Manuscript Title: The Immediate Effects of Virtual Reality Training on Dart Throwing Accuracy and Concentration: A Pilot Study

Reviewer's Comments: Is the manuscript important for the scientific community? (Few sentences)

Yes, the objective of this study could be important in the scientific (i.e. applied/translational/ healthcare) community. Specifically, the authors suggest that the use of virtual reality training devices may serve as a cost-effective alternative to training for a sports activity, in this case for improving dart-throwing accuracy. Studies investigating the efficacy of using focused recreational sports activities to improve motor control, coordination, and motor learning are well documented in the peer-reviewed literature. The idea of using "play" in many different populations are also well described to enhance motivation and compliance by participants.

Strengths:

1. **Clear Research Objective:** The authors state (last sentence of introduction): "*Based on the above research, this study uses Xbox 360 to conduct VR training to preliminarily try to understand whether this training has an immediate impact on the throwing accuracy and concentration of dart throwers and veterans."*
   This statement is understandable, although the meaning of the word "immediate" is not clearly indicated. What hypothesis are they proposing that this study will answer (based upon their outcomes of accuracy and concentration)?

2. **Methodological Rigor:** The apparent use of SSRD is an experimental design that is appropriate, however use of two subjects for comparisons necessitates very careful interpretation particularly with an acknowledgement that there is a lack of ability to make "between group" statistical inferences between novice and expert users.

3. **Clinical Relevance:** The objective for this study proposed a clinically relevant area of study; specifically they wished to identify whether a cost-effective, technologically innovative resource could be used to help individuals improve their performance in the throwing accuracy of darts with time needed for change a covariant to developing competition standard improvement.

4. **Comprehensive Literature Review:** The amount of literature review for this study appears to be comprehensive and relevant. Authors discussed background information thoroughly in introduction.

5. **Well-Organized Structure:** This manuscript was structured in an acceptable fashion but requires several areas to be clarified and expanded upon (please see below).

Area of improvement:

1. **Title:** The word "Immediate" is somewhat confusing in this title context. Removing and explaining in the text is better (or revise wording).
2. **Abstract:** Wording is sometimes awkward, such as "Accuracy declined un-significantly" is peculiar and unclear. Since "immediate" is an important aspect of this study, please clarify what you mean by this term. How does the testing of "novice" versus "expert" fit in to your hypothesis?

3. **Introduction:** More clarity as it relates to the timeframe of the intervention should be expanded upon, since subjective wording like "short-term" is not clear.

6. **Methodology Clarity:** The methodology used is not at all clearly stated in this manuscript. It appears that the investigators wished to use an experimental design, specifically a single subject research design due to their use of only two individuals – but this is not clearly identified. Furthermore, SSRD do not traditionally use two groups as comparison as the statistical strength of such methods are not considered of sufficient rigor to make well-supported conclusions. Rather, one (or a few of the same) individuals are noted under baseline (pretest), during intervention, and at follow up (post-test) and then examines the subject across the different phases. This section should be made much more clearly as to the intent of design and comparisons that are to be made.

    - **Subject selection:**

    Authors note reference by Harris et al., 2020 (#20) as defining the terms of "novice" (subject A) compared to the "veteran" (subject B). However in that paper, they are referring to participants in a different sport (golf) as in, novices are "based on having no official golf handicaps or prior formal golf putting experience" whereas the contrasted "experts" are "expert amateur golfers… recruited from three competitive golf teams". Given the small sample size in the current study, one might rightly expect a more defined characterization of who are representatives of these two "groups".

    Data collected The two outcomes noted in this study are referred to as "accuracy" and "concentration", with units of measurement cited as "cm" for each (Table 1).
        - Can you explain more clearly what the concentration index that is cited entails? The cited paper by Mousavi et al. (2019) does not appear to have this metric (at least with "cm") and therefore this was not clear.
        - What was the timeframe for data collected at pre-, during, and post-test intervals? SSRD usually requires that each phase should consist of several measurements taken over the course of different days (not all in one day). Was this done? Was the measurements taken so that an accurate baseline was established for each phase (for example, take the

baseline measures until each measure didn't change drastically).

- **Statistical methods**
  The authors assessed within group changes (pre-test compared to post-test) for the novice data; and then again separately for the expert data.
    - The use of visual analysis is much more common when using SSRD methods although the use of statistical testing is becoming more common. Adding a figure for this data will enhance the understanding of what you have found (assuming this study was done using the SSRD methodology of taking measurements over different times/sessions)
    - Table 1 data should include sample size (n) for each periods (Also, the "P<.05" would more appropriately be noted as "p<0.05").
    - The use of parametric independent t-tests (cited in the Bates 1996 study) is not necessarily supported, particularly in time-series or autocorrelated data. As noted above, visual analysis is much more common; if t-tests are to be included more recent support for the independent use should be included. Alternatively, the reporting of effect sizes (i.e. Cohen's d for interval/ratio data such as yours) is increasingly more common. This will indicate the treatment magnitude of the intervention and one that is more appropriate for this type of research design method.
    - The data section of the methods do not state if parametric criteria were assessed to be met prior to using parametric testing- was this done?

Summary: The inherent issues with sample size and comparison group structure make any conclusions drawn from a study using a SSRD approach not strongly backed in the manner currently provided in this study. Be cautious in your conclusions and be sure to be clear that you understand the limitations of using this type of research design (although completely appropriate for a pilot study).

Instead, one might possibly describe this as a SSRD for one person (novice) and then compare them to the SSRD results of the other person (expert) <u>but hold any findings to the descriptive level only</u>. For each individual SSRD, one might conduct within group comparisons (e.g. use ES as mentioned above) but if you do, be cautious in any summary findings are reported. Any comments should make clear that these are exploratory and descriptive and not based upon between-group inferential statistical assumptions.

4. **Figures:** Not applicable
5. **Tables and Study Data Integration:**
    - What does the title "Statistics" in Table 1 refer to? Are these the means <u>+</u>

standard deviations (cm) taken for "n" number of throws?

- Why does Table 2 accuracy pre-test values for the veteran listed only as 0? Is this correct? If it is, this should be clearly noted since the narrative seems to imply otherwise. (See *"The mean post-test accuracy value for the veteran (3.84) was significantly higher than the pre-test value (2.09), indicating that post-test accuracy performance was significantly worse than pre-test performance.")*
- When discussing table data, be sure to include not only means but also standard deviations (if that is what that is) in your discussion.

6. **Discussion:**

7. **Limitation:** Limitations of your study (not of the VR training) are not addressed well (if at all) in your Discussion section and should be discussed, particularly as it relates to the limitations of using a SSRD.

8. **Conclusion: "***The results of the present study indicate that VR dart training did not produce a significant immediate improvement in throwing accuracy within a short period."*

   The use of the words "short period" is confusing. Do you mean short period from pre-test to post-test (i.e. entire study), or short as it relates to the time of a single session itself, or over the course of several sessions grouped together in a short phase of time?

9. **References:**

Minor Suggestions:

1. Clarity & Conciseness Please see above comments.

2. Minor Typographical & Formatting Issues

Some typographical corrections appear to be needed. (For example, please see Gawron et al.)

Recommendations for Publication: Although this study has some positive merits, several points need to be revised as indicated in comments above.

Guideline:

- Accept:
- Minor Revision:
- **Major Revision: Several corrections with respect to the methodology (including being clear as to the type of research design methods used; data analysis methods, and data interpretation is needed.**
- Serious Major revision:
- Rejected (with repairable deficiencies and may be reconsidered):
- Strongly rejected (with irreparable deficiencies.):