Manuscript Title: *The Effects of a Single Session 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)?

Reply: In response to the reviewer's comment, the term immediate impact used in the original introduction may indeed cause ambiguity in meaning. Based on the reviewer's suggestion, the term immediate impact in the manuscript has been revised to immediate effect in order to more accurately reflect the research design and purpose.

In this study, the term "immediate effect" refers to changes in dart-throwing performance that occur over a short period following a single virtual reality training session.

The actual hypothesis of this study is to examine whether a single VR training session can produce effects on dart-throwing performance and concentration within a short period (measured immediately after training), and whether such effects occur separately in novice and veteran dart throwers with different levels of experience.

Based on the above reviewer's comments, this paragraph has been revised. The revised statement appears on page 4, line 12, as follows:

Based on previous research, this study uses Xbox 360 to conduct VR training to preliminarily examine whether a single session of VR training has an immediate effect on dart-throwing accuracy and concentration in a novice and a veteran dart thrower. In this study, the term "immediate effect" refers to changes in dart-throwing performance that occur over a short period following a single virtual reality training session. Specifically, this study aims to determine whether a single VR training session produces

short-term effects on dart-throwing accuracy and concentration performance separately in novice and veteran dart throwers.

Reviewer (2nd response): Thank you for that clarification in your revision. This change will make more sense to the reader as to what the intent of this study is. (Slight note: Grammatical correction recommended to add the word, "to" in line 9 on p. 9- as in, "this study uses Xbox 360 to conduct VR training...".

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.

Reply: The purpose of this study is to describe the individual performance changes of the novice and the veteran dart thrower in the pre-test and post-test, rather than to conduct between-group comparisons between the novice and the veteran. Therefore, the results of this study cannot and were not intended to be used for between-group statistical inference.

To avoid misunderstanding caused by wording that may imply between-group comparisons, the term In contrast that might suggest a comparative relationship in the original text has been deleted, and the description of the results has been revised to present only the pre-test and post-test changes of each participant.

Based on the above, this paragraph has been revised. The revised results statement appears on page 7, line 22, as follows:

The results in the above indicate that, in the accuracy test, no significant improvement was observed in results of the novice, and the accuracy in post-test was even slightly lower than that in the pre-test. The accuracy in post-test of veteran decreased significantly compared with the result of pre-test.

Reviewer (2nd response): Thank you for that clarifying revision. There still remains the concern that the limitation of this study in terms of the statistical rigor of drawing inferences when only one person per condition (novice and expert) is being compared. This should be clearly discussed in the Discussion section.

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).

Reply: Revised according to the reviewer's suggestion.

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).

Reply: Taking the reviewer's comments into consideration and in order to avoid potential misunderstanding by readers, the term Immediate has been removed from the title, and the intervention method and measurement time points of this study have been clearly described in the main text, emphasizing that the research design involves conducting pre-test and post-test comparisons after a single VR training session followed by a short rest period.

Based on the above, this paragraph has been revised. The revised title appears on page 1, line 2, as follows:

The Effects of a Single Session Virtual Reality Training on Dart Throwing Accuracy and Concentration: A Pilot Study

Reviewer (2nd response): Thank you for that revision.

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?

Reply: The phrase "Accuracy declined un-significantly" has been revised to "No significant change in accuracy was observed in the novice."

The part regarding Immediate has been explained in the main text.

This study did not perform statistical tests to compare the novice and the veteran. The novice is just beginning to learn the skill, whereas the veteran is already familiar with the skill. In motor skill performance, these two groups are completely distinct populations and are often the focus of research. This study is a pilot study. In order to preliminarily examine the effects of a single session of training on dart throwers, one novice and one expert were selected as participants.

Based on the above, the method section in the abstract has been revised. The revised statement appears on page 2, line 11, as follows:

The subjects included one inexperienced novice and one veteran, as novice and veteran represent distinct populations in motor skill performance.

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.

Reply: In the introduction of this study, subjective temporal terms such as short-term were not used; instead, the intervention time frame was clearly defined by the study design itself, emphasizing that this study involves a single virtual reality (VR) training session.

To make the intervention time frame clearer, the study flow has been further described in the introduction and methods sections, explicitly stating that the pre-test and post-test comparisons were conducted after a single VR training session followed by a short rest period, in order to describe the immediate performance changes after that session, rather than long-term or multiple-session training effects.

Through the above revisions, the intervention time frame has been clearly defined, avoiding the use of vague or subjective temporal descriptions.

Based on the above, this paragraph has been revised. The revised content appears on page 4, line 15, as follows:

In this study, the term "immediate effect" refers to changes in dart-throwing performance that occur over a short period following a single virtual reality training session.

This sentence should be sufficient to address the reviewer's concern. Additionally, occur over a short period following a single virtual reality training session refers to participants having a short rest after experiencing a single VR training session. The duration of this short rest has been explained in the methods section, which is 10 minutes.

Reviewer (2nd response): Thank you for this clarification. Both the initial and current manuscript indicated "Tirp et al. (2015) pointed out that a one-week simulation dart training on Xbox 360 for novices reduced the throwing error from 5.5 cm to 5.0 cm, indicating an improvement in accuracy in the short term." Your additional narrative now makes clearer what was indicated by the phrasing "short-term" . Thank you.

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.

Reply: This study did not perform statistical tests to compare the novice and the expert. To avoid misunderstanding, in contrast has been deleted from the text. In addition, this study is a pilot study. In order to preliminarily examine the effects of a single session of training on dart throwers, one novice and one expert were selected as participants.

Reviewer (2nd response): Thank you for this clarification. On p. 5, line 22-23, there is discussion of "with veteran subjects throwing three darts first" then, "followed by novice subjects". This phrasing appears to imply there were more than on veteran subject, and more than one novice subject. Please revise this passage to clearly indicate that only one novice and one veteran expert was involved in this methodology.

- **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".

Reply: This study is a pilot study, and the participants do not constitute a group; only one novice and one expert were selected.

As previously explained, the novice is just beginning to learn the skill, whereas the veteran is already familiar with the skill. In motor skill performance, these two are completely distinct populations and are often the focus of research. This study is a pilot study. In order to preliminarily examine the effects of a single session of training on dart throwers, one novice and one expert were selected as participants.

The novice in this study is described as: The novice (subject A) is defined as someone who has not received systematic darts training in the past and has no actual throwing experience.

The veteran in this study is a member of a darts team. To clarify the representativeness of this participant, the description has been revised as follows:

Based on the above, the main text has been revised. The revised content appears on page 4, line 26, as follows:

The veteran (subject B) is a subject with more than four years of competitive experience, having competed in international and national tournaments, including a top 32 finish in the International Darts Asian Cup and a championship title in the Taiwan National Darts League.

Reviewer response: Thank you for that clarification.

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.

Reply: Regarding the concentration index used in this study, its calculation follows the algorithm of Mousavi et al. (2019), but in this study, the unit was converted to centimeters (cm) for measurement and calculation. Although the unit is different, the calculation principle and method are consistent with the original study, and therefore it is considered reasonable and applicable to the dart-throwing context of this study.

Reviewer (2nd response): Thank you for that clarification.

- 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).

Reply: As this study is a pilot study of a single VR training session, it aims to preliminarily examine whether a single VR training session can produce effects on dart-throwing performance and concentration over a short period (measured immediately after training), and whether such effects occur separately in novice and veteran dart throwers with different levels of experience.

To avoid misunderstanding, the study flow has been further described in the introduction and methods sections, explicitly stating that the pre-test and post-test comparisons were conducted after a single VR training session followed by a short rest period, in order to describe the immediate performance changes after that session, rather than long-term or multiple-session training effects.

Reviewer (2nd response): Thank you for that clarification. Given that the interest is to identify if a single VR training session was adequate to elicit a change, it appears that the authors feel that collection of data could only be achieved during this same post-training session. Therefore, it cannot be deemed to be using an experimental SSRD design but one that is more of a descriptive comparative design approach. If this is the case, this would also explain why between-group statistical analysis was not performed (as noted below).

- **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.

o 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)

Reply: The study flow has been further described in the introduction and methods sections, explicitly stating that the pre-test and post-test comparisons were conducted after a single VR training session followed by a short rest period, in order to describe the immediate performance changes after that session, rather than long-term or multiple-session training effects.

o Table 1 data should include sample size (n) for each periods (Also, the "P<.05" would more appropriately be noted as "p<0.05").

Reply: The revisions have been made on page 2, line 23; page 6, line 31; page 7, line 3; and page 7, line 14.

o 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.

Reply: This paragraph has been revised. On page 6, line 27, a more recent reference supporting t-tests has been added: Knief, U., & Forstmeier, W. (2021). Violating the normality assumption

may be the lesser of two evils. Behavior Research Methods, 53(6), 2576–2590.

Reviewer (2nd response): Thank you for that clarification. To avoid confusion and misinterpretation to the reader, please clarify what the "N" sample refers to as it is not clearly articulated and could easily be mistaken as noting 15 individuals (for example) who were at the novice or expert level, rather than 15 repetitions from a single (n=1) individual. "N" usually refers to the total number of participants in a study or the entire population being surveyed. Was the N=15 referring to the total number of dart throws (pre- and post- test combined) metrics taken or 15 throws for each time (pre-, n=15; post-, n=15)? This should be corrected in your Table 1 notations as well as in your narrative (i.e. line 4, p. 7).

- 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?

Reply: No

Reviewer (2nd response): That is unfortunate as this makes it difficult to know the statistical validity of this reported data ($p<0.05$). Also, you do not report if the "p" value (not "P", as noted) was done using 1-way or 2-way methods. This should be corrected and discussed in some manner.

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.

Reply: It is indeed necessary to keep the results of the two participants at the descriptive level of comparison, and it should also be emphasized that it must be clarified to the participants that the exploratory and descriptive results for these two participants may not generalize to others.

Reviewer (2nd response): Agreed. Is this noted somewhere in your paper (e.g. in Discussion section)?

4. **Figures:** Not applicable
5. **Tables and Study Data Integration:**
    - What does the title "Statistics" in Table 1 refer to? Are these the means ± standard deviations (cm) taken for "n" number of throws?

Reply: Yes (N = 15), and it has been added to the content of Table on page 6, line 33.

- 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.

Reply: The value has been corrected to 2.09 ± 0.94, not 0, and has been added to the content of Table 2 on page 7.

Reviewer (2nd response): Thank you for that correction.

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.

Reply: The discussion of the novice and the veteran has been revised and clarified; the two were not compared with each other. The results of the two participants remain at the descriptive level of comparison, and in the future it can also be clarified to participants that the exploratory and descriptive results for these two participants may not generalize to others.

Reviewer (2nd response): Thank you for that correction. Please see comment above regarding limitations of study.

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?

Reply: To make the intervention time frame clearer, the study flow has been further described in the introduction and methods sections, explicitly stating that the pre-test and post-test comparisons were conducted after a single VR training session followed by a short rest

period, in order to describe the immediate performance changes after that session, rather than long-term or multiple-session training effects.

Reviewer (2nd response): Thank you for these revisions.

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.)

Reply: After confirmation, the article by Gawron has only one author.

Reviewer (2nd response): Thank you for that correction.

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

Reviewer (2nd response): This manuscript has been improved considerably due to the revisions made. However, some additional corrections are needed as noted above. Thank you .

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.):