

# Which global tennis rating better measures player skill? Evidence from the 2022 USTA Junior National Championships

Rebecca L. Mayew<sup>1</sup>, William J. Mayew<sup>2\*</sup>

April 10, 2023

## Abstract

Assessing relative player skill is important in many aspects of tennis. In 2008, the Universal Tennis Rating (UTR) was introduced as a global tennis player skill rating that put all players, regardless of gender, age or geographic location, on a common scale. The International Tennis Federation (ITF) recently launched a competitor rating called the World Tennis Number (WTN). We examine whether UTR or WTN ratings better predict head-to-head match success using 1,532 matches played by 870 participants at the 2022 United States Tennis Association (USTA) Junior National Championships. We observe overall classification accuracy of 73.9% and 70.4% for UTR and WTN ratings, respectively. Both exceed chance levels statistically and approximate the level observed for professional tennis bookmakers. UTR and WTN ratings are very highly correlated and classification accuracy does not statistically differ between ratings in the sample overall, by age division, by gender, by match format, or by the magnitude of player rating differences. These findings provide initial empirical evidence important to tennis organizations making rating adoption decisions, tennis coaches seeking play parity, tournament directors seeding players and college coaches screening potential recruits. We provide mapping functions between UTR and WTN ratings for situations where players have one rating but not the other.

**Keywords:** Universal Tennis Rating (UTR), World Tennis Number (WTN), Tennis player skill, Match forecasting, Classification accuracy

---

<sup>1</sup> United States Professional Tennis Association (USPTA) Professional Member, Cary, North Carolina, United States of America

<sup>2</sup> Fuqua School of Business, Duke University, Durham, North Carolina, United States of America

\* Corresponding author  
Email: [william.mayew@duke.edu](mailto:william.mayew@duke.edu) (WJM)

# Introduction

Assessing relative player skill is important in many aspects of sports administration. Player rankings serve as one popular tool for assessing whether a given player is more skilled than another. In professional tennis, rankings are used for tournament selection and seeding to help ensure competitive balance. Competitive matches are more entertaining because the outcome is more uncertain, and spectators are known to have preferences for uncertainty in match outcomes [1]. Player rankings for touring tennis professionals are provided by the Association of Tennis Professionals for men [2] and Women's Tennis Association for women [3] based on match histories, and these rankings have been extensively studied in the literature [4].

At the junior (i.e. pre-professional 18 years old and under) level, player rankings also exist, but not for the purpose of maximizing spectator enjoyment. Rather, at the junior level rankings serve the purpose of ensuring fairer matchups and facilitating player skill development. For junior players in the United States wanting to play competitively, the United States Tennis Association (USTA) offers gender and age specific rankings at the district, sectional and national level, where ranking points accumulate from playing USTA sanctioned events [5]. An inherent limitation of player rankings at the junior level, however, is that relative skill comparisons can only be made among players that are part of the same ranking. Consider a tennis coach interested in maximizing the development of a 16-year-old boy with a USTA national ranking of #50 (where 1 is the highest) and is seeking a comparably skilled practice partner. Locally available practice partners include a 16-year-old girl with a national USTA ranking of #5, an 18-year-old boy with a USTA national ranking of #100, a 16-year-old boy with a USTA national ranking of #200, and a 16-year-old boy without a USTA national ranking at all. Which player should the coach choose?

This question is difficult to answer because rankings are not comparable across genders (the 18-year-old girl) or age divisions (the 18-year-old boy). The 16-year-old boy with a national ranking of #200 may not necessarily be a less skilled practice partner. If he does not exclusively play USTA events to obtain ranking points, but rather plays both USTA events and high school tennis where no USTA ranking points are available, the relatively low national ranking may be deceiving. Further still, the 16-year-old boy without a USTA national ranking may be an exceptionally skilled player but not have a USTA ranking because he exclusively plays International Tennis Federation (ITF) tournaments globally rather than USTA events in the United States.

To overcome these inherent limitations of rankings and facilitate skill comparisons more universally, a junior tennis coach invented and introduced the Universal Tennis Rating (UTR) in 2008. UTR ratings are designed to promote fair and competitive play by providing a measure of player skill on a common scale regardless of age, gender or geographic location [6]. Owned by Universal Tennis, LLC, UTR ratings have been widely adopted, with more than 2.3 million UTR ratings for players spanning roughly 125 countries [7]. In 2020 Universal Tennis, LLC, announced its own world tour where players qualify and are seeded based on their UTR rating [8].

Despite the existence and popularity of UTR ratings, we know of no research investigating how well UTR ratings capture player skill. Moreover, in 2019, the ITF announced it was developing a competitor rating called the World Tennis Number (WTN) [9]. In 2022, the ITF began using WTN as part of its official acceptance criteria for tournaments on the ITF junior world tour circuit [10]. Additionally the USTA officially adopted WTN, as opposed to UTR, for use in junior tournaments [11].

Given two competing global rating systems now exist, how does each rating perform as a measure of junior tennis player skill and is one rating superior to the other? The purpose of this paper is to provide initial empirical evidence to answer this question. Universal Tennis, LLC, refers to its own UTR rating as “the world’s most accurate tennis rating system” in press releases [12] but we are aware of no existing empirical evidence that documents superiority of UTR ratings over WTN ratings. It is not possible to make theoretical predictions regarding how well UTR ratings or WTN ratings capture player skill based on the underlying algorithms as they are proprietary. Superiority of one rating over another cannot easily be inferred from adoption by tennis organizations thus far either. On the international front, Tennis Australia has adopted UTR as its official rating system [13] in contrast to the USTA’s adoption of WTN [14]. Adoption also varies domestically. Within the United States, junior players competing as part of their high school team use UTR as the official rating [7], which contrasts with junior players participating in USTA tournaments that use WTN. For United States college players, the Intercollegiate Tennis Association (ITA) partnered with Universal Tennis in 2017, referring to the partnership as affirming the UTR rating as “the one reliable metric anyone can use to gauge a tennis player’s skill” [15]. Nonetheless the ITA switched its partnership to the ITF in 2023, making WTN the new official rating system for college tennis players in the United States [16].

Answering our research question is important to organizations that oversee junior tennis, junior tennis tournament directors, junior coaches interested in skill development and college coaches who recruit junior tennis players. For our empirical strategy, we assume higher skilled players are more likely to win a match in head-to-head competition and assess whether UTR or WTN ratings more accurately identify match success.

## **Materials and methods**

## Sample

Our analysis is based on match results from the 2022 USTA Junior National Championships. We choose this particular tournament for five reasons. First, we choose a junior tournament as opposed to a professional tournament to best capture the demographic over which ratings would be particularly useful for identifying skill differences. At the professional level, bookmaker odds data from the gambling market is readily available to assist in differentiating player skill in head-to-head competition. Such data is not available at the junior level. Additionally, junior players occupy a much wider range of the UTR and WTN rating distributions which facilitates more generalizable inferences. The extensive existing literature on tennis match outcomes focuses primarily on Grand Slam events [17-19], but Grand Slam participants are a small set of the very best players in the world and have ratings that reside in the extreme tails of overall UTR and WTN rating distributions.

Second, among junior tournaments, the Junior National Championships is a national level tournament as opposed to a district or sectional tournament. National tournaments allow for a more representative sample of junior tennis players in the United States. Additionally, qualifying for national tournaments requires players to have played and won more matches than are typically required for district or sectional tournaments. Because WTN for players in the United States rely in part upon past USTA matches, using a national tournament helps ensure a critical mass of matches is available for constructing the WTN rating.

Third, among all national tournaments held by the USTA, the Junior National Championships arguably offer the highest incentives for players to give their best effort because this is the only national tournament where the winner of the 18 and under (16 and under) division receives entrance into the main draw of the US Open (Junior US Open). Fourth, over 1,500

matches are available for analysis in this tournament, which should facilitate appropriate statistical power for our analysis given other published research studying professional Grand Slam matches has used samples as small as 999 [20] to 1,081 matches [17]. Finally, from a practical standpoint, WTN ratings were introduced in the United States in June of 2022 and this tournament offered one of the first opportunities to empirically examine WTN ratings.

The 2022 Junior National Championships were played in Kalamazoo, Michigan from August 5-14, 2022 for boys (USTA Tournament ID 22-59861) and in San Diego, California from August 6-14, 2022 for girls (USTA Tournament ID 22-97375). Each location hosted hard court matches in both a 16 years and under division (16u) and an 18 years and under division (18u). A 256 player draw is used for all age divisions, which implies 256 matches in the main draw inclusive of a match to determine third place played by the losers in the semifinal round. The consolation draw is comprised of 251 matches. Each age division at each location therefore has 507 matches, implying 2,028 matches overall. Additionally, while a 256 player draw is used, only 224 players are accepted for participation because the top 32 players are given first round byes. This implies the full sample of participants is 896 (224x4). To identify individual participant names and who won each match, we review the publicly available draw sheets by entering the Tournament ID into the USTA tournament website [21].

UTR ratings were obtained from player profile pages via a power subscription to the UTR App [22]. WTN ratings are obtained from player pages on the WTN rating website [23]. We collect UTR and WTN ratings as of the morning of August 5, 2022 Eastern Standard Time prior to any matches starting. The WTN rating was missing for 26 players. The missing players were comprised of two boys from the 16u division, 16 boys from the 18u division and 8 girls from the 18u division. Missing values pertain primarily to top ranked players who do not have

extensive USTA match histories because they have already begun playing professional level tournaments that are not operated by the USTA. We require non-missing UTR and WTN ratings for each player in order to make comparisons. We remove 6 matches where there is no determinable skill differential because either the WTN rating or UTR rating between players is exactly equal. Overall, these criteria result in a final sample of 1,532 matches and 870 unique players as summarized in Table 1.

## **UTR Ratings**

UTR ratings are owned and operated by Universal Tennis, LLC, and were originally developed by Dave Howell, a tennis coach interested in facilitating more level-based play to maximize junior tennis player development. A proprietary algorithm rates players on a scale between 1.00 (lowest skill) and 16.50 (highest skill) based on up to 30 match results from the last 12 months [24]. Players with values from 1-4 are referred to as beginners, 5-8 as having intermediate skills, 9-12 as having advanced skills and greater than 13 as professionals [25]. Match results considered when deriving a player's UTR rating can come from many sources and include, but are not limited to, USTA matches, ITF matches, high school matches and matches played as part of Universal Tennis events [26]. UTR ratings are available in real time and are updated daily. A player's UTR rating can increase by winning more games than expected in a given match. Given the pre-match UTR rating differential between players, the UTR algorithm expects a certain percentage of total games to be won by each player. The player who performs better than the algorithm's expectation will experience an increase in their UTR rating while the opposing player UTR rating will decrease. The algorithm's expected value is not disclosed to the players.

## **WTN Ratings**

WTN ratings are owned and operated by the ITF. Like UTR, WTN is a skill rating system based on a proprietary algorithm and is available in real time. Unlike UTR, WTN ratings range from 40.00 (lowest skill) to 1.00 (highest skill). Players with WTN ratings from 40-30 are considered novices, 30-20 as having intermediate skill, 20-10 as having advanced skill, 10-3 as top level amateurs, and 3-1 are considered professionals [27]. Also in contrast to UTR ratings, WTN ratings are updated weekly, and are based on match data from ITF events in addition to match data shared with the ITF by a player's national association from 2016 onward [28]. For example, in the United States, USTA match results are provided for inclusion in WTN ratings and in Great Britain, Lawn Tennis Association matches are provided for inclusion. Compared with UTR ratings, WTN ratings are derived using a longer match history, but the types of matches considered are more narrowly restricted as they come from the player's national association and ITF events only. A player's WTN rating can increase by obtaining set scores better than expected in each set played during a match. For a given difference in WTN ratings prior to the match, the WTN algorithm expects a certain match outcome. The WTN rating will increase for the player who performs better than the algorithm's expectation and decrease for the other player. Players do not observe the WTN algorithm expectations.

## **Results**

### **Descriptive statistics**

We provide descriptive statistics of the 870 sample players in Panel A of Table 2. We find the average sample UTR rating is 9.83 and ranges from 5.59 (lowest skill) to 13.82 (highest skill). The average WTN rating is 15.63, with a range between 26.06 (lowest skill) and 4.37 (highest skill). Together, under either UTR or WTN, the lowest skilled player in our sample is above the beginner/novice category, as one would expect for a prestigious national tournament.



The highest skilled players border on the high end of the amateur level and the low end of the professional level, which again is expected given the winner of the 18u division will be provided a main draw entrance into the US Open, which is a professional tournament. Overall, our sample can be characterized as capturing the entire rating range above novice and below professional for both UTR and WTN. The average player in our sample would be characterized as advanced by UTR standards and intermediate by WTN standards. Whether the subjective term “advanced” has a systematically different meaning than “intermediate” is not something we can assess. However, we can statistically examine the association between ratings. We find UTR and WTN ratings are highly negatively correlated ( $\rho = -0.92$ ,  $p\text{-value} < 0.001$ ). This extremely high correlation is suggestive evidence that UTR and WTN ratings will likely perform similarly as predictors of match outcomes.

In Panel B of Table 2, we provide descriptive statistics at the match level. We find that the favored player based on UTR ratings wins the match 75% of the time, very similar in magnitude to the 76% we observe for WTN ratings. These percentages fall within the range documented in the literature using professional players. The favored player won 71.2% of the time among Grand Slam matches studied in del Corral and Prieto-Rodriguez [18], and between 75.8% and 81.8% of the time in Boulier and Stekler [19].

## **Logistic regression analysis of match outcomes**

An extensive literature examines match outcomes of professional tennis players using paired comparison models. These studies commonly utilize regression analysis to estimate the likelihood that the favored player wins as a function of the skill difference between the two players competing [17-20,29]. The favored player and skill differences at the professional level are assessed using official rankings. We follow this literature but use ratings instead of rankings

and estimate logistic regressions as outlined below for match  $m$  with robust standard errors via the statistical software STATA/SE 17.0:

$$FAV\_UTR\_WIN_m = \alpha_0 + \alpha_1 * UTR\_DIFF_m + \varepsilon_m \quad (1a)$$

$$FAV\_WTN\_WIN_m = \beta_0 + \beta_1 * WTN\_DIFF_m + \gamma_m \quad (1b)$$

The dependent variable in equation 1a (1b),  $FAV\_UTR\_WIN$  ( $FAV\_WTN\_WIN$ ), is an indicator variable that equals 1 if the player favored to win per their UTR (WTN) rating wins the match and 0 otherwise. The independent variable in equation 1a (1b),  $UTR\_DIFF$  ( $WTN\_DIFF$ ), is the difference in the UTR (WTN) rating between the favored player and the unfavored player.

Recall that higher (lower) UTR (WTN) values indicate more skill. Therefore, as  $UTR\_DIFF$  ( $WTN\_DIFF$ ) becomes more positive (negative) the skill difference between the favored and unfavored player increases, which should translate into a higher likelihood of the favored player winning the match. As a result, observing  $\alpha_1 > 0$  ( $\beta_1 < 0$ ) when estimating equation 1a (1b) would indicate that UTR (WTN) rating differences can help predict match outcomes. However, the magnitude of the  $\alpha_1$  ( $\beta_1$ ) coefficient in equation 1a (1b) does not indicate how accurate UTR (WTN) ratings are, nor is  $\alpha_1$  comparable to  $\beta_1$  because UTR and WTN ratings are on difference scales.

As a result, we assess rating classification accuracy via the area under the receiver operator characteristic curve (AUC). In signal detection theory, the receiver operator characteristic (ROC) curve is designed to assess how well a given continuous predictor variable can discriminate between outcomes as well as facilitate a comparison among alternative predictors [30]. A ROC curve captures the probability of detecting a true signal against a false signal. Deriving the curve requires estimation of the true positive rate (i.e. sensitivity) and the false positive rate (i.e. 1- specificity) for a continuum of cutoff points,  $c$ , applied to the fitted

value from an estimated logistic regression. Fitted values from logistic regressions, such as those outlined in equations 1a and 1b, are bounded between zero and one by construction. Suppose  $c = 0.75$ . All observations where fitted values are equal to or above (below) 0.75 are considered situations where the higher rated player is predicted to be the winner (loser). A 2x2 contingency table can be constructed at  $c = 0.75$  comparing the prediction of whether the higher ranked player won or not versus whether the higher ranked player actually won or not. With these data, sensitivity and specificity can be calculated at  $c = 0.75$  and the process can be repeated for all values from  $c = 0$  to  $c = 1$ . Plotting sensitivity against 1-sensitivity for all possible cutoff points generates the ROC curve and the area under this curve is the AUC.

AUC values equal to 0.50 represent classification accuracy at random chance levels and an AUC equal to 1.00 indicates perfect classification accuracy. Intermediate AUC values are commonly characterized as acceptable when between 0.70 and 0.80, excellent when between 0.80 and 0.90 and outstanding when over 0.90 [30]. AUCs have been used in the medical sciences to assess how well diagnostic tests predict whether a patient is diseased or not [31]. In the social sciences, AUCs have been used to assess how well financial statement measures can predict fraud in equity security markets [32] and how well post-conviction risk assessments predict future arrests in the criminal justice system [33]. We use the AUC to ascertain how well UTR and WTN rating differences predict match outcomes in their own right and also to assess whether one rating is more accurate than the other.

Estimation of 1a is provided in column 1 of table 3 panel A. We find a positive and statistically significant coefficient on *UTR\_DIFF* ( $\alpha_1 = 2.066$ , p-value  $< 0.001$ ), suggesting that as the difference in UTR rating between the favored and unfavored player increases, the likelihood the favored player wins increases. Estimation of 1b is provided in column 1 of table

3, panel b. We find a negative and statistically significant coefficient on WTN rating differences ( $\beta_1 = -0.474$ , p-value  $< 0.001$ ). Because smaller WTN values indicate higher skill, as *WTN\_DIFF* becomes more negative, the skill difference of the favored player increases as does the likelihood that the favored player wins. While these coefficient estimates establish both UTR and WTN ratings are statistically significant predictors of match outcomes, the coefficients cannot be directly compared to draw an inference on which is a better predictor. To draw such an inference requires the AUC derived from the regression estimations.

The first two rows of Fig 1 display the AUC for UTR and WTN ratings respectively, along with black bars designating the related 95% confidence intervals. We find the classification accuracy of UTR rating differences is 73.9% (95% CI, 71.2%-76.6%) compared with 70.4% for WTN (95% CI, 67.5%-73.2%). The 95% confidence intervals in both cases do not overlap with chance AUC levels of 0.50, but do overlap with each other. This implies that overall, UTR and WTN rating differences between competitors can predict match outcomes at better than chance levels, but UTR and WTN ratings exhibit statistically equivalent classification accuracy.

While the AUC serves as our measure of classification accuracy and enables statistical tests of our observed AUC both between ratings and against chance, we note that AUC is not the only way to assess classification accuracy. An alternative measure is a Brier score, which is popular in the extant literature studying professional tennis players. A Brier score,  $B$ , is defined as  $B = (\sum_{m=1}^N (P_m - X_m)^2) / N$  where  $P$  is the predicted probability that the favored player wins match  $m$ ,  $X$  equals one if the favored player actually wins and zero otherwise and  $N$  is the total number of matches. With Brier scores, a value of 0.00 represents perfect accuracy and 0.25 represents random chance. We report the Brier scores for each estimation in Table 3, and find

that the overall sample Brier score is 0.162 for UTR ratings and 0.167 for WTN ratings. Both Brier scores imply UTR and WTN ratings exhibit classification accuracy rates above chance levels and very similar to one another. We therefore reach the same conclusion with Brier scores as we do with AUC.

Because Brier scores have been reported in research analyzing professional tennis players, we can calibrate our observed Brier scores estimated from our analysis of junior tennis matches. In an analysis of Grand Slam match outcomes from 2005-2008, del Corral and Prieto-Rodriguez [18] report Brier scores ranging from 0.167 to 0.205. We note that the regression models used in del Corral and Prieto-Rodriguez [18] to predict match outcomes incorporate differences in player rankings as well as a host of additional covariates including differences in player height, age, and handedness. Such player characteristics are not publicly available for the junior players we study and so we cannot include them in equation 1a or 1b. However, that our models using a single covariate achieve Brier scores similar in magnitude to models using multiple covariates suggest UTR and WTN ratings are powerful classifiers.

As another point of reference for our Brier scores, consider the work of Lisi and Zanella [17] that uses bookmaker odds as a predictor of match outcomes at Grand Slam events. They document a Brier score of 0.166 for bookmakers, which implies classification accuracy of both UTR and WTN ratings at the junior level is nearly identical to the classification accuracy of bookmakers at the professional level.

## **Subsample analysis**

Both UTR and WTN ratings are marketed as skill measures that do not depend on gender or age. Nonetheless, physical maturation and tennis play can vary by both gender and age through adolescence [34-35] and our sample is comprised of adolescents. Gender has also been

shown to impact match competitiveness at the professional level [1,36-37]. To assess whether the overall findings vary by gender or race, we re-estimate equations 1a and 1b separately for boys and girls matches and separately for 16u and 18u matches. Results are reported in columns 2 through 5 in table 3. With respect to gender, we find the classification accuracy of WTN ratings for girls (boys) is 69.4% (71.4%), while it is 73.0% (74.9%) for UTR ratings. Fig 1 reveals the 95% confidence interval overlaps for all four estimates, implying classification accuracies are equivalent statistically. Similar inferences are observed when partitioning by age. The AUC ranges from 69.6% for 16u matches using WTN ratings to 74.8% for 16u matches using UTR ratings. All confidence intervals overlap, implying no statistically significant differences in classification accuracy.

While the results thus far suggest no differences in the classification accuracy of UTR relative to WTN ratings, it is conceivable differences exist on the basis of measurement precision. WTN ranges from 40 to 1 while UTR only ranges from 1 to 16.5. Without knowing the actual algorithm that generates these values it is not possible to make formal predictions about differential precision in the measures. However, if these differences in scale imply differences in precision, it may be the case that WTN ratings outperform UTR ratings only in situations where the prediction task is especially difficult. We consider two situations.

The first situation we consider is based upon whether the match was played as part of the main draw or the consolation draw. Winners are determined in main draw matches by the best of three tiebreak sets. In the consolation draw, winners are determined by the best of three tiebreak sets, but if a third set is reached a ten point tiebreaker is played instead of a full third set. As a result, there is relatively less tennis played to determine match outcomes in the consolation draw, making the outcome more stochastic and the prediction task more difficult. WTN ratings may

therefore exhibit superiority over UTR ratings when we focus on consolation draw matches. Re-estimation of equation 1a and 1b by whether the match was part of the main draw or consolation draw is presented in columns 6 and 7 of table 3. The AUC for main draw UTR and WTN ratings are 75.4% and 72.6%, respectively. Consolation draw AUC values are lower at 71.9% and 67.6% for UTR and WTN ratings, respectively. However, Fig 1 reveals the 95% confidence intervals overlap for WTN and UTR ratings in both the main draw and the consolation draw, suggesting no superiority in WTN rating classification accuracy.

The second situation we consider is whether there is a large or small difference in the ratings between players. We define large and small differences by reference to the sample median differences reported in panel B of table 2. Matches where *UTR\_DIFF* is greater than (less than or equal to) the median of 0.63 are considered large (small) differences. Matches where *WTN\_DIFF* is less than (greater than or equal to) the median of -2.20 are considered large (small) differences. Larger rating differences should better distinguish the more skilled player relative to the less skilled player. Smaller differences, however, make it more difficult to ascertain with certainty who the better player is, making the prediction task more difficult. If WTN provides a more precise measure of player skill, WTN may exhibit superior classification accuracy when player rating differences are relatively small. Re-estimation of equation 1a and 1b by the magnitude of the rating differential is provided in the final two columns of Table 3. We find that that when the skill difference is large, the AUC equals 69.5% and 66.7% for UTR and WTN ratings, respectively and Fig 1 reveals these reject equivalence to chance levels. For small differences the AUC values decrease to 59.8% and 54.8% for UTR and WTN ratings respectively, and Fig 1 reveals both values also reject chance levels. However, the 95% confidence intervals overlap for matches with large skill differences and matches with small skill

differences, suggesting that UTR and WTN ratings provide statistically equivalent classification accuracy regardless of magnitudes in skill rating differences.

## Discussion

This study provides, to our knowledge, the first empirical evidence comparing UTR and WTN ratings as measures of player skill. Under the assumption that more skilled players are more likely to win a match in head-to-head competition, we assess the classification accuracy of UTR and WTN ratings for predicting match outcomes. Differences in competitor UTR ratings predict match outcomes with 73.9% accuracy, compared with 70.4% for WTN ratings.

Classification accuracy exceeds chance levels of 50% in both cases, but the accuracy rates are not statistically different from one another. Moreover, we find no statistical differences in classification accuracy between UTR and WTN ratings when we investigate match outcomes by age division, gender, match format, or by the magnitude of the difference in player ratings. The lack of superiority in classification accuracy of one rating over another is consistent with both ratings capturing player skill to a very similar extent. Overall, we conclude that UTR and WTN ratings are equivalent measures of player skill for the sample analyzed here.

To execute our analysis comparing UTR and WTN ratings, we required our sample observations to have both ratings. We acknowledge, however, that situations can arise where a player does not have both a UTR and WTN rating. For example, consider a junior player in the United States who currently only plays high school tennis but wishes to play a USTA event that will seed players based on WTN ratings. Such a player will have a UTR rating because high schools utilize UTR ratings [7] but will not have a WTN as a result of not playing USTA matches previously. How should a tournament director handle seeding such a player? From our data, we can estimate mapping functions between UTR and WTN ratings. By estimating an



Ordinary Least Squares regression where the dependent variable is the WTN rating and the independent variable is the UTR rating, we find that the predicted WTN rating is equal to  $39.1836 - 2.3973 \times \text{UTR rating}$ . Applying this function to our average sample UTR rating in Table 2 of 9.83 we uncover the average sample WTN rating. We expect situations where a player has a WTN rating but not a UTR rating to be relatively rare in the United States at this point because USTA matches are inputs to both UTR and WTN rating algorithms. For completeness, we nonetheless regress UTR ratings on WTN ratings and the predicted UTR rating is equal to  $15.3567 - 0.3539 \times \text{WTN rating}$ .

Our findings are subject to the following limitations. First, how our results generalize outside the United States is unclear. As more and more countries adopt WTN ratings, repeating the analysis presented here outside of the United States will help assess whether classification accuracy varies in different geographic locations. Second, our sample of junior national level tennis players does not allow us to speak to whether rating-based classification accuracy differs for matches between novices and professionals. Repeating our analysis for local and sectional USTA tournaments would facilitate an analysis of more novice players. Analysis at the professional level will be useful for assessing not only classification accuracy of UTR and WTN ratings, but also whether these ratings add any predictive power above the growing list of specific player and match characteristics that are associated with winning tennis matches [38]. For example, Ma, Liu and Ma [39] establish the importance of player specific performance on match outcomes including the number of aces, first serve percentages and break points converted. Such rich point level data is, however, not available for amateurs making such an analysis for junior tournaments impossible. Third, Ma, Liu and Ma [39] conduct a longitudinal analysis highlighting that certain determinants of winning have changed over time at the

professional level. In the context of our study, examining future instances of the USTA Junior National Championships will help establish whether the classification accuracy of UTR and WTN ratings vary over time.

With these limitations in mind, we believe our findings provide useful initial information for tennis associations, tennis coaches, tournament directors, college coaches and academics who study tennis players. For tennis associations considering adopting an official rating, on the dimension of assessing player skill of junior competitors, we find one rating does not dominate the other. For tennis coaches interested in organizing practices to ensure play parity, tournament directors accepting and seeding players, and college coaches screening potential recruits our evidence suggests that either measure is equivalent in terms of assessing player skill. For academics who use player ratings when studying the determinants of tennis player skill [40] or developing tennis handicaps in junior tennis as a function of player skill differences [41], our findings imply that using either UTR or WTN ratings as a proxy for player skill is appropriate.

## **Conclusion**

Given two global tennis player skill ratings now exist, we assess whether one rating is superior to the other by comparing their classification accuracy for match outcomes. Using a sample of matches from a junior national tournament, we find both UTR ratings and WTN ratings exhibit classification accuracy levels that exceed chance levels and are similar in magnitude to the accuracy levels using bookmaker odds at the professional level. These findings suggest both UTR and WTN ratings are useful for measuring relative player skill. UTR and WTN rating classification accuracy levels do not differ from one another, suggesting use of either rating is appropriate for assessing relative player skill. Our evidence using junior level tennis matches extends a large literature studying how relative player ratings predict match

outcomes at the professional level [4]. For tennis administrators and coaches interested in assessing relative player skill among junior tennis players in the United States, our findings are of practical importance given the number of junior tennis players recently topped 6 million [42].

## Acknowledgements

We received excellent research assistance from Roshan Banapuram Raja and Qiru (Vincent) Zhang. We appreciate helpful comments from Kyle Bunds, Jon Casper, Julio del Corral, Will Mayew and Ian Mayew. The views expressed in this paper are those of the authors and do not represent positions of the United States Professional Tennis Association.

## Author contributions

Conceived and designed the empirical tests: RLM and WJM. Collected data and performed empirical analysis: WJM. Wrote the paper: RLM and WJM.

## References

1. Gomez-Gonzalez C, del Corral J. Professional tennis in the twenty-first century: Hawk-eye on competitive balance. In: Rodriguez P, Kesenne S, Humphreys BR, editors. Outcome Uncertainty in Sporting Events, Northampton: Edward Elgar Publishing; 2020. p. 27-43.
2. ATP. Association of Tennis Professionals Singles Rankings. 2022 [cited 2022 December 19]. Available from: <https://www.atptour.com/en/rankings/singles>
3. WTA. Women's Tennis Association Singles Rankings. 2022 [cited 2022 December 19]. Database. Available from: <https://www.wtatennis.com/rankings/singles>.
4. del Corral J, Gomez-Gonzalez C. Hitting the Ball Forward: The Economics of Racquet Sports. In: Downward P, Frick, B, Humphreys BR, Pawlowski T, Ruseski JE Soebbing BP, editors. The SAGE Handbook of Sports Economics, Sage Publications Ltd; 2019. p. 27-43. DOI: 10.4135/9781526470447.
5. USTA. United States Tennis Association Junior Tournaments Rankings. 2022 [cited 2022 December 19]. Database. Available from: <https://www.usta.com/en/home/play/rankings.html>.
6. Universal Tennis. How UTR works. 2022. Available from: <https://www.universaltennis.com/how-utr-works>.
7. NFHS. NFHS announces Universal Tennis as new official partner. 2021;1: 28. Available from: <https://nfhs.org/articles/nfhs-announces-universal-tennis-as-new-official-partner/>.

8. Rossingh, D. Larry Ellison-backed Universal Tennis invests \$20 million in three-year global tour. Forbes. 2020 December 7. Available from: <https://www.forbes.com/sites/daniellrossingh/2020/12/07/ellison-backed-universal-tennis-invests-20m-in-three-year-global-tour/>.
9. ITF. ITF and national associations announce world tennis number project. 2019;7: 8. Available from: <https://www.itftennis.com/en/news-and-media/articles/itf-and-national-associations-announce-world-tennis-number-project/>.
10. ITF. ITF World Tennis Number FAQs for Junior World Tennis Tour. 2022;5. Available from: <https://www.itftennis.com/media/7971/wtn-faqs-updated-0605.pdf>.
11. USTA. USTA launches ITF World Tennis Number widget online. 2022;6: 9. Available from: <https://www.usta.com/en/home/stay-current/national/usta-launches-itf-world-tennis-number-widget-online.html>.
12. Universal Tennis. Universal Tennis launches into pickleball with event management software and rating. 2022;11: 18. Available from: <https://www.tennis.com/baseline/articles/universal-tennis-expands-into-pickleball-with-new-rating-system-event-management>.
13. Sanford, J. Tennis Australia becomes first governing body to adopt UTR as official rating system nationwide. Tennis.com. 2022 January 30. Available from: <https://www.tennis.com/baseline/articles/tennis-australia-partners-universal-tennis-rating-system-first-governing-body>.
14. USTA. ITF World Tennis Number FAQs. 2022 [cited 2022 December 19]. Available from: <https://customercare.usta.com/hc/en-us/articles/4414716969492-ITF-World-Tennis-Number-FAQs>.
15. ITA. Intercollegiate Tennis Association and Universal Tennis announce five-year partnership agreement. 2017;4: 5. Available from: [https://www.itatennis.co/ita-archives/AboutITA/News/Intercollegiate\\_Tennis\\_Association\\_And\\_Universal\\_Tennis\\_Announce\\_Five-Year\\_Partnership\\_Agreement.html](https://www.itatennis.co/ita-archives/AboutITA/News/Intercollegiate_Tennis_Association_And_Universal_Tennis_Announce_Five-Year_Partnership_Agreement.html).
16. ITA. Intercollegiate Tennis Association Adopts ITF World Tennis Number as Exclusive Official Rating for College Tennis. 2023;1: 5. Available from: <https://www.wearecollegetennis.com/2023/01/05/wtn-named-official-rating-of-college-tennis/>.
17. Lisi F. Tennis betting: can statistics beat bookmakers?. Electronic Journal of Applied Statistical Analysis. 2017 Nov 15;10(3):790-808. DOI: 10.1285/i20705948v10n3p790.

18. del Corral J, Prieto-Rodríguez J. Are differences in ranks good predictors for Grand Slam tennis matches?. *International Journal of Forecasting*. 2010 Jul 1;26(3):551-63. DOI: 10.1016/j.ijforecast.2009.12.006.
19. Boulrier BL, Stekler HO. Are sports seedings good predictors?: an evaluation. *International Journal of Forecasting*. 1999 Feb 1;15(1):83-91. DOI: 10.1016/S0169-2070(98)00067-3.
20. Klaassen FJ, Magnus JR. Forecasting the winner of a tennis match. *European Journal of Operational Research*. 2003 Jul 16;148(2):257-67. DOI: 10.1016/S0377-2217(02)00682-3.
21. USTA. Tournaments Lookup. 2022 [cited 2022 December 19]. Database. Available from: <https://playtennis.usta.com/tournaments>.
22. Universal Tennis. Universal Tennis App. 2022 [cited 2022 December 19]. Available from: <https://app.universaltennis.com/>.
23. WTN. WTN Search players. 2022 [cited 2022 August 5]. Database. Available from: <https://worldtennisnumber.com/eng/player-search>.
24. Mali, M. UTR rating basics and how it works. 2021;1: 22 Available from: <https://blog.universaltennis.com/back-to-basics-what-is-utr-and-how-does-it-work/>.
25. Universal Tennis. How UTR Rating Works. 2018 July 3. Available from: <https://blog.universaltennis.com/how-utr-works/>.
26. Bader, S. What events count towards my UTR rating? 2022 [cited 2022 December 19]. Available from: <https://support.universaltennis.com/en/support/solutions/articles/9000138934-what-events-count-towards-my-utr-rating->
27. USTA. ITF World Tennis Number. 2022 [cited 2022 December 19]. Available from: <https://www.usta.com/en/home/play/itf-world-tennis-number.html>.
28. WTN. ITF World Tennis Number frequently asked questions. 2022. [cited 2022 December 19]. Available from: <https://worldtennisnumber.com/eng/faq>.
29. Clarke SR, Dyte D. Using official ratings to simulate major tennis tournaments. *International transactions in operational research*. 2000 Nov;7(6):585-94. DOI: 10.1016/S0969-6016(00)00036-8.
30. Hosmer, D and S. Lemeshow. *Applied Logistic Regression*. New York: Wiley-Interscience Publication, 2000.

31. Zou KH, O'Malley AJ, Mauri L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation*. 2007 Feb 6;115(5):654-7. DOI: 10.1161/CIRCULATIONAHA.105.594929.
32. Hobson JL, Mayew WJ, Venkatachalam M. Analyzing speech to detect financial misreporting. *Journal of Accounting Research*. 2012 May;50(2):349-92. DOI: 10.1111/j.1475-679X.2011.00433.x.
33. Lowenkamp CT, Holsinger AM, Cohen TH. PCRA revisited: Testing the validity of the Federal Post Conviction Risk Assessment (PCRA). *Psychological Services*. 2015 May;12(2):149.
34. Ucuz I, Ucar C, Gurer H, Yildiz S. Hypothalamo-pituitary-adrenal axis activity and related factors in adolescent during a tennis tournament. 2020. DOI: 10.5455/annalsmedres.2020.08.801.
35. Martínez-Gallego R, Villafaina S, Crespo M, Fuentes-García JP. Gender and age influence in pre-competitive and post-competitive anxiety in young tennis players. *Sustainability*. 2022 Apr 20;14(9):4966. DOI: 10.3390/su14094966.
36. Krumer A, Rosenboim M, Shapir OM. Gender, competitiveness, and physical characteristics: Evidence from professional tennis. *Journal of Sports Economics*. 2016 Apr;17(3):234-59. DOI: 10.1177/1527002514528516.
37. del Corral J. Competitive balance and match uncertainty in grand-slam tennis: effects of seeding system, gender, and court surface. *Journal of Sports Economics*. 2009 Dec;10(6):563-81. DOI: 10.1177/1527002509334650.
38. Yue JC, Chou EP, Hsieh M, Hsiao L. A study of forecasting tennis matches via the Glicko model. *PLOS ONE*. 2022 Apr 8. DOI: 10.1371/journal.pone.0266838.
39. Ma SM, Liu CC, Tan Y, Ma SC. Winning matches in Grand Slam men's singles: An analysis of player performance-related variables from 1991 to 2008. *Journal of Sports Sciences*. 2013 Jul 1;31(11):1147-55. DOI: 10.1080/02640414.2013.775472.
40. Kurtz JA, Grazer J, Alban B, Marino M. Ability for tennis specific variables and agility for determining the Universal Tennis Ranking (UTR). *The Sports Journal*. 2019.
41. Chan TC, Singal R. A Bayesian regression approach to handicapping tennis players based on a rating system. *Journal of Quantitative Analysis in Sports*. 2018 Sep 1;14(3):131-41. DOI: 10.1515/jqas-2017-0103.
42. Racquet Sports Industry. 2021 TIA Tennis Industry Participation Report. *Tennis Industry Magazine*. 2021 September/October [cited 2023 January 7]. Available from: <http://www.tennisindustrymag.com/issues/202109/industry-research.pdf>.

## Tables and Figures

**Table 1. Sample Selection**

	# Matches	# Players
USTA 2022 Junior National Championships	2,028	896
Less: First Round Byes	-128	0
Less: Unplayed Matches (Withdrawals)	-239	0
Less: Players Missing WTN rating	-123	-26
Less: Matches with identical UTR or WTN ratings	-6	0
Equals: Final Sample	1,532	870

**Table 2. Descriptive statistics**

<b>Panel A. Player level (N=870)</b>					
Variable	MEAN	STD DEV	MEDIAN	MIN	MAX
<i>UTR</i>	9.83	1.37	9.80	5.59	13.82
<i>WTN</i>	15.63	3.55	15.77	4.37	26.06
<b>Panel B. Match level (N=1,532)</b>					
Variable	MEAN	STD DEV	MEDIAN	MIN	MAX
<i>FAV_UTR_WIN</i>	0.75	0.43	1.00	0.00	1.00
<i>FAV_WTN_WIN</i>	0.76	0.43	1.00	0.00	1.00
<i>UTR_DIFF</i>	0.75	0.58	0.63	0.01	5.37
<i>WTN_DIFF</i>	-2.61	1.95	-2.20	-10.66	-0.01

Table 2 presents descriptive statistics for 870 sample players in Panel A and 1,532 sample matches in Panel B. In Panel A, *UTR* is the UTR rating, which uses a scale from 1.00 (lowest skill) to 16.50 (highest skill). *WTN* is the WTN rating, which uses a scale from 40.00 (lowest skill) to 1.00 (highest skill). In Panel B, *FAV\_UTR\_WIN* equals one if the player favored based upon the UTR rating wins the match, and zero otherwise; *FAV\_WTN\_WIN* equals one if the player favored based upon the WTN rating wins the match, and zero otherwise; *UTR\_DIFF* equals the UTR rating of the favored player (i.e. the player with the larger UTR rating) minus the UTR rating of the unfavored player; *WTN\_DIFF* equals the WTN rating of the favored player (i.e. the player with the smaller WTN rating) minus the WTN rating of the unfavored player.

**Table 3. Logistic Regression Results**

**Panel A: UTR ratings (Dependent variable = *FAV\_UTR\_WIN*)**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>UTR_DIFF</i>	2.066*** (0.179)	2.187*** (0.263)	1.957*** (0.243)	1.907*** (0.228)	2.242*** (0.279)	2.128*** (0.238)	1.979*** (0.273)	1.913*** (0.456)	1.927*** (0.425)
<i>Intercept</i>	-0.121 (0.108)	-0.16 (0.159)	-0.086 (0.146)	-0.083 (0.151)	-0.169 (0.155)	-0.119 (0.156)	-0.112 (0.150)	0.063 (0.454)	-0.094 (0.146)
N	1532	739	793	713	819	786	746	758	774
Pseudo R <sup>2</sup>	0.124	0.135	0.114	0.119	0.129	0.140	0.102	0.062	0.021
Brier Score	0.162	0.158	0.166	0.164	0.160	0.148	0.177	0.093	0.229
AUC	0.739	0.749	0.730	0.731	0.748	0.754	0.719	0.695	0.598
Sample	Overall	Boys Only	Girls Only	18U Only	16U Only	Main Draw	Consolation Draw	Large Difference	Small Difference

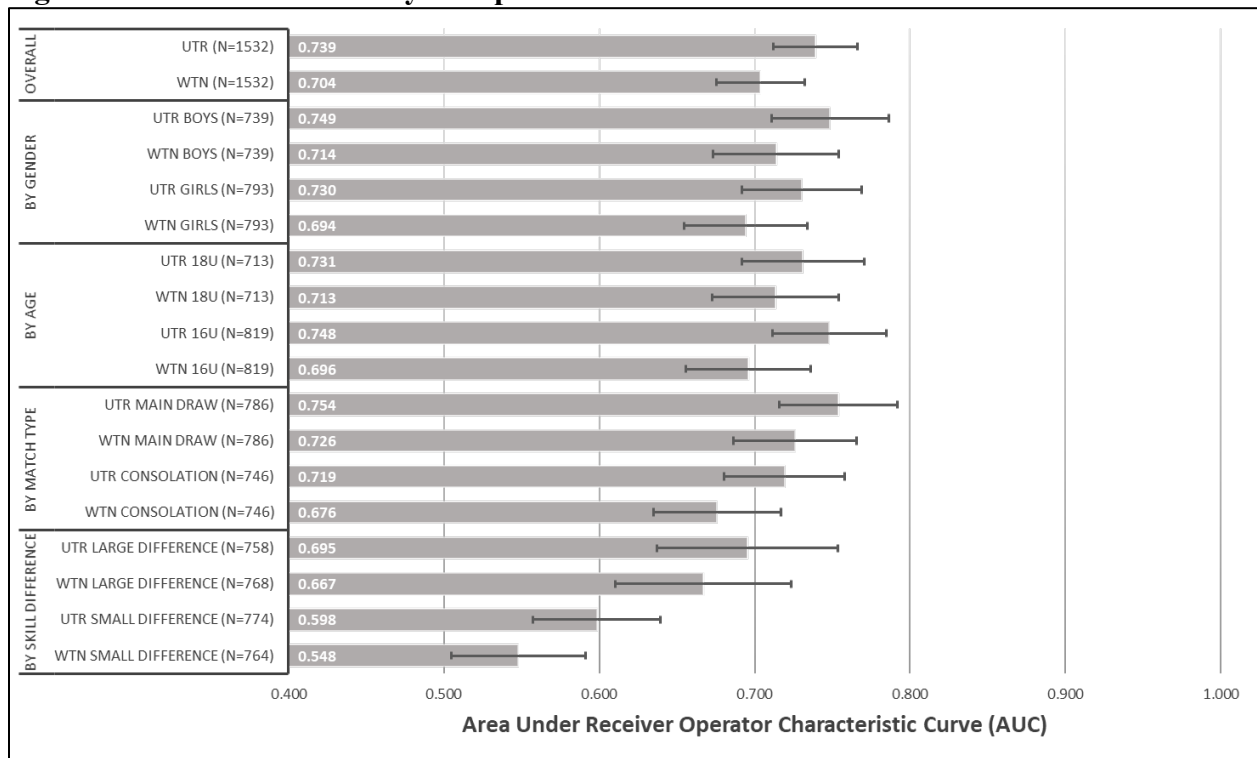
**Panel B: WTN ratings (Dependent variable = *FAV\_WTN\_WIN*)**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>WTN_DIFF</i>	-0.474*** (0.044)	-0.516*** (0.068)	-0.439*** (0.058)	-0.472*** (0.060)	-0.482*** (0.066)	-0.501*** (0.060)	-0.440*** (0.068)	-0.456*** (0.107)	-0.258** (0.121)
<i>Intercept</i>	0.104 (0.103)	0.028 (0.152)	0.167 (0.139)	0.034 (0.150)	0.150 (0.142)	0.064 (0.150)	0.148 (0.143)	0.249 (0.387)	0.286* (0.147)
N	1532	739	793	713	819	786	746	768	764
Pseudo R <sup>2</sup>	0.091	0.099	0.084	0.101	0.082	0.110	0.067	0.048	0.005
Brier Score	0.167	0.165	0.169	0.166	0.168	0.153	0.182	0.104	0.230
AUC	0.704	0.714	0.694	0.713	0.696	0.726	0.676	0.667	0.548
Sample	Overall	Boys Only	Girls Only	18U Only	16U Only	Main Draw	Consolation Draw	Large Difference	Small Difference

The logistic regression dependent variable, *FAV\_UTR\_WIN* (*FAV\_WTN\_WIN*) equals one if the player favored based upon the UTR (WTN) rating wins the match, and zero otherwise. The independent variable *UTR\_DIFF* equals the UTR rating of the favored player (i.e. the player with the larger UTR rating) minus the UTR rating of the unfavored player. The independent variable *WTN\_DIFF* equals the WTN rating of the favored player (i.e. the player with the smaller WTN rating) minus the WTN rating of the unfavored player. Robust Huber-White standard errors are in parentheses beneath the logistic regression coefficients. \*\*\*, \*\*, \* indicates statistical significance at less than the 0.01, 0.05 and 0.10 levels. N is the number of observations. AUC is the area under the receiver operator characteristic curve.



**Fig 1. Classification Accuracy Comparisons.**



This figure reports classification accuracy of WTN and UTR ratings for match outcomes using the area under the receiver operator characteristic curve (AUC). AUC values of 0.500 represent chance levels and 1.000 represents perfect classification accuracy. Each gray bar in each row graphically depicts the AUC, with the actual AUC value displayed in white font and the 95% confidence interval error bars displayed in black. N is the number of matches used in the estimation of the AUC. Rows 1 and 2 display AUC estimates for both ratings using the overall sample, rows 3-6 display AUC estimates separately by rating and player gender, rows 7-10 display AUC estimates separately by rating and age group (18 and under or 16 and under), rows 11-14 display AUC estimates separately by rating and whether the match format dictates a 10-point tiebreaker (consolation) or a full set is played (main draw) in the event a third set is reached, and rows 15-18 display AUC values separately by rating and player skill difference where UTR rating differences above the sample median and WTN values below the sample median are considered large differences and small differences otherwise. The number observations are not equal in rows 15 and 16 and rows 17 and 18 because of multiple rating value differences exhibiting a tie with the sample median.