Why ranking systems in pen spinning fail

Varoza, Fen Spinner white papers

September 8, 2026

Abstract

In pen spinning, there are numerous ranking and level systems, but they struggle to provide an accurate assessment of the spinning level or competitive ranking. We analyze these existing systems and describe alternative approaches to evaluating the level of an independent spinner's abilities.

1 Introduction

There are a number of different systems to determine the skill level of spinners in the current pen-spinner hobby. The contemporary start of the trend begins with a blog post by RPD describing a system to rank spinners by 'belt' levels consisting of a series of requirements that must all be met in order to achieve the colored belt level. This system was inspired by the system that the lockpicking community uses, which in itself is derived from the system used in martial arts.

As of recent, these level systems have further expanded into challenge ranking, tournament ranking, points scoring systems, which in themselves have different challenges. These are expanded upon in the next section. The proliferation of these systems has become an increasing trend, with the vast majority of the English speaking pen spinning community at the very least currently engaging with at least one of these systems in recent years.

2 Existing systems

2.1 Tournament result systems

The most basic system to rank spinners is by results in tournaments. This was initially put in place by the Spindys role system in which a rank could instantly be obtained by having won battles or achieved a specified ranking in a tournament. Currently, the only way to achieve the highest level in the Spindys role system is through winning a Pen Spinning World Tournament.

There are a number of challenges with this system. The first is that the level that is represented is much more honorary than representative, as tournaments extend far into the past and continue to be formed into the future, and the landscape of the upper levels changes dramatically with each passing year. Is winning World Cup 2020 the same as winning World Cup 2022 or winning World Tournament 2017 the same as winning World Tournament 2025?

The next major problem is that tournament results often are not representative of skill level, as there have been many upset battles as a result of judging, as in WT15, or from spinners not submitting or struggling to land a good draft late in the event, or from tactical decisions in team tournaments. Is Sutomo deserving of a level higher than Fel2Fram because of an upset battle? Many would disagree.

Tournaments are also affected heavily by rules around pen mods and filming rules, which may allow spinners to use mods that are dramatically superior to their opponents or use large filming times to gain a significant advantage, but for practical reasons only a few spinners make use of this. In the case of filming this is self evident, the top levels of World Tournament events are full of spinners who film combos even over a year before the event starts, but only a handful of people have the time to adhere to such a filming schedule, putting fairness at question.

2.2 Challenge systems

Challenge systems are another common approach that appears in the pen spinning ranking scene. Challenge setters create a list of tricks, sequences or combos to either perform or beat in a comparison contest, and then submissions are reviewed and roles awarded, with levels given to each role.

Challenges are inherently bound to the tastes and biases of the challenge setters, and rarely are chosen based on data. For instance, in Spindys challenges where the format started, the level 5 combo challenges are essentially only ever held by the person who is in the example used to set the challenge. This leads to a rather perverse landscape where the levels concentrate on people who are close to the person who sets the challenge.

The distribution is also often uneven. Challenges for technical spinning that require participants to achieve degrees in structure, execution and difficulty at the same time are almost never passed above an intermediate level, whereas power challenges are drastically overrepresented in number of ranked roles given.

2.3 Points systems

Points systems have been theorized and recently implemented. Namely the main points systems exist in Planet Pen Spinning with regards to their points and Elo systems.

2.3.1 Elo, Glicko and similar

For points based rating systems such as Elo, these systems are entirely unsuitable for pen spinning and generally misapplied. In Planet the Elo system is particularly egregious, as your initial Elo score is given directly to you upon review. The maximum that you can gain and lose score in the Elo system is very low, which is why it is designed to operate with low score numbers and high volumes of battles to adjust the score. In pen spinning there are far fewer battles and it is unlikely that a competitive scoring system will be configured correctly for a few reasons.

The first is that smart systems need to allow decay to rack up significantly in a person's rating. The best example of this is the Glicko system. However, due to the low frequency of battles and the long duration it takes for them to be performed, the uncertainty number that affects the maximum amount of score that can be switched between spinners would trend towards the maximum at all times, leading to tremendous swings in rating. At the same time, a cutoff is needed for score to decay entirely back to unranked because of large breaks, which would have to be constantly applied.

As people wish to have ranks and hold them for a significant period of time, these systems do not get implemented correctly, leading to the current system in Planet where the rank that you are granted at the start is so far away from the other ranks that participating becomes pointless.

Additionally, Elo points can also be awarded through mechanisms other than battles, which has led to situations where performing tasks for the community (such as making a submission to a video that the challenge setter is going to post to their youtube channel for views) results in gaining competitive rating points. This defeats the purpose of the rating in representing degrees of skill.

As a real example of how the current rating systems play out, at the time of analysis, the highest ranked spinner in planet, Mumm3y (who is responsible for the initial Elo scores being granted), has a score above 11000. Elo is not designed to work at these levels of score. He has one battle, which he lost to DumBananaMike, who was rated around 6000. The score adjustment was insignificant, to the extent that if DBM wished to reach Mumm3y's rank, he would need to defeat him no fewer than 200 more times in a row.

It should also be noted that due to other challenges giving Elo score, Mumm3y actually ended his only battle at the time with a higher rating than when he started. In essence, the Planet implementation in particular is a case of being given a rank that a given clique decides you should have, and it is extremely impractical to try and move up ranks.

2.3.2 Other points systems

Other points systems are not being used for rankings, but it is likely that this will be attempted in the future. The main issues with score systems that endlessly accumulate is that they do not decay, with the alternative being to allow decay of score but greatly harm the motivation to participate in them, and that they are entirely at the whims of the one setting the score challenges that grant score.

3 Problems associated with multiple systems

3.1 Rank rot

For the purposes of this document, "Rank rot" refers to the situation where a rank that a spinner holds begins to lose legitimacy as a result of time, and is eventually removed. This tends to happen with new developments in pen spinning, switches to different trends or the levels being readjusted to new challenges.

Rank rot causes a major issue in that it erodes the incentive to pursue the ranks themselves if the ranks are removed from individuals, since achieving these ranks becomes increasingly time consuming and difficult at higher levels. It is not as simple as never removing them either, as it does not make sense to have KTH on the same level as Criswea!! on a ranking system merely by nature of both winning World Tournament, or a power spinner at level 5 if they have not trained the new material in the past 5 years.

3.2 Rank toxicity

Ranks lead to clear divisions between the ranks that people hold in the server. It becomes extremely easy to dismiss someone when they are marked as a lower level than you, particularly if you're ranked many levels higher than they are. While having to not answer every question diligently that a beginner asks you is a topic of its own, it remains also true that discoveries and important interactions do not always happen from the highest ranked, most skilled members of a community, and rank is insufficient justification to dismiss.

Another tension between the ranks is in the obtaining of ranks. Those who obtain ranks tend to be those who spend a lot of time in a gamified system and not always those who are best capable to obtain the rank in that system. Many spinners who can achieve high ranks simply choose not to bother with the systems. It also introduces tension in verification and challenge setting, as the challenges have to be continually updated and the submissions verified, but these are subjective and so biases can be both demonstrated and accused. Can you tell the difference between an approval or denial being valid or gatekeeping for illegitimate reasons?

It also doesn't do to make the rank barriers blurry, since it starts to lead to controversy even within the ranks themselves. Spindys recently introduced a system that splits the ranks in two with "combo certified" being harder challenges that essentially rank people within a specific rank. Prior to that there was a lot of discussion as to whether people deserved to be in the rank at all due to the high variance in level 4 rank holders.

3.3 Rank attention

Often in the case of high rank holders in these systems, the attention from being given these ranks tends to be greatly increased. This often manifests as being nagged by newer members of the community, regardless of the rank holder's viewpoint on their community contributions or status as a mentor for the community or not.

As an example, the author of this article was until fairly recently the highest rank holder in Spindys for several years. The ranks were used to adjust members on the user list of the Discord server, leading to so many notifications and pings, varying from basic questions to outright toxicity, that all notifications on the server have had to be muted for years. The high rank holders were also confused for server staff, which they were not, further complicating matters.

4 Alternatives to ranks

The original goal of ranks in the community was to provide a system that allows spinners to see those of a higher level for advice. At least, this was the goal that RPD was looking towards in implementing this for SPSC. Unfortunately, this is not really how they played out in the end, with them acting much more like a game to be figured out and solved and not really encouraging either as much community participation as hoped or the same quality of community participation as hoped.

It should be noted that in his original suggestion, RPD made explicit requirements that people who wanted the ranks need to provide value to the community by writing articles, entering into collaborations and so forth, and these were not optional. There are issues with this suggestion too, particularly when it comes to individuals handling systems in a gamified way, but they are small compared to the larger issues outlined in this article.

The solution to ranks is to instead mark users apart in the community by the role that they play within that community. For instance, people who are skilled and willing to take on the task of helping new spinners should be marked "Mentors", who perform research should be marked "Researchers", and who perform clerical tasks such as wiki management "Curators" or similar. These are much more clear roles with responsibilities. In particular, they are responsibilities in the community outside of the typical User/Staff divide that separates regular users from those with moderator powers.