



FernUniversität
in Hagen

Analysing Collaborative Writing Activities

divAdapt <> LA DIVA

2022/10/18

CATALPA

Niels Seidel & Marc Burchart

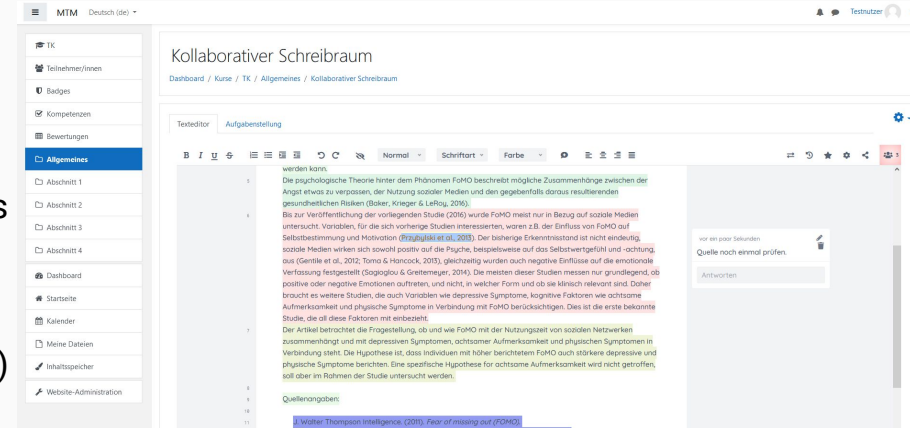
HUMBOLDT-UNIVERSITÄT
ZU BERLIN



FernUniversität in Hagen

Current status of the collaborative writing infrastructure

- Basic infrastructure to support and study collaborative writing
 - scalability for a large number of students
 - increased reliability and security
 - archive with read access for former semesters
 - peer review functionality
- data collection for Learning Analytics
- already in used by CATALPLA (e.g. divAdapt, APLE II)



Current datasets:

- Summer semester 2021: 415,892 entries (N=229, K=55)
- Winter semester 2021/22: 5,462,109 entries (N=2025, K=296)
- Summer semester 2022: ?

Input	Behavior	Product
session	login, logout, idle time	-
entered text	character change set, formatting, scroll	text artifact
chat messages	submit, scroll	message content
comments incl. replies	submit/edit/accept, text anchor, document	message content
system	errors, cpu, network and memory usage	-
self-reports	-	response items, text inputs

Data analysis

Goals

- Identify functional and dysfunctional groups.
- Support teachers to manage large-scale writing groups.
- Support students and student groups in writing.

Individual measures

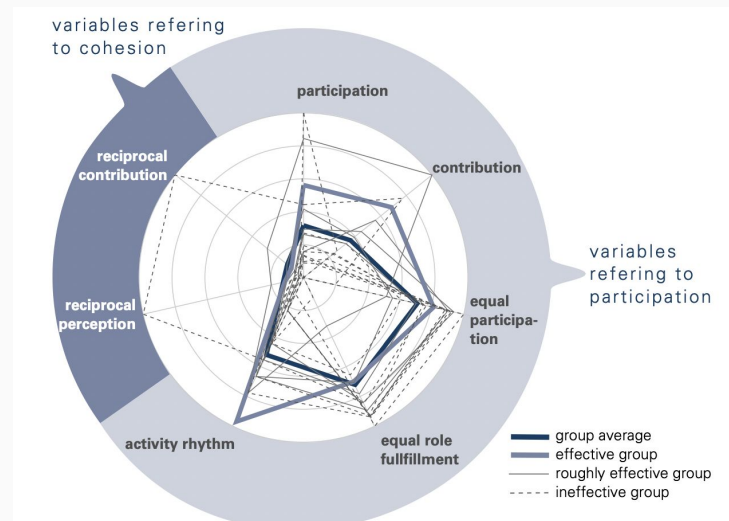
- writing process > writing patterns
- writing outcome

Group measures (Calvani, Fini & Molino, 2010; Martinez et al., 2012)

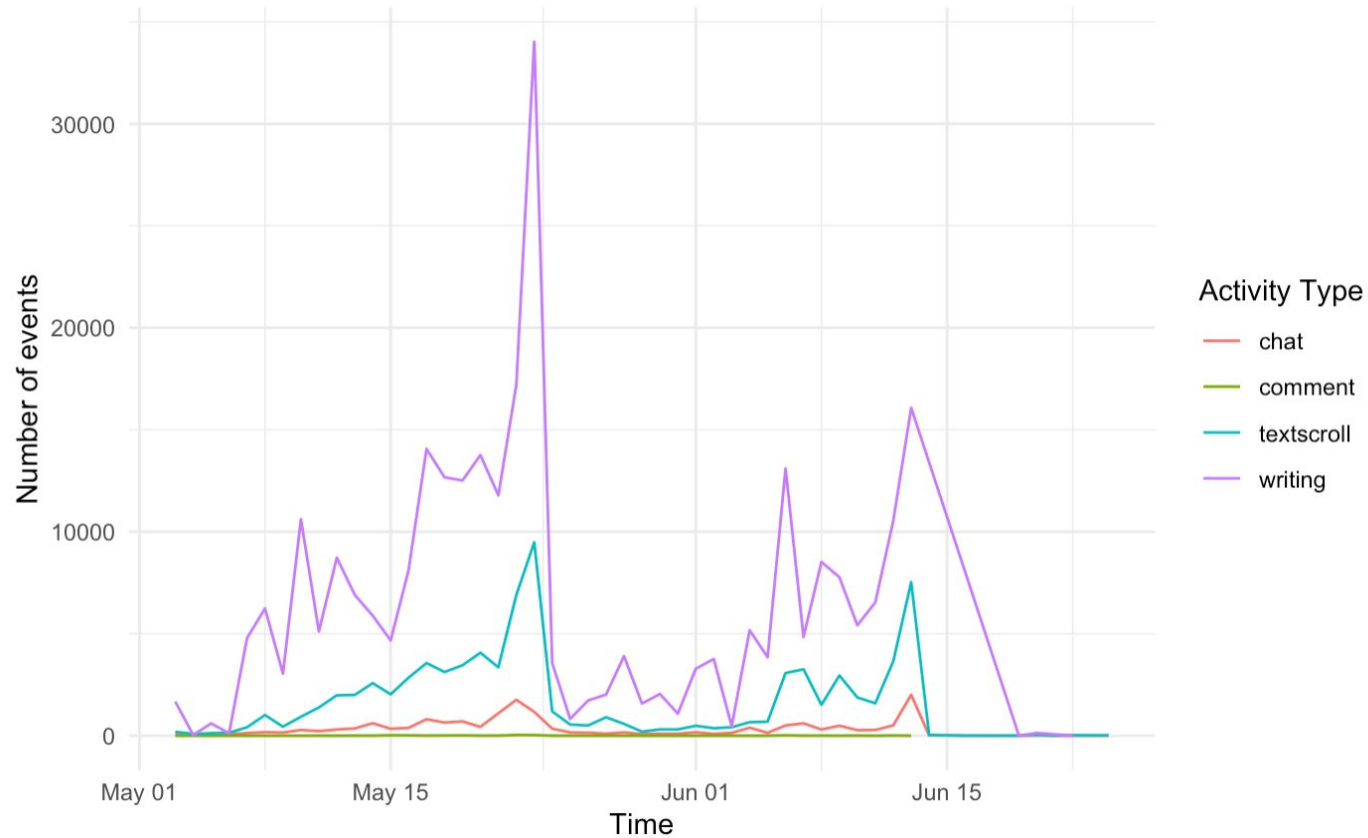
- 3C: communication, coordination, collaboration
- participation
- cohesion

Processing steps

- user > user features > clusters of users > group compositions
- user > user sessions > session features > session clusters > user behavior over time > group behavior clusters
- ...

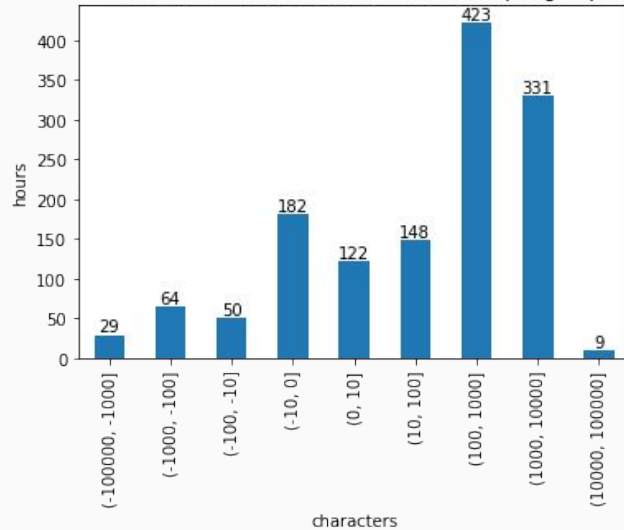


Overview, summer semester 2021

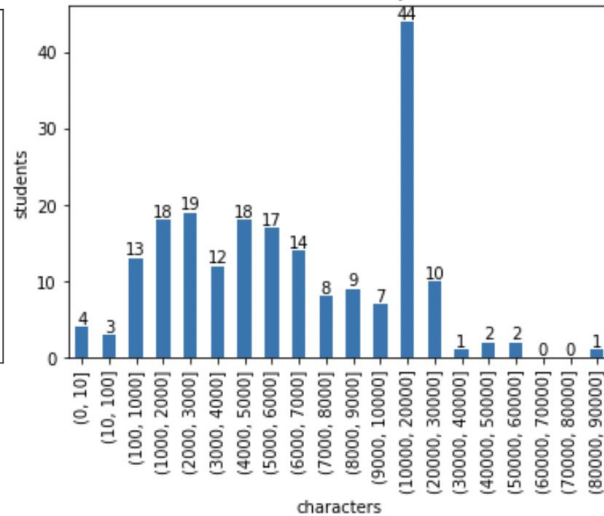


Overview, summer semester 2021

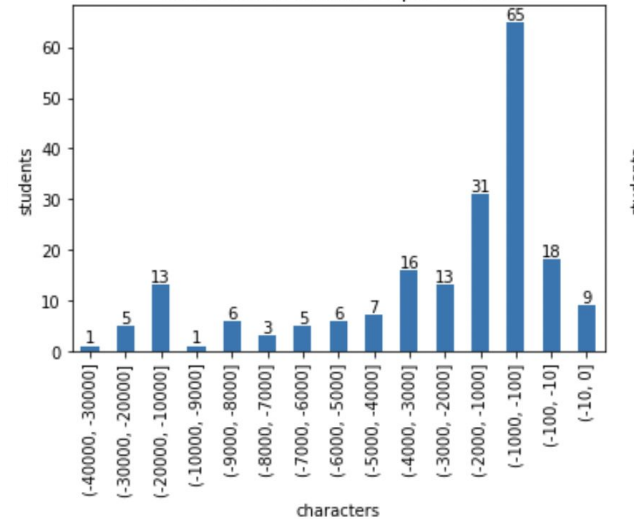
number of characters added in an hour per group



total added characters per student



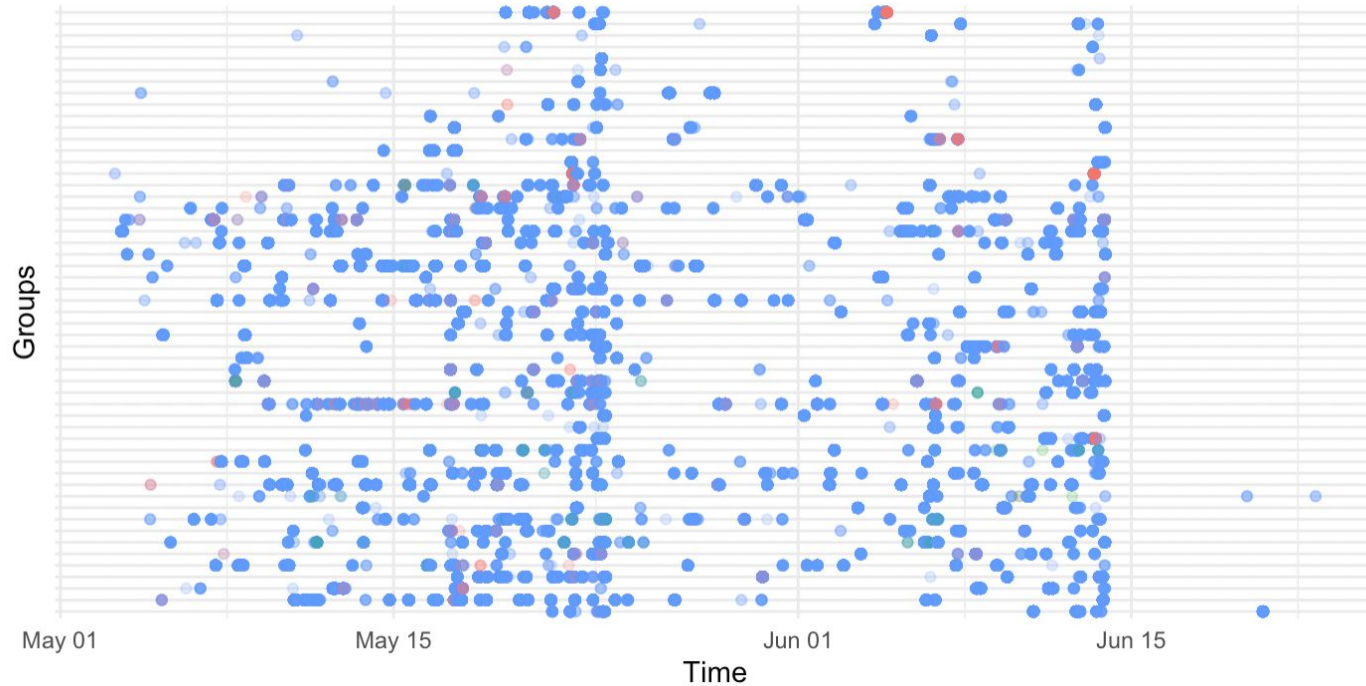
total deleted characters per student



Overview, summer semester 2021

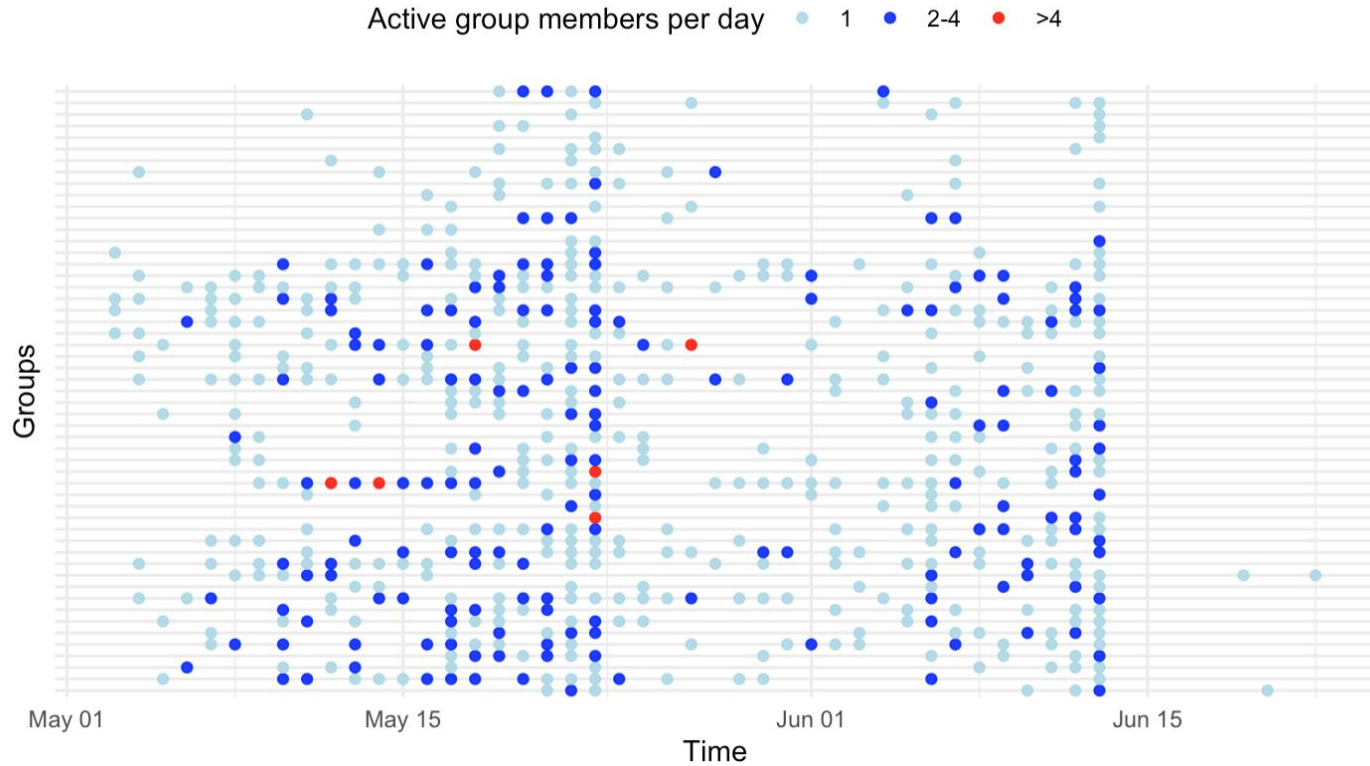
Group activity over time and by type

Activity: chat comment writing



Overview, summer semester 2021

Concurrent group activity per day



Example Analysis: User Session Clusters

Data

- FernUniversität in Hagen, PSY M1, summer semester 2021
- N = 229 participants, K = 55 groups, 415,892 log entries

Methods

- reconstruct user sessions (inactivity threshold 1h)
- extracted features from user sessions:
 - session: length
 - text contributions: count, per minute
 - scrolled lines: count, per minute
 - comments: count, length
 - chat: count, length
 - ~~text formatting~~
 - ~~text restructuring~~
 - ~~←you name it→~~
- clustering with kmeans, gap analysis

Results

- 4 session clusters:
 - 1: quick paste
 - 2: long time writing
 - 3: medium time writing
 - 4: revision

session cluster	1	2	3	4
cluster size	19	360	2268	239
session length (min)	2.1	122.1	12.7	0.9
text_contribution	3529.7	1568.0	178.9	-4.5
text_contribution/min	1920.5	17.1	13.1	0.3
scroll_length/min	138.7	34.4	75.0	897.9
comment count	0.105	0.531	0.028	0.004
chat_count	0.050	2.340	0.120	0.000

Draft: Mining patterns of collaborative writing

Method

Given: Multi-line document

1. index the lines and the words in the lines (maybe also sentences), e.g.

line 13: The quick brown fox jumps over the lazy dog.
1 2 3 4 5 6 7 8 9 10

2. Attribute the words to its authors, e.g.

a1: (13, 1), (13, 2), (13, 3), (13, 4), (13, 5)

a2: (13, 6), (13, 7), (13, 9)

a3: (13, 8) (13, 10)

3. At each document revision do:

- a. update word-author attributions

- b. identify changes:

- addition of a word (or character set)
- deletion of a word (or character set)
- edition of characters of a word
- formatting of a word
- add/remove line

- c. describe the 2-neighborhood for each change, e.g.

- within one's own words: These quick green fox
- between someone else's words: the lazy dog
- after/before of some one else word: jumps over the dog
- within someone else's word: dog.

- d. detect text restructuring

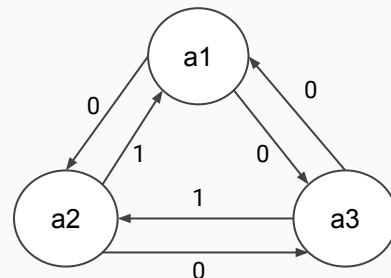
- copy & paste of words/sentences

Measures

Manhattan distance between changes of one author to another author, e.g.

	a2>a1	a3>a2	a3>a1
distance = 0:	0	1	0
distance = 1:	1	1	0
distance = 2:	1	0	0
distance = 3:	1	0	0
distance = 4:	0	0	0

- by type of change (add/delete/edit)
- consider distance changes over time
- representation as directed graph, e.g. for distance = 1:

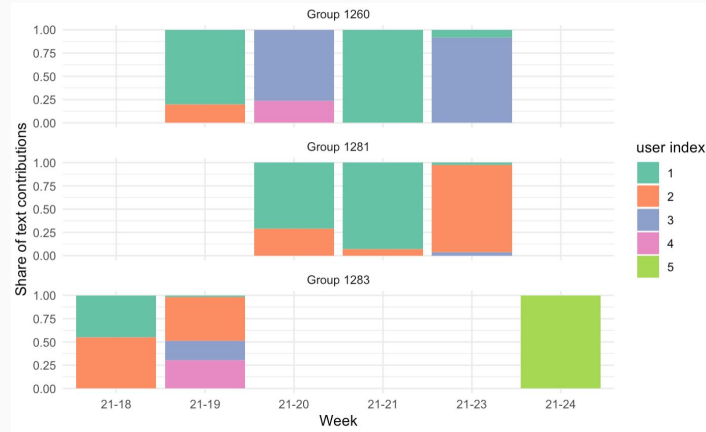


Step aside: Monitoring Collaborative Writing

Indicators for effective groups (Calvani, Fini & Molino, 2010) adapted to collaborative writing

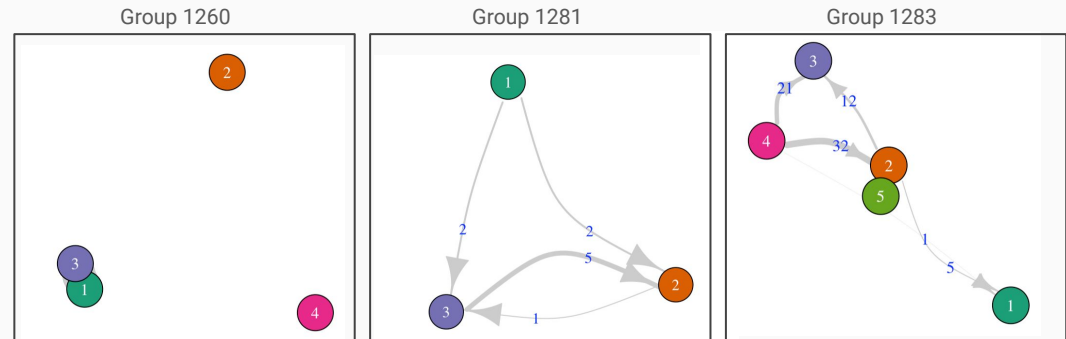
Participation

- participation in online learning
- text contributions
- equal contributions
- activity rhythm



Cohesion

- reciprocal perception of contributions
- mutual assistance
- joint writing efforts



Related works

Writing process analytics

- Wininger et al. (2014): sequence homology analysis, comparison of serial drafts, graph representation
- Shibani et al. (2018): analytical instrument of individual revision activities
- Southavilay et al. (2013): revision maps, topic evolution charts, and topic-based collaboration networks to examine collaborative writing activities.
- Hoppe et al. (2021): sequential alignment of writing (minor/major contributions), monitoring, and communication activities
- Key-level log data: Lindgren, Leijten & Van Waes, 2011; Leijten & Van Waes, 2013; Ahlsén & Strömqvist, 1999

Writing outcome analytics

- quantitative measures, e.g. word count, citations
- content-related features, e.g. used domain concepts, named entities, keyword extraction
- linguistic parameters, e.g. argumentation, vocabulary, language style, spelling

Group interactions

- network analysis: within-group conversations (=> divAdapt), paragraph-level writing (?)
- asynchronous/synchronous collaboration

Writing support

- emphasizing mistakes (Shibani et al., 2017a, Shibani et al., 2017b)
- constructive feedback (Gibson et al., 2017, Shibani et al., 2017a, Shibani et al., 2017b, Shibani et al., 2019)
- formative automated writing instructions (Allen et al., 2016, Roscoe et al., 2013)

Next steps

1. Identify user session clusters across multiple cohorts
2. Analyse distribution of session clusters over time (Menze, et al., 2022)
3. Implement the method to identify patterns of collaborative writing
4. Relate 1., 2. and 3. with other measures e.g. self-reported diversity data (Voltmer et al., 2022)