We would like to thank the reviewers for their comments. Below we provide a response to their specific remarks.

1 Reviewer 1

1. In my opinion, the weakest part of the paper is the lack of a more detailed analysis of the pros and cons of the suggested benchmarking scheme and its thorough comparison with other existing NISQ benchmarking schemes. Thus, I recommend a minor revision where I would ask the authors to better describe advantages and disadvantages of the suggested NISQ benchmarking scheme with respect to other existing NISQ benchmarking techniques. Although the focus of the paper is the PyQBench software, I find it important for the reader to have a clear understanding where and how the underlying NISQ benchmarking technique and corresponding software are superior to other available techniques and software.

   Response We improved the manuscript, which now includes a short section explaining the pros and cons of our method in the context of other existing benchmarking methods. Due to the length limitations in the journal, we do not think this section can become any longer, but we are open to further improvement suggestions if the reviewer finds the current analysis lacking.

2. When discussing the existing quantum computing frameworks, the Introduction should also mention the XACC framework which is known for its cross-architectural as well as cross-language portability (https://iopscience.iop.org/art 9565/ab6bf6). Additionally, the recent CUDA-Quantum framework backed by NVIDIA (https://github.com/NVIDIA/cuda-quantum) is worth mentioning as well. **Response** Both XACC and CUDA Quantum have now been mentioned in the manuscript.

3. The cross-entropy benchmarking technique used in the validation of the Sycamore-53 QPU in quantum supremacy experiments (`https://www.nature.com/articles/s41586-019-1666-5`) should also be mentioned in the Introduction (`https://www.nature.com/articles/s41567-018-0124-x`).

Response We modified manuscript to include mention of cross-entropy benchmarking as suggested by the referee.

4. The generalization of the suggested benchmarking techniqiue to multiple qubits should be described in more detail, explicitly mentioning whether the complexity of the derivation and actual benchmarking depends on the number of qubits.

   Response We thank the referee for this suggestion. We extended the manuscript with a paragraph describing differences between a single–qubit and higher-dimensional cases.

2 Reviewer 2