

# **An updated assessment of the predictive accuracy of World Tennis Number and Universal Tennis Ratings**

Lucas A. Kiely  
University of Chicago

Rebecca L. Mayew  
RSPA Tennis Coach

William J. Mayew\*  
Duke University

**April 10, 2025**

**ABSTRACT:** In 2024, the World Tennis Number (WTN) algorithm was updated to enhance its predictive accuracy. This study analyzes 4,174 matches involving world-level junior development players in the 12-and-under (12U) and 14-and-under (14U) categories, as well as Intercollegiate Division 1 players, from tournaments sanctioned by the Intercollegiate Tennis Association (ITA), United States Tennis Association (USTA), United Kingdom Lawn Tennis Association (LTA), Tennis Europe (TE). We compare the updated WTN and Universal Tennis Rating (UTR) classification accuracy in predicting match winners across these competitions. Our findings reveal no consistent evidence that WTN outperforms UTR in any category. UTR maintains superior predictive accuracy across junior and collegiate divisions, particularly in early-stage competitions where players transition into new competitive environments. This study provides a comprehensive evaluation of WTN and UTR for junior and collegiate talent development, updating prior literature on rating classification accuracy and establishing a new benchmark for assessing future algorithm changes.

**KEYWORDS:** World Tennis Number (WTN), Universal Tennis Rating (UTR), Junior Tennis, Match Forecasting

*\*Corresponding Author:* Fuqua School of Business, Duke University, 100 Fuqua Drive, Durham, NC, USA Phone: 919-660-7781, email: wmayew@duke.edu

## INTRODUCTION

On September 11, 2024, the International Tennis Federation (ITF) introduced enhancements to the World Tennis Number (WTN). This WTN update incorporated match results from over 80 nations, drawing on more than 30 million recorded matches, alongside feedback from players, coaches, and national federations (ITF, 2024). The purpose of this paper is to evaluate the ability of the newly updated WTN to predict match outcomes relative to the Universal Tennis Rating (UTR).

Prior research investigating the ability of WTN and UTR to predict match outcomes has generally documented that WTN and UTR are equivalent predictors of match outcomes (Mayew & Mayew, 2023; Krall et al., 2024; Im & Lee, 2023). However, these studies focus exclusively on matches from the United States Tennis Association (USTA) Junior National Championships prior to the 2024 WTN algorithm change, making it unclear whether existing insights regarding match prediction are generalizable. The USTA Junior National Championships is the premier junior tournament for players in the United States, spanning ages 16U and 18U for both boys and girls. To increase external validity, we analyze premier non-professional tournaments beyond the USTA Junior National Championships to assess WTN and UTR predictive accuracy in samples that are not confined to the U.S. players and in age ranges that come both before and after the age ranges of the USTA Junior National Championship.

For players older than the USTA Junior National Championship sample, we study the 2024 International Tennis Association (ITA) Division 1 Men's and Women's All-American Championships. For players younger than the USTA Junior National Championship sample, we study tournaments across USTA, Lawn Tennis Association (LTA), and Tennis Europe (TE) sanctioned events including the 2024 Lexus British Indoor National Championships, 2024 Junior

Orange Bowl, 2025 Les Petits As Mondial Wilson, and 2025 Open Super 12 Auray. These tournaments are highly selective which helps ensure the players have extensive match histories upon which a WTN or UTR could be based.

Updating the academic record regarding the predictive accuracy of WTN and UTR is important as tennis organizations around the world debate whether to adopt WTN or UTR as the preferred rating system and as WTN becomes more widely used for tournament entry and seeding criteria. Whether WTN or UTR will predict match outcomes differentially in our sample is an open empirical question. Given the ITF’s stated expectation that WTN algorithm change effects should be most pronounced in junior divisions (ITF, 2024), we may observe superior WTN classification accuracy in 12U and 14U events. At the collegiate level, the Intercollegiate Tennis Association (ITA) adopted WTN as its official rating system in 2023 (ITA, 2023) but the Hall of Fame former coach of the six-time national champion Georgia Bulldogs stated the adoption was a mistake and UTR should be used in college (Diaz, 2024). This may imply that UTR is superior at the college level.

## METHODS

### Sample

We obtain match and player information by reviewing publicly displayed match draws on the ITA, USTA, LTA and TE websites for the tournaments summarized in Table 1.

**Table 1: Sample Tournaments**

<b>Federation</b>	<b>Age Group</b>	<b>Tournament</b>	<b>Category</b>	<b>Gender</b>
ITA	College	2024 ITA All-American Championships	NCAA Div. 1	M / F
USTA	12U, 14U	2024 Junior Orange Bowl	Level 2	M / F
USTA	12U, 14U	2024 IMG Academy International Championships	Level 2	M / F
LTA	12U*, 14U	2024 Lexus Junior Indoor National Championships	Grade 1	M / F
LTA	12U*	2024 Lexus Junior National Tour	Grade 2	M / F
LTA	14U	National Age Group Matchplay	N/A	M

LTA	12U*, 14U	2025 Lexus Junior National Tour	Grade 2	M / F
LTA	14U	2025 Lexus Clay Court Championships	Grade 1	M / F
TE	14U	2025 Lexus Junior International Bolton	Cat 1	M / F
TE	14U	2025 Les Petits As Mondial Wilson	Super Cat	M / F
TE	14U	Kungens Kanna & Drottningens Pris	Super Cat	M / F
TE	14U	Tim Essonne	Cat 1	M / F
TE	14U	Lexus Junior International Liverpool	Cat 3	M / F
TE	12U	Open Super 12 Auray	Cat 1	M / F
TE	12U	Lexus Junior International Nottingham	Cat 2	M / F

\*This tournament also has an 11U division which is combined with 12U for analysis.

These tournaments result in a sample of 4,174 potential matches, including main draw, qualifying, and consolation events for both male and female players, where applicable. After excluding matches affected by byes, pre-match withdrawals, retirements, missing UTR or WTN values for either player, and matches where both players had identical WTN or UTR ratings, the final sample consisted of 3,727 matches involving 3,020 players.

Player WTN and UTR ratings were recorded on the day before the tournament began. UTR values were sourced through a power subscription to the UTR Sports App (UTR, 2023), while WTN ratings were obtained from the ITF WTN rating website (ITF, 2023) or official tournament federation websites. The collected WTN values reflect the updated 2024 algorithm.

### **Statistical Analysis**

For each match played, we follow Im & Lee (2023) and randomly choose one player from the match as the reference player. We then regress an indicator variable for whether the randomly chosen reference player won the match on the difference between the reference player's rating versus the rating of the other player. Separate bivariate logistic regressions are estimated for UTR and WTN ratings using STATA/SE 18.5 statistical software. From each logistic regression we obtain two measures of classification accuracy utilized in the tennis match forecasting literature: the area under the receiver operator characteristic curve (AUC) (Im & Lee 2023; Mayew and Mayew 2023, Krall et al 2024), and the Brier score (Boulier and Stekler,

1999; del Corral and Prieto-Rodriguez, 2010; Mayew and Mayew 2023, Krall et al 2024). An AUC (Brier score) value of 0.50 (0.25) represents random chance levels of classification accuracy, and an AUC (Brier score) of 1.00 (0.00) represents perfect classification accuracy. We statistically test for equality between WTN AUC ( $AUC_{WTN}$ ) and UTR AUC ( $AUC_{UTR}$ ), as well as WTN Brier score ( $BRIER_{WTN}$ ) and UTR Brier score ( $BRIER_{UTR}$ ). We consider p-values  $< 0.05$  to be statistically significant.

## RESULTS

In Panel A of Table 2, we provide WTN and UTR values for players in the 2024 ITA All-American Championships overall, followed by a decomposition by both gender and draw stage. Consistent with prior research for similarly aged players, males have higher (lower) UTR (WTN) values than females (Krall et al. 2024; Mayew and Mayew 2023). Additionally, the main draw players have higher (lower) UTR (WTN) than the pre-qualifying and qualifying draw, consistent with higher quality players making it further in the tournament than lower quality players.

In Panel B we tabulate how often the favored player wins the match. Overall, we find the favored player based on WTN wins 59.83% of the time. This is statistically smaller than what is observed for UTR, where the favored player wins between 66.50% of the time. These proportions differ from the existing literature examining the USTA Junior National Championships, where the favored player won about 75% of the time regardless of whether WTN or UTR was used (Krall et al. 2024; Mayew and Mayew 2023). These proportions are also smaller than the proportions ranging from 71.2% and 81.8% in professional tournaments (del Corral and Prieto-Rodriguez, 2010; Boulier and Stekler, 1999).

**Table 2. Analysis of 2024 ITA All-American Championships**

**Panel A. Player Level Descriptive Statistics**

	Overall		Men		Women		Main Draw		Pre-Qual/Qual	
	(N=438)		(n=220)		(n=218)		(n=125)		(n=346)	
VARIABLE	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD
<i>WTN</i>	9.97	3.80	6.62	1.91	13.35	1.61	8.28	3.75	10.49	3.66
<i>UTR</i>	11.51	1.45	12.88	0.44	10.12	0.45	11.94	1.42	11.38	1.43

**Panel B. How Often Favored Players Win Matches**

	Overall		Men		Women		Main Draw		Pre-Qual/Qual	
	(N=585)		(n=298)		(n=287)		(n=205)		(n=380)	
	#	%	#	%	#	#	#	%	#	%
<i>FAVORED PLAYER WINS<sub>WTN</sub></i>	350	59.83%	175	58.72%	175	60.98%	127	61.95%	223	58.68%
<i>FAVORED PLAYER WINS<sub>UTR</sub></i>	389	66.50%	193	64.77%	196	68.29%	128	62.44%	261	68.68%
<i>P-value test of equal proportions<sup>+</sup></i>	0.0003		0.0133		0.0125		1.0000		<0.0001	

**Panel C. Match Outcome Prediction Analysis Based on Logistic Regression Results**

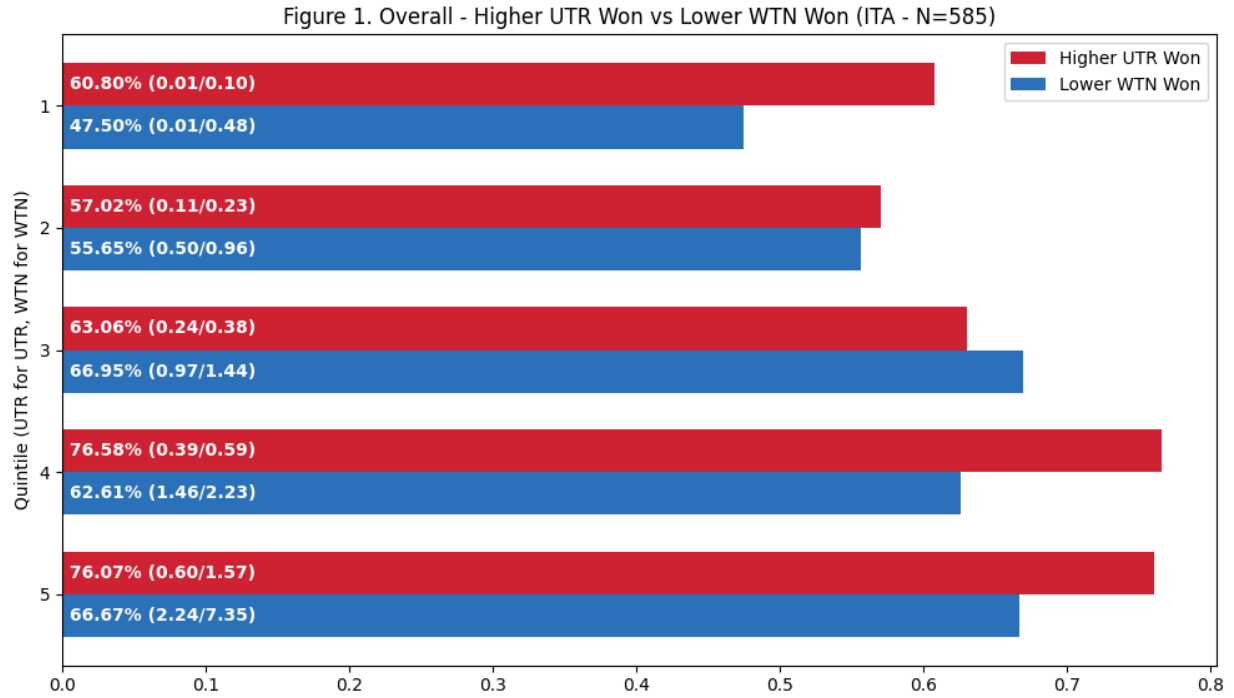
	Overall		Men		Women		Main Draw		Pre-Qual/Qual	
	(N=585)		(n=298)		(n=287)		(n=205)		(n=380)	
VARIABLE	EST	95% CI	EST	95% CI	EST	95% CI	EST	95% CI	EST	95% CI
<i>AUC<sub>WTN</sub></i>	0.6362	0.5917 0.6807	0.6346	0.5718 0.6975	0.6380	0.5749 0.7012	0.6554	0.5805 0.7303	0.6282	0.5714 0.6849
<i>AUC<sub>UTR</sub></i>	0.7038	0.6633 0.7444	0.6901	0.6281 0.7520	0.7150	0.6575 0.7725	0.6688	0.5938 0.7438	0.7272	0.6769 0.7776
<i>P-value of test<sup>++</sup>: AUC<sub>WTN</sub> = AUC<sub>UTR</sub></i>	<0.0001		0.0053		0.0008		0.5218		<0.0001	
<i>BRIER<sub>WTN</sub></i>	0.2361	0.2273 0.2449	0.2352	0.2225 0.2478	0.2364	0.2239 0.2489	0.2316	0.2154 0.2479	0.2382	0.2277 0.2487
<i>BRIER<sub>UTR</sub></i>	0.2183	0.2062 0.2305	0.2180	0.2001 0.2359	0.2183	0.2010 0.2355	0.2265	0.2078 0.2453	0.2106	0.1940 0.2272
<i>P-value of test<sup>+++</sup>: BRIER<sub>WTN</sub> = BRIER<sub>UTR</sub></i>	<0.0001		0.0019		0.0028		0.3267		<0.0001	

Table 2 presents data for 438 sample players in Panel A and 585 sample matches in Panels B and C from the 2024 ITA All-American Championships. UTR is the UTR value as of the start of the tournament, which ranges from 1.00 (lowest skill) to 16.50 (highest skill). WTN is the WTN value as of the start of the tournament and ranges from 40.00 (lowest skill) to 1.00 (highest skill). FAVORED PLAYER WINS<sub>WTN(UTR)</sub> indicates the player with the lower (higher) WTN (UTR) won the match. AUC<sub>WTN</sub> and BRIER<sub>WTN</sub> represent the area under the receiver operator curve and Brier score, respectively, derived from a bivariate logistic regression where the dependent variable is an indicator for whether a randomly chosen reference player from the pair wins the match and the independent variable is the reference player WTN minus the WTN of the other player. AUC<sub>UTR</sub> and BRIER<sub>UTR</sub> represent the area under the receiver operator curve and Brier score, respectively, derived from a bivariate logistic regression where the dependent variable is an indicator for whether a randomly chosen reference player from the pair wins the match and the independent variable is the reference player UTR minus the UTR of the other player. AUC (Brier score) values of 0.50 (0.25) represent chance levels of classification accuracy and 1.00 (0.00) represent perfect classification accuracy. EST is the derived estimate, and 95% CI indicates the 95% confidence interval using bootstrap method. All reported p-values are two tailed. <sup>+</sup>McNemar test of equal proportions; <sup>++</sup>DeLong et al. (1998) test of equal areas called via the roccomp command in STATA; <sup>+++</sup>Paired t-test.

We observe similar inferences as the overall sample when we consider matches by gender. The favored player based on UTR wins more often than the favored player based on WTN, suggesting the overall sample proportion differences do not appear to be driven by gender.

However, when we partition by draw stage, we find the overall difference in predictive accuracy is driven by the pre-qualifying and qualifying matches, not main draw matches. In the main draw, the favored player based on WTN and UTR wins 61.95% and 62.44% of the time respectively and these proportions do not statistically differ ( $p\text{-value} = 1.00$ ). However, in the pre-qualifying and qualifying draw, the favored player based on WTN wins the match 58.68% of the time, which is statistically smaller than the 68.68% based on UTR ( $p\text{-value} < 0.0001$ ).

Proportions only consider the sign of the rating difference between players, not the magnitude. To ascertain whether these patterns hold when we formally consider the magnitude of the difference between player ratings, we perform two different analyses. First, we sort the entire sample by the absolute difference in UTR and partition the sorted sample into quintiles from smallest differences to largest differences and report the percentage of the time the favored player wins. We repeat the exercise using WTN and display the results in Figure 1. Each quintile therefore is comprised of approximately 117 matches. We find that in four of the five quintiles, UTR outperforms WTN, suggesting that the overall outperformance of UTR over WTN is consistent across most quintiles, with UTR outperforming WTN in the first, second, fourth, and fifth quintiles, while WTN slightly outperforms UTR in the third quintiles.



In Figure 1 the x-axis shows the proportion of matches won by the player with the higher (lower) UTR (WTN). Each quintile of the absolute difference in rating between players is displayed from lowest (quintile 1) to highest (quintile 5) on the y-axis, with the minimum and maximum values of the absolute rating difference range indicated on each bar.

A second method to consider the magnitude of the difference in rating is to estimate logistic regressions where the dependent variable equals 1 if a randomly chosen player from the pair wins the match as a function of the magnitude of the rating difference between that randomly chosen player and the opponent. Such regressions allow for the derivation of AUC and Brier scores to assess classification accuracy. In Panel C, overall, we observe AUC values ( $AUC_{UTR} = 0.704$  versus  $AUC_{WTN} = 0.636$ , p-value =  $<0.001$ ) and Brier scores ( $BRIER_{UTR} = 0.218$  versus  $BRIER_{WTN} = 0.236$  p-value =  $<0.001$ ) that statistically differ in a manner consistent with UTR offering better classification accuracy than WTN. As with the proportions in Panel B, both AUC and Brier scores show statistically stronger classification accuracy for UTR over WTN for both males and females. Also consistent with the proportions in Panel B, we find WTN and UTR exhibit statistically equivalent classification accuracy in the main draw using either AUC ( $AUC_{UTR} = 0.669$  versus  $AUC_{WTN} = 0.655$ , p-value =  $0.522$ ) or Brier scores



( $BRIER_{UTR} = 0.227$  versus  $BRIER_{WTN} = 0.232$ ,  $p\text{-value} = 0.327$ ). In pre-qualifying and qualifying matches, however, UTR exhibits superior classification accuracy over WTN using both AUC ( $AUC_{UTR} = 0.727$  versus  $AUC_{WTN} = 0.628$ ,  $p\text{-value} = <0.001$ ) and Brier scores ( $BRIER_{UTR} = 0.211$  versus  $BRIER_{WTN} = 0.238$ ,  $p\text{-value} = <0.001$ ).

Thus far, we have only considered gender and draw stage separately. To ascertain whether there are interactive effects we examine subsamples conditioned by both draw stage and gender in Table 3. Panel B reveals that the proportion of matches won by the favored player based on WTN and UTR is statistically identical for both men and women in the main draw but statistically differs in the pre-qualifying and qualifying draw. The difference in proportions is similar in magnitude at 9.23% for men and 10.81% for women. Similar inferences are obtained using AUC and Brier Scores in Panel C. We therefore conclude that for the ITA sample, draw stage but not gender is responsible for the overall superiority of UTR over WTN for predicting match outcomes.

**Table 3. Analysis of 2024 ITA All-American Championships (Draw Stage x Gender)**

**Panel A. Player Level Descriptive Statistics**

	MD Men		MD Women		PQQ Men		PQQ Women	
	(n=63)		(n=62)		(n=174)		(n=172)	
VARIABLE	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD
<i>WTN</i>	4.89	1.68	11.72	1.35	7.16	1.62	13.86	1.29
<i>UTR</i>	13.30	0.37	10.56	0.35	12.76	0.36	9.99	0.38

**Panel B. How Often Favored Players Win Matches**

	MD Men		MD Women		PQQ Men		PQQ Women	
	(n=103)		(n=102)		(n=195)		(n=185)	
	#	%	#	%	#	%	#	%
<i>FAVORED PLAYER WINS<sub>WTN</sub></i>	63	61.17%	64	62.75%	112	57.44%	111	60.00%
<i>FAVORED PLAYER WINS<sub>UTR</sub></i>	63	61.17%	65	63.73%	130	66.67%	131	70.81%
<i>P-value test of equal proportions+</i>	1.0000		1.0000		0.0039		0.0045	

**Panel C. Match Outcome Prediction Analysis Based on Logistic Regression Results**

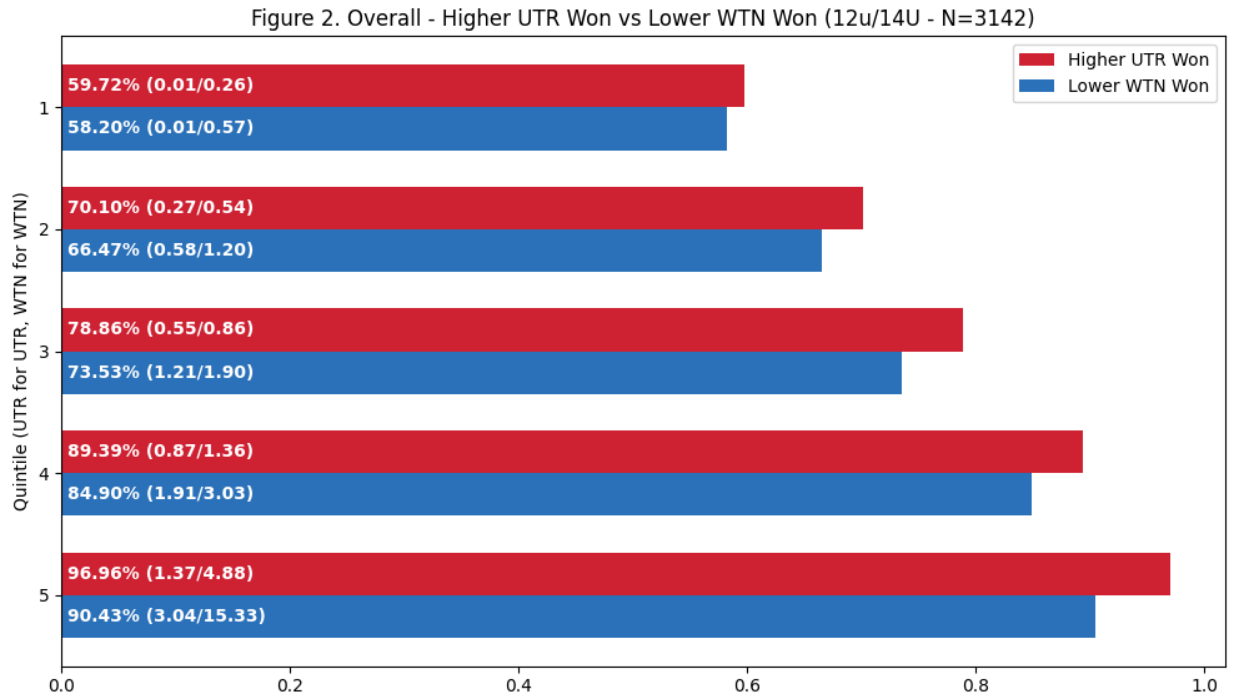
	MD Men		MD Women		PQQ Men		PQQ Women	
	(n=103)		(n=102)		(n=195)		(n=185)	
<b>VARIABLE</b>	<b>EST</b>	<b>95% CI</b>	<b>EST</b>	<b>95% CI</b>	<b>EST</b>	<b>95% CI</b>	<b>EST</b>	<b>95% CI</b>
$AUC_{WTN}$	0.6538	0.5412 0.7663	0.6576	0.5504 0.7648	0.6272	0.5486 0.7058	0.6301	0.5484 0.7118
$AUC_{UTR}$	0.6701	0.5625 0.7777	0.6502	0.5434 0.7570	0.7036	0.6304 0.7767	0.7506	0.6809 0.8203
$P\text{-value of test}^{++}: AUC_{WTN}=AUC_{UTR}$	0.5795		0.799		0.0048		0.0003	
$BRIER_{WTN}$	0.2331	0.2094 0.2569	0.2302	0.2055 0.2548	0.2363	0.2207 0.2518	0.2391	0.2245 0.2536
$BRIER_{UTR}$	0.2205	0.1907 0.2503	0.2317	0.2081 0.2553	0.2153	0.1929 0.2376	0.2048	0.1799 0.2296
$P\text{-value of test}^{+++}: BRIER_{WTN}=BRIER_{UTR}$	0.1506		0.7963		0.0045		0.0006	

Table 3 presents data for 438 sample players in Panel A and 585 sample matches in Panels B and C from the 2024 ITA All-American Championships. UTR is the UTR value as of the start of the tournament, which ranges from 1.00 (lowest skill) to 16.50 (highest skill). WTN is the WTN value as of the start of the tournament and ranges from 40.00 (lowest skill) to 1.00 (highest skill). FAVORED PLAYER  $WINS_{WTN(UTR)}$  indicates the player with the lower (higher) WTN (UTR) won the match.  $AUC_{WTN}$  and  $BRIER_{WTN}$  represent the area under the receiver operator curve and Brier score, respectively, derived from a bivariate logistic regression where the dependent variable is an indicator for whether a randomly chosen reference player from the pair wins the match and the independent variable is the reference player WTN minus the WTN of the other player.  $AUC_{UTR}$  and  $BRIER_{UTR}$  represent the area under the receiver operator curve and Brier score, respectively, derived from a bivariate logistic regression where the dependent variable is an indicator for whether a randomly chosen reference player from the pair wins the match and the independent variable is the reference player UTR minus the UTR of the other player. AUC (Brier score) values of 0.50 (0.25) represent chance levels of classification accuracy and 1.00 (0.00) represent perfect classification accuracy. EST is the derived estimate, and 95% CI indicates the 95% confidence interval using bootstrap method. All reported p-values are two tailed. +McNemar test of equal proportions; ++DeLong et al. (1998) test of equal areas called via the roccomp command in STATA; +++Paired t-test.

Table 4 extends this ITA analysis to elite junior development tournaments (LTA, USTA and TE). Similarly to the ITA data and consistent with prior research for similarly aged players, panel A shows that males have higher (lower) UTR (WTN) values than females (Krall et al. 2024; Mayew and Mayew 2023). We find in Panel B that across 3,142 junior matches, the WTN favored player wins 74.57% of the time, while UTR favored players win 78.87% (p-value < 0.001). The magnitudes of these proportions are in a range similar to the USTA Junior National Championships and to professional tennis results (Mayew & Mayew 2023).

Figure 2 displays the percentage of the time the favored player wins by quintile of the absolute difference ratings between opponents. Each quintile contains approximately 628 matches. In all quintiles, UTR outperforms WTN, suggesting the overall superiority of UTR is not concentrated in any particular part of the rating difference distribution. Moreover, the

proportion of the time the favored player wins generally decreases as the magnitude of the difference in rating decreases, similar to the results documented in Im & Lee (2023).



In Figure 2, the x-axis shows the proportion of matches won by the player with the higher (lower) UTR (WTN). Each quintile of the absolute difference in rating between players is displayed from lowest (quintile 1) to highest (quintile 5) on the y-axis, with the minimum and maximum values of the absolute rating difference range indicated on each bar.

Considering the magnitude of rating differences via logistic regression (Table 4, Panel C) also suggests UTR superiority. UTR significantly outperforms WTN overall (AUC: 0.868 vs. 0.817,  $p$ -value  $< 0.001$ ; Brier Score: 0.148 vs. 0.175,  $p$ -value  $< 0.001$ ). We also find UTR statistically outperforms WTN in each subsample, regardless of whether we partition by gender or by age.

**Table 4. Analysis of 2024/25 World Junior Development Tournaments**

**Panel A. Player Level Descriptive Statistics**

VARIABLE	Overall		Boys		Girls		12U		14U	
	(N=2,582)		(n=1,329)		(n=1,253)		(n=1,045)		(n=1,537)	
	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD
WTN	28.52	3.17	28.01	2.90	29.05	3.36	30.79	2.34	26.97	2.70
UTR	7.03	1.74	7.56	1.66	6.46	1.65	5.72	1.44	7.91	1.32

**Panel B. How Often Favored Players Win Matches**

	Overall		Boys		Girls		12U		14U	
	(N=3,142)		(n=1,615)		(n=1,527)		(n=1,357)		(n=1,785)	
	#	%	#	%	#	#	#	%	#	%
<i>FAVORED PLAYER WINS</i> <sub>WTN</sub>	2,343	74.57%	1,188	73.56%	1,155	75.64%	1,050	77.38%	1,293	72.44%
<i>FAVORED PLAYER WINS</i> <sub>UTR</sub>	2,478	78.87%	1,279	79.20%	1,199	78.52%	1,096	80.77%	1,382	77.42%
<i>P-value test of equal proportions</i> <sup>+</sup>	<0.0001		<0.0001		0.0028		0.0021		<0.0001	

### Panel C. Match Outcome Prediction Analysis Based on Logistic Regression Results

	Overall		Boys		Girls		12U		14U	
	(N=3,142)		(n=1,615)		(n=1,527)		(n=1,357)		(n=1,785)	
<b>VARIABLE</b>	<b>EST</b>	<b>95% CI</b>	<b>EST</b>	<b>95% CI</b>	<b>EST</b>	<b>95% CI</b>	<b>EST</b>	<b>95% CI</b>	<b>EST</b>	<b>95% CI</b>
<i>AUC</i> <sub>WTN</sub>	0.8167	0.8019 0.8315	0.8096	0.7896 0.8295	0.8243	0.8044 0.8442	0.8467	0.8263 0.8671	0.7953	0.7753 0.8153
<i>AUC</i> <sub>UTR</sub>	0.8677	0.8552 0.8803	0.8669	0.8498 0.8840	0.8687	0.8520 0.8855	0.8882	0.8710 0.9054	0.8510	0.8339 0.8681
<i>P-value of test</i> <sup>++</sup> : <i>AUC</i> <sub>WTN</sub> = <i>AUC</i> <sub>UTR</sub>	<0.0001		<0.0001		<0.0001		<0.0001		<0.0001	
<i>BRIER</i> <sub>WTN</sub>	0.1748	0.1679 0.1817	0.1780	0.1690 0.1871	0.1713	0.1617 0.1808	0.1591	0.1485 0.1698	0.1845	0.1760 0.1931
<i>BRIER</i> <sub>UTR</sub>	0.1481	0.1409 0.1554	0.1487	0.1389 0.1586	0.1474	0.1377 0.1571	0.1353	0.1245 0.1461	0.1575	0.1483 0.1667
<i>P-value of test</i> <sup>+++</sup> : <i>BRIER</i> <sub>WTN</sub> = <i>BRIER</i> <sub>UTR</sub>	<0.0001		<0.0001		<0.0001		<0.0001		<0.0001	

Table 4 presents data for 2,582 sample players in Panel A and 3,142 sample matches in Panels B and C from the Junior Development tournaments. UTR is the UTR value as of the start of the tournament, which ranges from 1.00 (lowest skill) to 16.50 (highest skill). WTN is the WTN value as of the start of the tournament and ranges from 40.00 (lowest skill) to 1.00 (highest skill). *FAVORED PLAYER WINS*<sub>WTN(UTR)</sub> indicates the player with the lower (higher) WTN (UTR) won the match. *AUC*<sub>WTN</sub> and *BRIER*<sub>WTN</sub> represent the area under the receiver operator curve and Brier score, respectively, derived from a bivariate logistic regression where the dependent variable is an indicator for whether a randomly chosen reference player from the pair wins the match and the independent variable is the reference player WTN minus the WTN of the other player. *AUC*<sub>UTR</sub> and *BRIER*<sub>UTR</sub> represent the area under the receiver operator curve and Brier score, respectively, derived from a bivariate logistic regression where the dependent variable is an indicator for whether a randomly chosen reference player from the pair wins the match and the independent variable is the reference player UTR minus the UTR of the other player. AUC (Brier score) values of 0.50 (0.25) represent chance levels of classification accuracy and 1.00 (0.00) represent perfect classification accuracy. EST is the derived estimate, and 95% CI indicates the 95% confidence interval using bootstrap method. All reported p-values are two-tailed. <sup>+</sup>McNemar test of equal proportions; <sup>++</sup>DeLong et al. (1998) test of equal areas; <sup>+++</sup>Paired t-test.

Table 5 further decomposes the sample by both age and gender, revealing that UTR consistently outperforms WTN in all categories. The largest disparity also emerges in 14U Boys, where UTR provides significantly higher classification accuracy (AUC: 0.858 vs. 0.785, p-value < 0.001; Brier Score: 0.154 vs. 0.189, p-value < 0.001), whereas 12U Girls (AUC: 0.900 vs. 0.849, p-value < 0.001; Brier Score: 0.128 vs. 0.157, p-value < 0.001) and 14U Girls (AUC: 0.844 vs. 0.807, p-value < 0.001; Brier Score: 0.161 vs. 0.180, p-value < 0.001) show smaller magnitude but still statistically significant differences.

**Table 5. Analysis of 2024/25 World Junior Development Tournaments (Age Group + Gender)****Panel A. Player Level Descriptive Statistics**

VARIABLE	12U Boys		12U Girls		14U Boys		14U Girls	
	(n=536)		(n=509)		(n=793)		(n=744)	
	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD
<i>WTN</i>	30.09	2.04	31.53	2.40	26.61	2.53	27.35	2.82
<i>UTR</i>	6.23	1.31	5.19	1.37	8.45	1.21	7.34	1.19

**Panel B. How Often Favored Players Win Matches**

	12U Boys		12U Girls		14U Boys		14U Girls	
	(n=694)		(n=663)		(n=921)		(n=864)	
	#	%	#	%	#	%	#	%
<i>FAVORED PLAYER WINS<sub>WTN</sub></i>	530	76.37%	520	78.43%	658	71.44%	635	73.50%
<i>FAVORED PLAYER WINS<sub>UTR</sub></i>	556	80.12%	540	81.45%	723	78.50%	659	76.27%
<i>P-value test of equal proportions<sup>+</sup></i>	0.0178		0.0619		<0.0001		0.0236	

**Panel C. Match Outcome Prediction Analysis Based on Logistic Regression Results**

VARIABLE	12U Boys		12U Girls		14U Boys		14U Girls	
	(n=694)		(n=663)		(n=921)		(n=864)	
	EST	95% CI	EST	95% CI	EST	95% CI	EST	95% CI
<i>AUC<sub>WTN</sub></i>	0.8436	0.8159 0.8713	0.8494	0.8201 0.8788	0.7850	0.7544 0.8147	0.8069	0.7780 0.8358
<i>AUC<sub>UTR</sub></i>	0.8779	0.8530 0.9029	0.8998	0.8772 0.9224	0.8581	0.8347 0.8816	0.8438	0.8177 0.8700
<i>P-value of test<sup>++</sup>: AUC<sub>WTN</sub> = AUC<sub>UTR</sub></i>	0.0037		<0.0001		<0.0001		<0.0001	
<i>BRIER<sub>WTN</sub></i>	0.1612	0.1467 0.1756	0.1570	0.1417 0.1724	0.1890	0.1768 0.2012	0.1796	0.1665 0.1926
<i>BRIER<sub>UTR</sub></i>	0.1420	0.1269 0.1570	0.1280	0.1130 0.1430	0.1537	0.1407 0.1667	0.1611	0.1474 0.1748
<i>P-value of test<sup>+++</sup>: BRIER<sub>WTN</sub> = BRIER<sub>UTR</sub></i>	0.0034		<0.0001		<0.0001		<0.0001	

Table 5 presents data for 2,582 sample players in Panel A and 3,142 sample matches in Panels B and C from the Junior Development tournaments. UTR is the UTR value as of the start of the tournament, which ranges from 1.00 (lowest skill) to 16.50 (highest skill). WTN is the WTN value as of the start of the tournament and ranges from 40.00 (lowest skill) to 1.00 (highest skill). FAVORED PLAYER WINS<sub>WTN(UTR)</sub> indicates the player with the lower (higher) WTN (UTR) won the match. AUC<sub>WTN</sub> and BRIER<sub>WTN</sub> represent the area under the receiver operator curve and Brier score, respectively, derived from a bivariate logistic regression where the dependent variable is an indicator for whether a randomly chosen reference player from the pair wins the match and the independent variable is the reference player WTN minus the WTN of the other player. AUC<sub>UTR</sub> and BRIER<sub>UTR</sub> represent the area under the receiver operator curve and Brier score, respectively, derived from a bivariate logistic regression where the dependent variable is an indicator for whether a randomly chosen reference player from the pair wins the match and the independent variable is the reference player UTR minus the UTR of the other player. AUC (Brier score) values of 0.50 (0.25) represent chance levels of classification accuracy and 1.00 (0.00) represent perfect classification accuracy. EST is the derived estimate, and 95% CI indicates the 95% confidence interval using bootstrap method. All reported p-values are two-tailed. <sup>+</sup>McNemar test of equal proportions; <sup>++</sup>DeLong et al. (1998) test of equal areas; <sup>+++</sup>Paired t-test.

**DISCUSSION**

This study evaluates the predictive accuracy of the WTN following its 2024 algorithm update versus UTR, using match data from the ITA All-American Championships and elite

junior development tournaments across USTA, LTA, and Tennis Europe circuits. Our findings indicate that WTN does not significantly outperform UTR in predicting match outcomes at either the collegiate or junior development levels. Instead, except for the main draw stage at the collegiate level, UTR exhibits higher predictive accuracy than WTN.

We can only put forward conjectures as to why WTN ratings exhibit less ability to predict match outcomes versus UTR. The results from the ITA All-American Championships may be driven by variation in the players country of origin. Collegiate tennis is comprised of a substantial fraction of international players (Grabb, 2023). The differential predictive accuracy may, therefore, be driven by collegiate players from countries that are not fully onboarded to WTN and may experience inferior WTN performance relative to UTR. Any such effects may also be exacerbated by class standing, whereby, underclassmen have more of their prior record tied to their historical origin country record, which dissipates as the player engages in more and more matches at the college level over time. Future research might therefore consider player country of origin and the role of collegiate class standing in explaining the predictive accuracy differences we observe at the collegiate level.

At the junior level, since the oldest age group we analyze is 14U, one possible explanation for the inferiority of WTN relative to UTR in all of our analysis is the extent of match histories. UTR analyzes matches on a rolling one-year basis. WTN, on the other hand, considers full match histories, and so for young players, it may be that enough time has not passed to accumulate a sufficient match history. This potential problem is, of course, exacerbated if a junior player is from a geographic location not onboarded fully to WTN.

The largest discrepancy occurs in the 14U boys' division, where UTR classification accuracy superiority is 7.4% based on AUC. WTN may underperform particularly in this setting

as this is the age elite boys increasingly transition from national to international competition, facing a broader range of opponents outside their domestic circuits. This change in player network may inhibit the ability of WTN to predict match outcomes much like new collegiate players entering the ITA setting. Moreover, puberty begins to play a role with larger serve speed changes for boys versus girls around this age (Fernandez-Fernandez et al., 2019). If puberty also plays a role in transitioning a player to a new player network due to increased size and skill, WTN may take longer to adapt because of the longer match histories it considers relative to UTR. Future research might begin to consider whether differences in competitor physical attributes that occur during puberty explain the UTR and WTN predictive accuracy differential we observe, as has been considered at the professional level (del Corral and Prieto-Rodriguez, 2010). To date public data on physical characteristics of junior players is, unfortunately, not readily available.

While we find evidence that UTR outperforms WTN in predictive accuracy in the sample of elite tournaments we analyze, we cannot ascertain how UTR and WTN will perform in other settings. For example, UTR and WTN have repeatedly performed equivalently well in the USTA Junior National Championship two years in row (Mayew & Mayew 2023; Krall et al, 2024). If the reason for this equivalent predictive accuracy is the homogenous nature of the players analyzed, it may well be the case that ITF tournaments containing players primarily from the same country will also exhibit UTR and WTN predictive accuracy that does not differ. Future research should extend our analysis to other settings. Moreover, we acknowledge that WTN has been in existence for a shorter period of time relative to UTR. As future algorithm enhancements occur and more countries onboard the accuracy differences we observe may dissipate.

## CONCLUSION

We provide insights on whether the classification accuracy of WTN improved after the 2024 algorithm change using 3,522 matches from tennis events sanctioned by ITA, USTA, LTA and TE. We find UTR generally outperforms WTN in predicting match outcomes, potentially due to our elite level tournaments containing players that are integrating into new competitive structures and due to players from countries who have not fully onboarded to WTN.

## CONFLICT OF INTEREST AND FUNDING

The authors declare no conflicts of interest and did not receive funding to conduct this research.

## ACKNOWLEDGEMENTS

We thank members of the LTA, ITF and USTA for helpful discussions and for facilitating access to the WTN data. We also thank UTR Sports for helpful discussions and for facilitating access to the UTR data. We appreciate helpful comments from Will Mayew and Ian Mayew and data collection efforts from Nate Krall and Nick Maroulis. The views expressed in this paper are those of the authors and do not represent positions of the United States Tennis Association, International Tennis Federation, Lawn Tennis Association, Tennis Europe or Racquet Sports Professionals Association. This paper is based in part upon Lucas Kiely's Master's thesis, Evaluating the Predictive Accuracy of Tennis Rating Systems: A Comparative Study of National Ranking, Universal Tennis Rating (UTR), and World Tennis Number (WTN) in Junior Tennis Events in the United Kingdom, submitted to the University of Chicago Graham School. We express our gratitude to Dr. Abid Ali, Millie Rey and Tim Murphy for their guidance and insights during the research process.

## REFERENCES

Boulrier, B.L. and Stekler, H.O. (1999). Are sports seedings good predictors?: an evaluation. *International Journal of Forecasting* 15(1), 83-91. [https://doi.org/10.1016/S0169-2070\(98\)00067-3](https://doi.org/10.1016/S0169-2070(98)00067-3).



- del Corral, J. and Prieto-Rodriguez, J. (2010). Are differences in ranks good predictors for Grand Slam tennis matches? *International Journal of Forecasting*, 26(3), 551-563  
<https://doi.org/10.1016/j.ijforecast.2009.12.006>
- Diaz, M. (2024). X post regarding WTN usage in college tennis  
<https://x.com/CoachMannyDiaz/status/1867335497535959245>
- DeLong, E.R, DeLong, D.M, and Clarke-Pearson, D.L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* 44: 837-845. <https://doi.org/10.2307/2531595>
- Fernandez-Fernandez, J., Nakamura, F.Y., Moreno-Perez, V., Lopez-Valenciano, A., Del Coso, J., Gallo-Salazar, C., Barbado, D., Ruiz-Perez, I., & Sanz-Rivas, D. (2019). Age and sex-related upper body performance difference in competitive young tennis players. *PLOS ONE* 14(9): e0221761  
<https://doi.org/10.1371/journal.pone.0221761>
- Grabb, M. (2023). For Many International Plyers, College Tennis Is A Passport To The Pros. *The Stanford Daily*. <https://stanforddaily.com/2023/06/07/for-many-international-players-college-tennis-is-a-passport-to-the-pros/>
- Im, S and Lee, C.H. (2023). World Tennis Number: The new gold standard, or a failure? *ITF Coaching & Sport Science Review*, 32(91), 6-12. <https://doi.org/10.52383/itfcoaching.v32i91.371>
- Intercollegiate Tennis Association (ITA) (2023, January 5). Intercollegiate Tennis Association adopts ITF world tennis number as exclusive official rating for college tennis.  
<https://wearecollegetennis.com/2023/01/05/wtn-named-official-rating-of-college-tennis/>
- International Tennis Federation (ITF) (2023, n.d.). World Tennis Number Player Search  
<https://worldtennisnumber.com/eng/player-search>
- International Tennis Federation (ITF) (2024, August 19). Enhancement to the ITF World Tennis Number calculation <https://worldtennisnumber.com/eng/news/enhancement-to-the-itf-world-tennis-number-calculation/>
- Krall, N., Maroulis, N., Mayew, R. L., & Mayew, W. J. (2024). Initial Evidence on the Impact of the 2023 World Tennis Number Algorithm Change for Predicting Match Outcomes. *ITF Coaching & Sport Science Review*, 32(94):52-58.  
<https://doi.org/10.52383/itfcoaching.v32i94.559>
- Mayew, R. L., & Mayew, W. J. (2023). Which global tennis rating better measures player skill? Evidence from the 2022 USTA Junior National Championships. *The Sport Journal*, 26(e08252023-2)1-9.
- Universal Tennis Rating (UTR) (2023, n.d.) <https://app.utrsports.net/home>