



Validity and Reliability of Interactive Virtual Reality in Assessing the Musculoskeletal System: a Systematic Review

Mohammed Gumaa^{1,2} · Alaaeldin Khaireldin^{3,4} · Aliaa Rehan Youssef^{1,5,6}

Accepted: 8 January 2021 / Published online: 29 January 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC part of Springer Nature 2021

Abstract

Purpose of Review Assessment of the musculoskeletal system requires consideration of its integrated function with the nervous system. This may be assisted by using valid and reliable methods that simulate real-life situations. Interactive virtual reality (VR) technology may introduce various auditory and visual inputs that mimic real-life scenarios. However, evidence supporting the quality and strength of evidence regarding the adequacy of its psychometric properties in assessing the musculoskeletal function has not been evaluated yet. Therefore, this study reviewed the validity and reliability of VR games and real-time feedback in assessing the musculoskeletal system.

Recent Findings Nine studies were included in quality assessment. Based on outcome measures, studies were categorized into range of motion (ROM), balance, reaction time, and cervical motion velocity and accuracy. The majority of the studies were of moderate quality and provided evidence of VR adequate concurrent and, in some cases, known-groups validity. Also, VR showed high intra-rater reliability for most of the measured outcomes.

Summary Based on the included studies, there is a limited promising evidence that interactive VR using games or real-time feedback is highly valid and reliable in assessing ROM in asymptomatic participants and patients with chronic neck pain and radial fracture. For the remaining outcomes, evidence is limited to draw a robust conclusion. Future studies are recommended to test VR psychometric properties in different patients' population using a rigor research methodology.

Keywords Exergaming · Psychometric properties · Musculoskeletal · Simulators

Introduction

COVID-19 pandemic and its associated lockdown have emphasized the importance of home-based approaches and telemedicine in delivering healthcare worldwide [1, 2]. There are several available technologies that can be employed to achieve this purpose, including serious gaming, gamification, or exergaming, using virtual reality (VR) [3]. VR may offer an attractive, economic telerehabilitation tool to assess and treat patients remotely, monitor their progress, and hence, customize treatment parameters accordingly [4, 5]. This technology has been used previously in the rehabilitation of different patients' populations [6•, 7, 8, 9•, 10]. Evidence support its non-inferiority to other treatment protocols, although it may be better in improving patients' satisfaction and adherence to rehabilitation as well as reduce treatment costs when used remotely [6•, 11]. VR has also been used as an assessment tool using various off-the-shelf and custom-made consoles; however, its psychometric property adequacy has not been established. As VR can simulate real-life scenarios, it may improve office-based assessment by considering various

✉ Aliaa Rehan Youssef
aliaa.rehan@gmail.com

Mohammed Gumaa
m.gumaa.bakry@gmail.com

Alaaeldin Khaireldin
alaaeldin.khaireldin@gmail.com

¹ Department of Physical Therapy for Musculoskeletal Disorders and Surgery, Faculty of Physical Therapy, Cairo University, Giza, Egypt

² TRUST Research Center, Cairo, Egypt

³ Department of Physical Therapy for Musculoskeletal Disorders and Surgery, Faculty of Physical Therapy, Deraya University, Minya, Egypt

⁴ Department of Physical Therapy, Banha University Hospitals, Banha University, Banha, Qalyubia, Egypt

⁵ Department of Physical Therapy for Musculoskeletal Disorders and Surgery, Faculty of Physical Therapy, Ahrum Canadian University, Giza, Egypt

⁶ Nasr City, Egypt

distractors that exist in real life, and hence, patients' outcome may reflect actual reality. In addition, it may be used to assess the integration between the neurocognitive and musculoskeletal systems, which is an important component in motor control that is claimed to play an important role in recurrence and chronicity of a few orthopedic disorders [12–15].

Previous systematic reviews evaluated the validity and reliability of VR consoles in assessing gait [16], balance [17], or as a clinical assessment tool in general orthopedic settings [18]. However, these reviews did not focus on VR interactive features. Thus, the aim of this study was to assess the quality of the studies that evaluated the validity and reliability of interactive VR games in assessing the musculoskeletal system.

Methods

This study was conducted according to the PRISMA guidelines. The protocol for this systematic review has been registered in the PROSPERO database (CRD42019121944).

Searching Strategy

Seven electronic databases (PubMed, CINAHL, Embase, Scopus, Web of Science, Cochrane, and IEEEExplore) were searched using different keywords and Boolean operators (Table 1). Searching was conducted from database inception until August 2019. An updated search was also conducted on the 29th of August 2020. Furthermore, all included article bibliographies were searched and forward snowballing was conducted in Scopus and Web of Science.

Article Selection and Eligibility

Three independent reviewers (MG, AK, and ARY) screened retrieved studies by title, abstract, and finally by reading through the whole article. Any disagreement was resolved by a consensus through discussion.

Observational or interventional studies were included if they investigated the validity and reliability of interactive VR games in assessing the musculoskeletal function in asymptomatic adults or patients with orthopedic disorders. Studies were excluded if the VR interactive features were not employed, patients had diseases other than orthopedic dysfunction, or if the articles were not in the English language. Furthermore, case reports and series as well as proof-of-concepts, reviews, and conference proceedings were excluded.

Data Extraction

For each eligible study, a standardized form was used to extract the objective, population, VR and reference instruments, and outcome measures as well as validity and reliability results (Table 2).

Quality Assessment

Three independent reviewers (MG, AK, and ARY) assessed studies quality using the “Critical Appraisal Tool”, which is composed of 13 items (five items to assess validity and reliability, four to assess validity only, and four to assess reliability only) [28]. Depending on information sufficiency, each item could be answered as “yes,” “no,” or “not applicable (N/A)”. Quality strength for validity studies was categorized as poor (0–2), fair (3–5), moderate (6–7), or high (8–9) (Table 3) [29]. For both intra- and inter-rater reliability, the maximum possible score was 8 points. If one type of reliability was not assessed, relevant items were rated as N/A and were not included in the total quality score calculation. Moreover, N/A was adapted as an answer for item 6, which inquired about varying the order of participant examination during reliability assessment, as this was not necessary and, hence, was not considered in most of the reviewed studies. Therefore, reliability study quality was scored as percentage rather than a total sum, in order to adjust for the N/A answers. Any quality scoring disagreement was resolved by consensus and discussion. Before the actual quality scoring, all reviewers completed a standardized training on using the tool on four manuscripts that were not included in this review.

Results

Search Results

After duplicate removal, electronic and manual search retrieved 4300 articles. After full screening by title, then abstract, and finally by reading through the full articles, five studies were found eligible for quality assessment. After snowballing duplicate removal and full screening, additional four articles were included. Thus, a total of nine articles were assessed for quality (Fig. 1). These articles were categorized based on the outcome measured into the following: range of motion (ROM) ($n = 3$) [19–21], postural sway and balance ($n = 3$) [22–24], reaction time ($n = 1$) [25], and velocity and accuracy ($n = 2$) [26, 27] (Table 2).

The most commonly used VR system was the Nintendo Wii and its accessory (the Wii balance board, WBB) [22–25]. Four studies used a head-mounted display [19, 20, 26, 27] and one study used a custom-made hardware [21]. All reviewed studies used custom software, except for two studies that used off-the-shelf games [22, 24].

Quality Assessment

The quality of the included studies ranged from fair to high. VR testing procedure was adequately described. Furthermore, the time interval between VR and reference standard testing was

Table 1 Keywords and Boolean operators of the used searching strategy

| Databases | Keywords and Boolean operators |
|---|--|
| PubMed Scopus Web of Science CINAHL Embase IEEEExplore | (virtual OR “virtual reality” OR “virtual environment” OR “computer-based” OR “computer-interface” OR cyberspace OR “artificial intelligence” OR “computer simulat*” OR simulator OR Exergam* OR “active video gam*” OR “interactive gam*” OR game OR gaming OR X-box OR Kinect OR Nintendo OR Wii) AND (orthopedic* OR orthopaedic* OR musculoskeletal) AND (posture OR deformity OR pain OR gait OR kinematics OR kinetics OR “motion analysis” OR function OR balance OR “range of motion” OR “ROM” OR move* OR motion) AND (validity OR reliability OR “gold standard” OR sensitivity OR specificity OR “ROC” OR “receiver operating characteristic” OR assess* OR evaluat* OR screening OR measurement* OR quantification OR tracking OR detect*) NOT (stroke OR cerebral palsy OR cancer OR tumor OR carcinoma OR oncology OR neurologic* OR dentistry OR obesity OR children OR pediatric) |
| Cochrane (with filter of clinical trials) | <p>#1 virtual OR virtual reality OR virtual environment</p> <p>#2 computer-based OR computer-interface</p> <p>#3 #1 OR #2</p> <p>#4 cyberspace OR artificial intelligence</p> <p>#5 #3 OR #4</p> <p>#6 computer simulat* OR simulator</p> <p>#7 #5 OR #6</p> <p>#8 Exergam* OR active video gam* OR interactive gam* OR game OR gaming</p> <p>#9 #7 OR #8</p> <p>#10 X-box OR Kinect</p> <p>#11 Nintendo OR Wii</p> <p>#12 #10 OR #11</p> <p>#13 #9 OR #12</p> <p>#14 orthopedic* OR orthopaedic* OR musculoskeletal</p> <p>#15 #13 AND #14</p> <p>#16 validity OR reliability</p> <p>#17 sensitivity OR specificity</p> <p>#18 gold standard</p> <p>#19 ROC OR “receiver operating characteristic”</p> <p>#20 assess* OR evaluat* OR screening OR measurement* OR quantification OR tracking OR detect*</p> <p>#21 #16 OR #17 OR #18 OR #19 OR #20</p> <p>#22 #21 AND #15</p> <p>#23 posture OR deformity OR pain OR gait OR kinematics OR kinetics OR motion analysis OR function OR balance OR range of motion OR ROM OR move* OR motion</p> <p>#24 #22 AND #23</p> |

sufficient, except for one study [24]. The VR instrument was part of the reference standard in two studies [19, 20]. For reliability, none of the studies blinded the findings among examiners or to prior findings. Only two studies stated examiners’ qualifications [21, 24]. In one study, information were insufficient regarding statistical analysis [24]. Four studies did not provide sufficient information on participant withdrawal [22–25] (Tables 4 and 5).

Range of Motion

Two studies assessed cervical ROM [19, 20] whereas one study measured wrist active ROM [21]. All studies

used customized games. The reference standard for cervical ROM was the electromagnetic tracking system, whether as a standalone instrument or combined with virtual environment projected using a head-mounted display [19, 20]. Cervical range was assessed in asymptomatic healthy volunteers only [19] or in a sample of patients with chronic neck pain and asymptomatic controls [20]. For wrist ROM assessment, CameraWrist captured wrist flexion-extension (F-E) movement in 15 patients with distal radial fracture and 15 healthy controls. The universal goniometer was used as a reference standard [21] (Table 2).

Table 2 Studies' characteristics

| Study | Objective | Population | VR instrument | Reference instrument | Outcomes | Validity | Reliability |
|---|---|--|--|--|---|---|--|
| <p>Range of motion</p> <p>Bahat et al. [19]</p> | <p>To compare the intra- and inter-tester reliability of cervical ROM VR-based and conventional assessment</p> | <p>34 asymptomatic healthy volunteers (age: 22 to 55; 28.6±7.5 years)</p> | <p>Head mounted display (I-glasses HRV Pro, Virtual Realities) combined with an electromagnetic tracking system (Fastrak, Polhemus)</p> <p>Customized game</p> | <p>Electromagnetic tracking system (Fastrak, Polhemus)</p> | <p>Cervical ROM in all directions</p> | <p>• VR rotation ROM was significantly greater than that recorded by conventional instrument.</p> <p>• Bland and Altman: - F-E mean difference = 7.2 (r95%=24.5).</p> <p>- Rotation mean difference = 16.1 (r95% = 23.7)</p> <p>• VR ROM in all directions was significantly greater than that recorded by conventional instrument.</p> <p>• Both instruments significantly predicted cervical pain, with each unit increase in ROM was associated with decreased probability of being symptomatic</p> <p>• VR instrument: - F-E: odds ratio=0.96 (0.93–0.99), cut off = 133.3, sensitivity = 0.88, specificity = 0.43, AUC = 0.70</p> <p>- Rotation: odds ratio = 0.92 (0.88–0.97), cut off = 146, sensitivity = 0.72, specificity = 0.79, AUC = 0.79.</p> | <p>• Number of testers: two</p> <p>• Test retest duration: two days</p> <p>• Bland and Altman plot: - Intra-rater reliability: F-E: r95%=22.6; Rotation: r95%=17</p> <p>- Inter-rater reliability: F-E: r95%=21.2; Rotation: r95%=15</p> <p>Not assessed</p> |
| <p>Bahat et al. [20]</p> | <p>• Compare cervical ROM measured in patients with chronic neck pain and asymptomatic volunteers using VR-based assessment.</p> <p>• Compare cervical ROM assessed with VR-based and conventional assessment</p> <p>• Establish the diagnostic value of Cervical ROM</p> | <p>67 participants:</p> <ul style="list-style-type: none"> • 25 patients with chronic neck pain (age: 39.0±12.7) years • 42 asymptomatic volunteers (age 35.3±12.4 years) | <p>Head mounted display (I-glasses HRV Pro, Virtual Realities) combined with an electromagnetic tracking system (Fastrak, Polhemus)</p> <p>Customized game</p> | <p>Electromagnetic tracking system (Fastrak, Polhemus)</p> | <p>Cervical ROM in all directions</p> | <p>• VR ROM in all directions was significantly greater than that recorded by conventional instrument.</p> <p>• Both instruments significantly predicted cervical pain, with each unit increase in ROM was associated with decreased probability of being symptomatic</p> <p>• VR instrument: - F-E: odds ratio=0.96 (0.93–0.99), cut off = 133.3, sensitivity = 0.88, specificity = 0.43, AUC = 0.70</p> <p>- Rotation: odds ratio = 0.92 (0.88–0.97), cut off = 146, sensitivity = 0.72, specificity = 0.79, AUC = 0.79.</p> | <p>• Number of testers: one</p> <p>• Test retest duration: within the same session (for the control group only)</p> <p>• F: r=0.488; CI: 0.003–0.791</p> |
| <p>Emi et al. [21]</p> | <p>Establish the psychometric properties of CameraWrist tracker as a dynamic assessment tool measure active wrist ROM</p> | <p>30 participants:</p> <ul style="list-style-type: none"> • 15 patients with distal radius fractures (age: 45.0±15.2 years) • 15 matched healthy participants (45.5±15.7 years) | <p>CameraWrist tracker</p> <p>Customized game</p> | <p>Universal wrist goniometer</p> | <p>Active wrist flexion and extension</p> | <p>• Wrist ROM was significantly less in the patients' group than in healthy individuals</p> <p>• Systemic error: - No significant differences in</p> | <p>• Number of testers: one</p> <p>• Test retest duration: within the same session (for the control group only)</p> <p>• F: r=0.488; CI: 0.003–0.791</p> |

Table 2 (continued)

| Study | Objective | Population | VR instrument | Reference instrument | Outcomes | Validity | Reliability |
|---|--|---|--|--|--|---|---|
| | | | | | | measuring wrist extension between the two instruments regardless to participants grouping - The two methods showed significant difference in wrist flexion in the patient group only with higher mean flexion range measured by the VR. | <ul style="list-style-type: none"> E: $r=0.924$; CI: 0.794–0.973 |
| | | | | | | <ul style="list-style-type: none"> Correlation: <ul style="list-style-type: none"> - In patients: both instruments significantly correlated for wrist flexion ($r=0.654$) and extension ($r=0.641$). - In healthy participants, significant correlation was found in wrist extension ($r=0.761$), whereas wrist flexion showed weak non-significant correlation ($r=0.311$) | |
| Balance and postural sway Wikstrom [22] | Establish the validity, intra- and inter-session reliability of the Wii Fit games in assessing balance | <ul style="list-style-type: none"> 45 participants aged: 27.0 ± 9.8 years 22 of the 45 participants had at least a history of one previous lower extremity disorder. | WBB Off the shelf games: basic balance, agility, steadiness, single leg stance, deep breathing, tree, standing knee, palm tree, single leg extension, single leg twist, sideways leg lift | <ul style="list-style-type: none"> Force platform (model 4060NC; Bertec Corporation, Columbus, OH) The star excursion test | <ul style="list-style-type: none"> Average postural sway in AP and ML directions COP excursion | <ul style="list-style-type: none"> Non-significant correlation between the VR and the other two instruments, except for: <ul style="list-style-type: none"> Walking activity with the resultant COP excursion ($r=0.32$) Single leg twist activity with the anterior star excursion balance test score ($r=0.21$) Palm tree activity with | <ul style="list-style-type: none"> Number of testers: one Test retest duration: within the same session and within 1 week Intra-session: <ul style="list-style-type: none"> No significant difference; ICC = 0.39–0.80 Inter-sessions: <ul style="list-style-type: none"> Non-significant differences in most of the games; ICC = 0.29–0.74 |

Table 2 (continued)

| Study | Objective | Population | VR instrument | Reference instrument | Outcomes | Validity | Reliability |
|--|---|--|------------------------------|---|-----------------------------|---|---|
| Jorgensen et al. [23] | Assess the validity and intra-rater reliability of WBB in assessing stillness and agility | <ul style="list-style-type: none"> • 30 community-dwelling elderly (age: 71.8±5.1 years) | WBB Customized game | Force platform (Good Balance, Metitur, Finland) | Stillness and agility | the posterolateral star excursion balance test score ($r=0.29$) <ul style="list-style-type: none"> • Stillness test: - COP velocity (mm/s): $r=-0.65$ - COP confidence ellipse (mm²): $r=-0.82$ - COP area (mm²): $r=-0.76$ - COP velocity moment (mm²/s): $r=-0.74$ • Agility test: - COP velocity (mm/s): $r=-0.29$ - COP confidence ellipse (mm²): $r=-0.23$ - COP area(mm²): $r=-0.29$ - COP velocity moment (mm²/s): $r=-0.29$ • No significant difference in the WBB scores between individuals with a history of concussion or lower limb injury and those without • Significant weak correlation between the two instruments ($r=0.35$) | <ul style="list-style-type: none"> • Number of testers: one • Test retest duration: 90±10 min • Stillness test: No significant differences - ICC=0.87 (95% CI: 0.75–0.94), - LoA (%)=17 • Agility test: Significant differences - ICC=0.69 (95% CI: 0.44–0.84) - LoA (%)=14.4 |
| Guzman and Aktan [24] | Establish the validity and reliability of WBB in assessing postural stability | 91 collegiate football players (age range: 18–25 years; mean age: 20 years) | WBB Wii training software | A modified balance error scoring system | Balance | <ul style="list-style-type: none"> • No significant difference in the WBB scores between individuals with a history of concussion or lower limb injury and those without • Significant weak correlation between the two instruments ($r=0.35$) | <ul style="list-style-type: none"> • Number of testers: one • Test retest duration: within the same session • Cronbach's alpha=0.96 |
| Reaction time Jorgensen et al. [25] | Establish age-specific discriminative validity of reaction time measured by WBB to distinguish between young and old age groups. Assess test-retest reproducibility of the WBB in measuring reaction time | 50 healthy participants: <ul style="list-style-type: none"> • 25 Young adults (age: 24.7±3 years) • 25 Elderly (age: 74.2±6.1 years) | WBB Customized game | N/A | Hand and foot reaction time | <ul style="list-style-type: none"> • Reaction time of hands and feet differed based on age. - Hand reaction time: -170.7 ms (95%CI: -209.4 to -132.0) - Foot reaction time: -224.3 ms (95%CI -274.6 to -173.9) | <ul style="list-style-type: none"> • Number of testers: one • Test retest duration: within the same session and within 1–7 days • Within day reliability: - Hand reaction time: Young: ICC=0.907 (0.806–0.962) - Old: ICC=0.841 (0.659–0.932) - Foot reaction time: |

Table 2 (continued)

| Study | Objective | Population | VR instrument | Reference instrument | Outcomes | Validity | Reliability |
|-------------------|--|---|---|----------------------|--|---|---|
| Bahat et al. [26] | <p>Accuracy and velocity</p> <p>Compare cervical kinematics between patients with chronic neck pain and asymptomatic individuals using an interactive VR system.</p> <p>To identify the sensitivity and specificity of the proposed system and clinically significant values.</p> <p>To investigate a novel VR based method to assess cervical motion accuracy</p> | <p>55 participants:</p> <ul style="list-style-type: none"> • 33 patients with chronic neck pain (age: 37.56±9.95 years) • 22 asymptomatic controls (age: 33.0±6.78 years) | <p>Head Mounted Display with a built-in tracker (Wrap™)</p> <p>1200VR by Vuzix, New York</p> <p>Customized game</p> | N/A | <ul style="list-style-type: none"> • Velocity: <ul style="list-style-type: none"> - Peak velocity - Mean velocity - NVP - TTTP • Head movement accuracy | <ul style="list-style-type: none"> • All velocity and accuracy measured variables were significant predictors; except for accuracy error X in extension, accuracy error Y in right rotation, TTTP and accuracy error $7Y$ in left rotation. • Peak velocity in all directions: Sensitivity: 0.76–1.00; Specificity: 0.65–1.00; AUC: 0.87–1.00. • Mean velocity in all directions: Sensitivity: 0.80–1.00; specificity: 0.65–1.00; AUC: 0.81–1.00. • NVP in all directions: sensitivity: 0.36–0.97; specificity: 0.45–1.00; AUC: 0.69–0.97. • TTTP in all directions: sensitivity: 0.65–0.95; specificity: 0.45–0.99; AUC: 0.60–1.00. • Accuracy error x in all directions: | <ul style="list-style-type: none"> • Young: ICC=0.919 (0.840–0.969) • Old: ICC=0.922 (0.842–0.968) • Inter-sessions: <ul style="list-style-type: none"> - Hand reaction time • Young: ICC=0.856 (0.663–0.935) • Old: ICC=0.777 (0.468–0.900) - Foot reaction time: <ul style="list-style-type: none"> • Young: ICC=0.870 (0.748–0.951) • Old: ICC=0.868 (0.742–0.952) |

Table 2 (continued)

| Study | Objective | Population | VR instrument | Reference instrument | Outcomes | Validity | Reliability |
|-------------------|---|--|---|----------------------|---|--|---|
| Bahat et al. [27] | Assess the inter-rater reliability of measuring cervical kinematics using VR based system in asymptomatic individuals. To identify the value of MDC of measuring cervical kinematics using the proposed system | 46 asymptomatic volunteers (age: 29.95±8.09, range: 23–49 years) | Head mounted display (I-glasses HR V Pro, Virtual Realities) combined with an electromagnetic tracking system (Fastrak, Polhemus) Customized game | N/A | <ul style="list-style-type: none"> • Peak velocity • Mean velocity • NVP • TTPP | Sensitivity: 0.37–0.93; specificity: 0.36–0.97; AUC: 0.55–0.93. • Accuracy error y in all directions: Sensitivity: 0.54–0.89; specificity: 0.60–1.00; AUC: 0.73–0.96 N/A | <ul style="list-style-type: none"> • Number of testers: two • ICC: - Peak velocity: 0.86–0.93 - Mean velocity: 0.67–0.84 - NVP: 0.39–0.78 - TTPP: 0.43–0.70 • MDC: - Peak velocity: 41.19–52.72 - Mean velocity: 19.73–24.97 - NVP: 0.74–2.26 - TTPP: 7.31–11.86 |

AP, anteroposterior; AUC, area under the curve; CI, confidence interval; COP, center of pressure; F-E, flexion-extension; ICC, intra-class correlation coefficient; LoA, limits of agreement; MDC, minimal detectable change; ML, mediolateral; ms, milliseconds; N/A, not applicable; NVP, number of velocity peaks; r95%, repeatability within 95% limits of agreement; ROM, range of motion; TTPP, time to peak percentage; VR, virtual reality; WBB, Wii balance board

Table 3 Quality scores categories. For validity studies, score points and its categories are shown. For reliability studies, score percentages are given. When a non-applicable answer was chosen, calculation was adjusted to exclude the relevant question

| Score category | Score points* | Score percentage** |
|----------------|---------------|--------------------|
| Poor | 0–2 | 0–22% |
| Fair | 3–5 | 23–56% |
| Moderate | 6 or 7 | 57–78% |
| High | 8 or 9 | > 78% |

*Score points of each category as proposed by Prowse et al. [29]

**Our new score scale calculated as percentage

In the three studies, customized games were used. For cervical ROM, each participant was instructed to target a fly that appeared at random locations using a spray canister that was controlled by head motion [19, 20]. For wrist ROM, the participant needed to navigate an airplane upward or downward by extending and flexing the wrist, respectively [21].

Validity

For cervical ROM, the mean difference between the two instruments was 7.2° (limits of agreement (LoA) = 24.5°) for F-E movement and 16.1° (LoA = 23.7°) for rotation [19]. VR distinguished patients with cervical pain from asymptomatic

volunteers, with each degree increase in F-E range (above 133°) associated with reduced odds of having neck pain by 4% on average. For rotation ROM, each degree increase above 146° was associated with an average reduction of chronic neck pain odds by 8% [20] (Table 2).

Wrist ROM measured by VR and goniometer was not significantly different, except for wrist flexion in patients with radial fracture. The two instruments correlated significantly for all measures ($r=0.64$ – 0.76), except for wrist flexion in control participants ($r=0.31$) [21] (Table 2).

Cervical ROM studies' quality [19, 20] was moderate (7/9), whereas that of wrist ROM was high (9/9) [21]. The quality of cervical ROM studies was limited by using the VR head-mounted display as a part of the reference test tool (Table 4).

Reliability

For inter-rater reliability, cervical range LoAs between the two examiners was 21.2° and 15° for F-E and rotation movements, respectively. For intra-rater reliability, the LoA was 22.6° for F-E and 17° for rotation ranges [19]. Wrist ROM inter-rater reliability was not examined, whereas its intra-rater reliability within the same session was low for flexion ($r=0.49$) and high for extension ($r=0.92$) movements in healthy individuals [21] (Table 2).

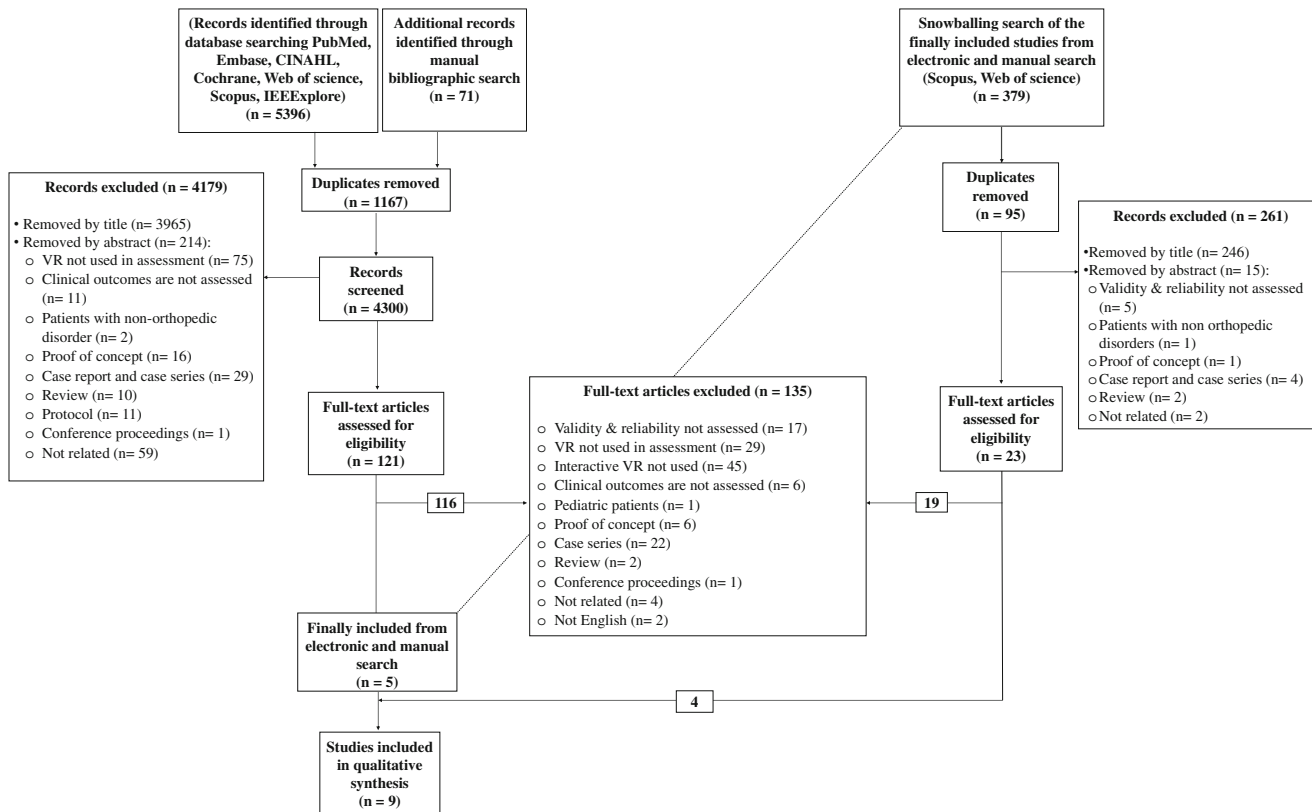


Fig. 1 A flow diagram showing article selection, screening, and exclusion. VR, virtual reality

Table 4 Detailed quality scores for the eligible validity studies

| | Sample characteristics described with details | Raters competence described | Index test sufficiently described | Study withdrawals explained | Appropriate statistical methods | Reference test explained | Adequate time between index and reference tests | Independence of the reference test from the index test | Reference test sufficiently described | Score (quality) |
|----------------------------------|---|-----------------------------|-----------------------------------|-----------------------------|---------------------------------|--------------------------|---|--|---------------------------------------|-----------------|
| ROM | | | | | | | | | | |
| Bahat et al. [19] | Yes | No | Yes | Yes | Yes | Yes | Yes | No | Yes | 7 Moderate |
| Bahat et al. [20] | Yes | No | Yes | Yes | Yes | Yes | Yes | No | Yes | 7 Moderate |
| Eini et al. [21] | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | 9 High |
| Balance and postural sway | | | | | | | | | | |
| Wikstrom [22] | Yes | No | Yes | No | Yes | Yes | Yes | Yes | Yes | 7 Moderate |
| Jorgensen et al. [23] | Yes | No | Yes | No | Yes | Yes | Yes | Yes | Yes | 7 Moderate |
| Guzman and Aktan [24] | Yes | Yes | No | No | No | Yes | No | Yes | Yes | 5 Fair |

Poor: 0–2; fair: 3–5; moderate: 6 or 7; high: 8 or 9

Table 5 Detailed quality scores for the eligible reliability studies

| | Sample characteristics described with details | Raters competence described | Index test sufficiently described | Study withdrawals explained | Appropriate statistical methods | Between observer blinding | Within observer blinding | Order of examination varied | Time between testing and theoretical stability | Score percentage (quality) |
|----------------------------------|---|-----------------------------|-----------------------------------|-----------------------------|---------------------------------|---------------------------|--------------------------|-----------------------------|--|----------------------------|
| ROM | | | | | | | | | | |
| Bahat et al. [19] | Yes | No | Yes | Yes | Yes | No | No | Yes | Yes | 6/9 (67%) Moderate |
| Eini et al. [21] | Yes | Yes | Yes | Yes | Yes | N/A | No | N/A | Yes | 6/7 (86%) High |
| Balance and postural sway | | | | | | | | | | |
| Wikstrom [22] | Yes | No | Yes | No | Yes | N/A | No | N/A | Yes | 4/7 (57%) Moderate |
| Jorgensen et al. [23] | Yes | No | Yes | No | Yes | N/A | No | N/A | Yes | 4/7 (57%) Moderate |
| Guzman and Aktan [24] | Yes | Yes | No | No | No | No | No | N/A | Yes | 3/8 (38%) Fair |
| Reaction time | | | | | | | | | | |
| Jorgensen et al. [25] | Yes | No | Yes | No | Yes | N/A | No | N/A | Yes | 4/7 (57%) Moderate |
| Accuracy and velocity | | | | | | | | | | |
| Bahat et al. [27] | Yes | No | Yes | Yes | Yes | No | N/A | Yes | Yes | 6/8 (75%) Moderate |

N/A, not applicable

Poor: 0–22%; fair: 23–56%; moderate: 57–78%; high: > 78%

Cervical ROM study [19] quality was moderate (67%), while that for the wrist ROM was high (86%) [21]. The two studies did not employ blindness between testers or to previous recordings by the same tester. The cervical ROM study [19] also did not report the qualification of the raters (Table 5).

Balance and Postural Sway

Three studies assessed balance and postural sway [22–24]: one study included 45 participants ($n = 22$ individuals with a history of a least one lower limb musculoskeletal condition) [22], the second study enrolled 30 elderly [23], while the third study assessed 91 college football players [24] (Table 2).

The three studies assessed WBB psychometric properties compared to force platform [22, 23], star excursion test [22], and balance error scoring system [24]. One study used off-the-shelf Wii Fit games [22], while the second study mentioned the use of Wii training software without giving sufficient details [24]. The third study used a custom software to perform two tests: (1) the stillness test that assesses static balance and (2) the agility test to assess dynamic balance [23] (Table 2).

Validity

Overall, VR correlated weakly with reference instruments. For agility test, VR results weakly correlated with that of the force platform ($r = -0.23$ to -0.29) [23] and walking activity ($r = 0.32$) [22]. The star excursion test showed significant weak correlation with the single leg twist ($r = 0.21$) and palm tree ($r = 0.29$) games [22]. The VR also showed weak correlation with the balance error scoring system ($r = 0.35$) [24]. One study assessed VR known-groups validity to distinguish between individuals with a history of concussion or lower limb injury and those without, and showed no difference in WBB scores between different groups [24] (Table 2).

Two studies were of moderate quality (7/9), as both lacked adequate information regarding participants' withdrawal from the study [22, 23], whereas the third study was of a fair quality (5/9) as it did not sufficiently describe the VR test, study withdrawals, statistical analysis methods, and the time interval between the VR and reference testing [24] (Table 4).

Reliability

All studies assessed the intra-rater reliability within the same session [22–24]. In addition, one study assessed the inter-sessions reliability within 1 week [22]. Intra-session reliability ranged between weak and excellent (intra-class correlation coefficient (ICC) = 0.39 to 0.96) [22–24], whereas inter-session reliability ranged between weak and moderate (ICC = 0.29 to 0.74) [22] (Table 2). Studies' quality ranged between fair (38%) [24] and moderate (57%) [22, 23]. None of the studies blinded the raters to their previous ratings

[22–24] and only one study provided information regarding the raters' competency [24] (Table 5).

Assessment of Reaction Time

Hand and foot reaction time was compared between young adults ($n = 25$) and elderly ($n = 25$) using WBB and a custom software (Table 2). In this study, the WBB was placed in front of the participant, either on the floor to step (foot reaction time measurement) or was placed on the top of a table to allow pressure by the fist hand (hand reaction time). A customized game was used and the participants were asked to step or press on the WBB as soon as a green light appears randomly on the screen [25].

Validity

Concurrent validity was not assessed as no reference standard was used. The WBB significantly distinguished between participants from the two age groups, with a mean difference of 170.7 milliseconds (ms) for hand reaction time and 224.3 ms for foot reaction time [25] (Table 2).

Reliability

Within session, the reliability of WBB in measuring reaction time ranged from moderate to high for the hand (ICC = 0.66–0.96) and foot (ICC = 0.84–0.97) measurements. Inter-session reliability ranged from weak to high for the hand (ICC = 0.47–0.94) and foot (ICC = 0.74–0.95) (Table 2) [25]. The quality of this study was moderate (57%) as the authors did not provide sufficient information about the study withdrawals nor raters' competency. Moreover, the rater was not blinded for previous own recordings (Table 5).

Assessment of Movement Accuracy and Velocity

Cervical movement velocity and accuracy were assessed in a sample of asymptomatic participants ($n = 22$) and patients with chronic neck pain ($n = 33$) [27] as well as in asymptomatic participants only ($n = 46$) [26]. Both studies used a custom software and a head-mounted display to project the VR game; however, one study used a head display with a built-in tracker [27], while the other study used an external magnetic tracking system [26]. The custom game simulated an airplane that was controlled by the participants' head movement in four directions (flexion, extension, right, and left rotation) [26, 27]. None of the two studies used other reference tests. The velocity was quantified in terms of peak and mean velocity, number of achieved peak velocity, and time to peak percentage [26, 27]. The accuracy of the cervical motion was assessed in terms of X and Y axes error (difference between the target position and the participants' head position) [27] (Table 2).

Validity

VR distinguished asymptomatic participants and patients with chronic neck pain. All measured velocity and accuracy outcomes were significant predictors for chronic neck pain, except for the accuracy error *X* (during extension motion), accuracy error *Y* (for right and left rotation), and time to peak velocity (%) for left rotation. Velocity predictor sensitivity ranged from 0.36 to 1.00, whereas the specificity ranged from 0.45 to 1.00, and the area under the curve (AUC) ranged between 0.60 and 1.00. The sensitivity of the significant accuracy outcomes was 0.37–0.93, whereas the specificity ranged between 0.36 and 1.00, and the AUC ranged from 0.55 to 0.96 [27] (Table 2).

Reliability

Reliability was assessed only in asymptomatic participants. The velocity inter-rater agreement ranged between weak and high (ICC = 0.39–0.93) [26] (Table 2). This study quality was moderate (75%) [26], as the two examiners were not blinded to each other's readings. Furthermore, information regarding raters' qualification was insufficient (Table 5). No intra-rater reliability was established.

Discussion

This study systematically reviewed and critically appraised the quality of the studies that assessed the psychometric properties of using interactive VR systems to assess the musculoskeletal system. In these studies, VR games or real-time feedback were delivered using off-the-shelf or customized systems. The majority of the studies were of moderate quality and provided evidence of VR adequate concurrent and, in some cases, known-groups validity. Also, VR showed high intra-rater reliability for most of the measured outcomes.

The quality assessment tool used in this review seemed to be the most appropriate to assess the study design and the nature of the VR measurements under investigation. Other tools, such as The Quality assessment Diagnostic Accuracy Studies (QUADAS) and the Quality Appraisal of Diagnostic Reliability Studies (QAREL) could not be used as they are designed to assess the diagnostic studies that evaluate the ability of a specific test to detect or predict a target condition [28]. In the included studies, the VR was used as an objective tool to assess one clinical outcome that cannot formulate a diagnosis. However, the quality assessment tool used in this review did not consider a few important points. For example, sample selection and recruitment, testing order randomization, and sample size calculation.

The sample selected was convenient in eight reviewed studies [19–22, 24–27], and random in one [23]. Although

convenient sampling aid in obtaining adequate numbers for proper study power, yet it does not accurately represent the population to whom the results should be generalized [18].

Only one study randomized the testing tool order [23]. Non-randomization of testing order could have resulted in variability due to fatigue or learning effect. None of the studies provided training for the examiners on the VR system which may have resulted in variability affecting reliability measurements. Also, none of the studies had calculated their sample size nor calculated the power of their studies.

Validity

Limited evidence supports VR concurrent validity for the assessed outcomes. It should be emphasized that only two studies [20, 21] (out of the nine that investigated concurrent validity) enrolled patients, while the remaining included only asymptomatic participants. The VR concurrent validity was investigated only in five studies against a valid reference standard tool [19–23]. It is worth noting that evidence was of fair to moderate quality regarding WBB validity in measuring balance compared to force platform [22, 23], star excursion test [22], and balance error scoring system [24]. Yet, the balance error scoring system is subjective and invalid in measuring small balance deficits [30]. In addition, although evidence supports VR high validity for the stillness test compared to force platform in assessing standing static balance, yet VR interactive features was not used in this test [23], and hence, its results were not discussed in this review. On the other hand, VR low validity during dynamic agility test could be attributed to the difference between the interactive nature of VR and the static nature of force platform [23].

Adequate known-groups VR validity was established through a simple comparison [21, 24, 25], or by investigating the diagnostic criteria as sensitivity and specificity [20, 27]. As the quality assessment tool used in this review required establishing the validity of the index test against a reference standard tool [28], thus, the quality of two studies [25, 27] was not assessed as they lacked this comparison.

Reliability

There is limited promising evidence to support VR high inter-rater reliability in assessing neck range and movement velocity in asymptomatic individuals [19, 26]. Furthermore, a moderate quality evidence supported VR high intra-rater reproducibility in measuring cervical and wrist extension range in asymptomatic participants, whereas its reliability in measuring wrist flexion was low [19, 21]. Authors attributed low reliability to the altered wrist mechanics caused by the tenodesis effect. Although, wrist flexion was assessed also using universal goniometer while the hand was not in a gripping position, its reliability was not reported, which may have served as a comparison for the hypothesized tenodesis effect.

VR reliability in assessing dynamic balance is inconsistent [22–24]. Wikstrom [22] used multiple Wii Fit games and reported a wide reliability range, which could be attributed to the difficulty of balance task demanded by each game. Furthermore, only dynamic balance was assessed in asymptomatic volunteers and elderly. Moreover, the variation in study quality and the software used impairs reaching a solid conclusion regarding its reliability in assessing this outcome. This agrees with the previously reported balance inconsistent intra-rater reliability in healthy and patients with neurological dysfunctions [17].

All reviewed studies had adequate time interval between testing sessions, with a maximum of 1-week interval that is not expected to influence the results due to changes in subjects' condition. On the other hand, all studies failed to blind the findings between examiners or the same examiner's previous findings. Although this could have increased the risk of inflating the reproducibility of the measurements [28], VR objective nature and automatic measurement procedures would be expected to reduce this risk of bias. Furthermore, varying the order of participant testing was fulfilled only in two studies [19, 26] out of eleven. This may have influenced the results as the examiner could recall the measurements of the participant [28]; however, due to the objective nature of VR assessment, testing order would not be expected to play a major role. Thus, in this review, this item was rated as N/A and the score calculation was adjusted.

Strengths and Limitations

This is the first review to evaluate validity and reliability of using interactive VR (games or real-time feedback) in assessing the musculoskeletal system, which may be an important feature to assess motor control during regular clinical assessment. This review included a reasonable number of biomedical databases and employed forward snowballing. Also, evidence strength categorization was done to increase the applicability of results. However, a few limitations exist: first, only articles in the English language were included. Second, a few studies were eligible, with a limited and heterogenous outcome measures, thus no meta-analysis was conducted to reach to solid conclusion. Third, quality assessment tool focused on concurrent validity and reliability, and did not account for other psychometric properties such as known-groups validity, sensitivity, and specificity. However, these properties were discussed in qualitative synthesis. Fourth, most of the reviewed studies included only asymptomatic volunteers; thus, result generalizability needs further validation in different patients' population. Finally, all included studies used interactive VR, except one study [24] that used Wii training software without providing sufficient information about employing its interactive features. Authors were contacted to obtain

detailed information, but no responses were received by the reviewers. To avoid excluding a potential study, this study was included in the review and VR was assumed to be interactive.

Conclusion

Based on the included studies, there is a limited promising evidence that interactive VR using games or real-time feedback is highly valid and reliable in assessing ROM in asymptomatic participants and patients with chronic neck pain and radial fracture. Evidence regarding VR validity in assessing balance and postural sway, reaction time, and movement velocity and accuracy is limited to draw a robust conclusion. Future high-quality studies in different patients' population are required to reach a rigid conclusion.

References

Papers of particular interest, published recently, have been highlighted as:

- Of importance
- Of major importance

1. Atherly A, Van Den Broek-Altburg E, Hart V, Gleason K, Carney J. COVID-19, deferred care and telemedicine: a problem, a solution and a potential opportunity. *JMIR public Heal Surveill.* Canada; 2020;
2. Haider Z, Aweid B, Subramanian P, Iranpour F. Telemedicine in orthopaedics and its potential applications during COVID-19 and beyond: a systematic review. *J Telemed Telecare.* England. 2020: 1357633X20938241 **A systematic review that provides evidence of telemedicine safety and cost-effectiveness in orthopedics. It highlights potential research areas.**
3. Ambrosino P, Fuschillo S, Papa A, Di Minno MND, Maniscalco M. Exergaming as a supportive tool for home-based rehabilitation in the COVID-19 pandemic era. *Games Health J United States.* 2020;9:1–3.
4. Riva G. Virtual reality. *Encycl Biomed Eng.* Grad: Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2006.
5. Weiss P, Kizony R, Feintuch U, Katz N. Virtual Reality in Neurorehabilitation. In: Selzer M, Cohen L, Gage F, Clarke S, Duncan P, editors. *Textb Neural Repair Neurorehabilitation.* New York: Cambridge Press; 2006. p. 182–97.
6. Gumaa M, Rehan Youssef A. Is virtual reality effective in orthopedic rehabilitation? A systematic review and meta-analysis. *Phys Ther [Internet].* 2019 [cited 2020 Jan 12];99:1304–25. Available from: <https://academic.oup.com/ptj/article/99/10/1304/5537309> **A systematic review that supports the use of virtual reality in the orthopedic rehabilitation as an alternative of traditional methods. Furthermore, it summarizes risks of bias in current study design that needs to be addressed in future research. It also provides a comprehensive list of off-the-shelf commercial exergame consoles that could be used in rehabilitation.**
7. Ferreira Dos Santos L, Christ O, Mate K, Schmidt H, Krüger J, Dohle C. Movement visualisation in virtual reality rehabilitation

- of the lower limb: a systematic review. *Biomed Eng Online* [Internet]. 2016 [cited 2020 Jan 22];15:144. Available from: <http://biomedical-engineering-online.biomedcentral.com/articles/10.1186/s12938-016-0289-4>
8. Morris LD, Louw QA, Grimmer-Somers K. The Effectiveness of virtual reality on reducing pain and anxiety in burn injury patients. *Clin J Pain* [Internet]. 2009 [cited 2020 Jan 22];25:815–26. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19851164>.
 9. Ravi DK, Kumar N, Singhi P. Effectiveness of virtual reality rehabilitation for children and adolescents with cerebral palsy: an updated evidence-based systematic review. *Physiotherapy* [Internet]. 2017 [cited 2020 Jan 22];103:245–58. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28109566>. **A systematic review that supports virtual reality efficacy in improving balance and motor skills of children with cerebral palsy. Furthermore, it provides a comprehensive list of different consoles and tools that could be used in the rehabilitation of these cases, which could be a rich resource for readers.**
 10. Laver KE, Lange B, George S, Deutsch JE, Saposnik G, Crotty M. Virtual reality for stroke rehabilitation. *Cochrane Database Syst Rev*. 2017;11:CD008349.
 11. Lloréns R, Noé E, Colomer C, Alcañiz M. Effectiveness, usability, and cost-benefit of a virtual reality-based telerehabilitation program for balance recovery after stroke: A randomized controlled trial. *Arch Phys Med Rehabil* [Internet]. 2015 [cited 2020 Jan 22];96:418–425.e2. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25448245>.
 12. Grooms DR, Page SJ, Nichols-Larsen DS, Chaudhari AMW, White SE, Onate JA. Neuroplasticity associated with anterior cruciate ligament reconstruction. *J Orthop Sport Phys Ther* [Internet]. 2017 [cited 2020 Jan 22];47:180–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27817301>.
 13. Grooms D, Appelbaum G, Onate J. Neuroplasticity following anterior cruciate ligament injury: a framework for visual-motor training approaches in rehabilitation. *J Orthop Sport Phys Ther* [Internet]. 2015 [cited 2020 Jan 22];45:381–93. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25579692>.
 14. Gandola M, Bruno M, Zapparoli L, Saetta G, Rolandi E, De Santis A, et al. Functional brain effects of hand disuse in patients with trapeziometacarpal joint osteoarthritis: executed and imagined movements. *Exp Brain Res* [Internet]. 2017 [cited 2020 Jan 22];235:3227–41. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28762056>.
 15. Tarragó M da GL, Deitos A, Brietzke AP, Vercelino R, Torres ILS, Fregni F, et al. Descending control of nociceptive processing in knee osteoarthritis is associated with intracortical disinhibition. *medicine* (Baltimore) [Internet]. 2016 [cited 2020 Jan 22];95:e3353. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27124022>.
 16. Springer S, Yogev Seligmann G. Validity of the Kinect for Gait Assessment: A Focused Review. *Sensors* (Basel) [Internet]. 2016 [cited 2020 Jan 22];16:194. Available from: <http://www.mdpi.com/1424-8220/16/2/194>
 17. Clark RA, Mentiplay BF, Pua Y-H, Bower KJ. Reliability and validity of the Wii Balance Board for assessment of standing balance: a systematic review. *Gait Posture* [Internet]. 2018 [cited 2018 Sep 24];61:40–54. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29304510>. **A systematic review that provides evidence supporting Wii Balance Board validity and reliability in assessing standing balance in a few neuromuscular disorders. The results of this review can be generalized to clinical settings and guide clinicians on the use of this feasible and readily accessible tool.**
 18. Ruff J, Wang TL, Quatman-Yates CC, Phieffer LS, Quatman CE. Commercially available gaming systems as clinical assessment tools to improve value in the orthopaedic setting: a systematic review. *Injury* [Internet]. 2015 [cited 2020 Jan 22];46:178–83. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0020138314004227>
 19. Sarig-Bahat H, Weiss PL, Laufer Y. Cervical motion assessment using virtual reality. *Spine* (Phila Pa 1976) [Internet]. 2009 [cited 2019 Dec 29];34:1018–24. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19404177>.
 20. Sarig-Bahat H, Weiss PL, Laufer Y. Neck pain assessment in a virtual environment. *Spine* (Phila Pa 1976) [Internet]. 2010 [cited 2019 Dec 29];35:E105–12. Available from: <https://insights.ovid.com/crossref?an=00007632-201002150-00023>
 21. Eini D, Ratzon NZ, Rizzo AA, Yeh S-C, Lange B, Yaffe B, et al. Camera-tracking gaming control device for evaluation of active wrist flexion and extension. *J Hand Ther* [Internet]. 2017 [cited 2019 Dec 29];30:89–96. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0894113016301132>
 22. Wikstrom EA. Validity and reliability of Nintendo Wii Fit balance scores. *J Athl Train*. 2012;47:306–13.
 23. Jorgensen MG, Laessoe U, Hendriksen C, Nielsen OBF, Aagaard P. Intrarater reproducibility and validity of Nintendo Wii balance testing in community-dwelling older adults. *J Aging Phys Act* [Internet]. 2014 [cited 2019 Dec 29];22:269–75. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23752090>.
 24. Guzman J, Aktan N. Comparison of the Wii Balance Board and the BESS tool measuring postural stability in collegiate athletes. *Appl Nurs Res* [Internet]. 2016 [cited 2019 Dec 29];29:1–4. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0897189715000932>
 25. Jorgensen MG, Paramanathan S, Ryg J, Masud T, Andersen S. Novel use of the Nintendo Wii board as a measure of reaction time: a study of reproducibility in older and younger adults. *BMC Geriatr* [Internet]. 2015 [cited 2019 Dec 29];15:80. Available from: <http://bmgeriatr.biomedcentral.com/articles/10.1186/s12877-015-0080-6>
 26. Sarig Bahat H, Sprecher E, Sela I, Treleaven J. Neck motion kinematics: an inter-tester reliability study using an interactive neck VR assessment in asymptomatic individuals. *Eur Spine J* [Internet]. 2016 [cited 2019 Dec 29];25:2139–48. Available from: <http://link.springer.com/10.1007/s00586-016-4388-5>
 27. Sarig Bahat H, Chen X, Reznik D, Kodesh E, Treleaven J. Interactive cervical motion kinematics: sensitivity, specificity and clinically significant values for identifying kinematic impairments in patients with chronic neck pain. *Man Ther* [Internet]. 2015 [cited 2019 Dec 29];20:295–302. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1356689X14001842>
 28. Brink Y, Louw QA. Clinical instruments: reliability and validity critical appraisal. *J Eval Clin Pract* [Internet]. 2012 [cited 2019 Dec 31];18:1126–32. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21689217>.
 29. Prowse A, Pope R, Gerdhem P, Abbott A. Reliability and validity of inexpensive and easily administered anthropometric clinical evaluation methods of postural asymmetry measurement in adolescent idiopathic scoliosis: a systematic review. *Eur Spine J* [Internet]. 2016 [cited 2019 Dec 31];25:450–66. Available from: <http://link.springer.com/10.1007/s00586-015-3961-7>
 30. Bell DR, Guskiewicz KM, Clark MA, Padua DA. Systematic review of the balance error scoring system. *Sports Health* [Internet]. *Sports Health*; 2011 [cited 2020 Jan 22];3:287–95. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23016020>.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.