

Reading Comprehension and Mental Model Development: A Cross-Validation of Methods and Technologies to Assess Student Understanding of the Text

Min Kyu Kim, Ali Heidari, and Kathryn Soo McCarthy
mkim120@gsu.edu, aheidari1@student.gsu.edu, kmccarthy12@gsu.edu
Georgia State University

Abstract: In this study, we developed a framework that conceptualizes two approaches to summary analysis (text-based and model-based) along three dimensions: Surface, structure, and semantic. We examined 47 cases in which students created an initial summary and a revision to explore how these two approaches track changes in students' summaries and mental models. The cross-validation between the two approaches demonstrated the power of assessing multiple dimensions of language to provide a more nuanced account of how students understand the text. Future studies should further investigate the relations between the two approaches and their theoretical and pedagogical meanings.

Introduction

Summary writing is essential because summaries act as an externalized re-representation of a student's mental model of the text (Johnson-Laird, 2005). Consequently, summaries can be used to assess students' understanding of the complex texts they encounter in their classes. In addition, writing a high quality summary of a text requires readers to identify main ideas as well as the connections between those ideas (Graesser, Singer, & Trabasso, 1994; Kintsch, 1988). Thus, helping students to construct a high quality summary of the reading material is also a common strategy for increasing students' understanding of the text. Given the important roles that summaries play in comprehension and learning, researchers have been working to develop effective ways to assess student-written summaries.

This body of research has, broadly, taken two different approaches: model-based and text-based. Model-based approaches elicit a concept model from the summary as an externally represented mental model of the individual's mind (Kim, Gaul, Kim, & Madathany, 2019; Kim, 2018). Text-based analytic approach quantifies multiple dimensions of language used in the summary (Crossley, Kyle & McNamara, 2016; Kyle, Crossley, & Berger, 2018). Though there are theoretical similarities and differences in these approaches, there have been few empirical evaluations of the relations between the two approaches in terms of analytical dimensions and assessment methods.

The current study

Though driven by similar theories, text-based and model-based indices approach the analysis of student's mental models in different ways. The purpose of this study was to examine how text-based indices are sensitive to changes in students' summaries from the initial and the final versions and to cross-validate these indices with indices from the model-based approach across dimensions. We focused on state-of-the-art methods and technologies that analyze summaries to evaluate students' reading comprehension. To calculate model-based indices, we used the Student Mental Model Analyzer for Research and Teaching (SMART; Kim et al., 2019), which uses natural language processing (NLP) to produce model-based indices at the three levels: surface, structure, semantic. In previous work (Kim et al., 2019), these SMART indices captured critical changes in student mental models as they revised their summaries. Thus, we used SMART as a benchmark upon which to compare text-based indices derived from NLP tools that assess various dimensions of language: Tool for the Automatic Assessment of Lexical Sophistication (TAALES, Kyle et al., 2018) and Tool for the Automatic Assessment of Cohesion (TAACO, Crossley et al., 2016).

Cross-validation framework

Both model-based and text-based approaches are grounded in theories that assume that readers produce mental models of texts and that these mental models are comprised of multiple levels of dimensions of representation. Theories of mental models (e.g., Kim, 2015; Ifenthaler, 2010; Spector & Koszalka, 2004) posit that mental models are comprised of three dimensions: surface, structure, and semantic dimensions. The surface dimension indicates the descriptive information of components of mental representation (e.g., number of concepts). The structural dimension relates to the way concepts are organized and connected in the mental model network (e.g.,

cohesiveness). The semantic dimension relates to the underlying ideas in the text, more specifically, individual concepts and propositional relations in a model. Similarly, theories of discourse comprehension (e.g., Graesser, Singer, & Trabasso, 1994; Kintsch, 1988) generally agree that mental models are comprised of three levels of representation: surface code, text-base, and situation model. The surface code includes the specific words in the text, conceptually corresponding to the surface dimension. Taken together with the text-based level (understanding propositional content), the situation level, the activation of higher-order mental representation of entities, characters, actions, and events, is associated with the structural and semantic characteristics of mental models. We summarized the theoretical dimensions, selected technologies, and measures across the two approaches (model-based vs. text-based) in Table 1, but describe them below.

Technologies such as SMART (Kim et al., 2019), GISK (Kim, 2018), and AKOVIA (Ifenthaler, 2014) build on the model-based framework. These technologies use a text input (i.e., summary) to elicit an underlying mental model in the form of a concept map from which various indices are generated. In this study, we used SMART analytics that provides the highest number of indices in the three dimensions (refer to the details of the SMART indices in Kim, 2015).

Technologies associated with the text-based approach, driven by theories of discourse comprehension, include the Tool for the Automatic Analysis of Lexical Sophistication (TAALES, Kyle et al., 2018) and the Tool for the Automatic Analysis of Cohesion (TAACO, Crossley et al., 2016). TAALES focuses on word-based metrics such as lexical frequency, psycholinguistic norms, and n-gram frequency. In contrast, TAACO evaluates discourse-level cohesion – that is, the degree to which information is connected across the text. TAACO calculates measures of overlap across sentence and paragraph, type-token ratio, and source-similarity. These tools derive hundreds of indices. We selected text-based indices that were theoretically-correspondent to dimensions of the model-based approach. These indices are:

- Content and Lemma TTR: Type-Token Ratio of content words (noun, verb, adjective, and adverb) and lemmas (e.g., measure is the lemma for measurement, measuring, and measurable); calculated by dividing the count of individual lexical items (types) by the total count of all words (Tokens).
- Overlap indices: Overlap between a given word (N, CN, or ARG) and its synonym set across an adjacent sentence (1S) or the next two sentences (2S).
- Source similarity: Source similarity indices provide the overall similarity between the words in a source text and a target text, building on Latent semantic analysis (LSA) – a statistical dimensionality reduction technique to locate extracted words in a multi-dimensional and Word2vec (W2V) – representing words in a vector-space model, putting words that occur in the same context as closer and those that occur in dissimilar contexts as more different in the vector space.

Table 1: Comparison of model-based and text-based approaches to summary analysis

Model-Based Approach			Text-Based Approach		
Dimension	Technology	Index*	Dimension	Technology	Index
Surface	SMART GISK AKOVIA	¹ Number of concepts; ² Number of <i>relations</i> ; ³ Density of the model	Surface Code	TAALES TAACO	Content TTR Lemma TTR
Structure		⁴ Average degree; ⁵ Mean distance; ⁶ Diameter	Text-based/ Situation model		Overlap N-1S; Overlap CN- 2S; Overlap N-2S; Overlap ARG.-2S
Semantic		⁷ Concept matching; ⁸ Propositional matching; ⁹ Recall-C; ¹⁰ Recall-P			Source Sim-LSA; Source Sim-W2V

Note. TTR (Type-Token Ratio), N (Noun), CN (Content words), ARG (Arguments), S (Sentence).

Method

Context and data source

We collected summaries from 38 students who used SMART in their graduate-level online course. Students wrote summaries of seven readings at various points throughout the semester (weeks 4, 7, and 11). In SMART, students

read a text, write a summary, and then received feedback in terms of missing concept and propositional relations along with an exemplary expert summary. They are then given the opportunity revise their summary and receive feedback as many times as desired. In the current study, we analyzed 47 cases in which students produced at least two revisions. Our analyses focus on the initial and the final version of the summaries.

Data processing and analysis

Students' summaries and the model-based indices were extracted from the SMART system. These summaries were then submitted to TAACO and TAALES to extract the text-based indices. We first, used a series of one-way repeated-measures Multivariate Analysis of Variance (MANOVA) to examine if the text-based indices changed from initial to final versions. We then examined the correlations between the text-based and model-based indices at each of the three dimensions for the final versions.

Results and discussion

Consistent with analysis of the model-based indices (Gaul, Bundrage, Madathany, & Kim, 2020), MANOVAs indicated significant changes in the text-based indices from initial to final summary across all three dimensions. Further analyses reveal interesting patterns of relations between text-based and model-based indices.

For the surface level dimension, univariate tests revealed a significant decrease in Content TTR and Lemma TTR (Table 2). Student summaries tended to decrease in lexical diversity over time (a decrease in TTR values) while they became more focused on essential concepts. However, these text-based indices were not consistently correlated with the model-based indices. Only Content TTR and Number of relations showed a significant (negative) correlation, indicating that students might connect more concepts with an optimal number of content words.

Table 2: Surface-related indices and correlations with the model-based indices

Index	Initial	Final	<i>F</i>	<i>p</i>	η^2	Correlation w/ Model-based Indices ^a		
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)				<i>N</i> Concepts	<i>N</i> Relations	Density
Content TTR	0.68 (0.10)	0.64 (0.07)	5.54	.023	0.11	-0.21	-0.48**	-0.05
Lemma TTR	0.55 (0.10)	0.51 (0.05)	6.18	.017	0.12	0.01	-0.28	-0.24

Note. a. * $p < .05$, ** $p < .01$.

All changes in the structure-related indices were statistically significant (Table 3). All the indices negatively related to Mean Distance and Diameter, meaning that students who used more overlapped nouns, content words, and arguments across the next two sentences built a concise and cohesive mental model (a shorter mean distance and diameter).

Table 3: ANOVA results of the structure-related indices and the correlations with the model-based indices

Index	Initial	Final	<i>F</i>	<i>p</i>	η^2	Correlation w/ Model-based Indices ^a		
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)				Avg. Degree	M Distance	Diameter
Overlap N-1S	0.14 (0.08)	0.15 (0.07)	1.43	0.24	0.03	0.55**	-0.38*	-0.43**
Overlap CN-2S	0.71 (0.26)	0.82 (0.17)	6.05	0.02	0.12	0.22	-0.38**	-0.34*
Overlap N-2S	0.62 (0.26)	0.72 (0.19)	7.96	0.01	0.15	0.27	-0.38*	-0.32*
Overlap ARG-2S	0.74 (0.18)	0.63 (0.25)	8.21	0.01	0.16	0.23	-0.34*	-0.33*

Note. a. * $p < .05$, ** $p < .01$.

Lastly, MANOVA revealed that the semantic-related indices improved from initial to final summary (Table 4). This suggests that students included more ideas related to the source text in their revisions. Importantly, these two similarity indices were correlated with the four model-based semantic indices, ranging from $r = 0.52$ and $r = 0.74$, with $p < .01$. These strong correlations indicate good validity across the text-based and model-based semantic indices.

Conclusion

This study cross-validated two approaches to summary analysis. These findings support the theoretical overlap of model-based and the text-based approaches. The findings of the study demonstrate that the text-based indices could be grouped into the surface, structure, and semantic dimensions, providing a further explanation of the relationships between summary writing and the embedded mental model.

This cross-validation study demonstrates the connection between the way to write a summary and the quality of a mental model of the text. Future studies should further investigate the relationships of the indices from the two approaches and assessment of reading comprehension in multiple dimensions to drive more specific feedback to learners.

Table 4: ANOVA results of the semantic-related indices and the correlations with the model-based indices

Index	Initial	Final	<i>F</i>	<i>p</i>	η^2	Correlation w/ Model-based Indices ^a			
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)				Concept Match	Prop. Match	Recall-C	Recall-P
Source Sim-LSA	0.77 (0.14)	0.85 (0.09)	14.26	0.00	0.23	0.74**	0.54**	0.70**	0.66**
Source Sim-Word2Vec	0.88 (0.08)	0.92 (0.05)	9.34	0.00	0.18	0.71**	0.52**	0.67**	0.63**

Note. a. * $p < .05$, ** $p < .01$.

References

- Carnine, L., & Carnine, D. (2004). The interaction of reading skills and science content knowledge when teaching struggling secondary students. *Reading & Writing Quarterly*, 20, 203-218.
- Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods* 48(4), pp. 1227-1237. doi:10.3758/s13428-015-0651-7
- Gaul, C., & Kim, M. (2020, April). Design research on learner participation feedback within peer moderated asynchronous online discussion. Paper presented at The 2020 American Educational Research Association Annual Meeting, San Francisco, California.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101(3), 371-395.
- Ifenthaler, D. (2014). AKOVIA: Automated Knowledge Visualization and Assessment. *Technology, Knowledge and Learning*, 19(1-2), 241-248.
- Ifenthaler, D. (2010). Relational, structural, and semantic analysis of graphical representations and concept maps. *Educational Technology Research & Development*, 58(1), 81-97.
- Johnson-Laird, P. N. (2005). Mental models and thoughts. In K. J. Holyoak (Ed.), *The Cambridge handbook of thinking and reasoning* (pp. 185–208). Cambridge: Cambridge University Press.
- Kim, K. (2018). An automatic measure of cross-language text structures. *Technology, Knowledge and Learning*, 23(2), 301–314.
- Kim, M. (2015). Models of learning progress in solving complex problems: expertise development in teaching and learning. *Contemporary Educational Psychology*. 42. 1-16.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological review*, 95(2), 163-182.
- Kim, M., Gaul, C., Kim, S., & Madathany, R. (2019). Advance in detecting key concepts as an expert model: Using Student Mental Model Analyzer for Research and Teaching (SMART). *Technology, Knowledge and Learning*, 1-24.
- Kyle, K., Crossley, S., & Berger, C. (2018). The tool for the automatic analysis of lexical sophistication (TAALES): version 2.0. *Behavior research methods*, 50(3), 1030-1046.
- Spector, J. M., & Koszalka, T. A. (2004). *The DEEP methodology for assessing learning in complex domains* (Final report to the National Science Foundation Evaluative Research and Evaluation Capacity Building). Syracuse, NY: Syracuse University.
- Van Dijk, T. A., Kintsch, W., & Van Dijk, T. A. (1983). *Strategies of discourse comprehension*.