A Data Mining Approach for Detecting Collusion in Unproctored Online Exams

J. Langerbein, T. Massing, J. Klenke, N. Reckmann, M. Striewe,
M. Goedicke, C. Hanck

University of Duisburg-Essen; Germany

Setting

- Data from the *Descriptive Statistics* course at U Duisburg-Essen, Germany
- Exams consist of arithmetical problems, programming tasks in `R`, and a short essay task
- Exams are conducted digitally with the e-assessment system JACK
  - Each student receives different randomized numerical values across all tasks
  - Event logs capture students' activities, time stamps, and points during the exams for every subtask

Table 1: Overview of the test and comparison group

	Comparison	Test
Year	18/19	20/21
N	109	151
Style	proctored	unproctored
Total points	60	60
Sub tasks	19	17
Duration	70	70

- The test group (2020/21) took the unproctored exam at home
  - The comparison group (2018/19) took a proctored exam at the university
- Data cleaning is conducted, removing students with minimal participation or achievement and students with internet problems

Aim of the Paper

Detecting potential collusion with a hierarchical clustering algorithm on event logs and strengthen the analysis with a proctored comparison group

Methodology

- The study utilizes an agglomerative (bottom-up) hierarchical clustering algorithm characterized by the following equation:

$$D(s_i, s_{i'}, v_i, v_{i'}) = \frac{1}{h}\sum_{j=1}^{h}\left(w_j^P \cdot d_j^P(s_{ij}, s_{i'j}) + w_j^L \cdot d_j^L(v_{ij}, v_{i'j})\right)$$

  - $D(s_i, s_{i'}, v_i, v_{i'})$ global pairwise dissimilarity
  - $d_j^P(s_{ij}, s_{i'j})$ points dissimilarity for each task j
  - $d_j^L(v_{ij}, v_{i'j})$ students event patterns dissimilarity for each task j
  - $\sum_{j=1}^{h} w_j^P + w_j^L = 1$ weight of each attribute h

- We reduce the weights for
  - `R`-tasks, as these tasks have more noise
  - Essay questions, as comparisons on these kinds of tasks are limited
  - Points achieved
- Dissimilarities in points achieved for each task j

$$d_j^P(s_{ij}, s_{i'j}) = |s_{ij} - s_{i'j}|$$

  - s_{ij} denotes the points achieved by student i in the j-th subtask
  - Manhatten metric

- Dissimilarities in the students event patterns (time of submission) for each task j

$$d_j^L(v_{ij}, v_{i'j}) = \sum_{m=1}^{K=70} |v_{ijm} - v_{i'jm}|$$

  - $d_j^L(v_{ij}, v_{i'j})$ students event patterns dissimilarity for each task j
  - Examination is divided into $m = 1, \ldots, 70$ time intervals
  - v_{ijm} denotes the number of answers of student i for task j in the m-th interval

Empirical Results

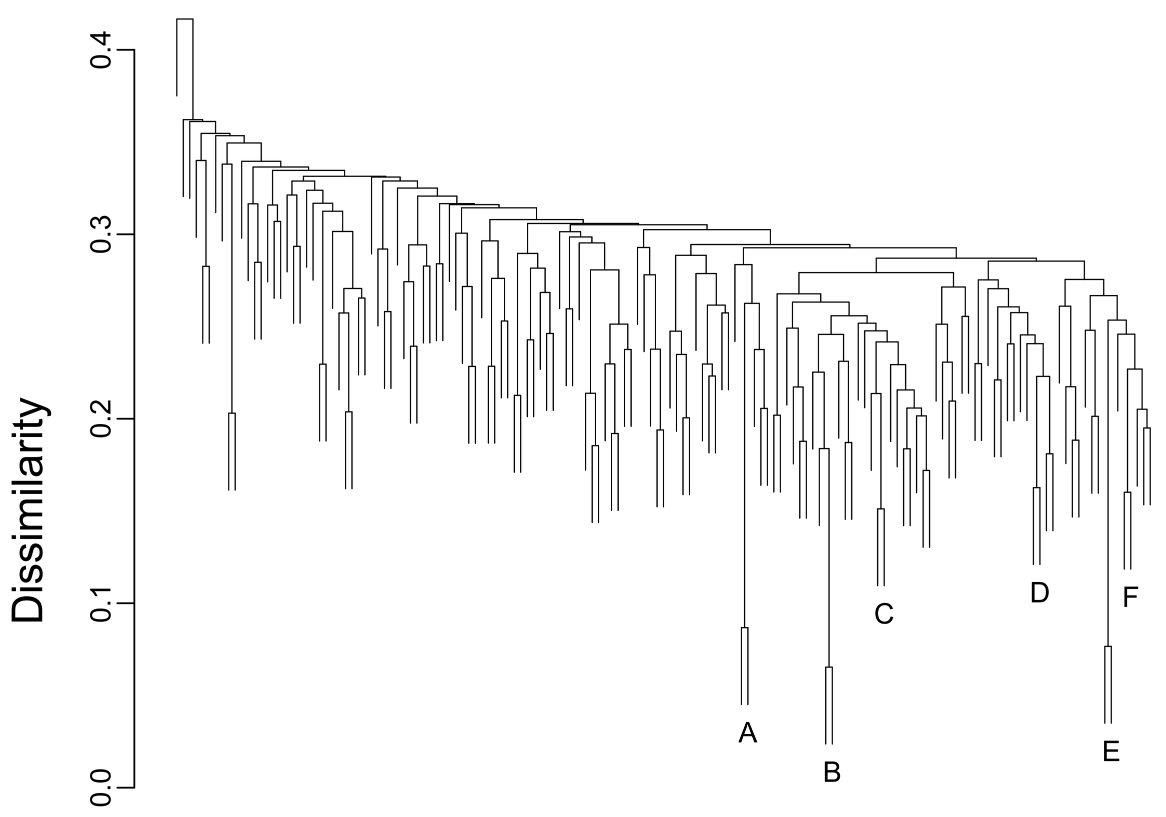

Figure 1: Dendogram produced by average linkage clustering of the unproctored test group (2020/21). **A-F** mark the clusters with the lowest dissimilarity.

- Figure 1 shows the dendrogram of the **test** group
  - Overall a lower level of dissimilarity compared to the comparison group
  - Six clusters (**A-F**) standing out noticeably from the rest of the cohort, suggesting potential collusion

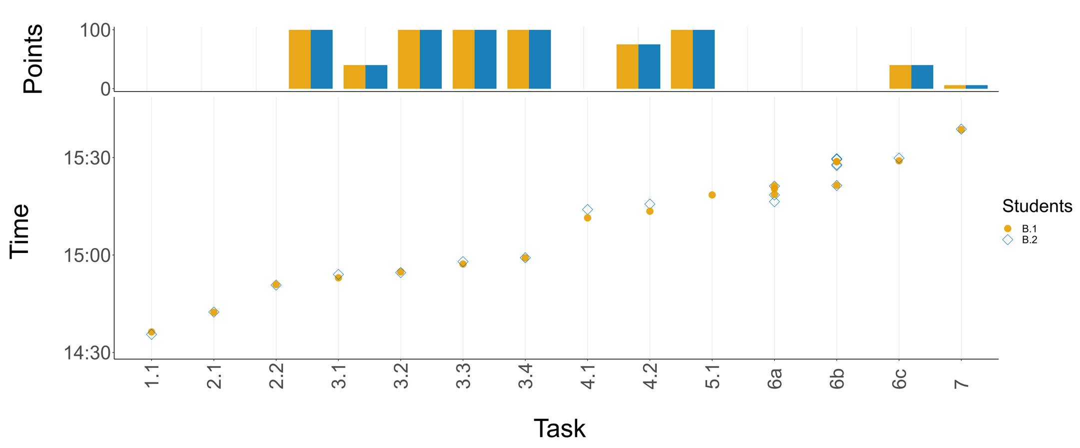

Figure 2: Event logs and achieved points of the cluster **B** from the test group (2020/21). Above the scatter plot, a bar chart is added to compare the points per subtask.

- Figure 2 illustrates the individual comparison of achieved points and event logs of the student cluster with the highest similarity
  - Similar time path and same points for each task

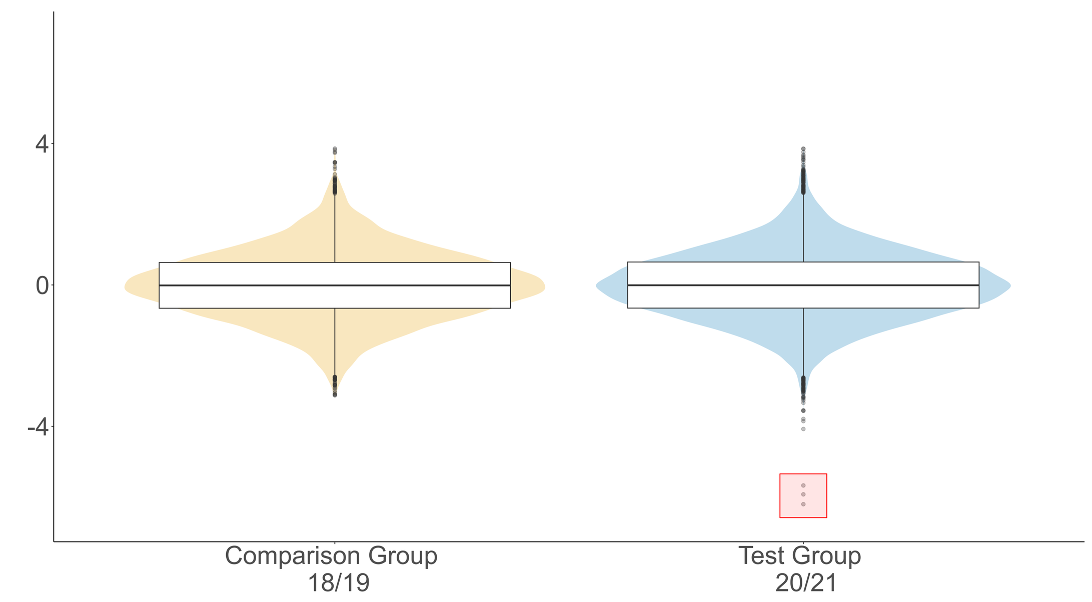

Figure 3: Comparison of the normalized distance measures.

- Figure 3 compares the normalized distributions of the dissimilarity measures between the comparison and test groups

Discussion

- Three notable clusters (**A**, **B**, and **E**) consisting of two students each
- Collusion in larger groups is not found
- Findings do not depend on linkage methods and parameter specifications
- The approach provides a basis for the examination of clusters based on comparison with a reference group