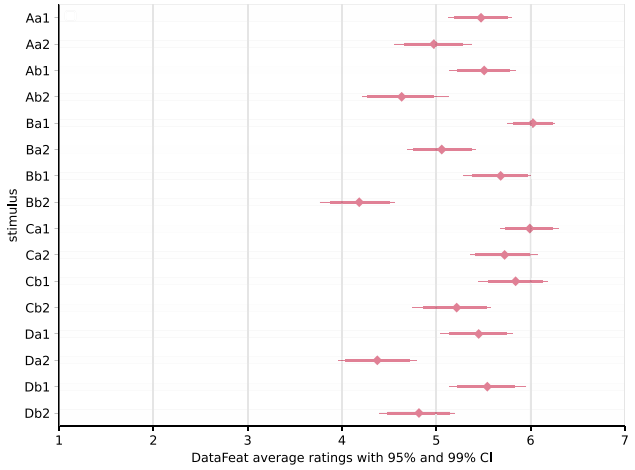


Figure 9: Point-estimate with 99 and 95% confidence intervals for all 16 stimuli on the Data Feature readability scale.

Table 5: ◆ **Understand** point estimates values with 95% Confidence Intervals.

Stimulus	Estimate	Lower CI	Upper CI
Aa1	6.183	5.986	6.356
Aa2	5.600	5.322	5.858
Ab1	5.739	5.480	5.992
Ab2	4.989	4.622	5.294
Ba1	6.183	6.011	6.361
Ba2	5.428	5.164	5.692
Bb1	5.633	5.344	5.895
Bb2	4.467	4.169	4.745
Ca1	6.289	6.119	6.444
Ca2	6.139	5.931	6.325
Cb1	6.328	6.097	6.508
Cb2	5.778	5.478	6.042
Da1	5.739	5.475	6.000
Da2	4.917	4.583	5.236
Db1	5.683	5.419	5.950
Db2	5.211	4.883	5.520

Table 6: Layout point estimates with 95% confidence intervals.

Stimulus	Estimate	Lower CI	Upper CI
Aa1	6.144	5.978	6.303
Aa2	4.583	4.261	4.903
Ab1	6.011	5.797	6.208
Ab2	3.978	3.644	4.356
Ba1	5.967	5.769	6.169
Ba2	3.933	3.622	4.250
Bb1	6.133	5.914	6.336
Bb2	3.300	3.011	3.584
Ca1	5.811	5.544	6.042
Ca2	4.956	4.597	5.311
Cb1	5.706	5.428	5.961
Cb2	4.211	3.903	4.531
Da1	5.606	5.333	5.861
Da2	3.333	3.022	3.678
Db1	5.944	5.711	6.161
Db2	4.522	4.186	4.833

Table 7: DataFeat point estimates with 95% confidence intervals.

Stimulus	Estimate	Lower CI	Upper CI
Aa1	5.475	5.208	5.729
Aa2	4.975	4.654	5.258
Ab1	5.508	5.192	5.783
Ab2	4.633	4.317	4.959
Ba1	6.025	5.817	6.217
Ba2	5.058	4.767	5.350
Bb1	5.683	5.396	5.950
Bb2	4.183	3.879	4.479
Ca1	5.992	5.717	6.217
Ca2	5.725	5.446	5.979
Cb1	5.842	5.550	6.125
Cb2	5.217	4.908	5.517
Da1	5.450	5.142	5.721
Da2	4.375	4.058	4.708
Db1	5.542	5.242	5.838
Db2	4.817	4.487	5.154

Table 8: DataRead point estimates with 95% confidence intervals.

Stimulus	Estimate	Lower CI	Upper CI
Aa1	6.175	5.958	6.358
Aa2	5.417	5.133	5.654
Ab1	6.108	5.896	6.304
Ab2	5.083	4.771	5.379
Ba1	6.142	5.979	6.308
Ba2	5.108	4.825	5.388
Bb1	5.750	5.450	6.042
Bb2	4.208	3.925	4.488
Ca1	6.067	5.862	6.250
Ca2	5.642	5.417	5.854
Cb1	6.308	6.117	6.475
Cb2	5.558	5.263	5.871
Da1	5.792	5.525	6.008
Da2	4.283	3.933	4.613
Db1	6.008	5.758	6.246
Db2	5.108	4.775	5.413