Missouri Assessment Program

ELA Grades 5 and 8

Spring 2015

Copying from Source or Copying from another Student Analyses Report

Data Recognition Corporation | CTB

October 1, 2015

2015 MAP English Language Arts / Literacy Grades 5 and 8

Copying from Source or Copying from another Student Response Analysis

Introduction

As a part of test security and data forensics, DRC/CTB has conducted text analysis on Writing Essay (WER) items administered as part of English Language Arts / Literacy (ELA) grades 5 and 8 tests to flag student essays under suspicion of copying or plagiarism. Specifically, the goals of the analysis were to identify the student responses that were copied from the sources and the responses that were copied from other students.

In order to identify the responses that were copied from the sources, a feature based classification rule was employed to identify student responses from condition code "B" score group. According to the scoring rubric, a condition code "B" will be awarded to a student response if the responses fall into any of the following category:

1. Un-scorable sentence fragments

2. Responses those are too brief

3. Copying or restating the prompt.

In an effort to understand the nature of the student responses, the current study helps to identify the students who may try to trick the system or a scorer by copying or restating the prompt and/ or instructions. Furthermore, the employed procedure also helps in identifying the student responses that were copied from the sources and were given a valid numeric score by the hand-scorer, when they were supposed to be given a condition code "B".

To identify the student responses that were copied from neighboring students at school level, a cosine similarity measure in conjunction with rule-based classification was used.

Methodology

<u>Identification of student responses copied from the sources</u>

The data analysis method described in this section was developed by DRC|CTB. In order to identify the responses that are copied from the sources, two features were computed based on the content of the student response and of the source. The two developed features are defined as:

$$All\ word\ ratio = \frac{total\ no.of\ unique\ words\ common\ in\ student\ response\ and\ source}{total\ no.of\ words\ in\ the\ student\ response}$$

$$Copy\ block\ ratio = \frac{\sum length\ of\ copied\ string > 20}{character\ count\ of\ student\ response}$$

The feature '*All word ratio*' ranges between [0, 1] and provides numeric insights into the amount of vocabulary that student uses from the source. For example if a student constructs his response completely with the vocabulary that is present in the source the resulting feature value is 1 and if the student did not use any of the vocabulary from the source the feature value is 0.

In the similar way, for each response, the feature 'Copy block ratio' ranges between [0, 1] and identifies the total length of the substrings that are common in both response and sources, which are greater than a string length of 20 characters, divided by the response character count. A perfect value of 1 indicates that the response is completely copied from the source and 0 value indicates that there is no substring in the response that is copied from the source that is of length greater than 20 characters. An arbitrary adjustable value of 20 is chosen so as not to accidentally flag any responses that quote a certain part of the source.

In the preprocessing stage, the student response and the sources were converted into the lower case characters and striped of the punctuations so only the alpha-numeric content of the essay is compared to the source. After computing the above two features from the lower case source and response content, and taking into consideration the hand-scoring information all student responses for each prompt were divided into two groups:

1. Condition code B group

2. Numeric score group

The "Condition code B" group is the responses that were given the condition code 'B' and the "Numeric score" group is the responses with scores of 1, 2, 3, or 4 awarded by a hand scorer.

In those two groups, the student responses were flagged based on features with 'All word ratio' > Th_1 and 'Copy block ratio' > Th_2. Th_1 and Th_2 are the arbitrary thresholds (0.85 and 0.70 respectively) selected based on careful manual review of 100 responses from each prompt.

The flag in the 'Condition code B' group indicates that the responses were copied from the sources and were awarded a condition code 'B'. The flag in the 'Numeric score group' indicated that the responses were also copied from the sources but were awarded a numeric score, when supposed to be scored with a condition code 'B'. The list of schools (in minimum of 5 districts) with most flags by grade and prompt in both groups are provided in a separate excel file MAP2015_RespCopyAnalysis.xlsx for DESE review.

<u>Identification of student responses copied from another student</u>

In order to identify the students who copied from their peers, a cosine similarity measure was employed (Manning, Raghavan, and Schütze, 2008; Singhal, 2001). Cosine similarity is a measure of similarity between two vectors of an inner product space that measures the cosine of the angle between them. Two vectors with the same orientation have a cosine similarity of 1.

$$similarity = \cos(\theta)\ \frac{A.B}{\|A\|\|B\|}$$

The student response data were grouped at the prompt level in each school (based on assumption that being in the same school may create an opportunity for a student to copy from another student). A unique set of all words that were used by the students in their responses to a given prompt in a given school was arranged in an ordered sequence for every student within this school who responded to the same prompt and the responses were converted into a vector based on the presence of the word in their response. If a given word was present in the student response, the vector value for this word was 1 and if not present then vector value for this word was 0. For each prompts in each school, the vector conversion process was repeated and pair-wise similarity measures were computed among the student responses. The students and their responses were flagged if the cosine similarity value was greater

than the predefined threshold Th_3. To find the extreme similar cases and after careful manual evaluation, the threshold value for flagging was selected to be 0.85 and any similarity value greater than the threshold was flagged as ‘copying from another student’.

Results

Responses flagged for copying from sources

The ‘copying from sources’ analysis included a total of 132050 student responses from 1576 schools a nd 555 school districts in Missouri. Tables 1 below provide grade and prompt level descriptive statistical information.

Table 1. Prompt level descriptive statistics

Grade	Prompt code	DRC item id	Genre	Number of Students	Total number of flags 'Copied from source'		Total number of flags 'Copied from source' and given condition code B		Total number of flags 'Copied from source' and given a score	
					Given condition code B	Awarded a score	Min in one school	Max in one school	Min in one school	Max in one school
5	55425	721666	Narrative	29161	323	231	0	7	0	6
5	55435	721668	Informational	18395	609	434	0	8	0	5
5	56365	721674	Opinion	18462	143	329	0	6	0	5
8	55415	721662	Narrative	18559	163	131	0	6	0	5
8	55419	721665	Explanatory	18663	93	545	0	4	0	10
8	56362	721673	Argumentative	28810	224	414	0	10	0	13

Responses flagged for copying from another student

When the data were analyzed for possible copying from another student, only one pair of students in one school was flagged for similarity of their responses for prompt 55425 (grade 5, form PA1) with a cosine similarity value of 0.9021. Manual review of the responses written by these students confirmed high level of response similarity. Information about these students will be provided to DESE upon request.

Final Remarks

This study examines potential copying of written responses from a source or from another student on the spring 2015 MAP Grade 5 and 8 ELA assessments. Note that only statistical criteria were used to flag the students for possible copying of their answers and no qualitative review of flagged responses was performed (except for the pair of responses flagged for copying from another student). It also should be noted that it is possible that responses flagged as having been copied and received a score instead of a Condition Code B were a combination of copied work and original student work. Such responses are typically considered scorable if there is sufficient original work in the response (even if portions of the response are copied). Therefore, in order to determine whether a student response should be considered as ‘copied from a source’ a qualitative review of such responses should be

performed. A statistical flag on its own does not necessarily indicate inappropriate behavior or unscorable response.

References

Manning, C.D., Raghavan, P., and Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press, Cambridge, England.

Singhal, A. (2001). Modern Information Retrieval: A Brief Overview. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering* 24 (4): 35–43