

# Using vector space models based on LDA and LSA for automatically grading exam questions

Johanna de Vos

Donders Institute, Radboud University, Nijmegen, The Netherlands

## KEY IDEAS

- Determining semantic similarity between student answer and 'perfect' reference answer in a vector space
- Varying what the dimensions in that vector space represent
- Varying on what data the LDA and LSA models are trained
- Correlating predicted grades and lecturer-assigned grades (Spearman's  $r_s$ )

## RESEARCH QUESTIONS

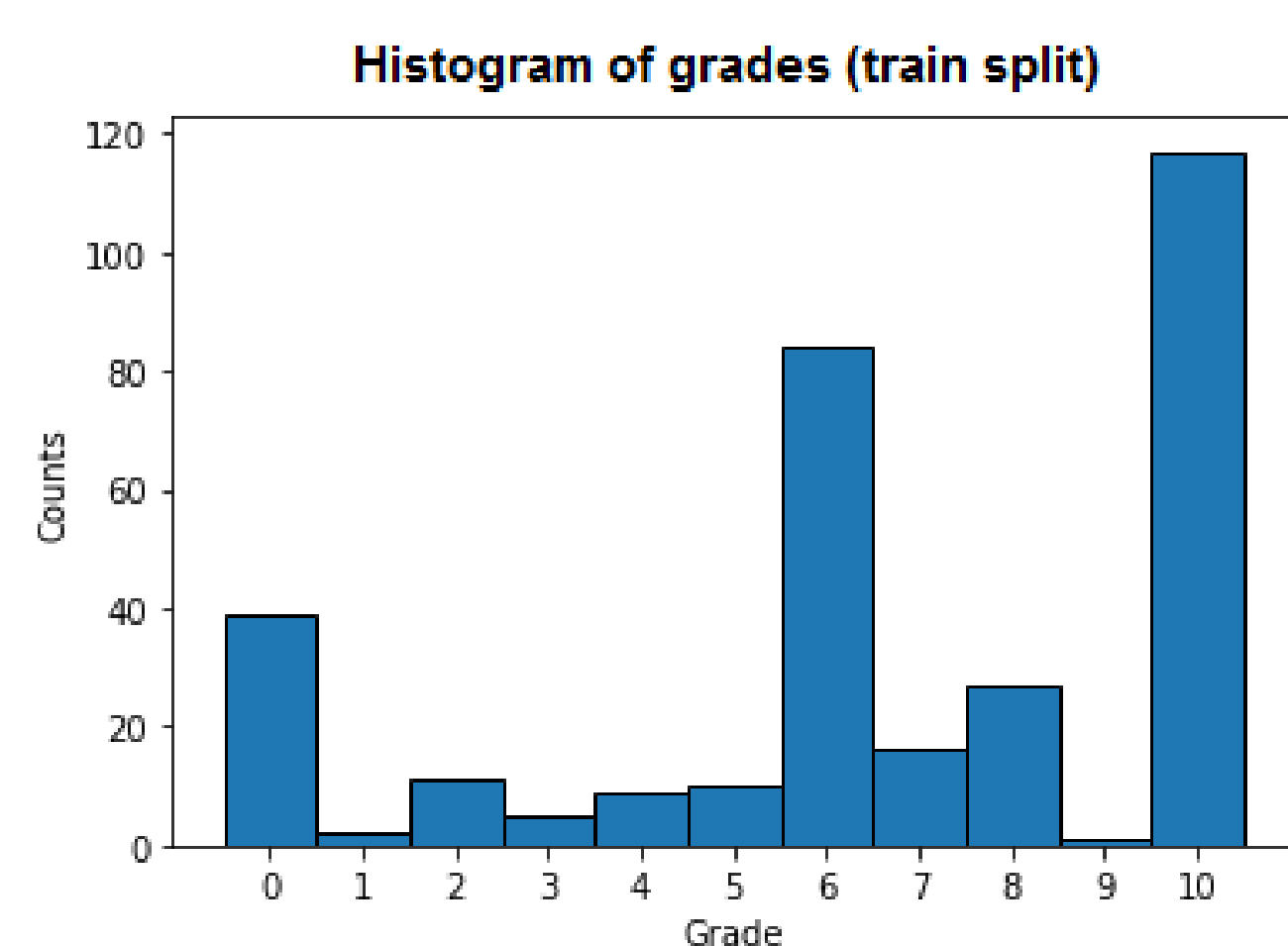
1. How effective are LDA-based and LSA-based vector space models for automatically grading students' answers to open exam questions?
2. Does the effectiveness of the above models depend on the training data that is used?
3. Are these topic models more effective than a simple vector space model based on vocabulary counts?

## APPROACH

### Data

402 graded answers to the following question:

"Discuss Whorf's language theory. Include the following terms in your answer: Strong and weak variations on the theory."



### Pre-processing

- Spelling correction, tokenizing, lemmatizing, stop word removal

### Training and testing

#### Training on:

- 321 student answers **versus** chapter from a psychology textbook

#### For both types of training data:

- Find optimal number of topics (LDA) or singular values (LSA)

#### Testing on:

- 81 student answers

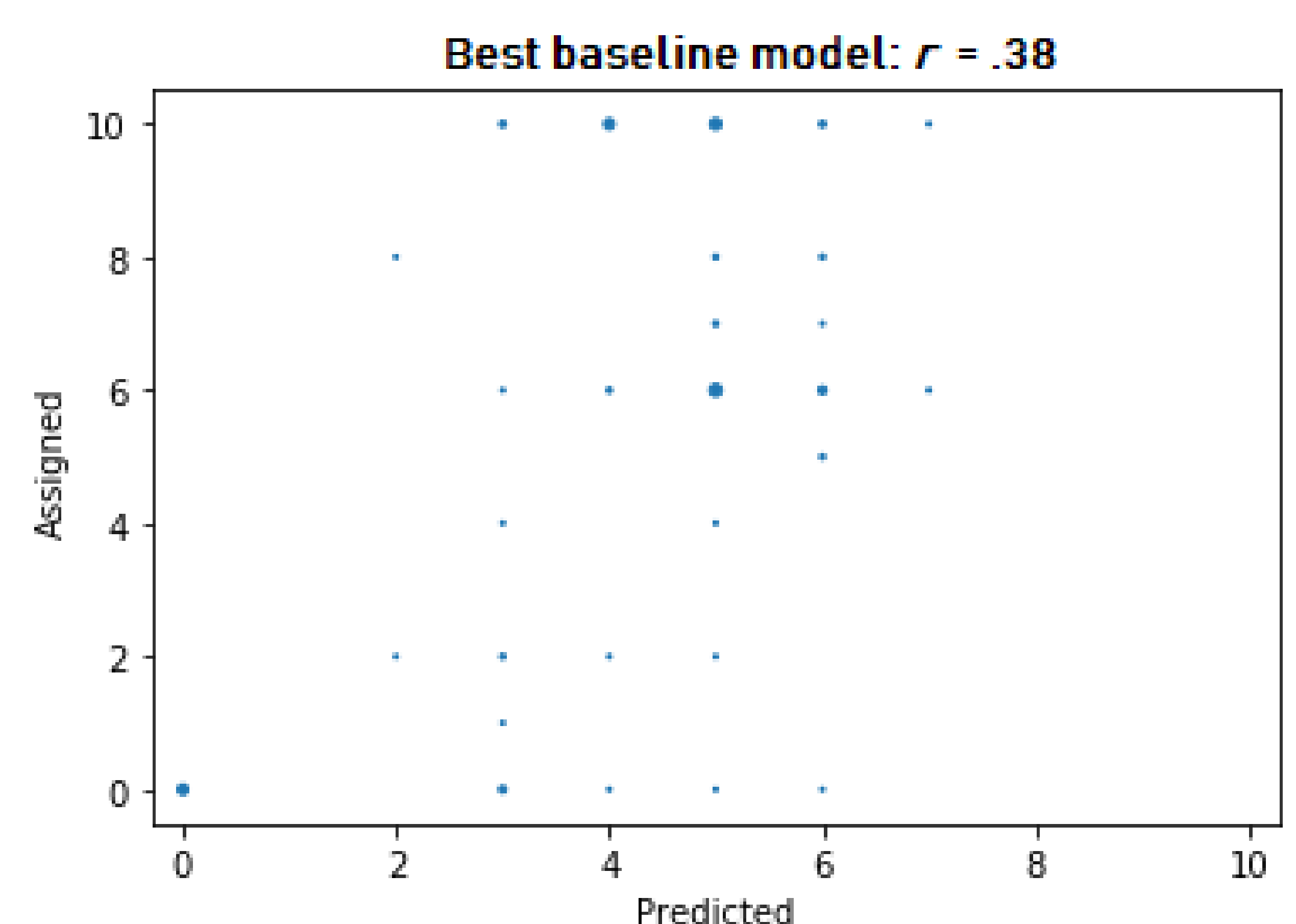
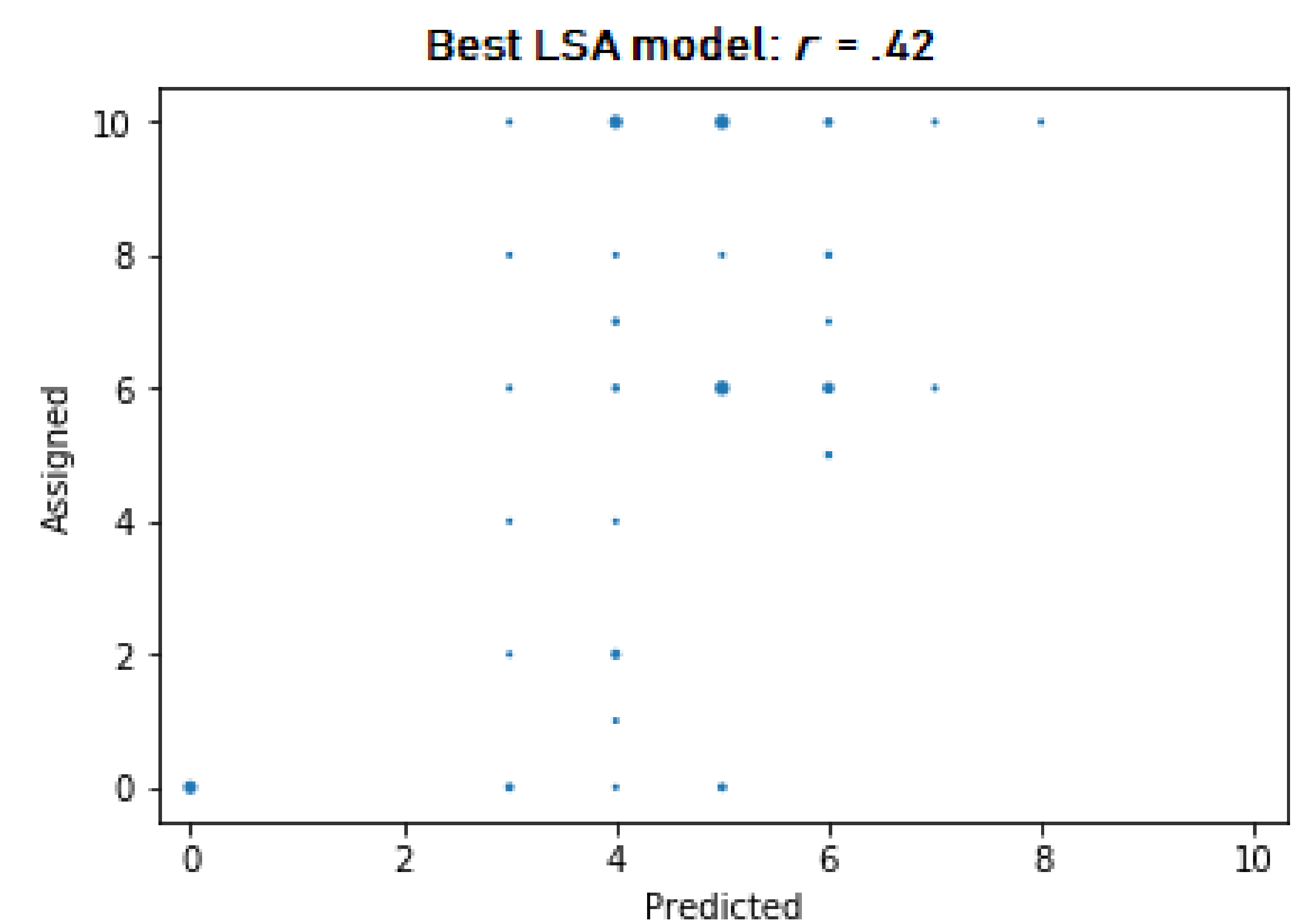
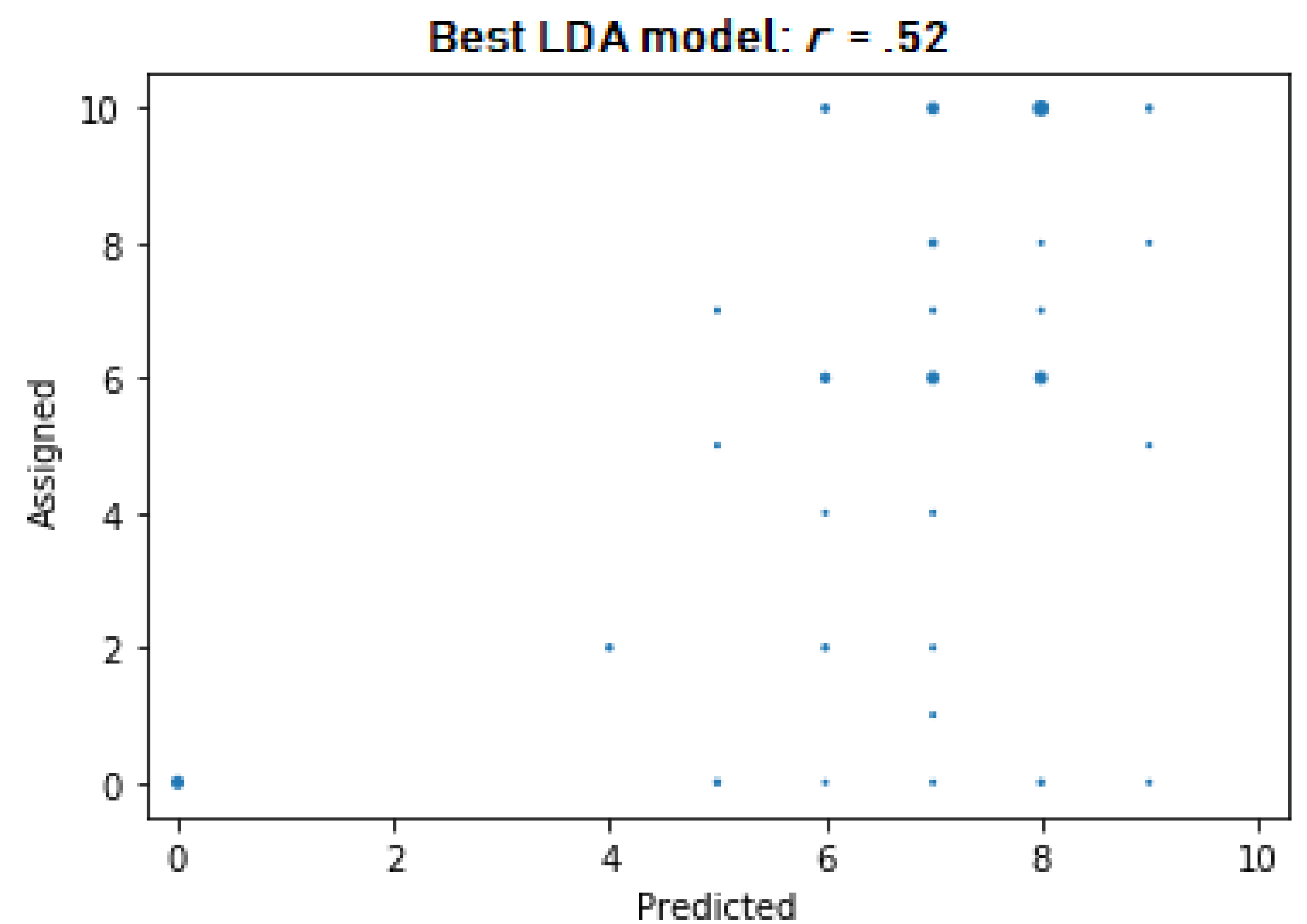
### Exploring variables

- Counting method: raw counts **versus** binary counts **versus** TF-IDF
- Spelling correction: yes **versus** no
- Mapping algorithm: x10 **versus** x10 but excluding 1 and 9

## RESULTS

		Baseline	LDA	LSA
Training data	Student answers	N/A	.36	.39
	Textbook chapter	N/A	.42	.32

## RESULTS



## DISCUSSION

Points for improvement:

- Assign the maximum grade (10) more often
- Don't punish creative answers

## CONCLUSION

- Going in the right direction, but not yet good enough for implementation in higher education