



# Eyes are the Windows to the Soul: Predicting the Rating of Text Quality Using Gaze Behaviour



Sandeep Mathias<sup>1</sup>, Diptesh Kanojia<sup>1,2,3</sup>, Kevin Patel<sup>1</sup>, Samarth Agrawal<sup>1</sup>, Abhijit Mishra<sup>4</sup>, Pushpak Bhattacharyya<sup>1</sup>  
<sup>1</sup>CSE Department, IIT Bombay; <sup>2</sup>IITB-Monash Research Academy; <sup>3</sup>Monash University, Australia; <sup>4</sup>IBM Research, India

## Introduction

### Aim:

To predict the rating of different properties of text quality using text and gaze features.

### Eye-Tracking Terminology:

1. Interest Area – A part of the screen which is of interest.
2. Fixation – When the reader focuses on the screen.
3. Saccade – Movement of the eye from one fixation to the next.
4. Regression – Saccade to an earlier fixation.

### Properties:

1. Organization – How well-structured the text is.
2. Coherence – How much sense the text makes.
3. Cohesion – How well-connected the text is.

Each of these are scored on a scale of 1 to 4.

### Text Quality Rating:

Sum of the organization, coherence and cohesion scores, scaled to a range of 1 to 10.

$$Quality = Organization + Coherence + Cohesion - 2$$

## Method

"Ronald Reagan was elected President in 1980. He defeated Jimmy Carter by winning 44 out of the 50 American states. During the Reagan Era, the country was facing through inflation, a bad economy, and the American foreign policy was not as good. When Ronald Reagan became president, he signed the Economic Recovery Tax Act of 1981 which helped the economy. During Reagan's presidency, he also helped expand the American military. This also created more jobs, but also raised the deficit.

In 1984, Reagan won in a major landslide by winning 49 out of the 50 American states. During his second term, Reagan focused on ending the Cold War. Reagan met four times with Soviet leader Mikhail Gorbachev, and their summit conferences led to the signing of the Intermediate-Range Nuclear Forces Treaty.

Also during his second term, Reagan's Invasion of Grenada and bombing of Libya were popular in the US, though his backing of the Contras rebels was mired in the controversy over the Iran-Contra affair that revealed Reagan's poor management style. Liberals are annoyed by the Reagan Era, while conservatives extremely like it and thought it was the best thing to happen in American history.

**Figure:** Sample text showing fixations, saccades and regressions. The circles denote fixations, and the lines are saccades. This is the output from SR Research Data Viewer software.

### Collection of Gaze Data:

1. The reader reads a text, and answers 2 comprehension questions about the text.
2. The reader then scores the text for organization, coherence and cohesion.
3. The quality score of the text is got by adding the scores of each of the individual properties, and subtracting 2 from the sum.

The gaze features are collected using the SR-Research Eye-Tracker.

## Results

Property	Text	Gaze	Text + Gaze
Organization	0.237	0.394	<b>0.563</b>
Coherence	0.261	0.285	<b>0.550</b>
Cohesion	0.120	0.229	<b>0.451</b>
Quality	0.230	0.304	<b>0.552</b>

**Table:** Results for the three feature sets on different properties.

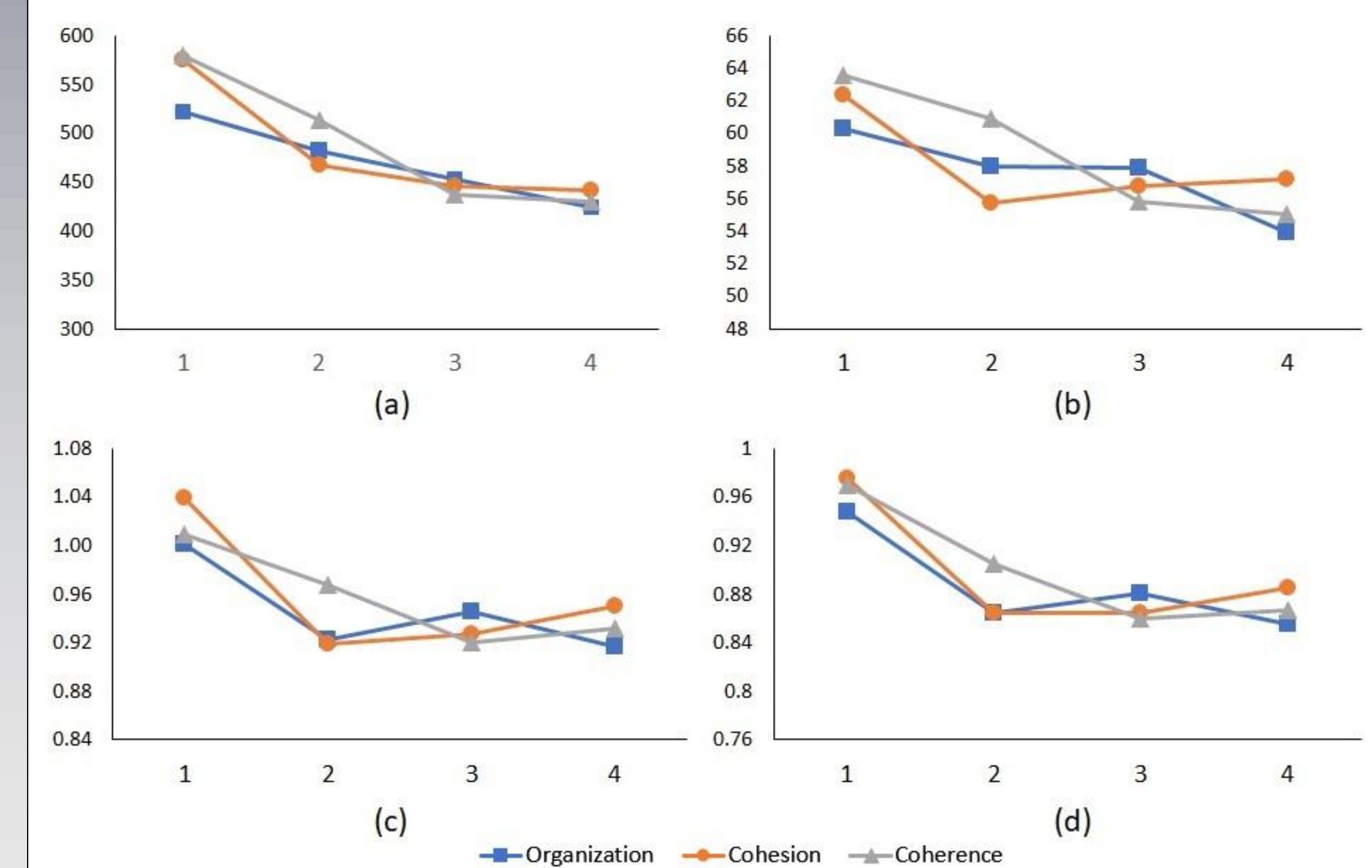
Property	Comprehension	Text	Gaze	Text + Gaze
Organization	Full	<b>0.319</b>	<b>0.319</b>	<b>0.563</b>
	Partial	0.115	0.179	0.283
Coherence	Full	0.255	<b>0.385</b>	<b>0.601</b>
	Partial	<b>0.365</b>	0.343	0.446
Cohesion	Full	<b>0.313</b>	<b>0.519</b>	<b>0.638</b>
	Partial	0.161	0.155	0.230
Quality	Full	<b>0.216</b>	<b>0.624</b>	<b>0.645</b>
	Partial	0.161	0.476	0.581

**Table:** Results for the three feature sets on different properties categorized on the basis of reader comprehension.

Property	Fixation	Regression	Interest Area
Organization	-0.102	-0.017	<b>-0.103</b>
Coherence	-0.049	-0.077	<b>-0.088</b>
Cohesion	-0.015	<b>-0.040</b>	0.037
Quality	0.002	0.016	<b>-0.056</b>

**Table:** Difference in QWK<sup>[3]</sup> scores when ablating each of the gaze behavior feature sets.

## Analysis



**Figure:** Relation between some of the different gaze features and the score. The gaze features are (a) Regression Duration, (b) Second Fixation Duration, (c) Fixation Count and (d) Run Count.

Texts with lots of fixations and regressions, as well as longer fixations and regressions tend to have lower scores, because the reader has to spend more time and effort in understanding it, compared to texts that are better written.

## Dataset Details

### Details of Texts:

1. No. of texts = 30
2. Size of texts = 200 words (approximately)
3. Source of texts: Simple English Wikipedia (10 articles), online news articles (12 articles), Wikipedia (8 articles)

### Details of Annotators:

1. Number of annotators = 20
2. Age of annotators = 20 to 25

### Scoring details:

1. Scoring Range: 1 to 4
2. Inter-Annotator Agreement Metric: Gwet's AC2<sup>[4]</sup>

Property	Full	Overall
Organization	0.610	0.519
Coherence	0.688	0.633
Cohesion	0.675	0.614

**Table:** Inter-Annotator Agreement (Gwet's AC2<sup>[4]</sup>) for the different properties – organization, coherence and cohesion. Quality score is calculated from these 3 properties. Full means participants who fully understood the text. Overall is without considering comprehension of the text.

## Features

Text-Based Features	Gaze-Based Features
1. Length-based features	1. Fixation features <ol style="list-style-type: none"> <li>1. First Fixation Duration</li> <li>2. Second Fixation Duration</li> <li>3. Last Fixation Duration</li> <li>4. Dwell Time</li> <li>5. Fixation Count</li> </ol>
2. Complexity features	2. Regression features <ol style="list-style-type: none"> <li>1. IsRegression</li> <li>2. Regression Count</li> <li>3. Regression Duration</li> </ol>
3. Stylistic features	3. Interest area features <ol style="list-style-type: none"> <li>1. Run Count</li> <li>2. Skip Count</li> </ol>
4. Word embeddings <sup>[5]</sup>	
5. Language modeling features	
6. Sequence features	
7. Entity grid <sup>[1]</sup> features	

Evaluation Method: 70% Training & 30% Testing data split  
Classifier Used: Feed-forward neural network<sup>[2]</sup>  
Number of Epochs: 10000

## Conclusions and Future Work

### Conclusions:

1. Gaze features help in better prediction of subjective properties of text, like organization, coherence, cohesion and quality.
2. Gaze features are more reliable if we take into account the reader's comprehension of the text.

### Future Work:

Using multi-task learning<sup>[6]</sup> in estimating gaze features and using those estimated features in our predictions.

## Bibliography

- [1] Regina Barzilay and Mirella Lapata. 2005. Modeling Local Coherence: An Entity-based Approach. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*. 141–148.
- [2] George Bebis and Michael Georgiopoulos. 1994. Feed-forward Neural Networks. *IEEE Potentials* 13(4):27–31.
- [3] Jacob Cohen. 1968. Weighted Kappa: Nominal Scale Agreement Provision for Scaled Disagreement or Partial Credit. *Psychological Bulletin* 70(4):213.
- [4] Kilem L Gwet. 2014. *Handbook of Inter-rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Raters*. Advanced Analytics, LLC.
- [5] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In *Advances in Neural Information Processing Systems*. 3111–3119.
- [6] Abhijit Mishra, Srikanth Tamilselvam, Riddhiman Dasgupta, Seema Nagar, and Kuntal Dey. 2018. Cognition-cognizant Sentiment Analysis with Multi-task Subjectivity Summarization Based on Annotators Gaze Behaviour. In *Proceedings of the 32nd Annual AAAI Conference on Artificial Intelligence*. 5884–5891.