ISSN: 3139-0854

# Impact of Virtual Reality Training on Motor Skill Transfer and Retention in Physical Education: Protocol for a Randomised Controlled Trial

[1*]Dhanaji Shivaji Jadhav, [2*]Dr. Rajeev Sharma

[1*] Phd. Research Scholar, Department of Physical Education, William Carey University, Shillong, Meghalaya, India
[2*] PhD Guide, Department of Physical Education, William Carey University, Shillong, Meghalaya, India

## Abstract

Virtual reality (VR) training can improve performance on virtual tasks, but its value for physical education depends on whether gains transfer to real movements and are retained. Existing studies often compare VR with no training rather than with equivalent conventional practice, or measure only short-term performance. This paper sets out the protocol for a two-group parallel randomised trial comparing a VR-supported programme with matched conventional instruction among school students aged 13–14 years. Both programmes provide 24 supervised 30-minute sessions over eight weeks. The primary outcome is dynamic balance, measured as the bilateral normalised lower-quarter Y-Balance composite score after training and adjusted for baseline. Secondary outcomes are retention of balance four weeks later, coordination, visual response time and accuracy, throwing accuracy and transfer to a novel physical task. For a standardised difference of 0.50 with 80% power and two-sided $\alpha = .05$, 64 analysable students per group are required; allowing for 15% loss gives a target of 152. Calculations show that dependence within teaching groups can raise this substantially: with 10 students per group and an intraclass correlation of 0.05, the target rises to 220. The analysis plan uses baseline-adjusted multilevel models with effect sizes and confidence intervals. The protocol separates virtual performance from real-world transfer and retention, and it follows CONSORT 2025 reporting guidance.

***Index Terms—****Virtual reality, physical education, motor skills, transfer, retention, randomised controlled trial, Y-Balance Test, sample size, study protocol*

## 1. Introduction

Motor competence supports participation in physical activity and sport through childhood and adolescence [1]. Physical education is the main setting in which students practise these skills, and teachers need methods that provide enough practice, feedback and challenge for students of differing ability. Virtual reality can offer repeatable, adjustable practice with immediate feedback. However, improvement within a virtual task does not necessarily mean that the real skill has improved. Transfer from virtual to physical performance depends on the similarity of task demands, the information available to the learner and the organisation of practice [2].

A systematic review of VR training in sport found that virtual environments can improve real-world motor skills, but that study quality was mixed and many comparisons were with no-training controls [3]. For schools, the relevant question is different: does VR-supported instruction improve real skills more than the conventional teaching it would supplement or replace, and do the gains last? This paper, derived from a doctoral thesis, contributes:

1) A trial design that compares VR-supported training with time-matched conventional instruction and separates virtual performance from real-world transfer and retention.
2) A specification of primary and secondary outcomes, assessment timing and measurement quality control.
3) A sample-size justification that includes the effect of clustering within teaching groups.
4) A prespecified analysis plan with effect sizes, confidence intervals and procedures for multiplicity and missing data.

## 2. Related Work

### *2.1 Evidence of Transfer from VR Training*

Several studies have shown transfer to physical performance. Gray trained 80 baseball players over six weeks and found that adaptive virtual batting practice improved real batting outcomes compared with other training conditions, including additional physical practice [4]. Michalski *et al.* found that VR table tennis training improved real table tennis performance in 57 participants compared with no training, with blinded assessment [5]. Markwell *et al.* found that commercially available VR putting practice and real putting practice both improved real putting in 68 participants, with no significant difference between groups [6]. In a school setting, an eight-week VR programme improved Y-Balance and coordination scores in middle-school students [7].

ISSN: 3139-0854

### *2.2 Evidence of Limited Transfer*

Other studies urge caution. Adaptations learned during VR obstacle avoidance transferred only partly to physical obstacles and were poorly retained after one week [8]. Experienced golfers showed increased error on their first real putt after VR practice [9], and VR dart training improved some perceptual-motor features while reducing performance [10]. Table I summarises these studies and the boundaries of their conclusions. Two gaps are clear: few studies compare VR with an equivalent amount of conventional practice, and few measure retention after training ends.

TABLE I SELECTED EVIDENCE ON TRANSFER FROM VR TRAINING

| Study | Task and sample | Comparison | Main finding | Boundary of interpretation |
|---|---|---|---|---|
| Gray [4] | Baseball batting; 80 players | Adaptive VR vs other VR, physical and control | Transfer to real batting | Specific task and adaptive system |
| Michalski *et al.* [5] | Table tennis; 57 adults | VR vs no training | Improved real performance | No matched physical practice |
| Markwell *et al.* [6] | Golf putting; 68 adults | VR vs real practice | Both improved; no group difference | Short duration; not equivalence |
| Weber *et al.* [8] | Obstacle avoidance | VR adaptation to physical task | Limited transfer and retention | Locomotor task, not school PE |
| Harris *et al.* [9] | Golf putting; experienced golfers | VR vs real environment | Initial real-world error after VR | Immediate transition effect |
| Yanar Tunçel and Avşar [7] | Balance and coordination; 39 students | VR vs control | Improved Y-Balance and KTK scores | Small sample; post-test only |

## 3. Objectives and Hypotheses

The general objective is to examine the impact of a structured VR-supported training programme on selected motor skills of PE students. The primary null hypothesis is that there is no difference between groups in post-training dynamic balance after adjustment for baseline; the research hypothesis is that the VR-supported group improves more. Table II lists the specific objectives and the outcomes that address them.

TABLE II OBJECTIVES AND CORRESPONDING OUTCOMES

| No. | Objective | Outcome and time point |
|---|---|---|
| O1 | Compare post-training dynamic balance between groups | Y-Balance composite at T1 (primary) |
| O2 | Determine whether balance gains are retained | Y-Balance composite at T2 |
| O3 | Compare coordination and visual response | KTK score; response time and accuracy at T1 and T2 |
| O4 | Compare movement accuracy and object control | Mean radial throwing error at T1 and T2 |
| O5 | Test transfer to an untrained physical task | Changed-target contacts at T1 and T2 |
| O6 | Describe engagement, confidence and experience | Questionnaire items and interview themes |
| O7 | Document exposure, fidelity, discomfort and feasibility | Attendance, active minutes, checklists, symptom logs |

## 4. Methods

### *4.1 Design and Timeline*

The study is a two-group parallel randomised educational trial with assessments at baseline (T0), after training (T1) and at follow-up (T2) (Fig. 1). T0 takes place within seven days before training and before allocation is revealed. T1 takes place two to five days after the final session, to avoid immediate fatigue, and T2 takes place 28 ± 7 days after T1, during which no organised study training is provided. The protocol and statistical analysis plan are finalised and registered before recruitment, and reporting follows CONSORT 2025 [11].

ISSN: 3139-0854

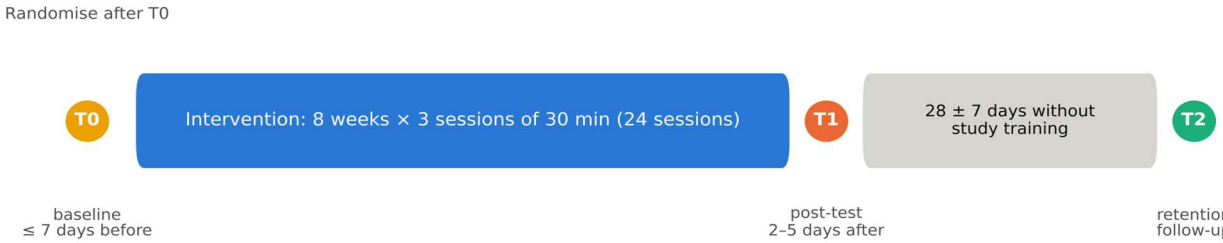



**Fig. 1.** Assessment timeline. Students are randomised after baseline (T0), complete eight weeks of training, and are reassessed after training (T1) and four weeks later (T2).

### *4.2 Participants and Allocation*

Participants are school PE students aged 13 years to 14 years and 8 months at baseline, with written parental permission and student assent. Low motor ability is not an exclusion criterion; students are excluded or deferred only when a condition or injury makes the tasks unsuitable, when equipment restrictions apply, or when they take part in another structured programme targeting the same outcomes. Students are allocated 1:1 within class strata using a concealed, computer-generated sequence. Assessors are unaware of group allocation wherever possible. Fig. 2 shows the planned participant flow.

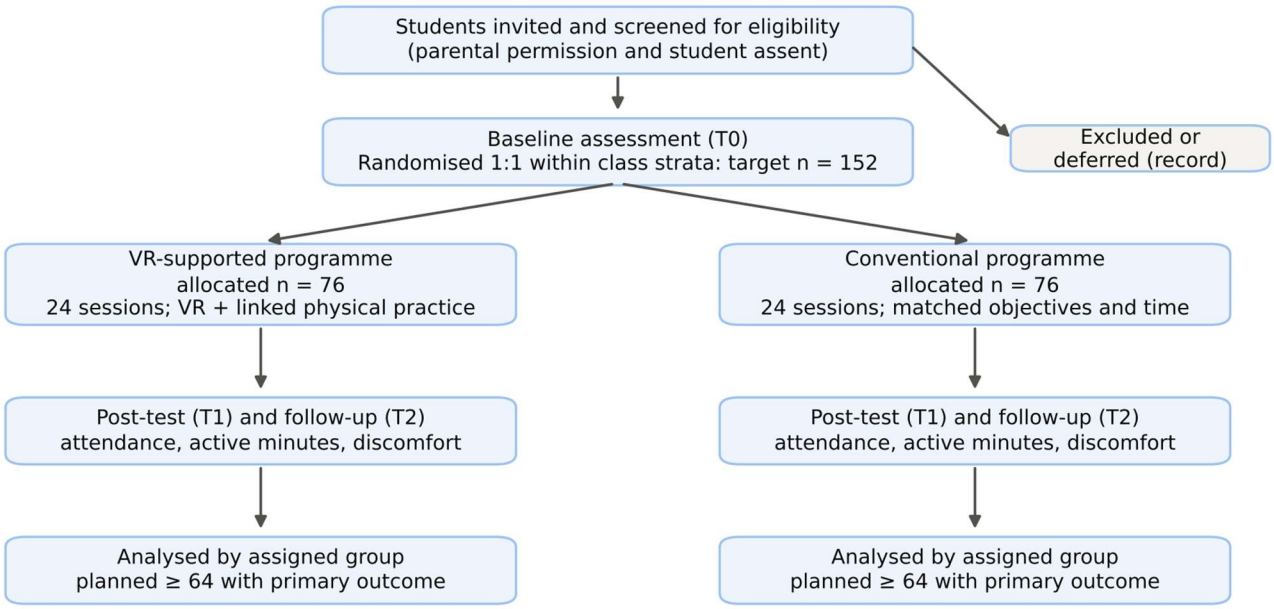



**Fig. 2.** Planned participant flow from screening to analysis. Numbers are recruitment targets, not achieved enrolment.

### *4.3 Interventions*

Both groups receive three supervised 30-minute sessions per week for eight weeks, in addition to the ordinary PE curriculum. Each session has a 5-minute warm-up, 20 minutes of targeted practice and a 5-minute cool-down. In the VR-supported group, the targeted period combines about 10 minutes of VR practice with 10 minutes of linked physical practice. The conventional group practises the same objectives for 20 minutes through physical targets, balance tasks, coordination exercises and teacher feedback. Equipment, software version, progression rules and permitted adaptations are documented using the TIDieR checklist [12].

### *4.4 Outcome Measures*

Table III lists the outcome measures. The primary outcome is the bilateral lower-quarter Y-Balance composite: for each leg, the sum of the maximum reaches in the anterior, posteromedial and posterolateral directions is divided by three times limb length and multiplied by 100, and the two legs are averaged:

$$\mathrm{YBT} = \tfrac{1}{2} \Sigma_{\mathrm{leg}} \left[ (A + PM + PL) / (3\,LL) \right] \times 100 \qquad (1)$$

ISSN: 3139-0854

Coordination is measured with the Körperkoordinationstest für Kinder (KTK), visual response with a two-choice reaction task, movement accuracy with mean radial error over ten throws, and transfer with a changed-target throwing task not practised in either programme. Scores from the VR software are recorded as supporting information only, because the trial's question concerns real-world performance.

TABLE III OUTCOME MEASURES AND SCORING

| Domain | Procedure | Reported score | Role |
|---|---|---|---|
| Dynamic balance | Instrumented lower-quarter Y-Balance; fixed familiarisation and valid-trial rules | Bilateral composite, % limb length (higher is better) | Primary at T1; secondary at T2 |
| Coordination | Authorised KTK battery, one edition | Overall score; subtests reported separately | Secondary |
| Visual response | 10 practice and 40 scored two-choice trials on fixed hardware | Median correct response time (ms); accuracy | Secondary if timing is validated |
| Movement accuracy | 3 practice and 10 scored soft-ball throws at 3 m | Mean radial error (cm; lower is better) | Exploratory |
| Physical transfer | 10 throws to changing left or right targets | Correct contacts out of 10 | Exploratory |
| Experience and exposure | Questionnaire, interviews, attendance and session logs | Engagement, confidence, active minutes, discomfort | Supporting |

### *4.5 Assessment Standardisation*

Assessments follow a fixed order with planned rest between domains (Fig. 3), and questionnaires follow the physical tests. The same equipment, instructions, limb-length method and scoring rules are used at every time point. Students are asked to avoid unusually strenuous activity beforehand and to wear comparable footwear; recent sleep, illness, pain and strenuous activity are recorded but are not used as unplanned reasons to remove results. Visits may be split into two appointments within the same window to prevent fatigue, using the same arrangement for each participant across visits.




**Fig. 3.** Fixed order of real-world assessments at T0, T1 and T2. Each domain is scored separately so that change in one skill does not conceal lack of change in another.

Assessors are trained and their agreement is checked in a pilot before the main study, with reliability reported as intraclass correlation coefficients using an explicitly stated model [13]. Discomfort and cybersickness symptoms are recorded at every session [14].

## 5. Sample Size

### *5.1 Individual-Level Calculation*

For two independent equal groups, a two-sided α of .05, 80% power and a standardised difference of 0.50, a noncentral *t* calculation requires 64 analysable students per group (Fig. 4a). Allowing for 15% loss gives 64/0.85 ≈ 76 per group, or 152 in total. The standardised difference of 0.50 is a planning assumption, not an observed effect; the meaningful difference will be translated into Y-Balance units before recruitment, as recommended for sample-size justification [15]. A difference of 0.40 would require 100 analysable students per group, and 0.60 would require 45. At 90% power, 0.50 requires 86 per group.

### *5.2 Effect of Clustering*

Students practise in classes and teaching groups, so their outcomes may be correlated. The design effect for an average cluster size $m$ and intraclass correlation ρ is

$$DE = 1 + (m - 1)\,\rho \qquad (2)$$

and the required number of students is multiplied by DE. Fig. 4b shows the total recruitment target when 64 analysable students per group are inflated by DE and by 15% loss. With 10 students per teaching group, the target

ISSN: 3139-0854

rises from 152 to 178 at ρ = 0.02, 220 at ρ = 0.05 and 288 at ρ = 0.10; with 15 per group and ρ = 0.10, it reaches 362. Individual randomisation within classes reduces but does not remove this dependence, so the final target will be recalculated with the confirmed teaching structure. If the required number cannot be recruited, the study will be reported as a feasibility trial [16].

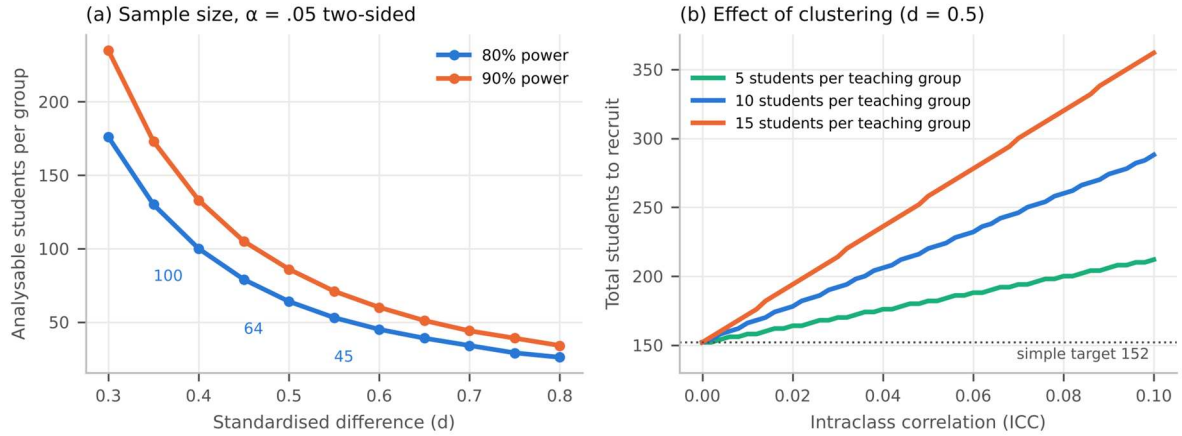



**Fig. 4.** (a) Analysable students per group required to detect a standardised difference *d* at 80% and 90% power. (b) Total recruitment target for *d* = 0.5 and 80% power as a function of intraclass correlation and teaching-group size, including 15% loss (Eq. 2).

## 6. Statistical Analysis

The primary analysis compares T1 Y-Balance scores between groups using a linear mixed model with group and baseline score as fixed effects, class stratum as a design term and teaching group as a random effect. Results are reported as the adjusted mean difference with a 95% confidence interval and a standardised effect size. Secondary outcomes are analysed with outcome-appropriate models: linear models for continuous scores, models for counts for the transfer task, and group-by-time models across T0, T1 and T2 for retention. Confirmatory secondary comparisons use Holm adjustment; exploratory analyses are labelled as such. The main analysis follows the assigned group. Missing outcomes are handled with methods valid under missing-at-random assumptions, with sensitivity analyses for departures from this assumption, and per-protocol analyses based on active practice time are supplementary. Table IV summarises the plan.

TABLE IV RESEARCH QUESTIONS AND PLANNED ANALYSES

| Question | Outcome and time | Planned analysis |
|---|---|---|
| Does VR-supported training improve balance more than conventional teaching? | Y-Balance at T1 | Baseline-adjusted mixed model; difference and 95% CI |
| Are balance gains retained? | Y-Balance at T2 | Adjusted comparison; group-by-time model |
| Do coordination and visual response differ? | KTK, response time and accuracy at T1 and T2 | Outcome-specific models; Holm adjustment |
| Does learning transfer to untrained physical tasks? | Throwing error; changed-target contacts | Continuous and count models; exploratory |
| Is the programme practical and acceptable? | Exposure, fidelity, experience, discomfort | Descriptive summaries; thematic analysis |
| Are conclusions robust? | Primary and selected outcomes | Missing-data sensitivity; adherence analyses |

## 7. Ethics and Safety

The study requires institutional ethical approval, school permission, written parental permission and student assent. Participation is voluntary and students may withdraw at any time without consequence for their ordinary PE. Sessions are supervised, practice areas are cleared and equipment fit is checked. Any discomfort, dizziness, eye strain or nausea leads to a pause or stop, and all such events are recorded. Data are stored under participant codes with restricted access. After the trial, the conventional group may be offered access to the VR programme if it proves beneficial and safe.

## 8. Interpreting Possible Outcome Patterns

Because the trial measures acquisition, transfer and retention separately, several result patterns are possible, and each supports a different conclusion. Table V sets out how each pattern will be interpreted and the evidence

ISSN: 3139-0854

required, so that interpretation is fixed before the data are seen. In particular, gains in virtual scores without gains in physical tests will be read as task-specific learning rather than motor-skill improvement [2], and a non-significant difference will not be presented as evidence that the programmes are equivalent [6].

TABLE V POSSIBLE OUTCOME PATTERNS AND THEIR INTERPRETATION

| Pattern | Interpretation to examine | Evidence needed |
|---|---|---|
| Both programmes improve similarly | Shared practice may explain the gains | Adjusted group difference and 95% CI |
| VR-supported advantage at T1 | Possible short-term benefit of VR | Primary estimate, dose and measurement quality |
| Advantage remains at T2 | Possible retention during school life | Follow-up contrast, timing and missing data |
| Virtual gains without physical gains | Software-specific learning; no transfer | Physical test results and their precision |
| Faster responses with more errors | Possible speed–accuracy trade-off | Response time, errors and omissions |
| Wide interval around a small effect | Firm conclusion is premature | CI width and educationally meaningful range |
| Conventional programme performs better | Review tasks and delivery | Group estimate, exposure and fidelity |

## 9. Discussion

The protocol addresses the main weaknesses of earlier VR motor-training studies. The comparison group receives the same amount of targeted practice, so the trial tests the added value of VR rather than of extra practice. The primary outcome is a real-world test rather than a virtual score, an untrained transfer task tests generalisation, and the four-week follow-up tests retention [2], [8]. The sample-size analysis shows that clustering can make a simple calculation misleading; with 10 students per teaching group and an intraclass correlation of 0.02–0.10, the required sample is 17–89% larger than the individual-level estimate. Limitations include the difficulty of blinding students and teachers, possible novelty effects of VR, a single age group and a short follow-up. The findings will apply to the selected students, programme and tests and should not be generalised to all VR applications.

## 10. Conclusion

This protocol describes a randomised trial comparing VR-supported training with matched conventional instruction in physical education, using real-world dynamic balance as the primary outcome and measuring transfer and retention. It specifies outcomes, assessment timing, a sample-size justification that accounts for clustering, and a prespecified analysis plan. The trial is designed to show whether VR adds value to physical education beyond the practice it provides, and whether any gains last after training ends.

## References

[1] D. F. Stodden *et al.*, "A developmental perspective on the role of motor skill competence in physical activity: An emergent relationship," *Quest*, vol. 60, no. 2, pp. 290–306, 2008, doi: 10.1080/00336297.2008.10483582.

[2] D. E. Levac, M. E. Huber, and D. Sternad, "Learning and transfer of complex motor skills in virtual reality: A perspective review," *J. NeuroEng. Rehabil.*, vol. 16, Art. no. 121, 2019, doi: 10.1186/s12984-019-0587-8.

[3] S. C. Michalski, A. Szpak, and T. Loetscher, "Using virtual environments to improve real-world motor skills in sports: A systematic review," *Front. Psychol.*, vol. 10, Art. no. 2159, 2019, doi: 10.3389/fpsyg.2019.02159.

[4] R. Gray, "Transfer of training from virtual to real baseball batting," *Front. Psychol.*, vol. 8, Art. no. 2183, 2017, doi: 10.3389/fpsyg.2017.02183.

[5] S. C. Michalski, A. Szpak, D. Saredakis, T. J. Ross, M. Billinghurst, and T. Loetscher, "Getting your game on: Using virtual reality to improve real table tennis skills," *PLOS ONE*, vol. 14, no. 9, Art. no. e0222351, 2019, doi: 10.1371/journal.pone.0222351.

[6] L. T. Markwell, K. Cochran, and J. M. Porter, "Off the shelf: Investigating transfer of learning using commercially available virtual reality equipment," *PLOS ONE*, vol. 18, no. 10, Art. no. e0279856, 2023, doi: 10.1371/journal.pone.0279856.

[7] N. Yanar Tunçel and Z. Avşar, "Virtual reality experience in physical education and sports lessons: An investigation into its effect on the balance and coordination skills of middle school students," *PLOS ONE*, vol. 21, no. 8, Art. no. e0354258, 2026, doi: 10.1371/journal.pone.0354258.

[8] A. Weber *et al.*, "Limited transfer and retention of locomotor adaptations from virtual reality obstacle avoidance to the physical world," *Sci. Rep.*, vol. 12, Art. no. 19655, 2022, doi: 10.1038/s41598-022-24085-w.

[9] D. J. Harris *et al.*, "The effect of a virtual reality environment on gaze behaviour and motor skill learning," *Psychol. Sport Exerc.*, vol. 50, Art. no. 101721, 2020, doi: 10.1016/j.psychsport.2020.101721.

[10] S. A. Drew *et al.*, "The trade-off of virtual reality training for dart throwing: A facilitation of perceptual-motor learning with a detriment to performance," *Front. Sports Act. Living*, vol. 2, Art. no. 59, 2020, doi: 10.3389/fspor.2020.00059.

[11] S. Hopewell *et al.*, "CONSORT 2025 statement: Updated guideline for reporting randomised trials," *BMJ*, vol. 389, Art. no. e081123, 2025, doi: 10.1136/bmj-2024-081123.

[12] T. C. Hoffmann *et al.*, "Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide," *BMJ*, vol. 348, Art. no. g1687, 2014, doi: 10.1136/bmj.g1687.

[13] T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *J. Chiropr. Med.*, vol. 15, no. 2, pp. 155–163, 2016, doi: 10.1016/j.jcm.2016.02.012.

[14] D. Saredakis *et al.*, "Factors associated with virtual reality sickness in head-mounted displays: A systematic review and

ISSN: **3139-0854**

meta-analysis," *Front. Hum. Neurosci.*, vol. 14, Art. no. 96, 2020, doi: 10.3389/fnhum.2020.00096.
[15] D. Lakens, "Sample size justification," *Collabra Psychol.*, vol. 8, no. 1, Art. no. 33267, 2022, doi: 10.1525/collabra.33267.
[16] S. M. Eldridge *et al.*, "CONSORT 2010 statement: Extension to randomised pilot and feasibility trials," *BMJ*, vol. 355, Art. no. i5239, 2016, doi: 10.1136/bmj.i5239.