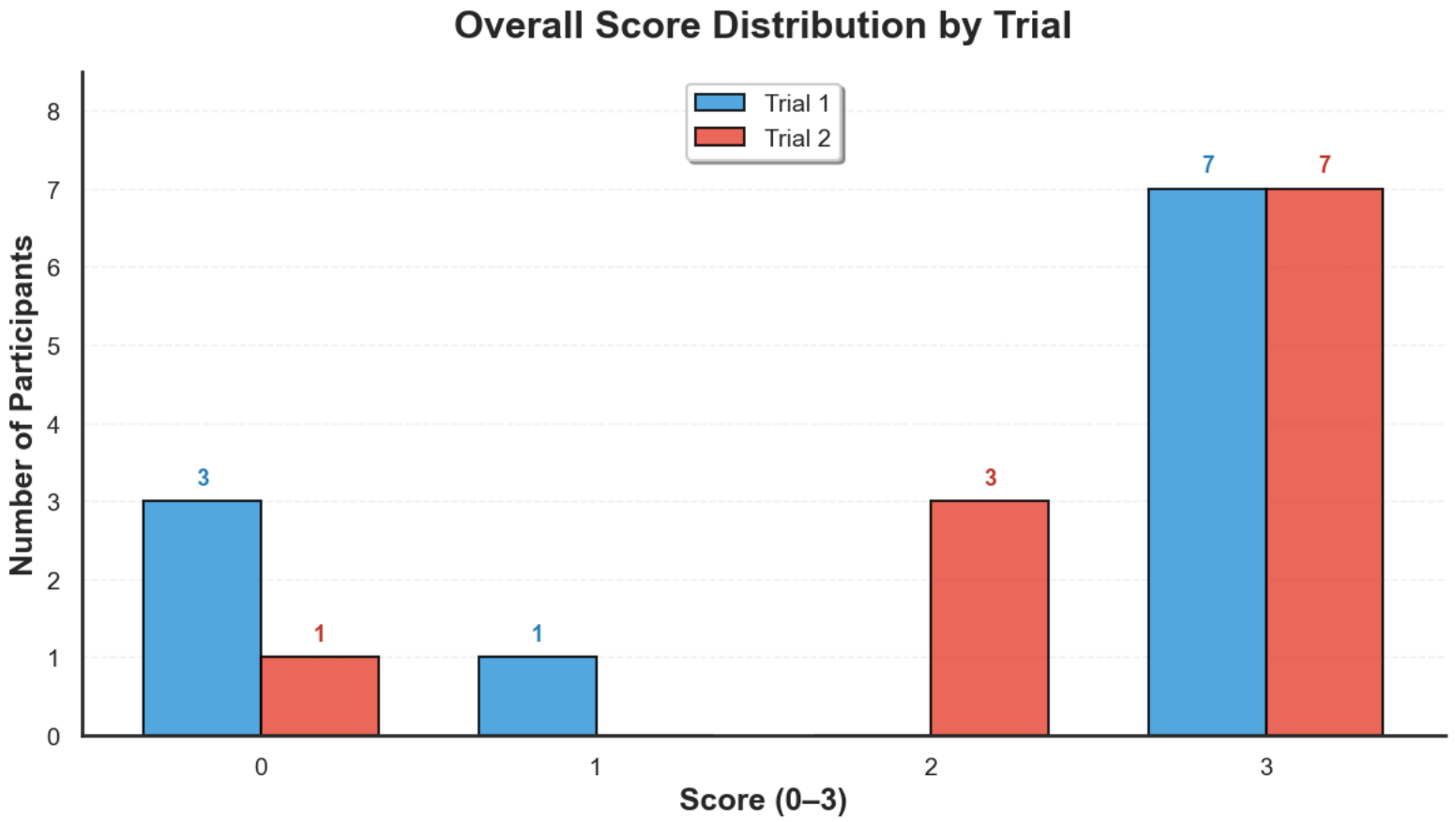
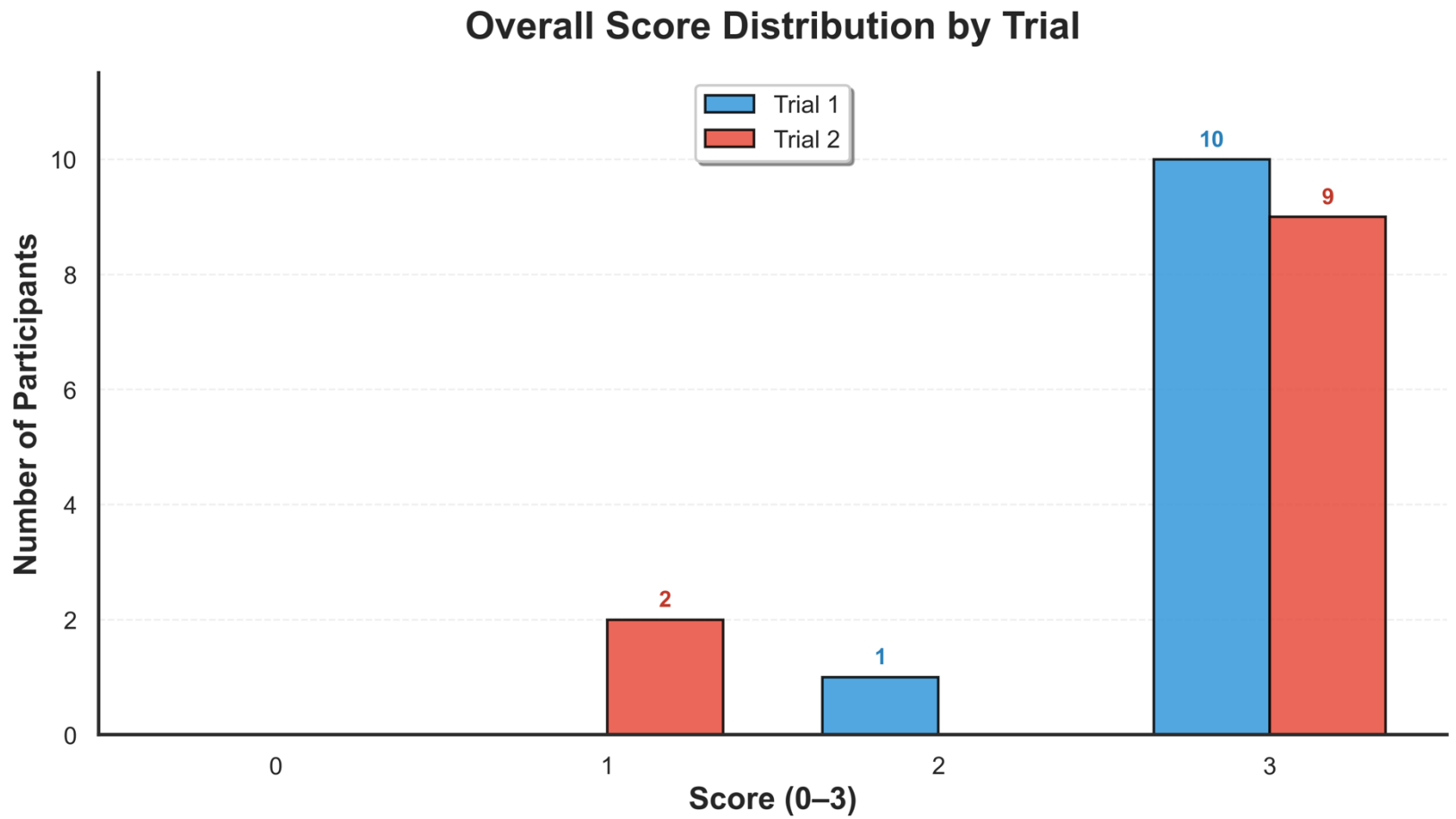
Pilot Data for Hanabi Task (Updated)

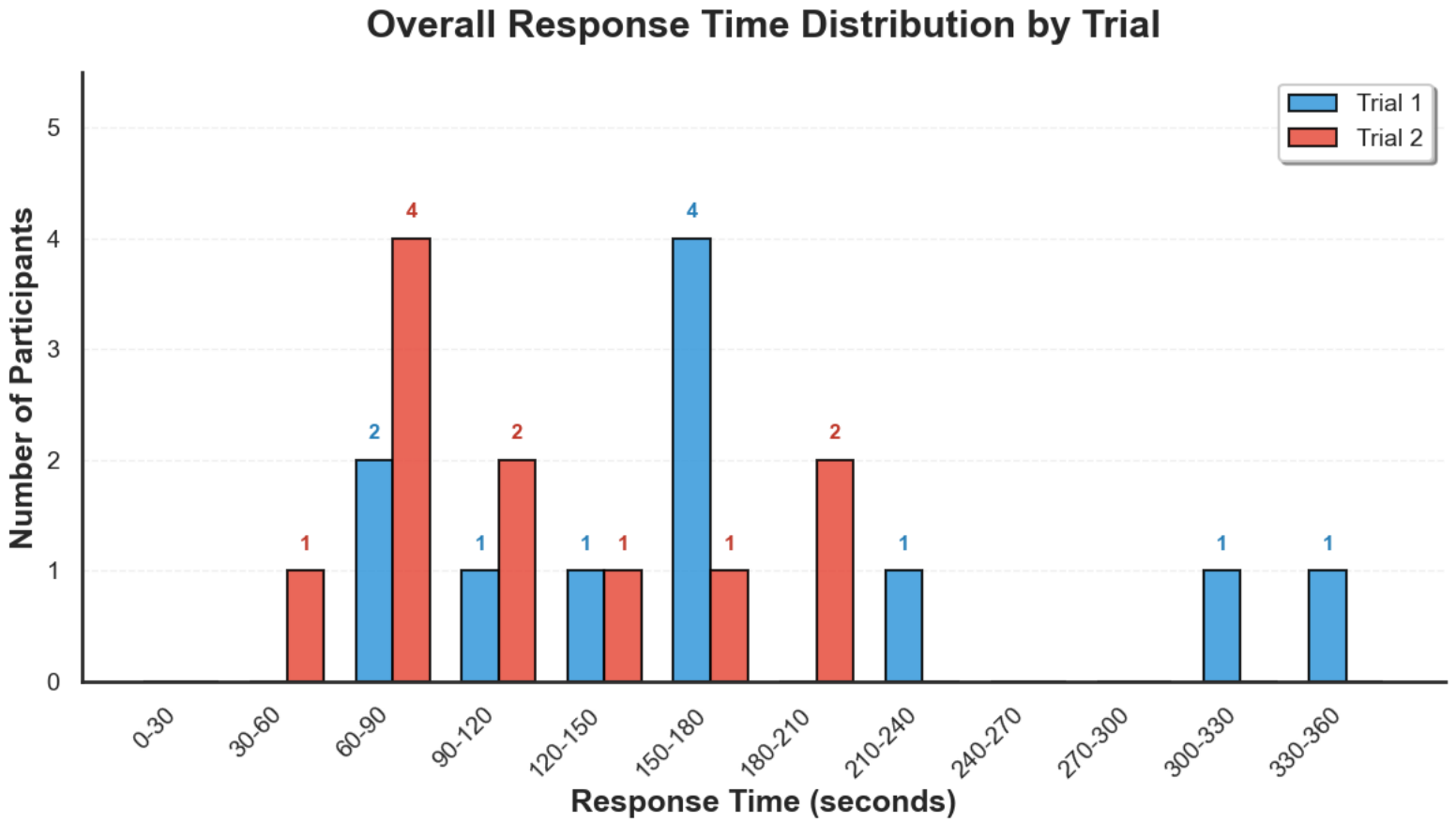


Trial 2 generally shows better performance.

Version 0.1

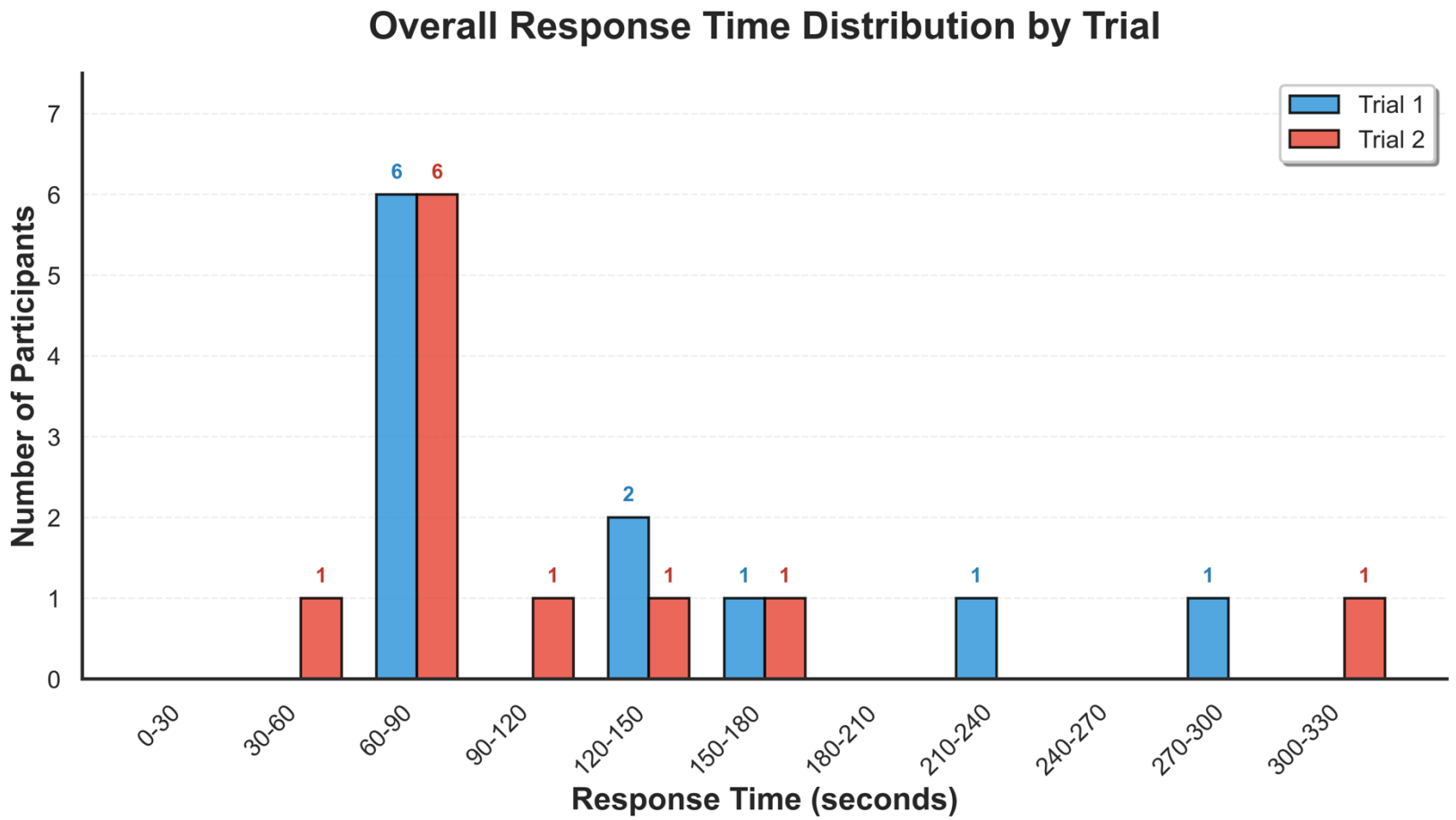


Generally better performance, trial number has no bearing.

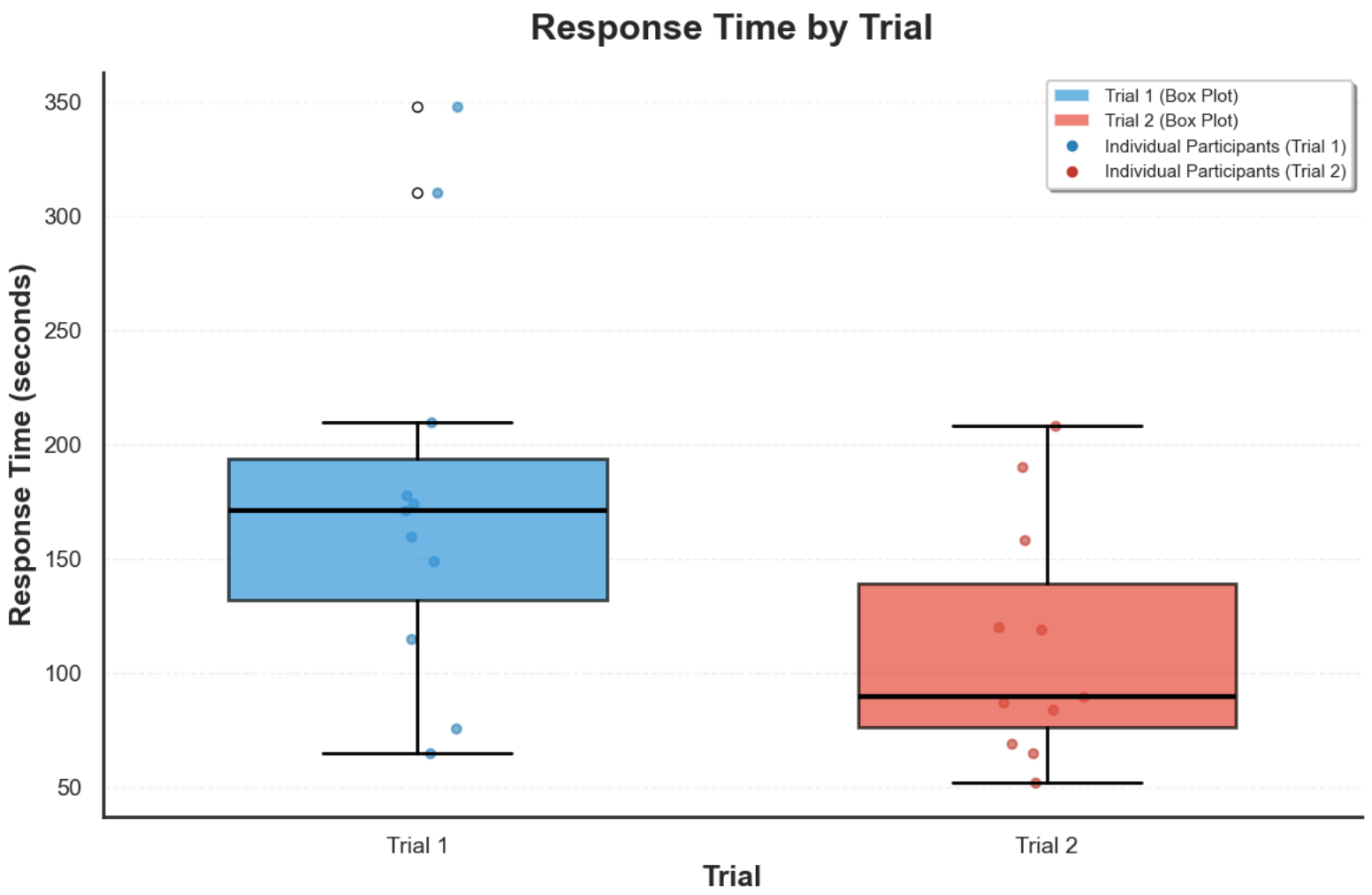


Trial 2 generally shows better performance.

Version 0.1

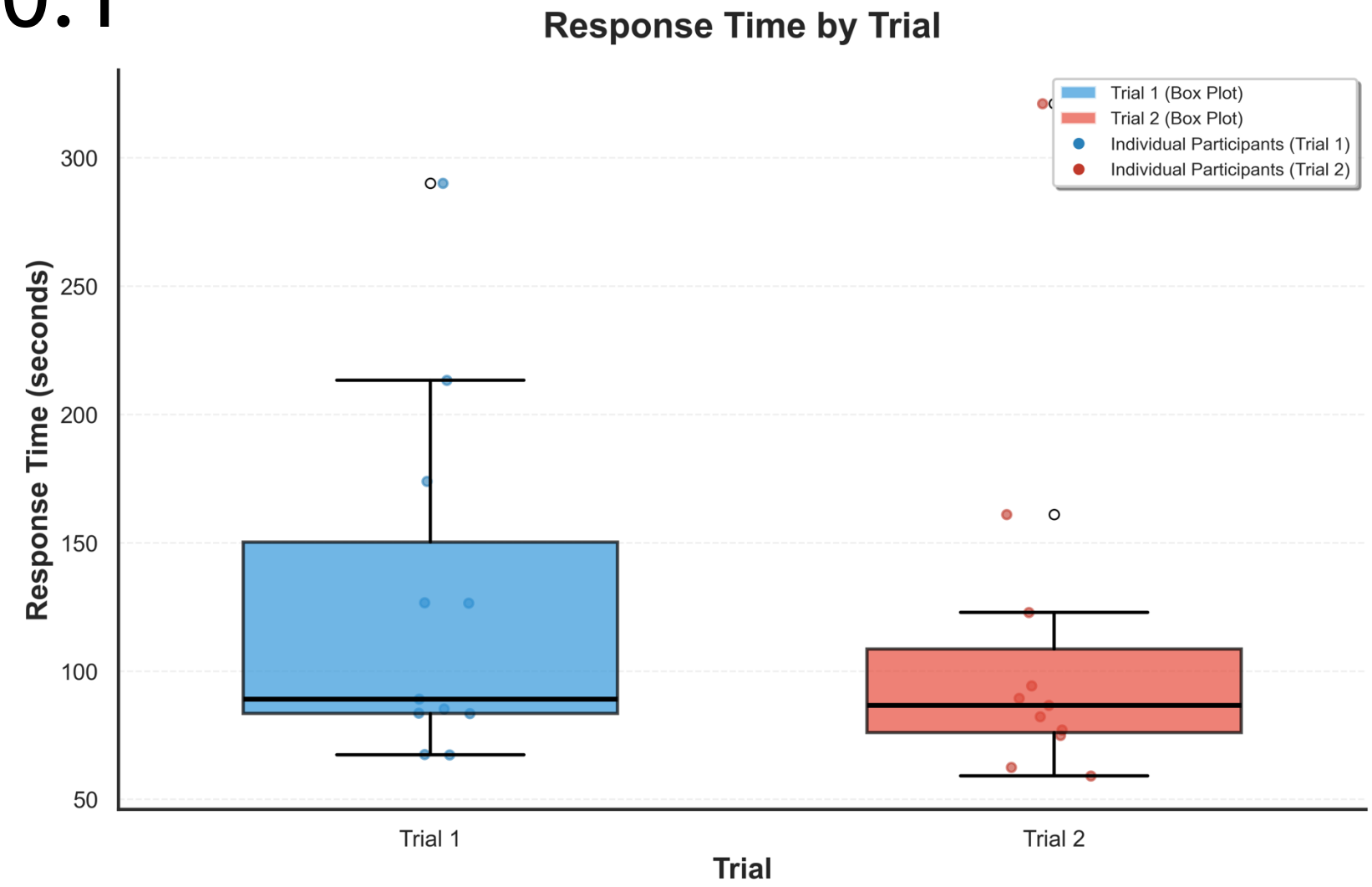


Not much of a difference.



Trial 2 generally shows better performance.

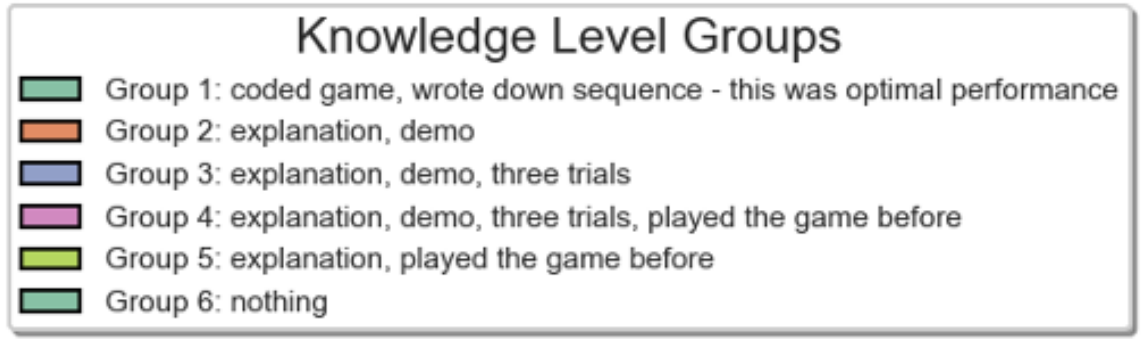
Version 0.1

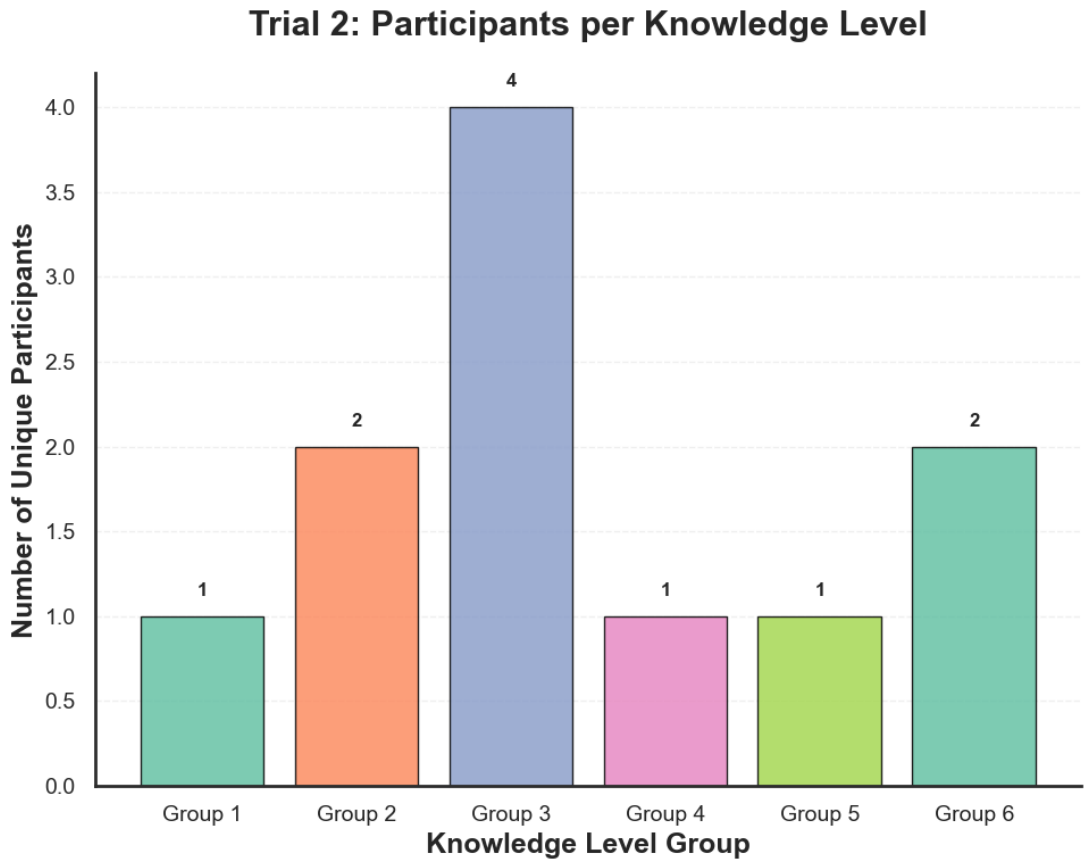


Trial 1 generally shows better performance.

Knowledge level descriptors

- **coded game, wrote down sequence - this was optimal performance** - Kahini
- **explanation, demo** - had a lengthy explanation where I drew things out, saw a demo of me playing the game
- **explanation, demo, three trials** - had a lengthy explanation where I drew things out, tried the first round with me explaining it to them/demoing, then retried the task
- **nothing** - were just given the task and nothing else
- **explanation, played the game before** - the task was simply explained to them, but they had played the Hanabi game before
- **explanation, demo, three trials, played the game before** - had a lengthy explanation where I drew things out, tried the first round with me explaining it to them/demoing, then retried the task, had also played the hanabi game before

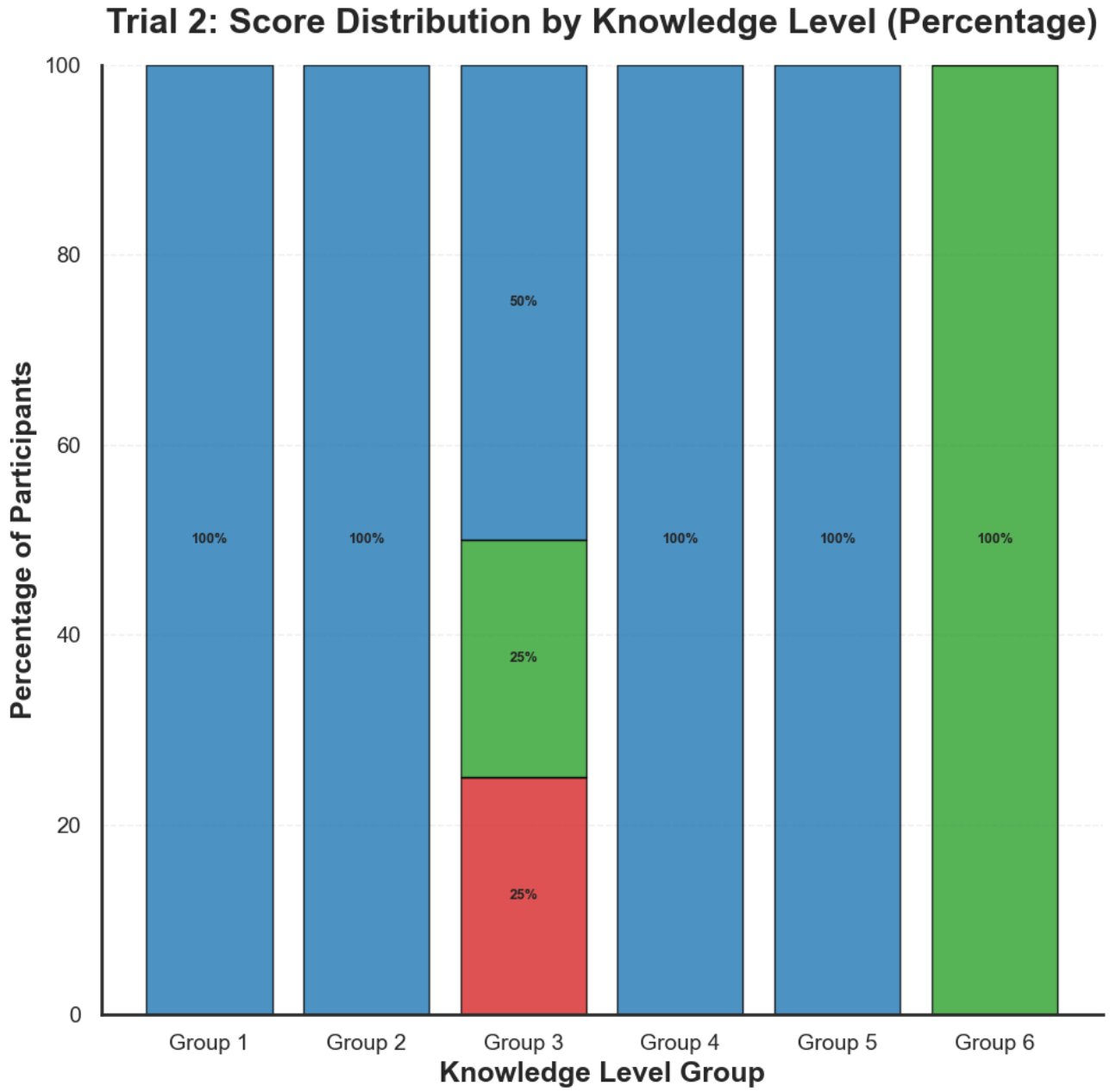


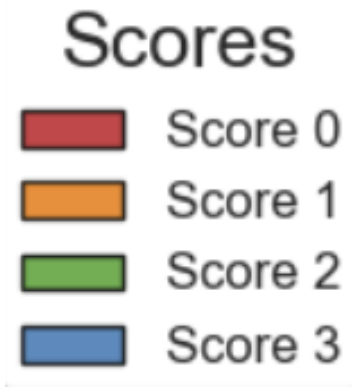


Knowledge level descriptors for v0.1

New – people playing for the first time (N=6)

Old – people playing after playing v0.0 (N=5)



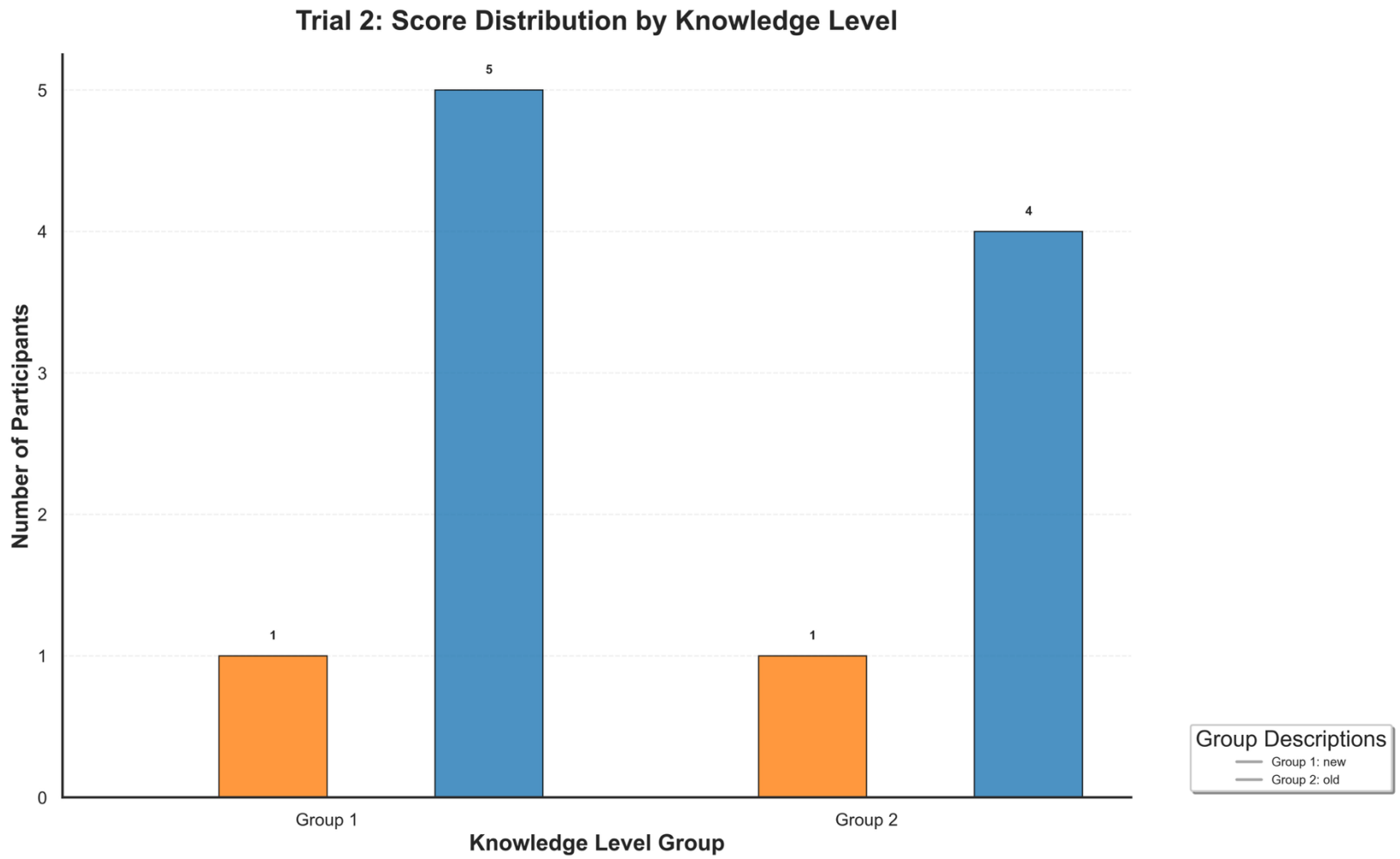


Knowledge Level Groups

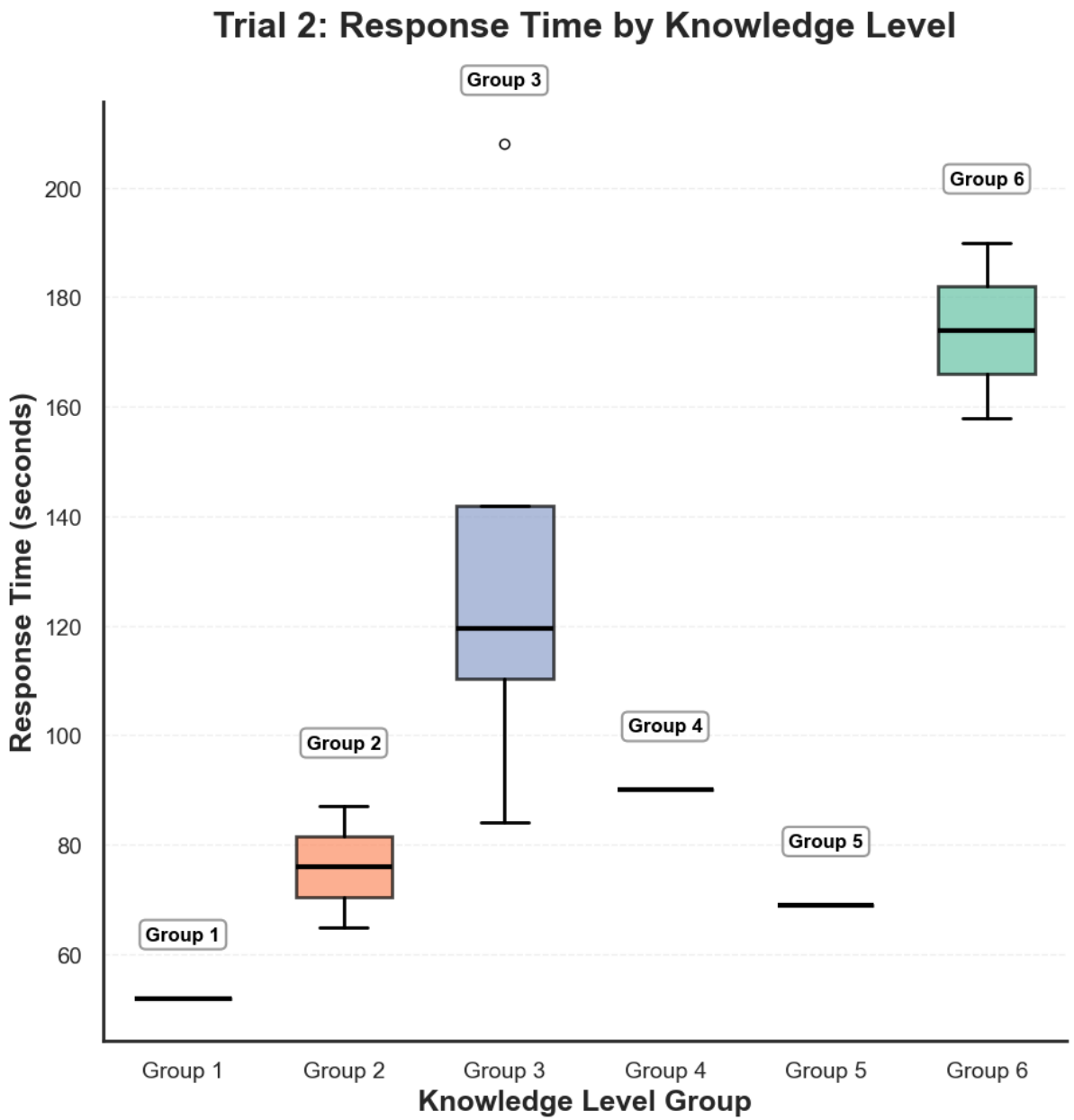
Group 1: coded game, wrote down sequence - this was optimal performance
Group 2: explanation, demo
Group 3: explanation, demo, three trials
Group 4: explanation, demo, three trials, played the game before
Group 5: explanation, played the game before
Group 6: nothing

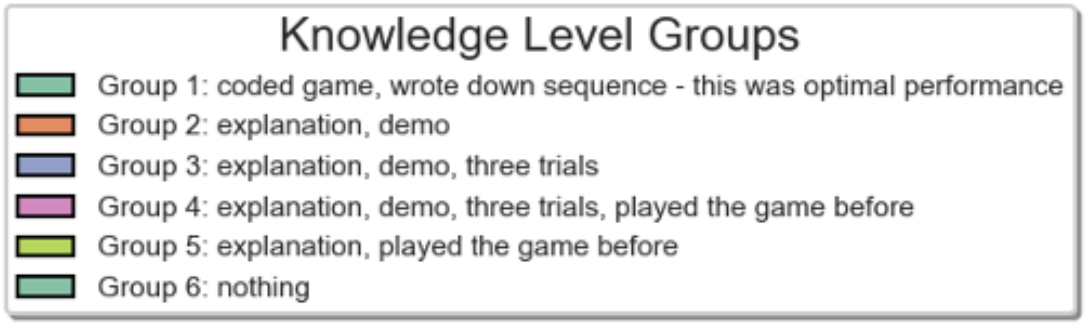
Group 3 had mixed results (took longer to understand the task), while group 2 (with no info) performed the worst.

Version 0.1



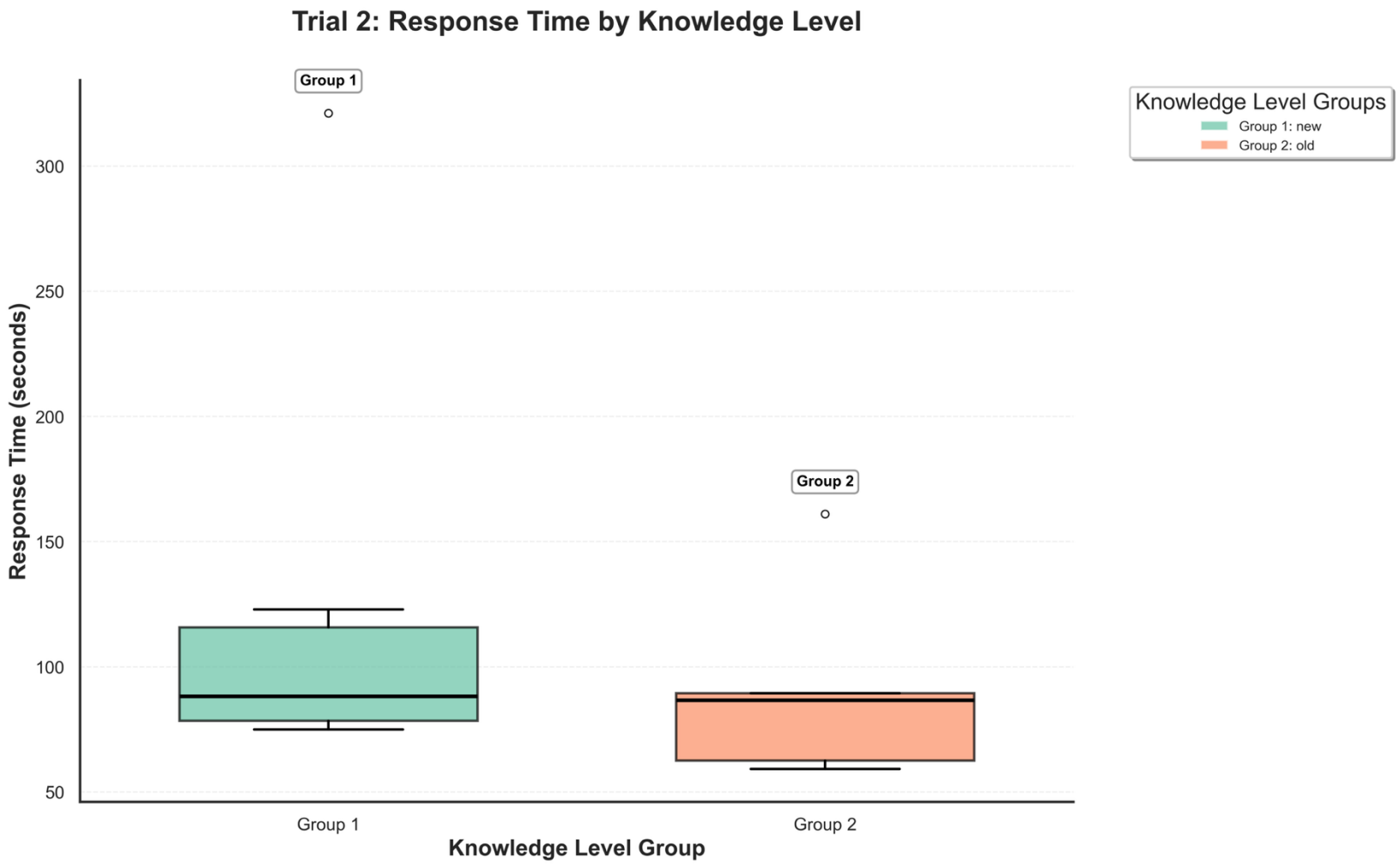
No difference regardless of whether they'd played before or not.





The group with no information performed the worst. Those with explanation/demo or who had played before did best.

Version 0.1



New group took slightly longer.

Common Complaints

- Complicated instructions
- Confusing task setup or position – how does the AI work?
- Difficult or annoying memory component
- Inconsistent task difficulty amongst people
- Frustrating but manageable overall
- Do we really need “replace”?
- Not enjoyable

Suggestions for future versions?

- Remove “replace” option
- Add logs for previous rounds
- Have AI with a better-designed strategy
- Keep sequence on the screen longer