Quality Question Means and Significances

Following are the detailed results of the quality questions described in "Sentence Simplification, Compression, and Disaggregation for Summarization of Sophisticated Documents." Evaluators were asked to rate the subjective quality of each summary using the seven DUC quality questions. We assigned each answer a numerical score from 0 to 4, with "a," the most positive evaluation, worth four points, "b" worth 3 points, and so on to a minimum of 0 for "e," the most negative evaluation. The following tables give the mean scores for summaries from each condition on each question. In addition, they indicate which means are significantly different. Overall, the Compressed condition was most often significantly different from the LexRank Only and Human conditions.

Condition	Mean	Significantly Different From
Simplified	1.7895	LexRank Only,* Human*
Compressed	1.3500	LexRank Only,** Human**
Disaggregated	1.5263	LexRank Only*, Human*
LexRank Only	2.9444	Simplified,* Compressed,** Disaggregated*
Human	3.2500	Simplified,* Compressed,** Disaggregated*

* $p < .05$; ** $p < .001$

Table 1: Mean scores for Quality Question 1: "Does the summary build from sentence to sentence to a coherent body of information about the topic?"
0 = Incoherent, 4 = Very coherently

Condition	Mean	Significantly Different From
Simplified	1.7895	Human*
Compressed	1.5000	LexRank Only,* Human**
Disaggregated	1.8421	Human*
LexRank Only	2.5000	Compressed*
Human	3.1250	Simplified,* Compressed,** Disggregated*

* $p < .05$; ** $p < .001$

Table 2: Mean scores for Quality Question 2 "If you were editing the summary to make it more concise and to the point, how much useless or confusing text would you remove from the existing summary?"
0 = Most of the text, 4 = None

Condition	Mean	Significantly Different From
Simplified	3.1250	
Compressed	2.6000	
Disaggregated	2.6842	
LexRank Only	3.3333	
Human	3.2500	

* $p < .05$; ** $p < .001$

Table 3: Mean scores for Quality Question 3, "To what degree does the summary say the same thing over again?"
0 = Quite a lot, 4 = None

Condition	Mean	Significantly Different From
Simplified	2.0000	Human**
Compressed	1.4500	LexRank Only,** Human**
Disaggregated	1.8421	LexRank Only,* Human**
LexRank Only	2.8889	Compressed,** Disaggregated*
Human	3.8750	Simplified,** Compressed,** Disaggregated**

* $p < .05$; ** $p < .001$

Table 4: Mean scores for Quality Question 4, "How much trouble did you have identifying the referents of noun phrases in this summary? ..."
0 = Severe problems, 4 = No problems

Condition	Mean	Significantly Different From
Simplified	2.6316	
Compressed	2.6000	
Disaggregated	2.5789	
LexRank Only	3.1111	
Human	3.1250	

* $p < .05$; ** $p < .001$

Table 5: Mean scores for Quality Question 5, "To what degree do you think the entities (person/thing/event/place/...) were re-mentioned in an overly explicit way, so that readability was impaired? For example, a pronoun could have been used instead of a lengthy description, or a shorter description would have been more appropriate?"
0 = A lot, 4 = None

Condition	Mean	Significantly Different From
Simplified	1.8947	Compressed,** Disaggregated,* LexRank Only,** Human**
Compressed	0.8500	Simplified,** LexRank Only,** Human**
Disaggregated	1.2632	Simplified,* LexRank Only,** Human**
LexRank Only	2.9444	Simplified,** Compressed,** Disaggregated**
Human	3.6250	Simplified,** Compressed,** Disaggregated**

* $p < .05$; ** $p < .001$

Table 6: Mean scores for Quality Question 6, "Are there any obviously ungrammatical sentences, e.g., missing components, unrelated fragments or any other grammar-related problem that makes the text difficult to read?"
0 = Too many problems, 4 = No noticeable grammatical problems

Condition	Mean	Significantly Different From
Simplified	2.2632	
Compressed	1.6316	Human*
Disaggregated	1.9474	Human*
LexRank Only	2.2778	
Human	3.3750	Compressed,* Disaggregated*

* $p < .05$; ** $p < .001$

Table 7: Mean scores for Quality Question 7, "Are there any datelines, system-internal formatting or capitalization errors that can make the reading of the summary difficult?"
0 = Many, 4 = No noticeable formatting problems