

# Visual Feedback for Players of Cybersecurity Games

(usability study and future work)

**Radek Ošlejšek**



# Cybersecurity games

- **Hands-on** training in the KYPO Cyber Range
- **Capture the Flag** games
  - Multi-level games
  - Flag = secret code obtained at the end of a successful level used to proceed to the next level
  - In each level, a specific task has to be solved.
    - Scan the network and find a vulnerable server.
    - Use SQL injection to get into the server
    - Break root's password
    - ...



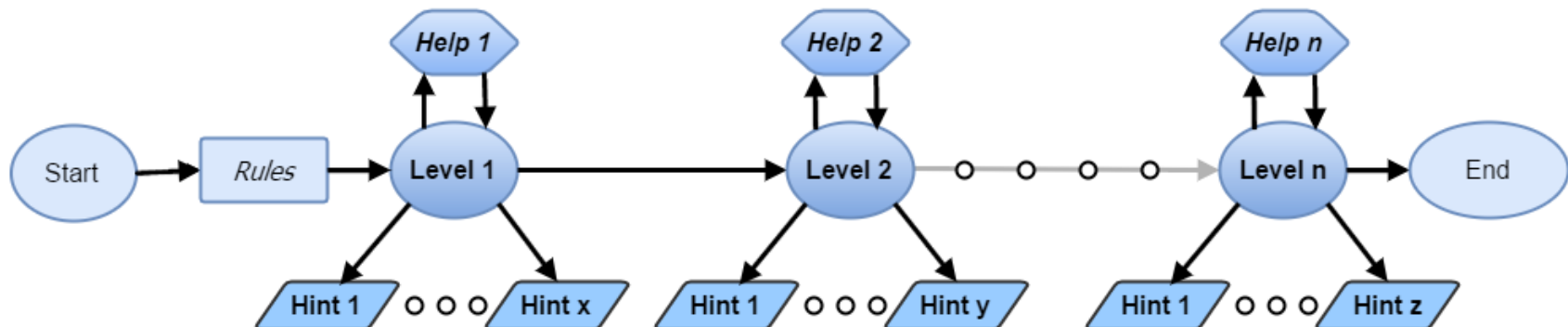
video

## Score

- Each level has assigned a maximal score.
- Trainees are penalized (the score obtained at the end of the level is decreased)
  - Taking hints, showing help/solution, providing an incorrect flag, ...

## Time

- Every event (taking a hint, providing a flag, ...) is recorded with a timestamp.
- => Time series



## Q: How to improve impact of the exercise?

- **Our goal:** Improve impact of CFTs by providing valuable feedback
- Learning from doing is not sufficient
  - Trainees do not see various ways how to do things, e.g., multiple ways how to find a vulnerable server in the network.
  - Feedback provided by supervisors is OK, but we want to support on-line plying without supervision at any time.
- **Challenge:** Provide **automated** (visual) feedback right after the exercise

# Requirements (types of interests)

We followed the principles of user-centered design and derived 3 requirements from the discussion with domain experts – CTF supervisors:

## R1: Provide **personalized feedback**

- Person-centric view. Identification of well-done and problematic parts of my gameplay.
- *"In which level did I lose the most points and why?"*

## R2: Provide **comparative feedback**

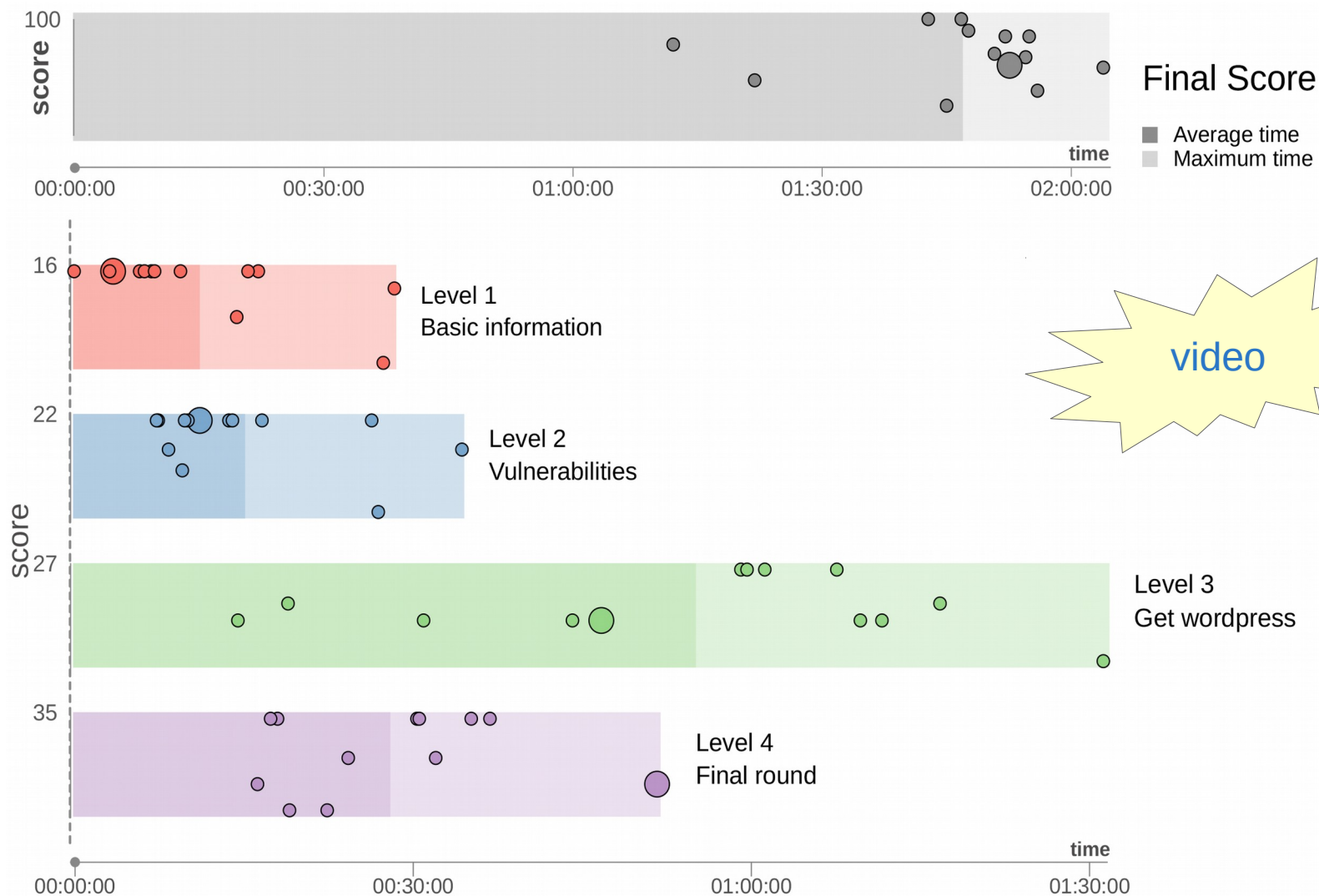
- *"Where I was better or worse than other players?"*
- Assessment of learner's abilities within a group, typically in a competition.

## R3: Provide a **brief overview** of the overall game results and features

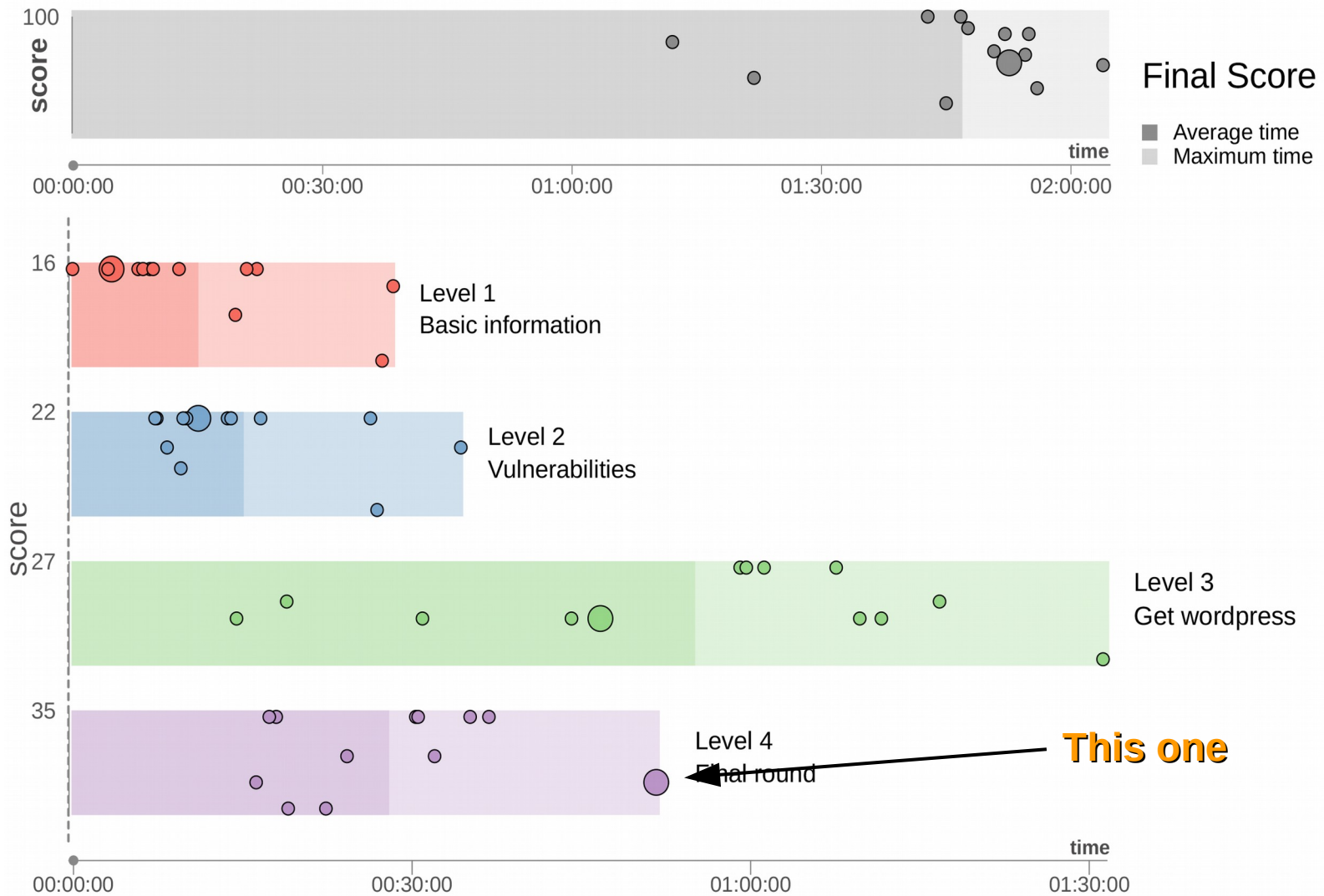
- Insight into the game difficulty and other aspects that enable trainees to put their personal and comparative findings into the context of this particular game.
- Useful mainly in situations when a user plays multiple games.

Cross-cutting concern: Intuitiveness and easy to use

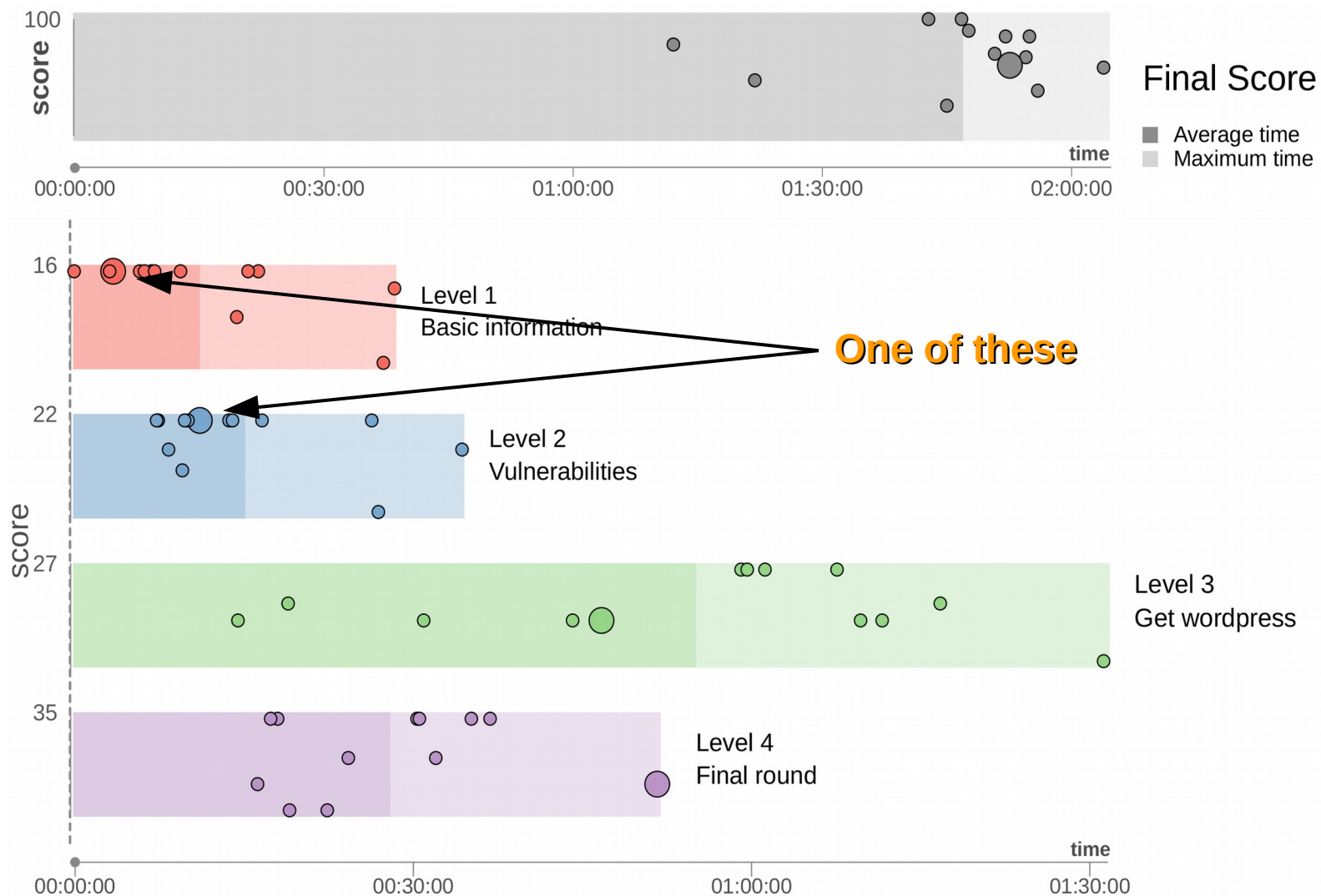
# An intuitive overview of the score and time



# Which level was my worst?

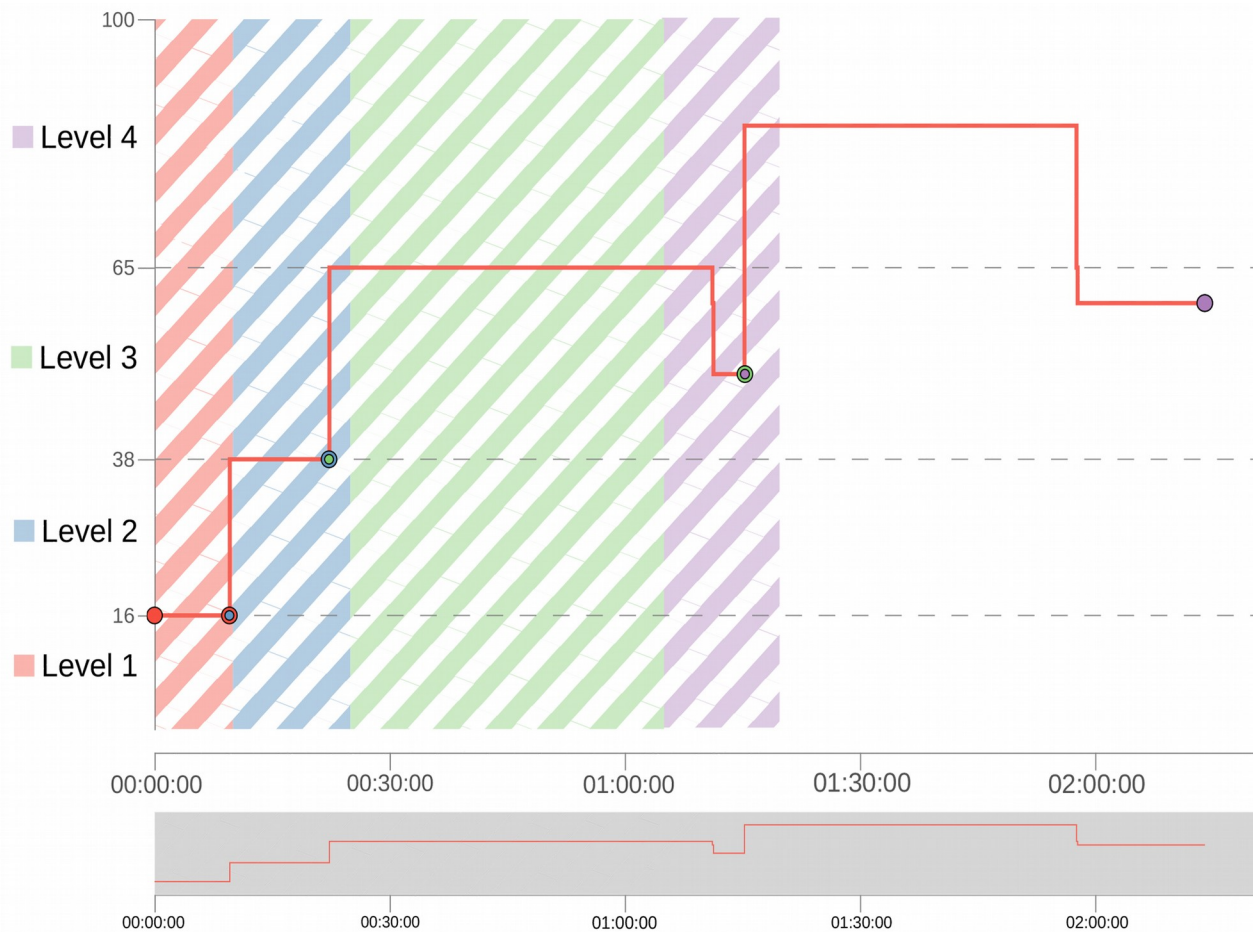


# Which level was my best?





# Detailed view on score development

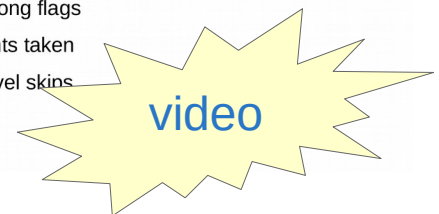


Player	Level 1 Score	Level 2 Score	Level 3 Score	Level 4 Score	Final Score
9003575	16	22	12	10	60
9003576	16	22	27	35	100
9003579	16	22	12	20	70
9003580	16	22	12	35	85
9003582	16	22	0	-	38
9003584	0	15	12	20	47
9003585	16	15	27	0	58
9003590	16	22	12	35	85
9003593	16	10	17	35	78
9003570	8	22	27	10	67
9003571	13	0	12	0	25
9003572	16	22	17	35	90
9003574	16	22	27	35	100

Estimated time

## Event filters

- ☒ Correct flags
- ☐ Wrong flags
- ☐ Hints taken
- ☐ Level skins



# Usability study – hypotheses

- H1: Requirements **R1 – R3 are meaningful** and useful for learners.
- H2: **The visual feedback is useful** in providing insight into the tasks of R1 – R3.
- H3: Some **visualizations or their parts are more useful for specific tasks** of R1 – R3 than others.
  - Both the view are complementary, enabling users to reach the same results in a different way.

# Usability study – results

- **H1: Design requirements are correct**, and the tasks reflect user interests. But ...
  - Trainees prefer exploration of *personal results* to the *overall game results* and *comparison with others*.
  - Unexpectedly, *comparative feedback* is the least useful.
- **H2: Visualizations support trainees** in the understanding of results.
  - Learners were able to complete given tasks correctly, and tasks were easy to solve with the visual feedback.
- **H3: We did not find that any of the visualizations would better support** the *personal results*, *comparative feedback*, or *overall game results*. But ...
  - We identified specific tasks (across the requirements) for which one of the visualizations might be more appropriate.
  - The results are uncertain due to data limitations, and further inspection is needed.

## Usability study – summary

- Our visual feedback is meaningful and worth of further development.
- Details of the study, its limitations, possible improvements of the feedback, and other lessons learned from the study can be found in

OŠLEJŠEK, Radek, Vít RUSŇÁK, Karolína BURSKÁ, Valdemar ŠVÁBENSKÝ and Jan VYKOPAL. ***Visual Feedback for Players of Multi-Level Capture the Flag Games: Field Usability Study***. In Proceedings of the IEEE Workshop on Visualization for Computer Security (VizSEC). Vancouver, Canada. October 2019.

- Capturing also bash history
- Reconstruction of user process from time series
  - Petri nets, process mining (Martin Macák)
- Run-time analysis of user behavior
  - Adaptive learning (Valdemar Švábenský)
  - Online adaptation of hints based on the learner's previous behavior.
- Visual analysis of learner's behavior

