We thank all the reviewers for their time and feedback.

R1. We enumerate to match R1's questions:
1. Every factorized CRS model class is a subset of the full CRS model class, and hence our upper bounds on the error of the full-CRS extend to the factorized-CRS. It is likely that low-rank models admit an improved rank-dependent convergence rate, and we agree that a simulation of the presumed gap would be a valuable addition. We plan to revise this.
2. Inferring the Mallows Model's parameters is NP-hard, and the Mallows Greedy Approximation (MGA) only provides an approximate solution. We believe at least part of the larger performance gap between CRS and Mallows (compared to CRS and PL) can be attributed to the shortcomings of the approximate solution.

R2. Regarding the role of rank in the learning rate of factorized CRS models: We agree this is a good point, also raised by R1; see our answer to their Q1. Regarding rank and identifiability, the factorized model is identifiable if the full model is. In Ref [48] on the CDM choice model, their Thm 4, they give a linear-algebraic sufficient condition for identifiability of the full CDM, in terms of the sets compared. It's possible that the condition could be turned into an improved rank-dependent sufficient condition for low-rank factorized models; we haven't tried to refine that aspect of their work. Regarding the necessity of $O(n \log(n)^2)$ samples: that is for the full CRS. A possible conjecture for the rank-r factorized case would be $O(r \log(n)^2)$, but we haven't tried to prove that.

R3. We enumerate to match R3's questions and comments:
1. For algorithmic results, we indeed rely on the machinery developed for CDM in [48]. More specifically, we rely on the well-established and general results that smooth and differentiable strongly convex functions in a bounded parameter space are guaranteed to efficiently converge using first order gradient methods. These methods are the ones we also use in the empirics section.
2. Good suggestion, we should include the discrepancies in optimization in the main text.
3. We strongly agree guarantees for partial rankings would strengthen the contribution, and found it difficult to resist the temptation to include such results. Due to space constraints, we ultimately prioritized clarity in development and exposition, and felt the results would find a home in a lengthier journal version of the contribution, also covering partial rankings and top-k rankings.
4. The quadratic parameter scaling is only for the full CRS model; the factorized model, used in our empirical work, scales linearly (at a fixed rank). Regarding Figure 2: it is confusing, we apologize. The point of the figure is to show that CRS is doing particularly well, relative PL, at predicting second and third position choices. We will re-consider the logL-based position measure, perhaps using probabilities or accuracy instead.

R4. We enumerate to match R4's questions and comments:
1. The CRS model is defined after line 175; we will try to make this definition clearer during revisions.
2. Regarding Thm 1, our use of $\lambda_2(L)$ in the result of Theorem 1 is deliberate, to characterize the intricate choice set structure-dependence on the rate of error convergence. It is also consistent with the literature; see [49] for examples of how $\lambda_2(L)$ can vary considerably with different choice set structures. From [49], $\lambda_2(L)$ also appears in the lower bound of the error, making it a fundamental quantity to the rate of error convergence. Our tight upper bound in Thm 1 improves upon multiple prior attempts [24, 49].
3. Regarding Thm 2, our tail bound is not trivial for large n; its claim is that so long as the number of samples ℓ grows linearly with the number of parameters (items) n, the error obeys an exponential tail. That is, it is bounded with exponentially high probability. Our result is in-fact tight, as demonstrated by a lower bound in [24]: no algorithm to infer PL parameters can do so without requiring ℓ on the order of n.
4. Thank you for raising this point. Our use of "identifiability" stems from discrete choice settings, where choices conditioned on choice sets are treated as the data, but choice sets define the support of the (model) distribution. Without conditions on the choice set structure (e.g., encoded by $\lambda_2(L)$) a choice model cannot be identified even with infinite data. Thought of in this manner, we believe identifiability is the correct term.

For repeated selection models of rankings like PL and CRS, however, the choice set distribution is also fixed by the ranking model parameters, and for the class of model parameters we consider, both PL and CRS are always identifiable. The claims we make about CRS and PL needing a particular number of rankings should be amended to be claims about *estimability*—the number of rankings before the parameters are no longer under-determined—and *not identifiability*. We thus apologize for our misuse of the word identifiability and the resulting confusion. We will amend it.