We appreciate the reviewers for reading our paper and their constructive comments. This response letter is to clarify our major claims according to the comments from reviewer 1, reviewer 2, reviewer 3 and reviewer 4. To save space, we first answer some shared concerns from reviewers and then answer their specific questions separately.

Shared Concerns:

1) **Reviewer 2:***"I would have liked to have seen comparisons to more fundamental baselines that didn't make the same assumptions, such as other recurrent models and other models meant for multi-agent modelling"*
Reviewer 3:*"socialGAN, SoPHie and other multi-agent representation learning approaches should be added..."*
Reviewer 4:"The paper mentions other approaches and it might be useful to see a comparison to other papers..."
Our comparison method includes MA-BE, which is a recently proposed multi-agent embedding model applied to sequential data (Line 226 in submission). SocialGAN and other multi-agent methods are designed for trajectory data and therefore not directly applicable to our event data. For example, SocialGAN describes a model to discriminate fake from real trajectories. We mentioned modelling player interactions for play-by-play data as a topic for future work in our conclusion.

2) **Reviewer 3:***"The shot quality prediction is similar to the results reported in ""Quality vs Quantity"... Can the authors provide some key insights from the proposed approach that was missing in this and other prior work on ..."*
Reviewer 4:*"It is unclear that the ladder aspect of the architecture is providing an improvement on this application."*
Prior work on ice hockey shot prediction does not take into account the identity of the shooter. Certainly not as part of a general player representation framework. For instance, the scoring chance is higher for a top player v.s., an average player under similar game context. Table 1 shows the benefits of modelling shooter-specific effects.

The ladder structure mitigates posterior collapse during training (Lines 148-156). We provide a detailed discussion and results is in C.3 of our Appendix.

Comments from Reviewer 1

1) *"I would have liked to see some analysis of all the latent variables, not just ones at the lowest level."*
We visualize only $\mathbf{z}_{\boldsymbol{r},t}$ (at the lower level of ladder structure) because it conditions on $\mathbf{s}_t, \mathbf{a}_t, \boldsymbol{r}_t$ and contains the most complete information about each player. The latent variables at higher levels, for example $\mathbf{z}_{\mathbf{s},t}$, have no access to $\boldsymbol{r}_t$ or $\mathbf{a}_t$ (This is where our contextualized model differs from the traditional ladder structure). We have visualized the $\mathbf{z}_{\mathbf{s},t}$ and $\mathbf{z}_{\mathbf{a},t}$, but found them less informative so we did not include them. Specifically, latent values from the higher levels distinguish players less, and show a smaller shrinkage effect: many points are smoothly distributed around the mean. (Similar results were observed in the ladder VAE paper [16]). We can discuss the higher levels in the final version.

2)*"The main takeaway for the embedding visualization in Figure 2 is also unclear...How do the embeddings compare with those from CVRNN, the best baseline? I suspect they might look similar to VaRLAE"*
Our main contribution is the idea of Player representation through Player Generation (Section 3). CVRNN and VaRLAE are different architectures for implementing this fundamental idea. Since both methods use the same general idea, we expected their visualization to look similar. In particular, both exhibit a shrinkage effect leading to similar T-SNE projections. The key point of Figure 2 is to show the difference of a model without a shrinkage effect, namely traditional auto-encoder (CAERNN).

3)*"The performance using VaRLAE player representations is on par with CVRNN player representations ... The effectiveness of the learned representations is unclear some more experiments (or domains)".*
Our paper covered three popular tasks in the Ice hockey domain. CVRNN is indeed the strongest ablation method implementing Representation-Through-Generation (Section 5.1). Our VaRLAE beats it by an average of 8% (over 12% for players with sparse participation) in player identification. Expected goals results are mixed: CVRNN has higher precision, VaRLAE has the second-best precision, and achieves overall best performance (Recall and F1-score).

Comments from Reviewer 2

1) *"The paper is very dense and at times lacking in clarity... The paper is well-written at a local level. However,..."*.
Thank you for your suggestions which will help us improve clarity.

2) *"I was somewhat disappointed by the broader impacts section..."*
We will make our code available, to help level the analytics playing field. While technical skills do require resources, professional scouts are even more expensive. Our model focuses only on a player's professional skills without considering race, gender, or age, which encourages fairness and reduces bias. Extending the model to capture player development over time is a great idea, thank you for the suggestion.

Comments from Reviewer 4 (also Reviewer 1)

1) "There is reason to believe that the VaRLAE architecture is applicable to more domains than just hockey...".
It is true that our VaRLAE can be applied to other team sports, as we mention in our conclusion. We thought it was important to provide a thorough in-depth evaluation of several tasks in one domain.