We thank the reviewers for their valuable suggestions. Please find our answers (**A**) for each reviewer (**R**) below.

R1, R2: *Formal definition of* VCD

A: We will add the following definition: $\mathsf{VCD}(X, H) = \max |X'|, \text{ s.t. } X' \subseteq X$ and $|H_{|X'}| = 2^{|X'|}$.

R1: *A more detailed explanation about the teaching models/protocol and intuition behind the main result*

A: As suggested by the reviewer, we will incorporate a detailed explanation of the existing teaching models and protocols in the updated version of the paper. In particular, we will clarify that the teacher knows the learner's preference function. This is the protocol used in existing teaching models for both the batch settings (e.g., as in RTD/PBTD [ZLHZ11, GRSZ17]) and the sequential settings (e.g., as in [CSMA+18]).

R2: *Insights on the proof of the main lemma (Lemma 4) and connection to the reference*

A: Thanks for pointing out the similarity between the proof of our Lemma 4 and the proof of Theorem 9 in [FW95]. We will acknowledge this connection and add a proper discussion in the revision. Concretely, the concept class $C^{\{x\}}$ of [FW95] is equivalent to our definition of H_x^0 (line 221). [FW95] then applied induction on $C^{\{x\}}$ and $C - x$ (which are represented as H_x^0, and $H \backslash H_x^0$ in our paper, respectively). As pointed out in the review, our novelty lies in that we introduced (i) "compact-distinguishable" set to ensure that each H^j is non-empty, and (ii) a recursive procedure for constructing the preference function.

R2: *Questions regarding Algorithm 2*

A: We realized that there were some notation issues with Algorithm 2, and we agree with the fix suggested in the review. We will incorporate the following updates which are related to Algorithm 2 as detailed below: In Algorithm 2, Line 8, we should have $H_x^y \leftarrow \{h' \in H : h' \triangle x_{|X_{\text{rest}}} \in H_{|X_{\text{rest}}}, h'(x) = y\}$, where X_{rest} (as described in Line 226–229) denotes the set of instances in Ψ_H that have not been traversed in the current for-loop. We will revise the algorithm accordingly to make sure the notations are consistent and self-contained. Furthermore, we will also update the Appendix, in particular, between Line 502–525, with proper conditions (e.g., among others, in Line 505, we will update $h' = h \triangle x_1$ into $h'_{|\Psi_H} = h \triangle x_{1|\Psi_H}$ to be more explicit about the instances being considered).

R2: *"Minor Comments": Typos, grammatical/spelling errors, and notation issues*

A: We greatly appreciate the time and effort spent by the reviewer in pointing us to the minor issues. We will thoroughly proofread the paper and fix all the minor issues pointed out in the reviews. Also, we will address the following definitions/notations as pointed out by the reviewer: [(3) *Page 3, definition of* U]—yes, the first condition should be $\exists z, \text{ s.t. } C_\sigma(H, h, z) = h^*$, and [(5)/(9) *definition of* $\Sigma_{\mathit{const}}, \Sigma_{\mathit{global}}, \Sigma_{\mathit{gvs}}, \Sigma_{\mathit{local}}, \Sigma_{\mathit{lvs}}$]—we will revise each of these definitions by moving the existential quantifier before the universal quantifier.

R2: Suggested improvements and regularity/non-regularity properties of the general teaching parameter

A: We will add a detailed discussion about these interesting questions and properties mentioned by the reviewer. Below, we share a few thoughts:

- First, after reading the review, we explored the question of finding upper/lower bounds on the Σ-TD parameter. We are able to show that for certain hypothesis classes, Σ-TD is lower bounded by a function of VCD. In particular, for the power set class of size d (which has $\mathsf{VCD} = d$), Σ-TD is lower bounded by $\Omega\left(\frac{d}{\log d}\right)$. We will further study whether this bound is tight.
- Regarding the additive/sub-additive property, we will continue to study this property and add a detailed discussion in the revised paper.
- Regarding extension to infinite VC classes, our current results (Lemma 4) is not directly applicable; however, we consider a generalization to the infinite VC classes as a very interesting direction for future work.

R4: *Notions of collusion-freeness in sequential models*

A: Collusion-freeness for the batched setting is well established in the research community. It remains an open question for the research community to agree on a well-accepted notion of collusion-freeness for the sequential setting. In this paper, we are introducing a possible notion of collusion-freeness for the sequential setting (Definition 1). As discussed in Section 6, a stricter condition is the "win-stay lose-shift" model, which is easier to validate without running the teaching algorithm. In contrast, the condition of Definition 1 is more involved in terms of validation and is a joint property of the *teacher-learner pair*. We will further add a discussion on this in the updated version of the paper.

R4: *Discussion on the "presumably increased complexity of sequential learners"*

A: Our model generalizes classical teaching models [ZLHZ11, GRSZ17, CSMA+18], and inherits the complexity results from all these settings. It is known that the optimal teacher achieving TD amounts to solving a set cover problem which is NP-hard; moreover, the complexity of the sequential teaching has been discussed in [CSMA+18] as a planning problem. It remains an open problem to understand the complexity of the general sequential teaching setting as a sequential optimization problem. We will add a discussion in the updated version of the paper.