Multidimensional success measurements in tournaments increase participation through hedonic framing

Marc Oliver Rieger*† Nicolas Diewald‡ Maximilian Lenz* Maolin Zhang*

March 5, 2018

1 Introduction

Tournament systems in professional tournaments on the one hand try to measure the relative strength of the participant as accurately as possible and on the other hand try to increase the experience of the audience. In amateur tournaments, particularly in mind games, however, often another goal is much more important: the enjoyment of the participants themselves and their willingness to participate.

While there are surely many ways in which to improve this enjoyment, in this article we focus on one particular factor: success measurements. Most tournament systems provide basically one measurement of success, e.g., in a KO-tournament, success may be measured by your final rank, the number of wins or the round up to which you proceeded, but all of these measures are obviously very highly correlated (if not identical). Similarly, in a Swiss system tournament, number of wins and final rank, again, are by definition mostly the same success measure.

Why would it be better to have more than one such measure? The psychological reason is that having two or more such measures allows you to do "hedonic framing", see, e.g., Thaler (1985). By this we mean that you are free to weight these measures *after the fact* in order to increase your subjective happiness about your results. If this indeed happens, the satisfaction of participants with their performance (and thus probably with the whole tournament) would on average be higher as compared to tournaments with only one success measure.

*Trier University, Germany

†Corresponding author: mrieger@uni-trier.de.

‡Polarixpartner, Saarburg, Germany

In this article we look at one such tournament system with (at least) two independent success measures, the McMahon system (British Go Association 2004). This system is a variant of the Swiss system that assigns starting points to each participant depending on their strength/rating. The system is most frequently used in Go where it is the standard in amateur tournaments in Europe and America. In this system, only the best players have a chance to win the tournament, while weaker players cannot even theoretically win it, since their starting points are too low. Weaker players in this system will play against other weaker players and have therefore, however, the chance to win many (or even all) of their games. This leads to two different success measures that are basically independent: final rank and number of wins.
In an empirical study with data from all European Go tournaments from 1996 to 2016 with around 40,000 participants (taken from the European Go Federaton (2018)), we find evidence that participants indeed show hedonic framing: the likelihood to participate again (and sooner) in a tournament is higher if the participant's rank was better and if their number of wins was higher, but both effects did not add up. This implies that multi-dimensional tournament systems (like McMahon) allow participants to see their own performance in a better light and thus be more satisfied.
This leads to a higher chance of participating again. The result has obvious implications to other mind sports and strongly suggests to use multi-dimensional success measures as often as possible in amateur tournaments.
The rest of this article is structured as follows: In Section 2, we describe the data and the variables of our analysis. In Section 3, we present the empirical results. Section 4 concludes.

2 Data and variables

We use data from the European Go Federaton (2018), an online database managed by the Association for Go in Italy on behalf of the European Go Federation, that includes all Go tournaments in Europe since 1996 with more than 40'000 players and more than 10,000 tournaments across all countries. It also provides up-to-date rankings of all players, summarized by the GoR ("Go ranking") that starts with a minimum of 100 points (corresponding to traditional Go ranks of 20 kyu or below) and increasing by 100 points for each kyu level. The highest ranked players in the database are professionals[1]
We excluded players with only one tournament and cases where players played only one game in a tournament. Besides the GoR and the total num-

[1]Due to the fact that East Asian professionals only play few tournaments in Europe, their ranking is of course not as precise, but still reasonable: at the time of writing, the highest ranked player is Gu Li (9p) with 2976 points. The strongest amateur player is already on rank 107. Given the small number of tournament participations in Europe of most professional payers (mostly one or two tournaments only), the sample is clearly dominated by amateur players.

ber of tournaments that a player played, we also recorded for each player's tournament participation his loss ratio LRT (losses divided by rounds played) and his placement ration PRT (placement divided by total number of participants). Both serve as success indicators (as our subsequent analysis will confirm). They are also relatively independent: a good placement in the McMahon system can coincide with a poor performance, since the former is mostly determined by the initial rank of the player as compared to the ranks of the other participants.

Let us illustrate this with a typical example: at the 15th Pisa Tournament, the player on the 9th place had lost four out of his five games, but since he started as 1 kyu, he was still scoring better than the player on the 17th place who had won four out of his five games, but only started as 7 kyu.[2] All in all, 49 players participated, so while the first of the above mentioned players could be happy with his placement (9th out of 49, i.e. $PRT = 0.18$), the second one could be happy with his score (4 wins out of 5 games, $LRT = 0.2$). The key idea is that both players will likely evaluate the tournament in terms of their success and will disregard the (not so successful) success criterion. Thus both can be happy about their result!

When looking at the data, the large degree of independence between LRT and PRT is confirmed: their correlation is only 0.35.[3]

Finally, we defined the target (dependent) variables in two ways: first, a dummy variable denoting whether or not a player dropped out of playing tournaments (i.e. whether this was his final participation within our dataset[4]) and second a variable that measured the time until he played the next time. (In this case, we obviously had to take out his last recorded participation from the data, since we then do not know when (and whether) he will play next time.)

The question that we want to study empirically is now whether (and how) LRT and PRT influence the chances of playing again in a tournament (after a short time), in other words, how general tournament participation is affected by them.

[2] We have to point out that the tournament system is still fair, since the weaker player had also weaker opponents and would have likely failed to show such a good performance against much stronger opponents.

[3] Not all tournaments in Europe are McMahon tournaments, so when disregarding Swiss system, round Robin and KO tournaments, the correlation should be even smaller. Given, however, that the tournament system is not recorded in the EGD, it was not possible to calculate this coefficient. For the same reason we could also not directly compare drop out rates after McMahon tournaments and other tournament types.

[4] Of course we cannot exclude the possibility that the player will indeed play again, but at least this serves as a good proxy for dropping out.

Figure 1: *Upper table:* Expected behavior if LRT and PRT were simply two independent factors in the evaluation of tournaments. *Lower table:* In reality, it is sufficient that one of the two factors LRT and PRT is high to observe the effect. This demonstrates that tournaments with several independent evaluation factors lead to a higher chance of a perceived success.

	PRT low	PRT high
LRT high	+	++
LRT low		+

	PRT low	PRT high
LRT high	++	++
LRT low		++

3 Empirical results

To measure the effect on the dropping out likelihood, we used a OLS regression with LRT and PRT as dependent variables. Additionally, we check for an interaction effect between both variables. – Why that?

Let us assume that a player has the two success criteria LRT and PRT, but his judgement on his success is simply the average of both. In this case, having two criteria instead of one will not increase the average subjective success: players who are successful in only one factor will somehow feel "half successful" in total.

If, however, a player decides after the tournament to mainly focus on the factor in which he was more successful, then the picture changes: the number of players who now feel positive about their performance increases substantially!

Fig. 1 illustrates this graphically. We see that in order to prove our hypothesis (that multiple success criteria in tournaments increase average subjective success and thus the tournament participation) we need not only PRT and LRT to be significant, but also their interaction term (PRT×LRT) significant with the opposite sign.

Table 1 shows that this is indeed the case, even when controlling for playing strength and total number of tournament participations.

Of course, one might say that this is some kind of "self-deceit". Therefore one would expect to see this effect to be weaker for more experienced players, i.e. for players with more tournament participations or with a higher rank. This is indeed the case as we can see by studying the interaction terms

between LRT×PRT and the total number of tournaments[5] or the GoR, see Table 2.
As a final robustness check we repeat the analysis of Table 1 with the duration to the next tournament participation as dependent variable (instead of the dropping out dummy). The results confirm that again PRT, LRT and their interaction term are significant explaining variables (Table 3).
Finally, we take a look at the effect size in Table 1: in both models, the probability to drop out increases by around 10-20% through each of the variables LRT and PRT and their interaction. This is a substantial amount, given that many other potential factors for dropping out (family or job reasons, e.g., or the general fun of participating an event, meeting friends etc.) are not controlled for.
The duration to the next participation is also substantially shortened by a good tournament result in at least one of the two success measurements (Table 3).

4 Conclusions and recommendations

The empirical findings of our analysis are clear: a major factor in the regular participation of players in amateur tournaments is success. If a tournament system allows the players to measure their success in more than one (independent) way, they use this chance and evaluate their participation according to whatever criterion they were better in. Therefore choosing a tournament system with several evaluation criteria, like the McMahon system, will enhance perceived success of participants and therefore leads to a higher long-term participation rate.
All of this strongly suggests to use multi-dimensional success measures as often as possible in mind game amateur tournaments. This makes in particular a strong case in favor of the McMahon tournament system. Tournaments where participants play in separate groups according to their playing strength can be a solution if the McMahon system is not feasible. Pure Swiss tournaments and of course KO tournaments should be avoided.

References

British Go Association (2004), 'McMahon pairing rules'.
URL: *www.britgo.org/organisers/mcmahonpairing.html*

European Go Federaton (2018), 'European Go Database'.
URL: *www.europeangodatabase.eu*

Thaler, R. H. (1985), 'Mental accounting and consumer choice', *Marketing Science* **4**(3), 199–214.

[5]Since its distribution is highly skewed, we take its logarithm.

Table 1: Joint effect of loss ratio (LRT) and placement (PRT) in tournaments on the likelihood to stop participating at Go tournaments.

	Model 1	Model 2
Constant	1.59***	1.85***
	(94.5)	(84.9)
LRT	0.11***	0.17***
	(3.5)	(5.6)
PRT	0.26***	0.15***
	(9.3)	(5.3)
Go Rating		0.00005***
		(6.9)
Log Tournaments		0.14***
		(-43.5)
LRT × PRT	-0.14**	-0.17***
	(-2.9)	(-3.9)
R^2 (in %)	1.6	11.8
F	87.0***	439.8***
N	16455	16455

Table 2: Changes of effect size depending on playing strength.

	Model 1	Model 2
LRT	3.49***	2.96***
	(97.7)	(55.3)
PRT	2.71***	2.08***
	(92.8)	(72.9)
Go Rating		0.0008***
		(78.4)
Log Tournaments	0.53***	
	(74.1)	
LRT × PRT	-4.57***	-3.32***
	(-80.1)	(-50.8)
LRT × Log Tournaments	-1.15***	
	(-61.3)	
PRT × Log Tournaments	-0.94***	
	(-45.3)	
LRT × PRT × Log Tournaments	1.60***	
	(-45.0)	
LRT × Go Rating		-0.001***
		(-40.7)
PRT × Go Rating		-0.0007***
		(-22.5)
LRT × PRT × Go Rating		0.001***
		(23.4)
R^2 (in %)	93.9	94.1
F	36278.2***	37686.2***
N	16455	16455

Table 3: Robustness check: effect on the time until the next tournament participation.

	Model 1	Model 2
Constant	230.78***	-178.81***
	(10.5)	(-7.9)
LRT	316.21***	274.27***
	(7.4)	(6.7)
PRT	96.49**	75.11*
	(2.63)	(2.2)
Go Rating		0.14***
		(20.5)
Log Tournaments		128.33***
		(29.1)
LRT × PRT	-295.34***	-261.33***
	(-4.7)	(-4.4)
R^2 (in %)	0.3	10.32
F	23.4***	426.8***
N	18598	18598