How to 4K Playtest

By [Crz]raber

Introduction & Convention

My experience in worldwide tournaments is limited. Before applying for MWC, I had only incompletely participated as a playtester for THMC4. However, I was still successfully recruited by MWC 2025, which might indicate that I have a certain understanding of playtesting. Given that I'm already 24, I cannot guarantee maintaining a long-term stable and sufficiently high skill level. Furthermore, there don't seem to be many Chinese players specializing only in playtesting, with mappers or players often taking on the role at the same time. Therefore, I am presenting some personal understandings and viewpoints here, hoping to serve as a reference for players in community tournaments who are interested in playtesting, as well as those wanting to join world cup teams. My technical skill is limited, and so is my level of education. Inevitable improper wording, I kindly ask for your forgiveness.

This guide is primarily aimed at pure playtesters for osu!mania 4k tournaments, i.e., players with basically negligible mapping experience. Players who can both map and play at a high level are also welcome to read, but do not need to truly reference the guide. If you already know how to map, why would you need to listen to someone like me, who doesn't, on how to evaluate?

The playtesting work for other games like Malody has commonalities and can be used for reference, but the specific terminology and numerical values involved may differ. Please adjust according to the actual situation.

This guide will not explain terms like SV.

If you need to read this, you should already have an understanding of these basic terms.

All Accuracy (Acc) values mentioned in this guide are based on Score V2. Phrases like "can acc," "cannot acc," etc., may be repeatedly mentioned. To minimize ambiguity, let's establish the following conventions:

- "Can acc" refers to above 99.70% accuracy. The specific value should be determined by your own ratio. Generally, a player who hasn't specifically

trained their ratio will score around 99.7% (990k) when SSing a map, so we use this as the boundary.

- "Can play" refers to above 98% accuracy. The specific value depends on your own situation (e.g., stamina, combo retention ability). For typical map pools, 98% (900k) means you can hit the vast majority of patterns accurately and generally won't need to manip in non-late-tournament difficulties. Hence, we use this as the boundary.
- Anything below falls into the "cannot acc" category. However, you are likely to offer valid suggestions even if you cannot play a map, the effectiveness and credibility of your test results should be considered lower than those from players who can play the map.

1 Map Pool Structure & Terminology (Based on MWC 4K)

1.1 Common Terminology

- **Map:** The osu! community's term for a beatmap.
- **Mappool:** A collection of selectable beatmaps for a specific round of a tournament. The gerund "mappooling" means the act of creating a mappool.
- **Note:** The objects that make up a beatmap.
- **RC:** Rice. Refers to notes that only require a single press. In tournaments, it refers to maps where rice is the absolute main component, primarily testing a player's ability to hit rice.
- **LN:** Long Note. Refers to notes that require holding and releasing. In tournaments, it refers to maps where LN is the absolute main component, primarily testing a player's ability to handle LNs.
- **HB:** Hybrid. Refers to maps with a balanced ratio of rice and LN, presenting a composite difficulty. Tests both rice and LN abilities simultaneously.
- **SV:** Slider Velocity. Refers to maps with scroll speed changes. Requires specialized training and beatmap understanding. Starting from MWC2025, the osu!mania 4K World Cup series no longer includes SV, we are not introducing SV slots accordingly.

- **TB:** Tiebreaker. Cannot be actively picked. Is forced to play only when both teams are one point away from winning, used to decide the final victory.
- **Winners' Bracket / Losers' Bracket:** In a double-elimination format, losing a match moves a team from the Winners' to the Losers' bracket. Losing another match in the Losers' bracket results in elimination.
- **Qualifier:** A preliminary stage used when there are many registered players/teams to filter a specified number into the main event. Generally contains maps ranging from early to mid-late tournament difficulty. The best score(s) achieved by a player/team within a specified time/number of attempts is used (specific scoring rules determined by the organizers).
- **Ro *x*:** Round of *x*, where *x* is a positive integer, denotes the tournament round. E.g., Ro32 is Round of 32, referring to the Winners' Bracket 32 to 16 matches and the corresponding Losers' Bracket matches.
- **QF:** Quarterfinals. Winners' Bracket 8 to 4 and corresponding Losers' Bracket matches.
- **SF:** Semifinals. Winners' Bracket 4 to 2 and corresponding Losers' Bracket matches.
- **F:** Finals. Winners' Bracket final and corresponding Losers' Bracket matches.
- **GF:** Grand Finals. The final match of the tournament. In a double-elimination format, this pits the last remaining team from the Winners' Bracket against the last remaining team from the Losers' Bracket to decide the champion.
- **Bo *x*:** Best of *x*, where *x* is a positive integer, denotes the match format. Under this rule, the first team to reach $(x + 1)/2$ points wins. E.g., Bo13 means first to 7 points.
- **Bracket Reset:** In Double Elimination Tournaments, the Losers' Bracket team must beat the Winners' Bracket team twice consecutively to win, whereas the Winners' Bracket team only needs to win once. If the Losers' Bracket team wins the first match under this rule, the scores are reset to 0-0 and another match (or set) is played, which is called Bracket Reset.
- **WBD:** Win by Default. Victory due to a walkover or opponent forfeiting.

1.2 Map Pool Structure Overview

In the MWC format, a Best of *x* mappool always featuring *x* + 4 maps. For example, a Best of 13 mappool consists of 17 maps. The specific structure depends on the tournament team (e.g., the well-known tournament JHC [Jack House Cup] obviously doesn't include LN and HB), but typically, the baseline structure for a comprehensive tournament is as follows:

- Bo9: 6 RC + 3 HB + 3 LN + TB
- Bo11: 7 RC + 3 HB + 4 LN + TB
- Bo13: 8 RC + 4 HB + 4 LN + TB or 8 RC + 3 HB + 5 LN + TB

For tournaments without SV starting in 2025 and beyond, the Qualifier typically consists of 7 or 8 maps. Among these, Rice maps account for 3 or 4, sharing 50% of the Qualifier weighting; 2 LN maps and 2 HB maps share the remaining 50% of the Qualifier weighting.

The above description outlines the general structure of mappools and qualifiers mappools. In practice, the structure of mappools may vary depending on the mappool selector (especially head mappoolers).

1.3 Detailed Mappool Structure

Based on MWC's composition, specific details are decided by the tournament organizers.

- **RC1:** Generally takes consistent single stream as main difficulty, which tests player's speed and overall finger independence.
- **RC2:** Generally high-BPM Jumpstream map.
- **RC3:** Generally lower BPM, mid density but extremely stamina-demanding Jumpsteam map, which might also feature intrusive burst patterns. (In early rounds, RC2 and RC3 might be merged into one Jumpstream map, with subsequent slots filled accordingly).
- **RC4:** Generally chordjacks. Differs from dan course chordjacks. Tournament chordjacks' main difficulty lies in awkward patterns and mixing non-chordjack patterns in between, forcing player to transition between the jack movements and burst movements.

- **RC5-8:** Tech. In principle, the higher the RC number, the harder it is to acc.
- **HB1:** Generally features jumpstream and LN density patterns, relatively structured.
- **HB2:** Generally focused on light chordjack and shield patterns, can also include bursty and/or technical patterns.
- **HB3:** Generally manifests as a Wildcard. Mapper's creative freedom, with bizarre and extremely hard-to-acc patterns.
- **HB4:** Generally extreme technical rice patterns featuring faster minijacks interrupted by comfy, flowy or straightforward LN, or a cut version featuring the first verse of a tiebreaker song.
- **LN1:** In early rounds, manifests as shieldy coordination with some releases; in later rounds, as timing hell. Featuring columnlock jack and brokenstream with releases, aiming at testing players' coordination, release timing and stability.
- **LN2:** LN density. High-density LN, usually featuring LN Jumpstreams and intuitive pseudo-inverse patterns. This slot aimed to have minimal release difficulty.
- **LN3-5:** Generally manifests as a Mixed / Technical / Wildcard. Mapper's creative freedom, with bizarre and extremely hard-to-acc patterns.
- **TB:** Contains everything. The overall difficulty is significantly higher than other maps in the pool.

2 Basic Qualities

2.1 The Basic Discipline of a Playtester

As a playtester, you primarily need self-awareness and a certain amount of experience playing tournament maps. You need a relatively objective understanding of your own technical characteristics and performance in different environments. Additionally, a certain accumulation of tournament map experience is necessary to find references or share your points when needed.

2.2 What Skill Level is Required to Undertake Playtesting?

It depends on the tournament's difficulty. It is recommended that players who can at least play most of the Qualifier maps for a given tournament consider participating in its playtesting. Usually, the difficulty increases each round. If you reach a round where you nearly cannot acc any maps, it is advisable to stop testing or indicate that your skill is limited, and your reviews and scores are for reference only. This is not giving up halfway, but a responsible attitude towards both the mappers and the players. If you have significant skill biases, you can adjust flexibly based on the specific situation, e.g., testing RC up to GF, but LN only up to SF.

2.3 There are ET Players in the Testing Team, I Feel Worthless. What Should I Do?

Top players may not have time to go through every map. Furthermore, tops might have a slightly weaker perception of the early-round easy pools. In such cases, more opinions are never a bad thing. Believe in your own value and work with the team to present the best possible tournament content for the players.

3 Self-Assessment

3.1 The Purpose of Self-Assessment

Before stating your opinions, you need to know about yourself. It's well known that everyone's real-time skill varies, which can significantly impact test results. Therefore, quick and accurate self-assessment can reduce the influence of fluctuating skill level on your judgment.

3.2 The Content of Self-Assessment

Everyone has a vague awareness of their current skill level. Self-assessment is essentially verifying whether you are currently meeting your own estimate. Experienced players can get a fairly accurate judgement just by feeling their physical status, so the entire self-assessment process is quick – maybe playing two or three maps can give a relatively accurate result. For players with less tournament map experience or those new to playtesting, you can approach understanding your current level from three aspects: Physicality, Stamina, and Accuracy.

3.3 Methods for Self-Assessment

- **Physicality:** The easiest thing to judge. Note that the concept here refers to the upper limit of being able to play tournament maps, not your dan course. Usually, after warming up, just play maps from the target round. Pay attention to Acc and whether obvious signs of insufficient ability appear, like finger/wrist stiffness. After gaining some tournament map experience, your physicality is generally stable unless you don't play for a long time. If your physicality fluctuates, it is recommended to increase the number of reference maps played in subsequent tests.
- **Stamina:** If time is enough, it's recommended to assess both stamina when playing properly and stamina when maniping. Stamina for hitting accurately can be assessed by playing maps like RC1, observing the UR and ratio for reference. Stamina for maniping can be assessed by playing maps like RC3/LN1, observing the UR bar and miss count for reference. Generally, if the UR bar shows a significantly larger positive (+) value on a stamina map, it indicates a stamina issue causing consecutive late hits later on; conversely, stamina is likely sufficient.
- **Accuracy:** Recommended to assess by playing RC Tech or HB/LN Wildcard maps within your comfort zone. What to reference specifically depends on your technical characteristics. If your playstyle is favoring combo, then reference the count of 200s and 100s. If your playstyle is favoring accuracy, then reference your ratio. The UR bar might not be effective enough here; if you're already in your comfort zone, the offset for most notes is very small, and a single small error might cause the UR bar to move unpredictably. If your accuracy is relatively good, you can reference UR.

4 Testing Process

4.1 Self-Assessment

Perform a quick self-assessment after fully warming up. It is generally not recommended to test uniformly without warming up. Even for easy-round mappools, not warming up can have a noticeable impact. You should roughly know the highest round you can play and what kind of scores you can achieve. Then you can start testing the required maps.

4.2 Playing the Map

Do not pause. Pausing directly disrupts your perception of stamina, and stamina is itself a component of the map's difficulty. Examining difficulty while disregarding stamina is, at least when relying solely on playing, unrealistic. For maps clearly within your ability range, you can play only once, paying attention to patterns, density, etc., during the play. The specific focus areas based on your score can be referred to in the table below. If you have spare capacity, you can try to consider more mapping-oriented aspects like "expression." If you don't find the map easy to play, it's recommended to play seriously a few times first to get a score that matches your self-assessment before evaluating.

Tier	Acc	Score	Key Focus Areas
Tier 0	99.7%+	990k+	No pressure playing maps of this difficulty. Understanding of the map's difficult sections might be biased. Should focus on acc-related difficulty. Can still offer opinions on difficult sections, but reference value may be lower.
Tier 1	99%~99.5%	970k~980k	Score is relatively consistent with normal tournament participants. Likely clearer awareness of miss points and hard-to-acc points. Recommend focusing on these two aspects.
Tier 2	98%~99%	920k~950k	This range usually involves FC-type plays or few misses. Sensitive to miss points, but possibly less sensitive to timing difficulty. Recommend focusing on miss points, difficulty of burst sections, and overall stability requirements.
Tier 3	96.5%~98%	850k~920k	This range is close to the tester's limit. Acc might drop sharply due to insufficient stamina or inability to handle burst sections. Already difficult to pursue precise hits like a tournament player. Recommend focusing on overall map difficulty and stamina requirements.
Tier 4	96.5%-	850k-	At or beyond the tester's limit. Not recommended to evaluate if in this range. If the team severely lacks testers, then recommend to compare horizontally with similarly

			positioned maps. If other maps can reach Tier 3 or above scores, it might indicate this map's overall difficulty is harder; or if other maps' scores are **extremely** lower than this map, it might indicate overall difficulty is easier. But in either case, reference value is low.

4.3 Verifying Suitability

Suitability simply means "Is this map suitable for this slot?" For example, placing "time to say goodbye" in RC3 would likely be unsuitable, as its difficulty lies entirely in bursts, requiring little stamina. For general tournaments, if the organizers haven't specified otherwise, you can refer to the map pool structure described earlier. More renowned tournaments with strong mappooling teams generally don't make errors in this regard, but it's still good to pay attention.

4.4 Verifying Difficulty

The most critical and fundamental task of playtesting: "Does the map's difficulty fit the current round?" If the organizers have explicit requirements, e.g., THMC requires being slightly easier than the same MWC round, you can directly reference maps from the corresponding round/difficulty, play them, and give your opinion. Note: **It is not recommended to use Acc as the ONLY evaluation standard.** Being inside or outside your comfort zone, personal skill biases, etc., can all affect Acc on specific maps. Comparative playing is, after all, a small-sample experiment. Increasing the number of evaluation indicators within your capability (e.g., how easy it is to manip, how hard it is to acc, difficulty distribution, drop(miss) point distribution and difficulty, stamina, etc.) is necessary to improve the stability of your results. Simply put, it's best not to just drop a score and leave (Anyway it is still way better than nothing lol). Describe your playing experience as much as possible. The tester's descriptions and suggestions won't necessarily be 100% referenced by the mapper, but from the perspective of improving the overall pool quality, suggesting more is better than suggesting less.

4.5 (Optional) Discussing Specific Patterns

In some situations, such as when testers are scarce or you find a map section particularly problematic, you can choose to discuss specific patterns. For example, stating that a certain JS/HS section is too difficult, a burst is too fast, or some patterns feel unintuitive, etc. This assists your own playing experience and provides more information to the mappooling team. But be aware that some mappers might consider excessive interference with specific patterns as overstepping. Gauge the communication appropriately. Generally, it's recommended to point out that a certain block of patterns has issues without explicitly stating the optimization solution, unless the mapper asks for it.

4.6 (Optional) Macro-Level Discussion

As everyone knows, difficulty is almost impossible to control accurately. No matter how well a round's pool is made, there will always be relatively easier and harder maps. As a tester, you can, if allowed, provide suggestions for the entire pool based on factors like the participant players, past tournaments, or your own experience. For example, there are too many easy-side maps, or the entire pool's stamina requirement is slightly too high, leading to situations where players might exhaust their stamina on one map and collapse subsequently, etc. Note that this aspect is not easy for mappers to notice, and the mappooling team might also overlook it due to busy schedules. So, if you identify and decide to raise macro-level issues, it's best to articulate them clearly, allowing the mappooling team and other testers to provide feedback quickly.

4.7 Credibility Statement

Based on your self-assessment and your achieved score, you can attach declarations like "low reference value" when stating your opinions, or differentiate through tone and wording.

5 Conflict Handling

5.1 Principle of Friendly Discussion

Testers do not directly create beatmaps, thus they are not really "creators". Therefore, regardless of the situation, please respect the mapper's wishes and discuss friendly.

5.2 Conflict of Opinion Among Testers

Testers generally don't have serious disagreements. If two testers aren't familiar with each other, they can simply state their own views—it's subjective experience, and completely opposite conclusions aren't strange. Let the mappooling team judge. If a disagreement requiring interaction does arise, it's recommended to focus on describing "the impact of the map's objectively existing characteristics on your own play" or "the impact of your own technical characteristics on your own score," etc. **Avoid evaluating others' skills** or discussing issues unrelated to the current map.

5.3 Conflict of Opinion Between Tester and Mapper

Generally, conflicts arise mostly when mappers ignore or do not adopt testers' suggestions. Here, I personally recommend respecting the mapper's wishes. Even in the worst-case, e.g., the mapper creates an completely unreasonable pattern and refuses to change it despite feedback, it is the mappooling team and the corresponding mapper who bear the responsibility. As long as you raised the issue, it is not your fault regardless of the outcome. Furthermore, unlike players, playtesters aren't forced to practice the map whether it's good or bad. If you think it's not good, you can choose not to play it or simply delete the map. Therefore, there's no need to stubbornly argue.

5.4 Conflict Arising from Map Pool Updates

This situation is relatively rare. For example, a map you tested extensively and communicated about is suddenly requested to be replaced by the mappooling team, or the pool structure changes temporarily, rendering previous work wasted. A mature tournament team should be able to avoid such situations. If such problems do occur, as a tester, you do not need to publicly express any

opinion. If really feel being offended, you can refuse to participate in subsequent testing for that slot by making excuses or simply not replying. Usually, no one will bother you about it, as such changes are visible to all. In short, absolutely do not publicly provoke conflict. In a team, overall work efficiency is the priority. There's no need to be the one who firstly harms the team's efficiency.

5.5 Self-Negation

Due to your own issues, such as inappropriate self-assessment, leading to different results in two separate tests, is a relatively common phenomenon. Don't worry. If it's only a difference in degree, e.g., "much harder" becomes "a bit hard," you might not even need to post a score; just send a couple of messages to compensate. The mappooling team will understand. If there is a change in conclusion, e.g., "harder" becomes "easier," it is recommended to attach the score screenshot, explain the reason for the difference, state which score you believe is more reliable, and apologize (optional).

Conclusion

This guide set out to translate hands-on experience as an osu!mania 4K tournament playtester into a practical framework for others—especially “pure” playtesters without substantial mapping background. It formalizes shared language, clarifies mappool structures, and outlines an actionable testing process centered on rapid self-assessment, disciplined playthroughs, and evidence-based feedback.

Key Insights.

1. A concise taxonomy of common slots (RC/HB/LN/TB) and their intended skill emphases, aligned to typical MWC-style pools.
2. A lightweight, repeatable self-assessment method (physicality, stamina, accuracy) to calibrate expectations before testing.
3. A score-tiered interpretation of test results that anchors what to focus on at different performance bands (from acc nuances to systemic stamina demands).
4. Clear guidance on suitability (slot fit) and difficulty verification (round fit) that encourages multiple indicators beyond raw Acc.

5. Practical norms for constructive communication—within tester groups, with mappers, and when pools change—so feedback improves pools without eroding team trust.

Implications for Practice.

- For playtesters: Treat self-assessment as part of testing, not a preface to it. Report how patterns affected your play (miss points, stamina decay, timing traps) and qualify your confidence based on tiered results.
- For mappoolers: Leverage tester diversity. Aggregate reports to detect macro imbalances (e.g., cumulative stamina tax, skewed difficulty spread) that are hard to spot from slot-by-slot polishing.
- For mappers: Expect specific, impact-focused feedback rather than prescriptive pattern edits; invite deeper pattern discussion only when useful.

Threats to Validity. Small-sample comparisons, day-to-day form variance, and personal skill biases can distort difficulty judgments. The proposed mitigations are: (i) quick re-calibration via self-assessment, (ii) triangulating multiple indicators (ratio, UR tendencies, miss distribution, manipulability), and (iii) transparent credibility statements.

Playtesting is most valuable when it is repeatable, humble, and specific. By pairing disciplined self-calibration with concrete, impact-oriented observations—and by maintaining respectful dialogue with mappers and mappoolers—individual testers can measurably raise the quality and fairness of tournament mappools, regardless of their absolute skill ceiling.