# Cookie Cats A/B Test Analysis

Impact of moving a mobile game progression gate on retention & engagement

## **Executive Summary**

This report analyzes an A/B test from the mobile puzzle game Cookie Cats. The experiment tested whether moving a level gate from  $30 \rightarrow 40$  affected player retention and engagement.

Key finding: Moving the gate to level 40 reduced 7-day retention by approximately 0.8 percentage points. While this result is statistically significant (p < 0.05), it falls below the pre-specified Minimum Detectable Effect (MDE) of 1.0 pp. Therefore, the change is not practically meaningful, and rollout is not recommended.

### Dataset

#### Source

| column         | data type | description   |
|----------------|-----------|---|
| userid         | int       | Unique player ID  |
| version        | str       | Experiment split (gate_30 = control, gate_40 = treatment) |
| sum_gamerounds | int       | Total rounds per player                                   |
| retention_1    | bool      | Active 1 day after installing the game                    |
| retention_7    | bool      | Active 7 days after installing the game                   |

The dataset contains approximately 90,000 unique players, evenly split across experiment groups.

# **Experiment Design**

Experiment groups: Players split into control (gate 30) and treatment (gate 40).

Primary metric: Day-7 retention (retention\_7). Tested with a two-proportion z-test.

## Guardrail metrics:

- Day-1 retention (retention\_1), tested with a two-proportion z-test.
- Engagement (sum\_gamerounds), tested using Mann-Whitney U and Welch's t-test on log-transformed values

**Decision rule:** Recommend rollout only if the primary metric is significant at alpha = 0.05 and the observed effect is > 1.0 pp, without any negative effect on guardrails metrics.

**Power / MDE:** The test targeted 80% power to detect a 1.0 pp change. Observed sample size achieves 0.74 pp MDE, a smaller detectable effect than the designed 1.0 pp MDE at 80% power.

## Results

| Metric          | Test             | Control | Treatment | Abs $\Delta$ (pp/unit) | p-value |
|-----------------|------------------|---------|-----------|------------------------|---------|
| Day-7 retention | Two-proportion z | 19.02%  | 18.20%    | -0.82 pp (-4.31%)      | 0.0016  |
| Day-1 retention | Two-proportion z | 44.81%  | 44.22%    | -0.59 pp               | 0.15    |
| Game rounds     | Mann-Whitney U   | 52.46   | 51.30     | -1.16                  | 0.15    |
|                 | Welch's t (log)  | 2.888   | 2.870     | -                      | 0.15    |

### Note:

• Guardrail p-values are Holm-adjusted.

• Cohen's h = 0.02 for the primary metric, which is negligible in standardized magnitude (below the conventional threshold of 0.20 for a small effect).

Interpretation: The difference in primary metric (retention\_7) is statistically significant (p=0.0016). The treatment group (gate\_40) shows a -0.82 pp decrease in day-7 retention compared to control (gate\_30) with 95% confidence that it falls between [-1.33, -0.31] pp, corresponding to a -4.3% relative drop. While the result is statistically significant, the effect is negative and below the practical threshold of 1.0 pp. Cohen's h = 0.02 indicates that the magnitude of this effect is negligible. Even though the negative effect is minimal, rollout is not recommended.

## **Business Impact**

Player impact: For every 100,000 new installs, the change corresponds to ~820 fewer retained players at day 7. The 95% confidence interval suggests the true impact is likely between 312 and 1,328 fewer players.

**Revenue impact:** Assuming an ARPU of \$0.50/month, this indicates approximately \$410 monthly revenue loss per 100,000 installs.

# Limitations & Next Steps

The dataset does not indicate whether users actually reached the level gate, so some players may not have been exposed. This may dilute experiment results.

No covariates or behavioral segments are available for adjusted analysis or potential confounders.

Future tests should:

- Track exposure (whether a player actually reached the gate).
- Extend analysis to 30-day retention curves.
- Include covariates (region, platform, spend) for adjusted models.

## Conclusion

The experiment demonstrates a statistically significant decrease in retention when moving the gate to level 40, though the decrease is minimal in size. Both statistical and business perspectives suggest not rolling out the change.