Reviewer #1. 1) Comparisons with "From PAC to Instance-Optimal Sample Complexity in the Plackett-Luce Model" by Saha and Gopalan (denoted as [SG]). We were not aware of this paper at the time of submission. Thanks for bringing it to our attention. After reading this paper, we found that our work is different from [SG] in the following aspects. (i) Our paper studies full ranking while [SG] focuses on maxing, which is a different problem. (ii) The algorithms in [SG] only work for the PL model, while our algorithm works for a larger class that strictly contains the pairwise PL model as a special case. The algorithms in [SG] depend on the special properties of the PL model, while ours does not. (iii) In [SG], the algorithm checks whether an item is (PAC-)less preferred than another one by comparing the empirical $p_{i,j}$-value and $(1/2 - \epsilon_s)$. In contrast, our algorithm needs to check whether an item can be inserted to a sorted list with enough confidence, which can hardly be done by simply comparing empirical means. For this purpose, we constructed a subroutine ATC to replace the comparisons in Binary Search in [14] and redesigned the random walk procedure to get an attempting version of Binary Search. (iv) We also proved several lower bounds that are significant for understanding the full ranking problem. Further, to the best of our knowledge, we are the first to introduce the $\Omega(\Delta_i^{-2} \log\log \Delta_i^{-1})$ lower bound to the field of active ranking. **2) PAC v.s. exact ranking.** (i) In some applications, we may want to find the exact order, especially in "winner-takes-all" situations. For example, when predicting the winner of an election, we prefer to get the exact result but not the PAC one, as only a few votes can completely change the result. (ii) Our algorithm IIR can utilize unequal noise levels, while previous PAC algorithms may not. For example, Refs [10,11,12] only depend on $n\epsilon^{-2}$. (iii) Studying exact ranking is theoretical significant and helps obtain insights on instance-wise upper and lower bounds. Specifically, the PAC bounds that depend on ϵ^{-1} will become vacuous as ϵ shrinks to zero, i.e., the PAC bounds do not reduce to meaningful instance-wise bounds when one pushes ϵ to 0.

Reviewer #2. 1) Independence. (i) We believe assuming independence is realistic. For example, for ranking movies, it is reasonable to define $p_{i,j}$ as the fraction of users that prefer movie i to j, and a comparison between i and j refers to sampling a user's preference. The comparisons can be assumed to be independent if the users are randomly sampled. (ii) Independence across time, items, and sets is a common assumption in this field. (iii) If the comparisons were not independent, we might not be able to recover the true preference of the users. For example, if all annotators compare the items based on the comparison results of some user, then the ranking got from these comparisons may not correctly represent the preference of all users. **2) Notation of** $p_{i,j}$. For a comparison over set S, we use $p_{i,S}$ to denote the probability that i wins this comparison. When $|S| = 2$, we write $p_{i,j} = p_{i,\{i,j\}}$ to simplify notation, and $\Delta_{i,j}$ is defined as $|p_{i,j} - 1/2|$. **3) Computational complexity.** (i) It is hard to quantify the actual computational complexity of a comparison as it can vary with applications. For example, a comparison may refer to asking an annotator to compare some items, which can take minutes or longer. Also, a comparison may refer to comparing random numerical values in computer simulations, which may only take milliseconds. Further, due to similar reasons, the difference between the computational complexities of pairwise comparisons and listwise comparisons is also hard to quantify. (ii) It is reasonable to assume that an m-wise comparison takes $\Omega(m)$ time to pass the items to compare. Thus, by Proposition 7, we get that ranking n items with m-wise comparisons takes $\Omega(mn \log_m n)$ time. From this perspective, pairwise ranking uses less time. **4) Worst case scenario.** (i) The MNL model is a widely used model. Thus, the lower bound for this model is significant in its own right. (ii) We analyzed the listwise ranking bounds for several cases not presented in this paper, and found that whether listwise ranking uses less samples depends on how we define the gaps of listwise comparisons. We believe the general listwise ranking problem deserves an independent paper. **5) Comparisons.** Comparisons may refer to querying users about their preferences, asking annotators to compare items, or retrieving data from a data center. We view comparisons as black-box procedures in this paper. **6) Assumptions.** For A1, please refer to "1) Independence" above. If A2 is not true, then the unique true ranking does not exist, rendering the problem unsolvable. A3 means that a more preferred item is more likely to win a comparison. If it is not true, the meaning of "more preferred" becomes vague and we may not have enough information to recover the ranking. **7) Minor.** Thanks. We will make the introduction of the shorthand MNL more properly in the final version.

Reviewer #3. Question 1. p1, p2) (i) Theorem 2 is not a corollary of Theorem 3. Theorem 2 states an instance-wise lower bound for all δ-correct algorithms and input instances satisfying A1-A3, while Theorem 3 only works for the MNL model. (ii) In Theorem 3, the algorithms know *a priori* that the input instances satisfy the MNL model, but in Theorem 2, the algorithms do not. Thus, Theorem 3 is also not a corollary of Theorem 2. **p3)** (i) Eqn(1) is no larger than Eqn(2). The right hand side of Eqn(1) (i.e., $\min\{...\}$) is no larger than $\sum_{i\in[n]} \tilde{\Delta}_i^{-2} \log n$ (by letting each $x_i = 1/n$). Summing up the left hand side of Eqn(1) (i.e., $\sum_{i\in[n]}[...]$) and $\sum_{i\in[n]} \tilde{\Delta}_i^{-2} \log n$, we get Eqn(2). Thus, Eqn(1) $\preceq$ Eqn(2). (ii) There are cases where Eqn(1) is strictly lower than Eqn(2). For example, if $\tilde{\Delta}_1 = n^{-2}$ and $\tilde{\Delta}_i = 0.1$ for all $i \geq 2$, then the right hand side of Eqn(1) is $\Omega(\tilde{\Delta}_1^{-2} + \sum_{i\geq 2} \tilde{\Delta}_i^{-2} \log n)$ (by letting $x_1 = 1/2$), lower than $\sum_{i\in[n]} \tilde{\Delta}_i^{-2} \log n$, making Eqn(1) $\prec$ Eqn(2). **Question 2.** (i) PLPAC-AMPR only works for the MNL model, so it only occurs in Figure 2(b) type-MNL. (ii) Active Ranking is for finding the Borda ranking [15]. For type-random instances, Borda ranking may not be the same as the exact ranking, so we do not compare Active Ranking in this case (so not present it in Figure 2(c) type-random). **Typos. p1)** Yes, we agree, and we will add the clarification in the final version. **p2)** The "2" near $O(...)$ in Line 87 is a footnote index, not "square". We will make it clearer in the final version.