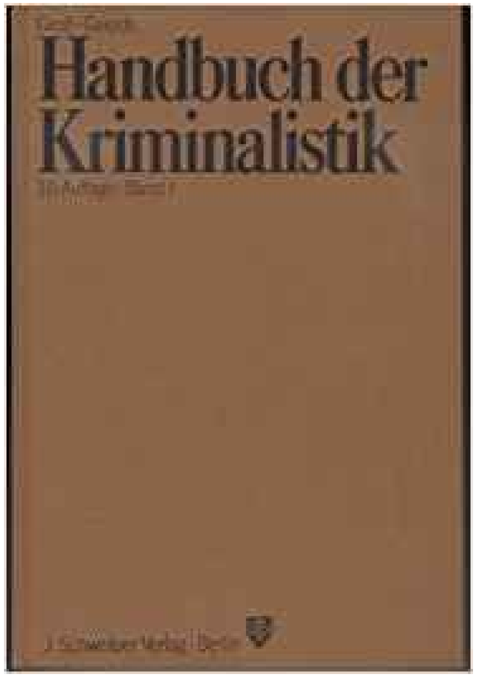
A Controlled Experiment in Digital Investigation

Felix Freiling   Christian Zoubek

FAU
FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

TECHNISCHE HOCHSCHULE NÜRNBERG
GEORG SIMON OHM

How do they (classical) crime investigators work?

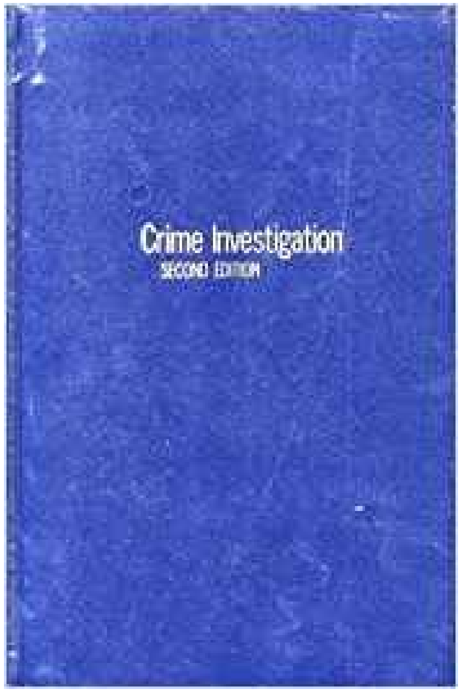


Groß-Geerds

Handbuch der Kriminalistik

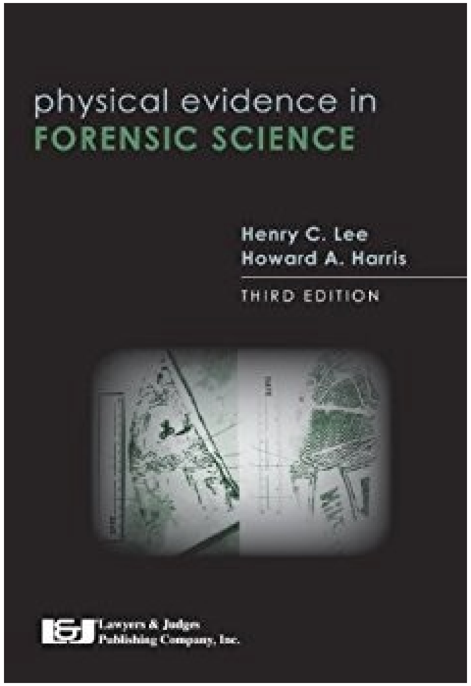
10. Auflage · Band I

J. Schweitzer Verlag | Berlin



Crime Investigation

SECOND EDITION



physical evidence in
FORENSIC SCIENCE

Henry C. Lee
Howard A. Harris

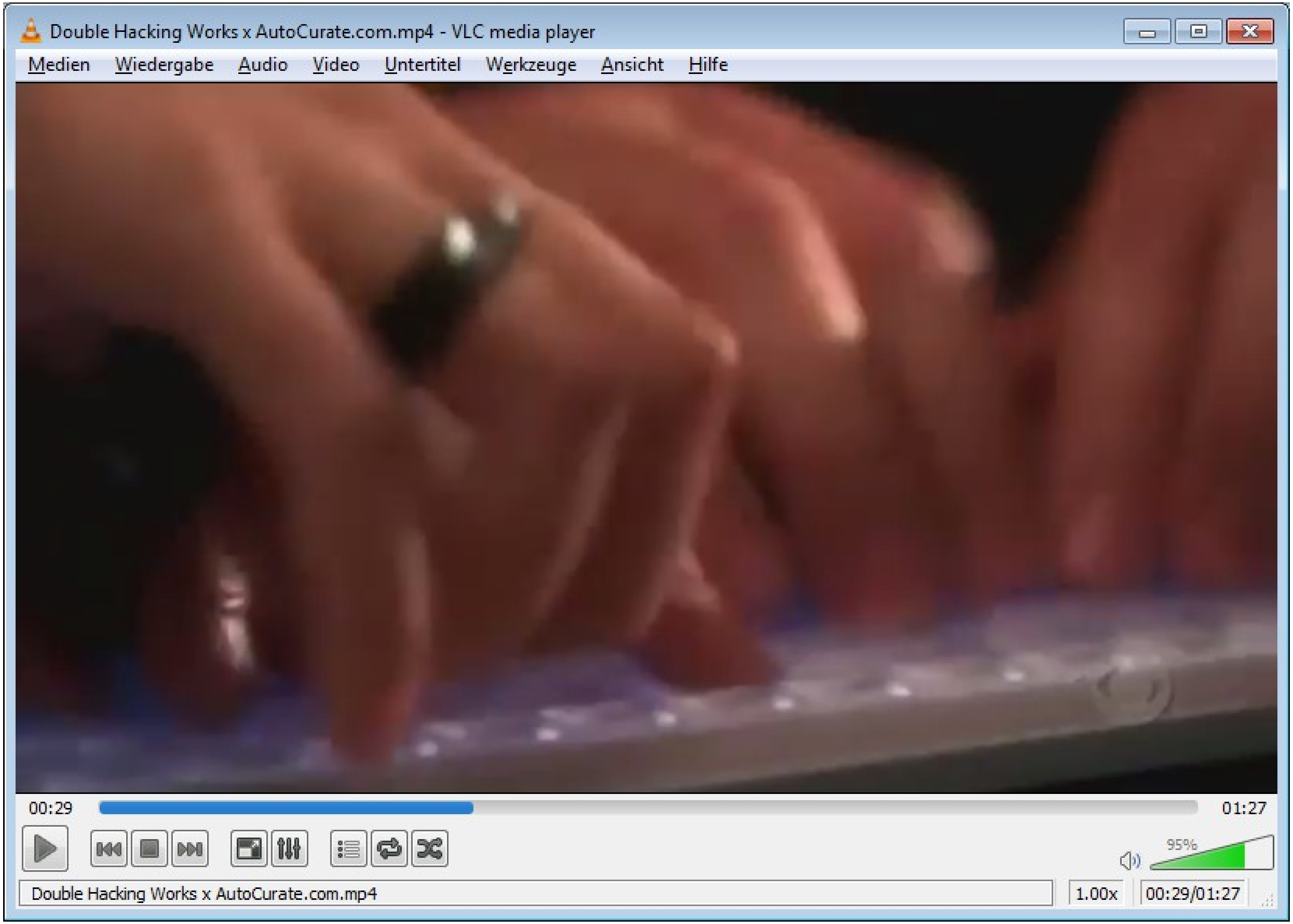
THIRD EDITION

Lawyers & Judges
Publishing Company, Inc.

How do they do it?

- Established inventory of methods and evidence types
- Clear separation of duties between investigator and forensic scientist
- Documented experience that is systematically used in criminalistics education

How do digital investigators work?



Double Hacking Works x AutoCurate.com.mp4 - VLC media player
Medien
Wiedergabe
Audio
Video
Untertitel
Werkzeuge
Ansicht
Hilfe
00:29
01:27
95%
Double Hacking Works x AutoCurate.com.mp4
1.00x
00:29/01:27

How do they do it?

- Unclear role of “digital forensic scientist”
- Hardly any (peer reviewed) literature on how digital investigators work
- We know how to teach technical skills, but how do we teach investigative skills?

Overview

1. Research questions
2. The experiment
3. Experimental results
4. Conclusions

1. Research Questions

Terminology

- “Case”
  - Description of case context and investigative goals
  - A collection of digital evidence
- “Participant” and “Group”
  - Human who participated in the experiment
  - Multiple participants
- “Effort”
  - Time in minutes spent on solving the case
  - “Individual effort” vs. “group/total effort”
- “Quality”
  - Percentage/amount of correctly interpreted digital evidence

Different Types of Work (Task Types)

- T1: conceptual work with pen and paper, including documentation
- T2: group meetings, discussion
- T3: programming new tools, interfacing with old tools, automating investigative/analysis steps
- T4: applying tools, doing the actual investigation

Research Questions

- Is there a difference between the total effort to solve different cases?
- Do groups use different strategies when trying to solve different cases?
- Is the distribution of task types different for different cases and groups?

- What factors correlate with total effort per case?
- What factors can predict total effort?
- What factors correlate with result quality?
- What factors can predict result quality?

2. The Experiment

The Setting

- Course „Forensik II", October 2015-February 2016
- Almost **40 students**, all of them with basic forensics education from earlier course
- Split up into **10 groups** of investigators
- 3 (arguably realistic) cases
- Pre-study questionnaire, final investigative report
- Mandatory documentation of effort by every participant

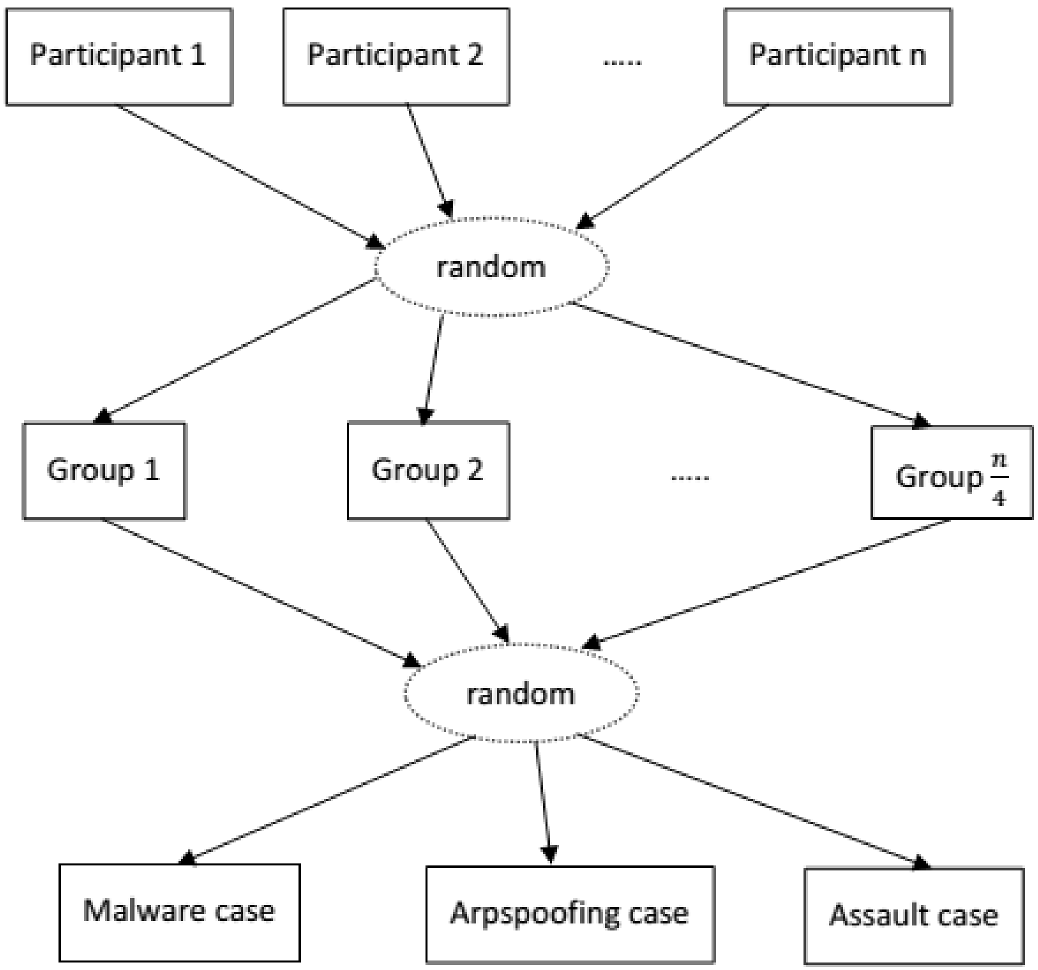
- In total we used data from **34 participants**

The Cases

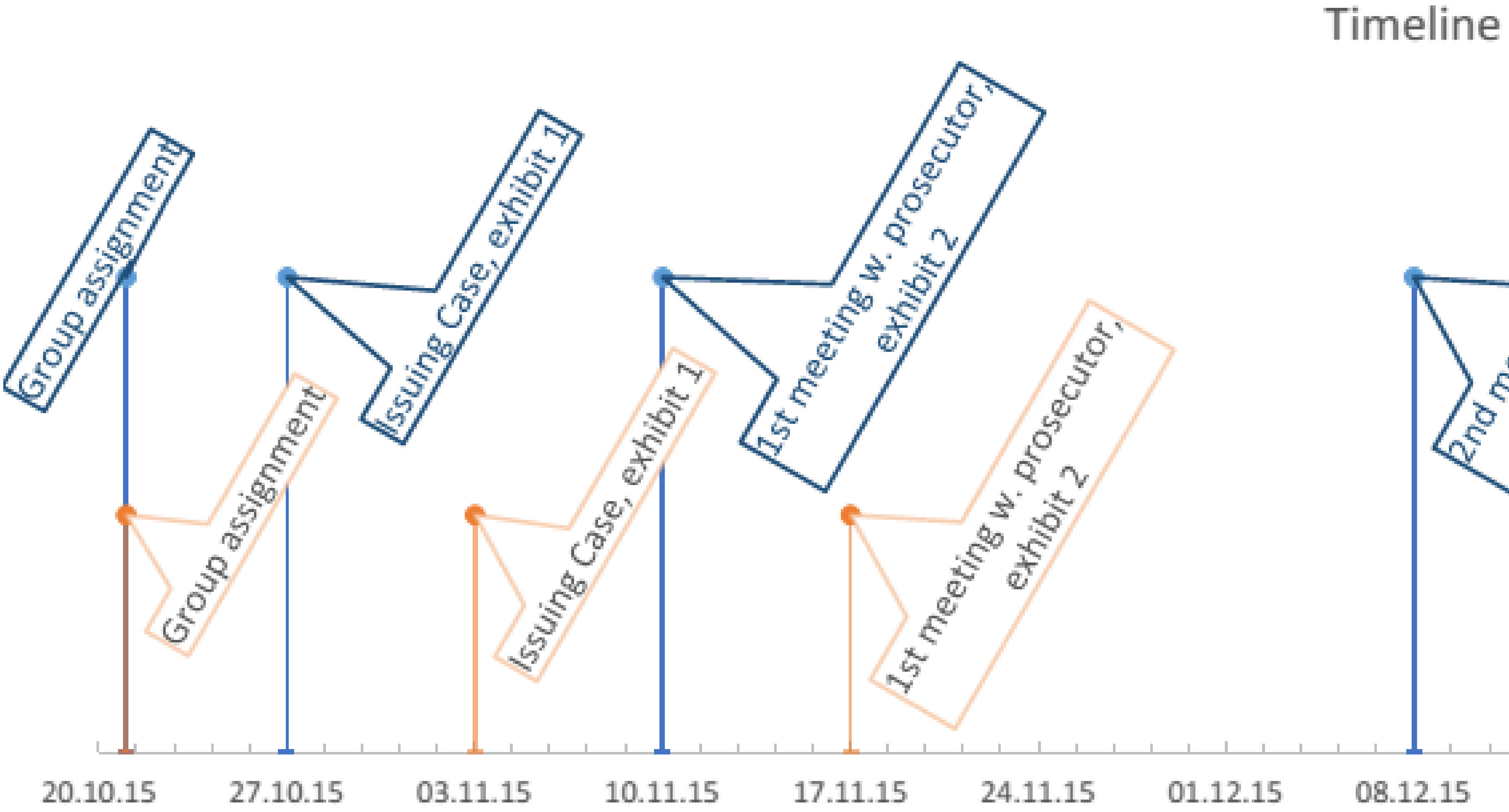
- ARPspoof
  - Sysadmin gets access to passwords via ARP spoofing
- Terror
  - Terrorists coordinate bombing attack on embassy in a web forum trying to hide their traces
- Malware
  - Distribution of malware over a an infected website, infection of clients, keylogging

- At least three disk images to analyse in a stepwise fashion
- One false false trail in each case description

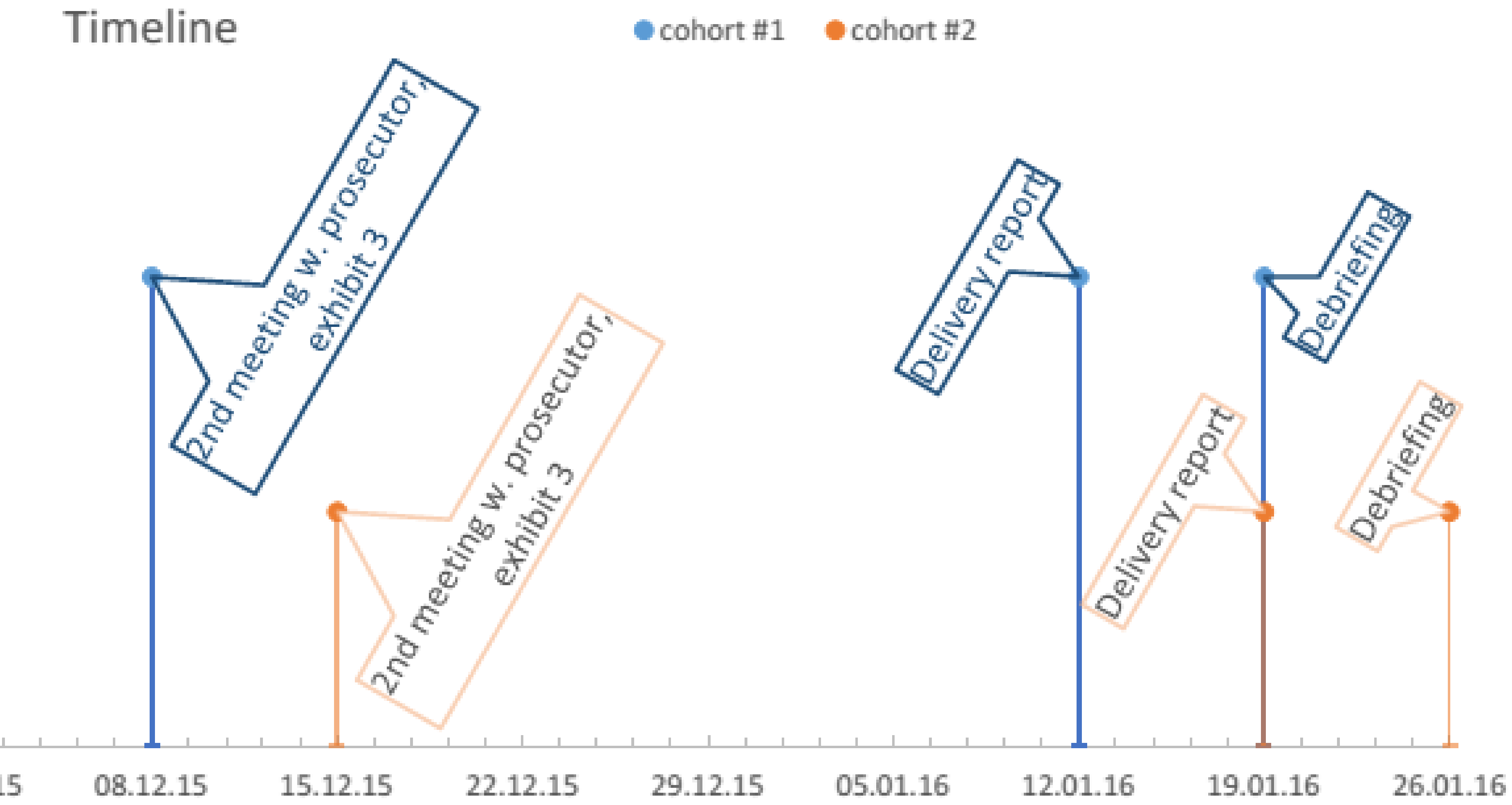
Experimental Design



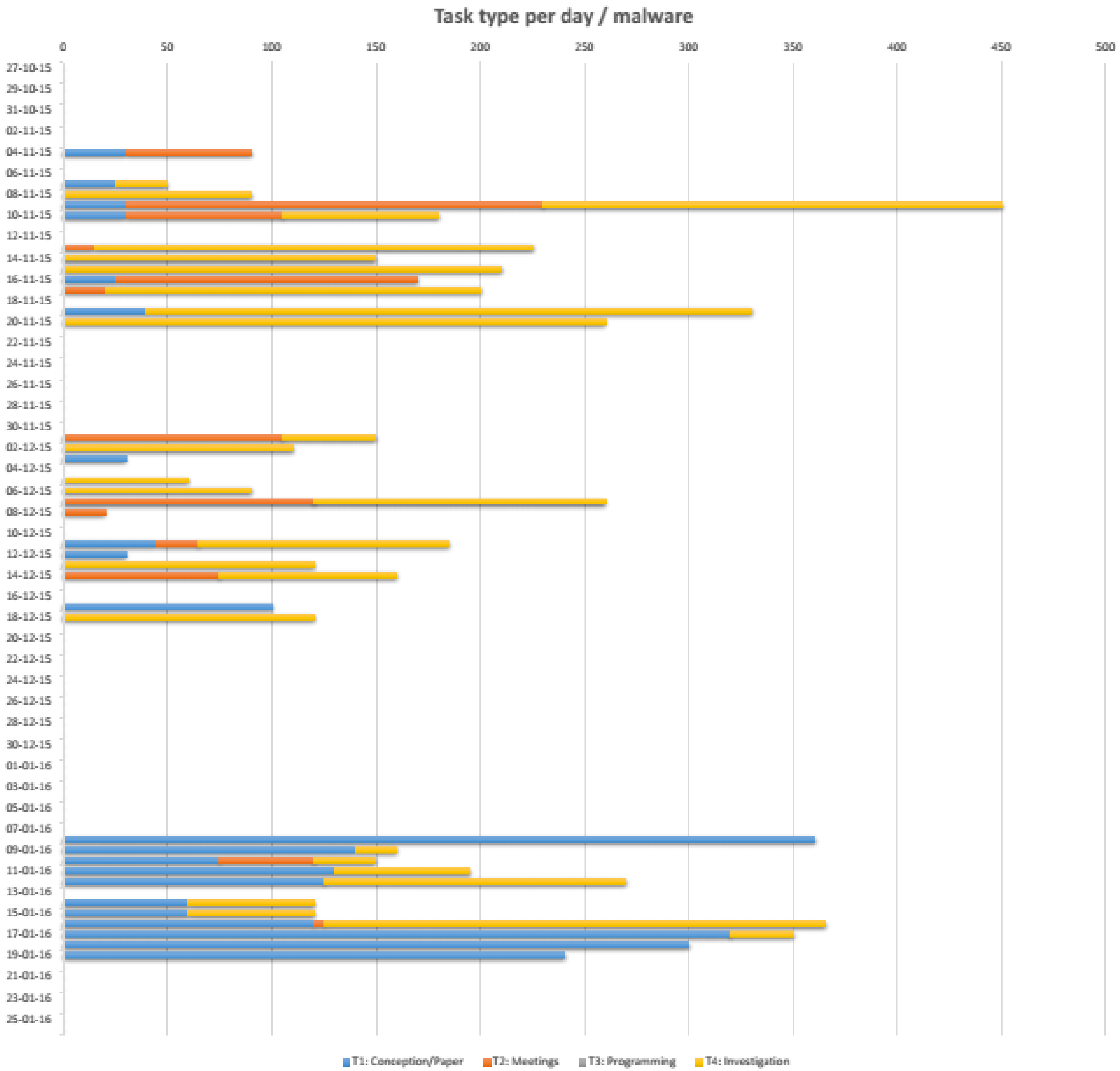
Timeline ...

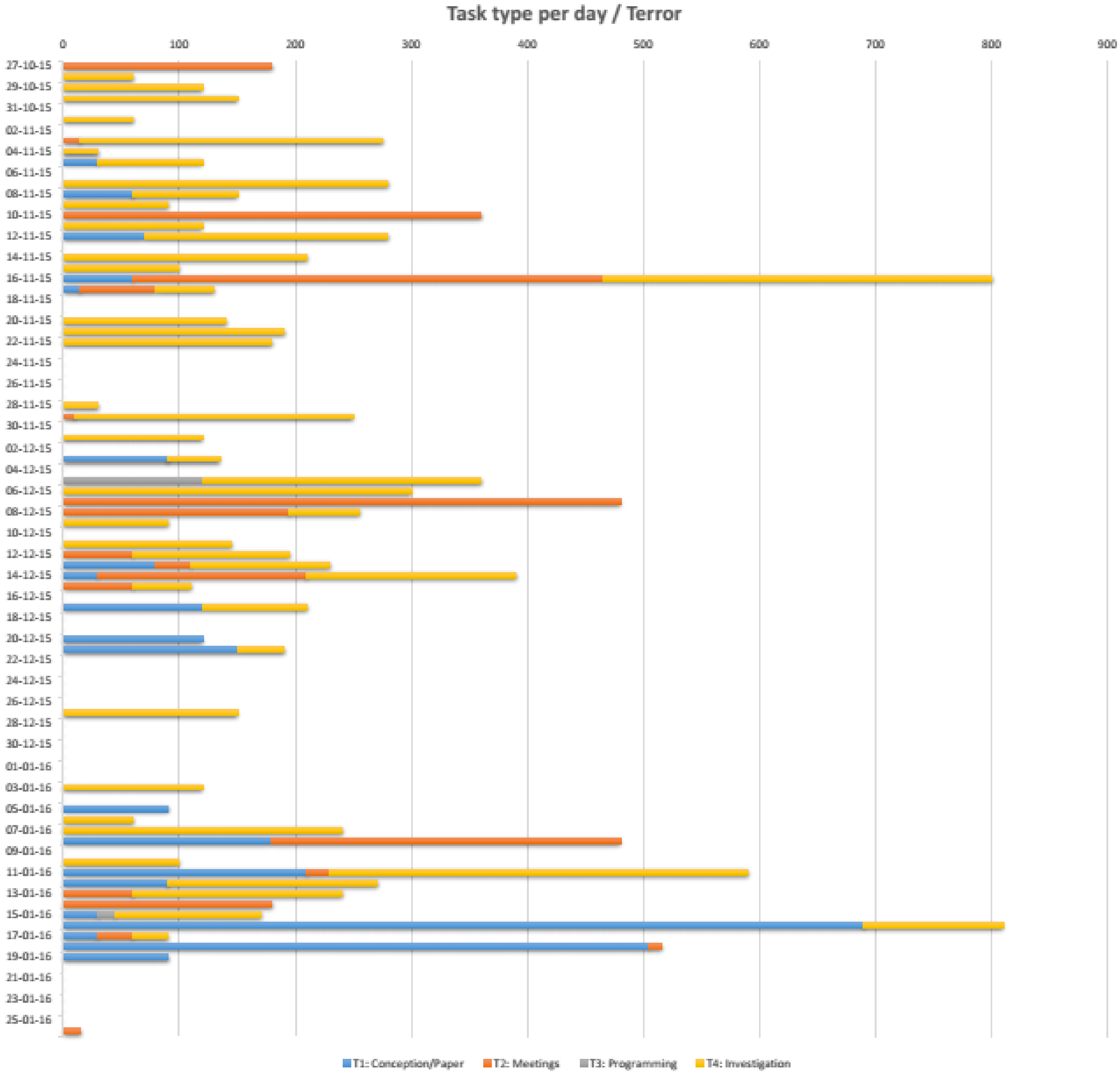

Timeline
Group assignment
Issuing Case, exhibit 1
1st meeting w. prosecutor, exhibit 2
2nd m
Group assignment
Issuing Case, exhibit 1
1st meeting w. prosecutor, exhibit 2
20.10.15
27.10.15
03.11.15
10.11.15
17.11.15
24.11.15
01.12.15
08.12.15

… Timeline

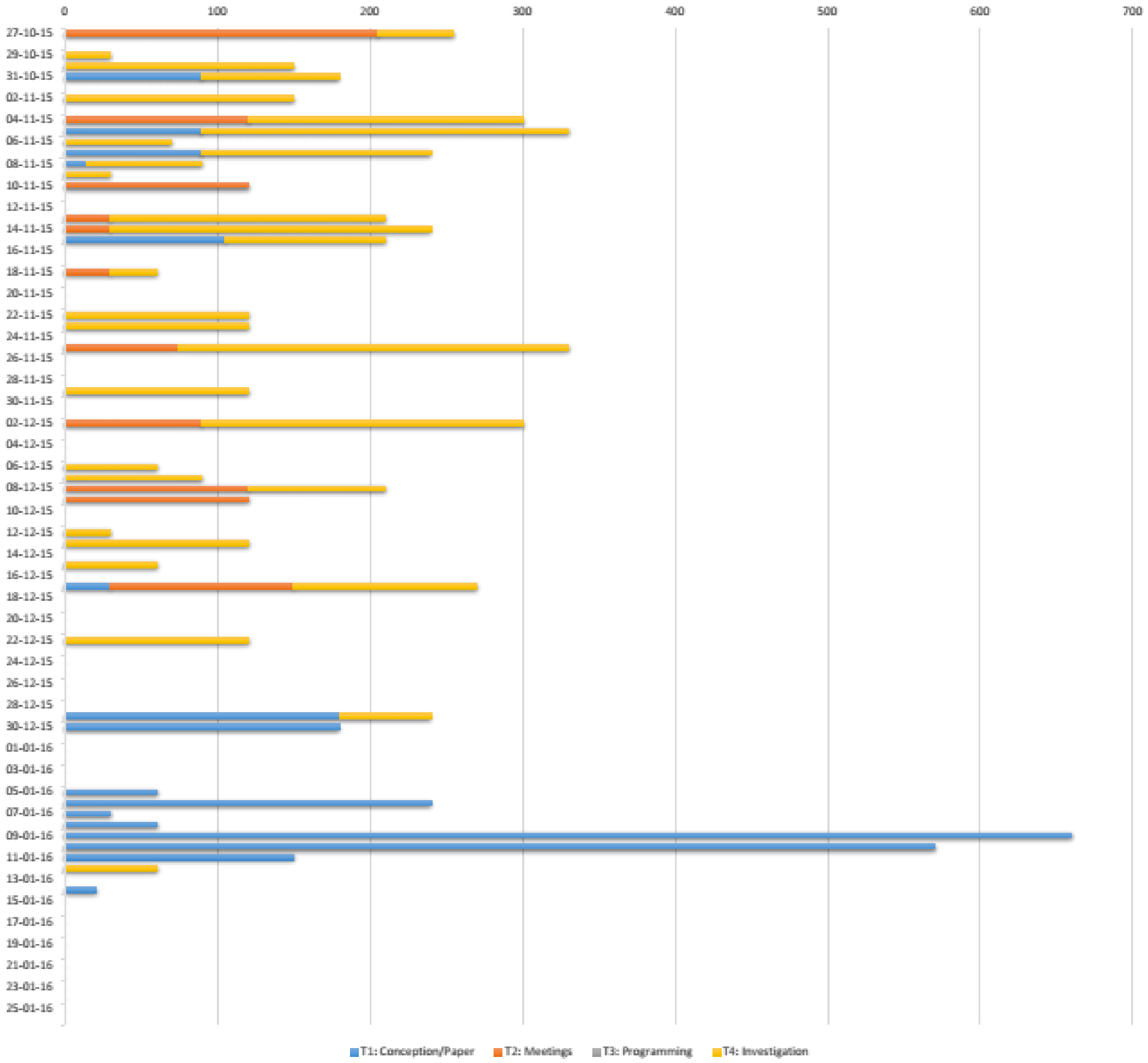

Timeline
● cohort #1
● cohort #2
2nd meeting w. prosecutor, exhibit 3
2nd meeting w. prosecutor, exhibit 3
Delivery report
Debriefing
Delivery report
Debriefing
15
08.12.15
15.12.15
22.12.15
29.12.15
05.01.16
12.01.16
19.01.16
26.01.16

3. Experimental Results


Task type per day / malware
0
50
100
150
200
250
300
350
400
450
500
27-10-15
29-10-15
31-10-15
02-11-15
04-11-15
06-11-15
08-11-15
10-11-15
12-11-15
14-11-15
16-11-15
18-11-15
20-11-15
22-11-15
24-11-15
26-11-15
28-11-15
30-11-15
02-12-15
04-12-15
06-12-15
08-12-15
10-12-15
12-12-15
14-12-15
16-12-15
18-12-15
20-12-15
22-12-15
24-12-15
26-12-15
28-12-15
30-12-15
01-01-16
03-01-16
05-01-16
07-01-16
09-01-16
11-01-16
13-01-16
15-01-16
17-01-16
19-01-16
21-01-16
23-01-16
25-01-16
T1: Conception/Paper
T2: Meetings
T3: Programming
T4: Investigation



Task type per day / Terror



Task type per day / Arpspoof-case

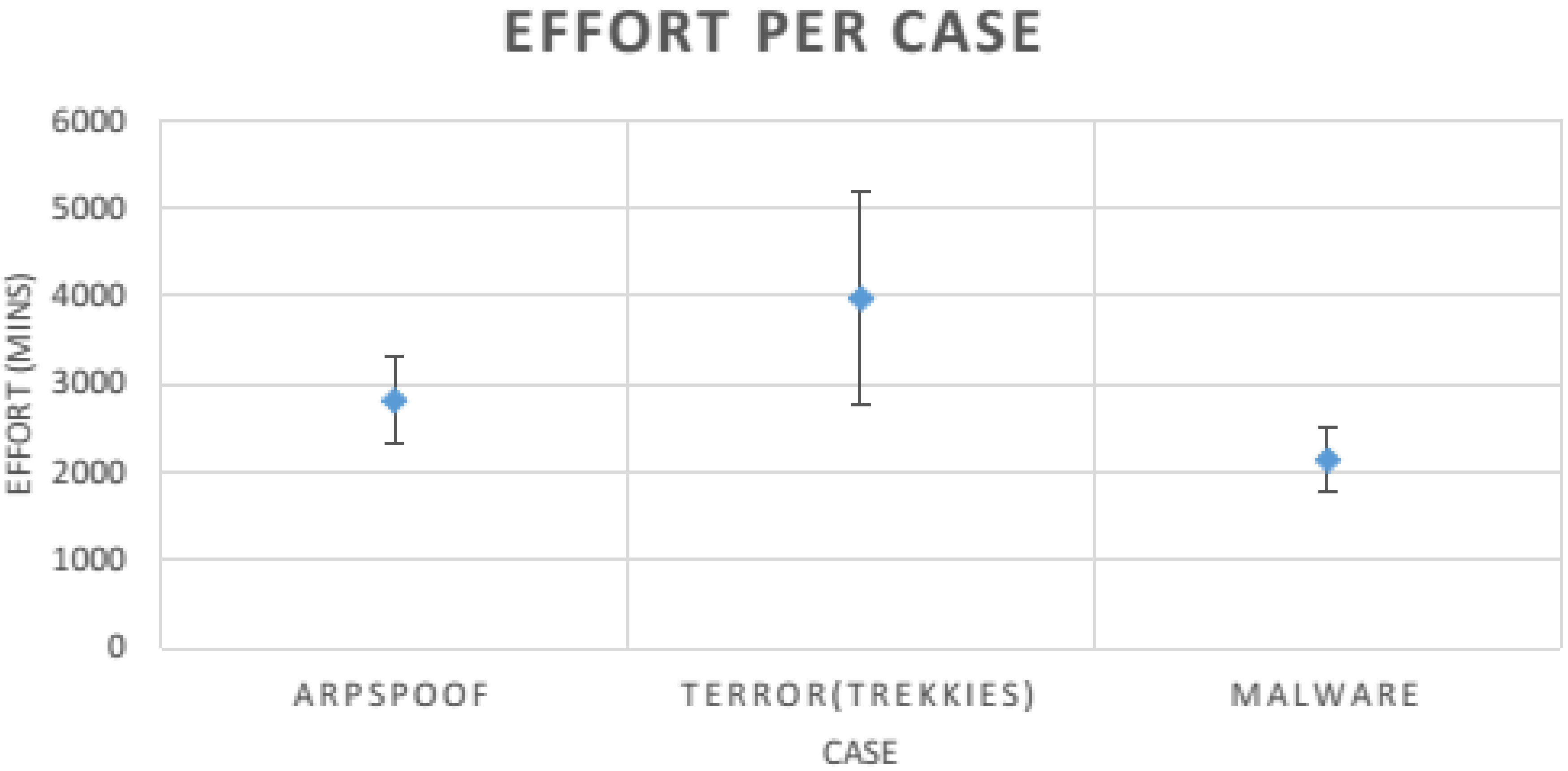


Figure 6: Total effort per case (plot of average and standard deviation).

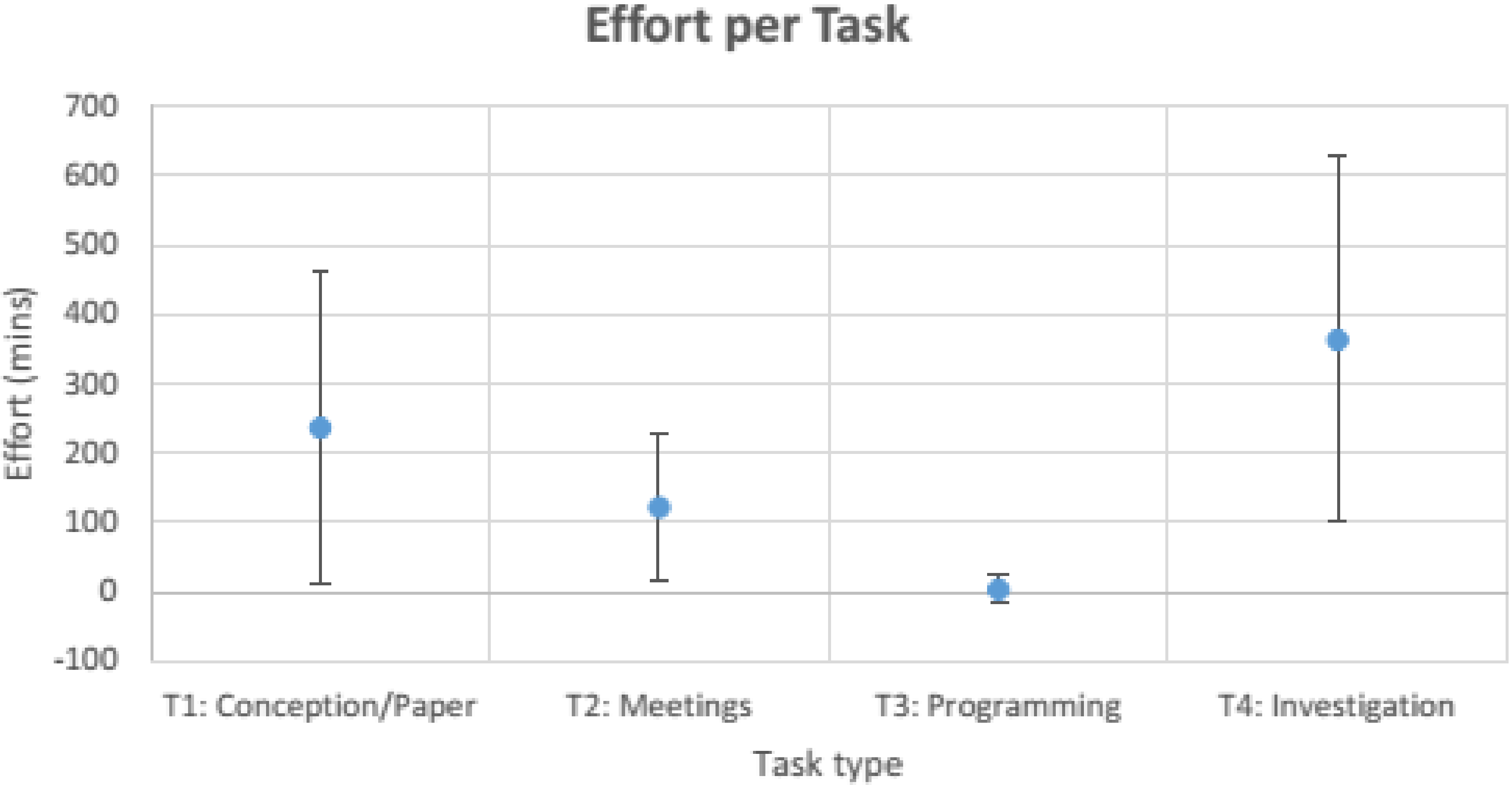


Figure 7: Total effort per task type (plot of average and standard deviation).

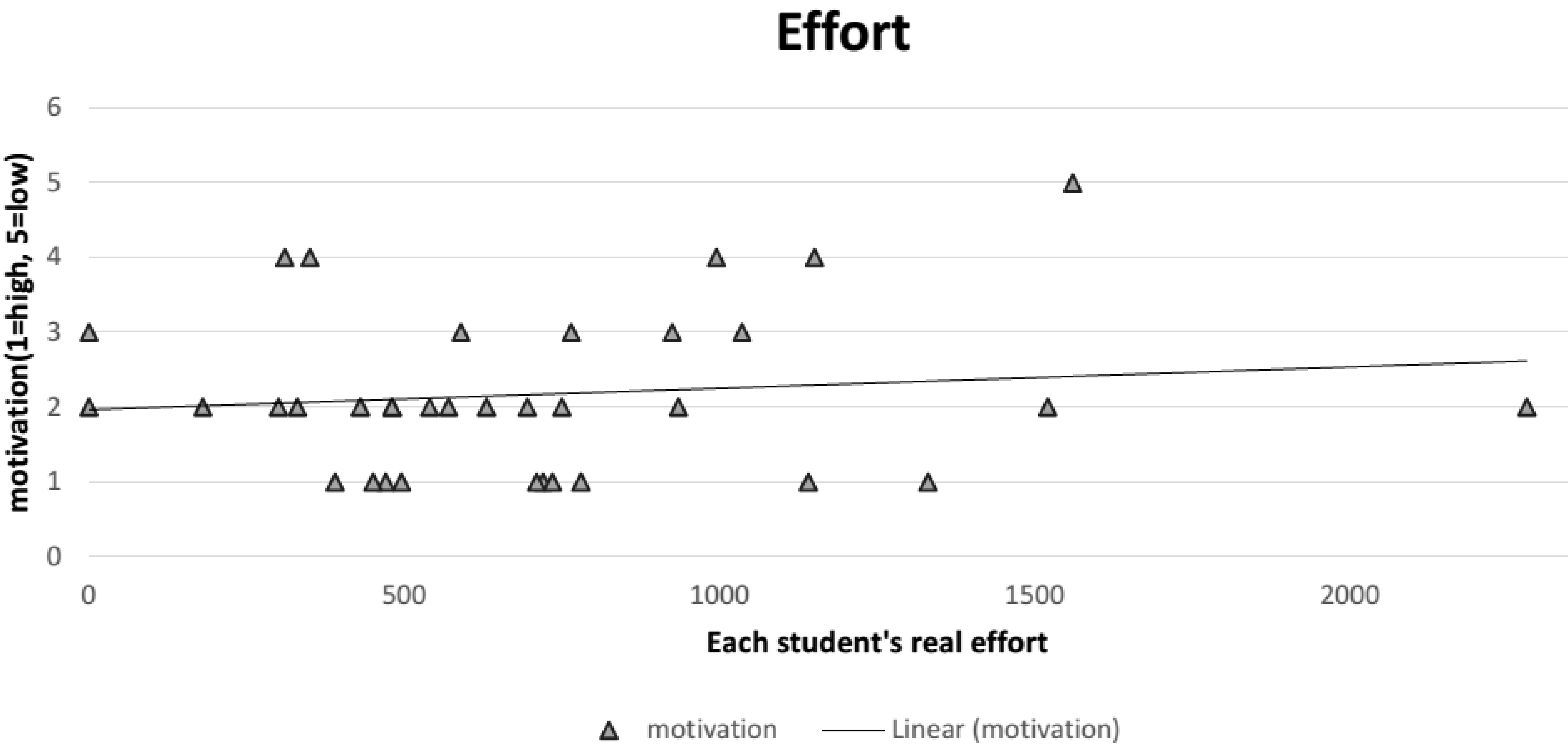


Figure 9: Real total effort per student vs. motivation.

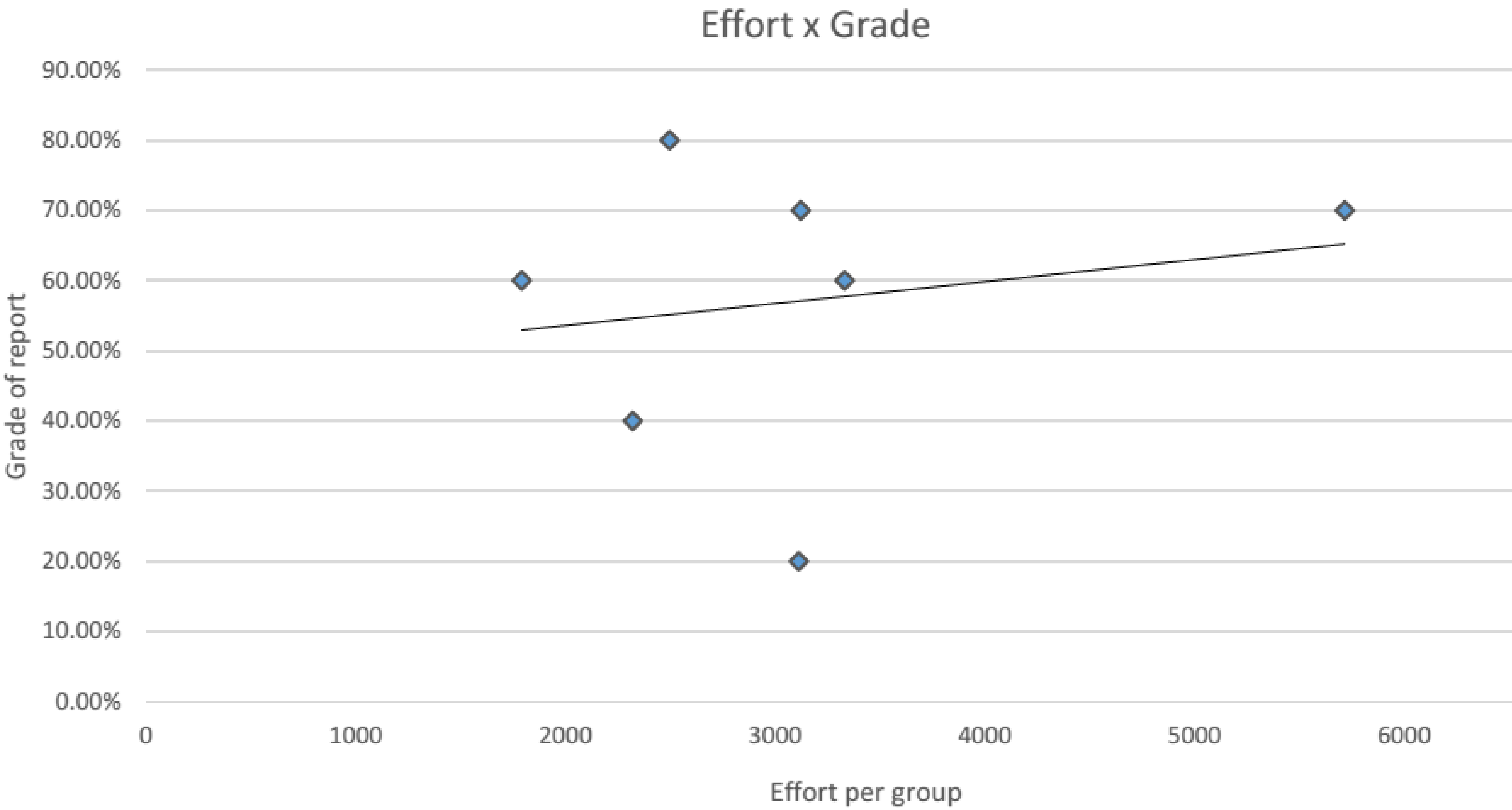


Figure 10: Grade vs. effort per group.

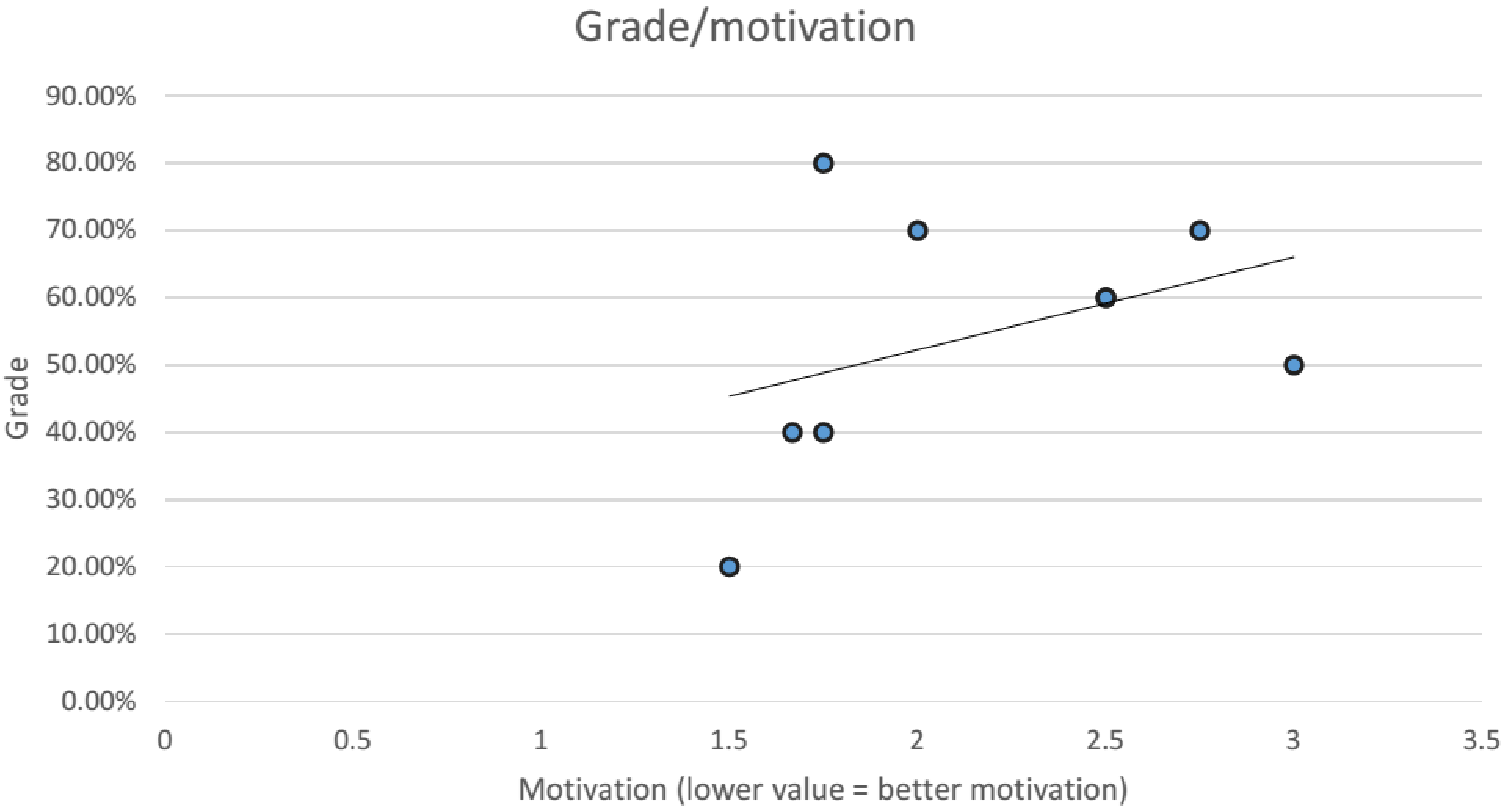


Figure 11: Grade vs. total motivation per group.

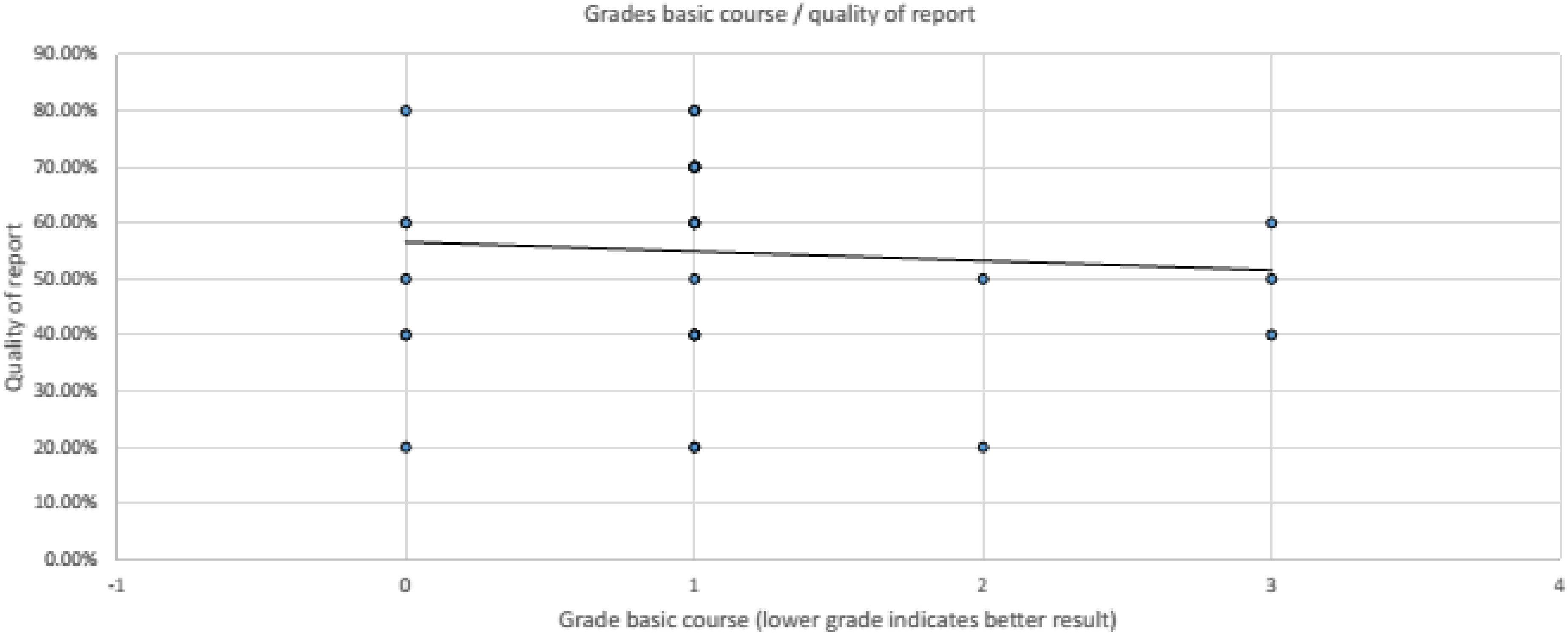


Figure 12: Result quality/grade vs. grade of basic course.

Previous Grades vs. Quality

- Quality correlates positively with grade in introductory forensics course

- Previous grades are a good predictor of future grades

4. Conclusions

Interpretation of Results

- Bounded (well-specified) investigation goals reduce effort
- Effort is more important than motivation for good quality
- Use quality of previous work to select good people

Future Studies

- Focus more on measurements of individuals than on groups
- Formulate precise hypotheses and calculate statistical significance with more (100+) participants
- Case comparison is hard, can this be done better?
- Data available online: **`https://www1.cs.fau.de/filepool/publications/freiling-zoubek-dfrws-eu-imf-2017-data.csv`**

A Controlled Experiment in Digital Investigation

Felix Freiling Christian Zoubek

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

TECHNISCHE HOCHSCHULE NÜRNBERG GEORG SIMON OHM