Validity and Reliability of Occupational Therapy Assessment for Telehealth:

A Systematic Review

Maya Aranda, OTS, Nathalie Resendis, OTS, Travis Robertshaw, OTS

Abstract

Importance: As telehealth becomes increasingly used to deliver occupational therapy services to underserved populations, ensuring the validity and reliability of remotely administered assessments is critical to maintaining quality care.

Objective: To identify, evaluate, and synthesize the current literature concerning the validity of assessments used during occupational therapy evaluations performed over the telehealth delivery mode.

Data Sources: A literature search occurred in May 2025. Databases included CINAHL, PubMed, and ScienceDirect using Hawai'i Pacific University's online library databases. Search terms included "occupational therapy", "telehealth", "remote assessment", "sensor technology", "wearable device", and "validity", as well as combinations of these terms.

Study Selection and Data Collection: This systematic review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. Published studies on the validity of assessments used during occupational therapy evaluations were included in the systematic review. Data from presentations, non-peer reviewed literature, and dissertations were excluded.

Findings: Five studies were included (four level III, one level IV) according to the American Occupational Therapy Association's Levels of Evidence. The outcomes of these studies indicate that the assessments were valid for use over a telehealth delivery model.

Conclusion and Relevance: The five assessments identified are effective and valid for use over telehealth and improves access to care for people receiving healthcare remotely.

What This Systematic Review Adds: This review highlights the limited high-quality evidence currently available on validated telehealth OT assessments. It provides a foundation for practitioners and researchers seeking to understand or build upon validated telehealth tools, while emphasizing the need for continued research to expand the repertoire of assessments suitable for remote delivery.

Key words: Assessment tools, frailty, occupational therapy, telehealth, validity, wearable device

Assessments are an important part of occupational therapy evaluation and intervention planning. Many traditional occupational therapy assessments have been validated, standardized measures, but most assessments have been validated for in person administration by an expert examiner (Yang, et al., 2023). The increasing use of telehealth as a delivery model for occupational therapy requires the validation of additional assessment tools for use over video conference and other telehealth technologies (Boone, et al., 2022).

Occupational therapy (OT) assessments can be used with a wide array of populations. With infant assessments over telehealth, the absence of an expert clinician in person requires the parent or caregiver to handle the infant, potentially changing the validity of an assessment (Davies et al., 2025). Motor function in adults can be provided over telehealth for patients living in rural areas or who are unable to attend in-clinic appointments due to health restrictions (Yang, et al., 2023). Elderly clients can be assessed for falls risk in home and community settings instead of a clinical setting (Garcia-Villamil, et al., 2021). Advances in internet access and adoption of telehealth technologies can allow for greater monitoring of health and detection and prevention of frailty in older addults (Bian, et al., 2022).

The COVID-19 pandemic highlighted benefits of telehealth as a delivery model for occupational therapy. Regardless of the service delivery model, evaluation and treatment needs to be effective and reliable (Yang, et al., 2023). This systematic review explored which OT assessments have been validated for use through telehealth or can be completed by the client without a therapist present.

Method

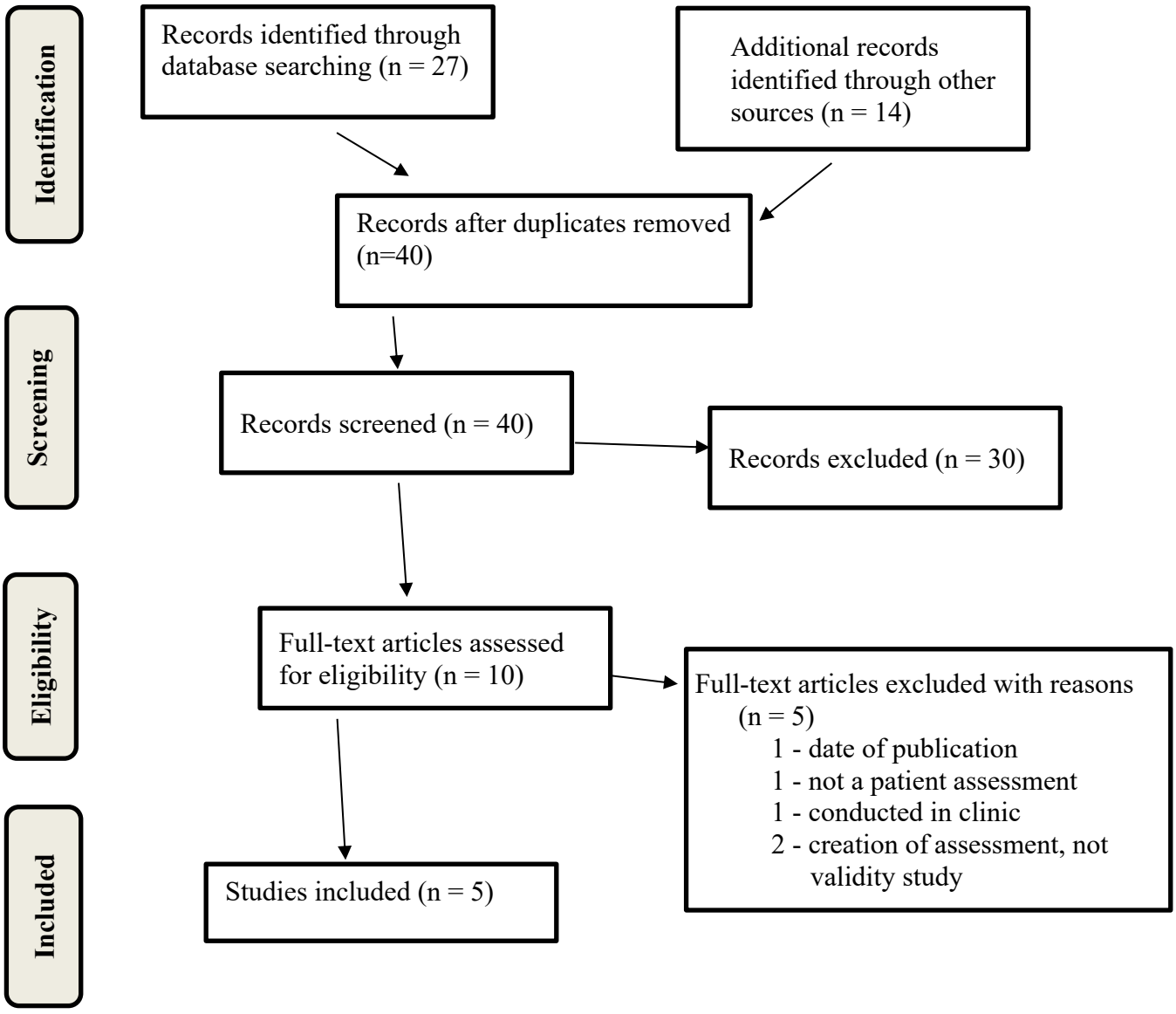
The systematic review adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines and incorporated recommended processes for conducting a systematic review. The guiding research question for this systematic review was: What is the validity and reliability of occupational therapy assessment tools when delivered via telehealth or remote technologies?

A broad search of the literature and follow-up search to ensure all relevant research was included occurred in May 2025. The inclusion criteria for studies in this systematic review were as follows: included an occupational therapy assessment, peer-reviewed, published in English, and dated between 2019-2025. Exclusion criteria, in addition to those studies that did not meet the inclusion criteria, included articles that were systematic reviews, scoping reviews, dissertations, and presentations. Home assessments without a telehealth component were also excluded. A search for relevant literature was completed using electronic databases: CINAHL, PubMed, and ScienceDirect through Hawai'i Pacific University's online library database. Search terms included "occupational therapy OR occupational therapist OR OT" AND "telehealth OR telemedicine" AND "assessment OR assessment tool OR assessment tools OR assessing OR assessment methods" AND "psychometrics OR validity OR validation OR reliability". Appendix A provides an extensive list of all search terms used for this systematic review. The initial search included 27 articles related to the research topic (Figure 1) and an additional 14 articles were identified by an experienced researcher through citation tracking and direct search of OT journals. One article was identified as duplicate and removed. Abstracts were reviewed and filtered for inclusion and exclusion criteria and 30 articles were excluded for not meeting inclusion criteria. An additional five articles were excluded after full-text review for not meeting

the inclusion criteria of being related to telehealth or completed independently by the client without the therapist present (e.g., asynchronous telehealth). Three independent reviewers completed the screening and selection of the studies, assessed their quality, and extracted the data.

Figure 1

PRISMA Flow Diagram



Results

Five studies met the inclusion criteria. The articles were assessed according to their risk of bias, level of evidence, and quality. The studies contained relevant information regarding occupational therapy assessment tools that have been validated and found effective for telehealth use. The information from these articles was divided into two themes related to the validity and reliability of occupational therapy assessments that can be administered by a therapist through telehealth and assessments using ambient and wearable sensor technologies. An evidence table is provided in Appendix B. The Cochrane risk-of-bias guidelines were used to assess each article and are provided in Appendix C.

Validity and Reliability of Occupational Therapy Assessments Via Telehealth

Three of the five studies on the topic discussed the validity and reliability of occupational therapy assessments delivered via synchronous telehealth. Two of these studies were level III studies and one was a Level IV study (see Appendix B). All studies provided evidence that occupational therapy assessments conducted via telehealth are effective and potentially beneficial.

Boone et al. (2022) evaluated the electronic version of the Activity Card Sort (eACS) that was created based on the paper version. Twenty adult participants were assessed using both the traditional assessment and the electronic version, with the results from both studies showing high concurrent validity because the scores were comparable with the original paper version of the Activity Card Sort. Participants reported that the eACS was easier to use, more visually appealing, and more inclusive than the traditional assessment.

Davies et al. (2025) evaluated administration of the Alberta Infant Motor Scale (AIMS) through telehealth for infants who were at high risk for developmental delays.z Ten raters, six experts and four novices, independently rated twenty-three infants who were neurodevelopmentally high-risk using recorded telehealth assessments. It was observed that interrater and intra-rater reliability were excellent, but some difficulties arose with specific tests (i.e., the standing tests) due to inadequate camera placement. There were some benefits to using the telehealth assessment, as opposed to in-person assessments, because the telehealth visit was recorded and the assessor could pause and rewind the video to assure accurate ratings.

Yang et al. (2023) used the ArmCAM tool remotely for upper extremity motor assessment post stroke. The ArmCAM is a set of ten motor tasks intended to be viewed remotely over video and arranged in an order that minimizes the need to reposition the camera or the client between tasks. Thirty-one adults with a history of a stroke greater than six months prior were included in the study. Two telehealth visits were conducted within one week to assess test-retest validity. The test was conducted utilizing resources that all participants reported were easily found within their home. Interrater and test-retest reliability were excellent (0.993 and 0.997, respectively) in all areas except for one test, holding a magazine, where participants reported difficulty with camera placement when using a smartphone or tablet device. The study authors suggested using an egg carton, an item commonly found in most homes, to hold the device.

Limitations of the studies on telehealth validity and reliability included small sample size (Boone et al., 2022; Davies et al., 2025), limited diversity of participants, and diverse home environments that might have influenced the accuracy of assessments (Yang et al., 2023).

Assessments Using Ambient and Wearable Sensor Technologies

Two of the five studies explored the efficacy of assessments using ambient and wearable sensor technologies. Both studies were Level III studies (see Appendix B). All studies provided evidence that these assessments are potentially beneficial and could be used in a telehealth service delivery model.

A study by Bian et al. (2022) evaluated the use of a Frailty Toolkit (FT) system with ambient sensors to assess the behavior and physical signs of frailty of at-risk populations using easily placed equipment and a smart speaker to ask simple questions. Frailty is an issue with older adults that can be remedied if it is identified early in the process. Nine healthy adults participated in the study because of restrictions at the time of the study due to COVID-19 precautions. The authors measured activity level and behavior by using a weight sensor mat placed on a chair to measure time spent sedentary, motion sensors to measure the time taken to climb stairs and time spent moving around the space, and a smart speaker to ask the participant to report fatigue levels. The study showed that by using multiple methods of assessment in the home, the FT is reliable for monitoring physical and behavioral signs of frailty in home settings.

García-Villamiland Neira-Álvarez (2021) conducted a study on a wearable inertia device to assess fall risk in elderly clients. Twenty-one elderly adults participated in the study, separated into two groups: fallers and non-fallers. By placing an inertial measuring device on each participant's ankle, they found that wearing the device accurately measures gait parameters in older adults. The study concluded that it provides assessment results in less time than conventional tests, and can analyze walking behavior on different real-world surfaces, which is difficult in most clinical settings.

Limitations of the studies on assessments using ambient and wearable sensor technologies included small sample size and not accounting for the biological sex of the participants, which is important when considering frailty and falls risk (García-Villamil, & Neira-Álvarez, 2021). Limitations for the FT study (Bian et al., 2022) included conducting the study on a younger population due to COVID-19 restrictions. Additional concerns related to external validity of the FT include possible privacy concerns due to using smart devices in home environments and only being applicable to single-person dwellings.

Discussion

The results of this systematic review suggest that specific telehealth assessments provide valid assessment data for clients seeking occupational therapy within a telehealth service delivery model. Occupational therapy, as an evidenced based practice, relies upon the validity of assessments to provide accurate data and reliable measurements of outcomes. By conducting studies to show that the assessments are valid and reliable, these studies increase confidence in using specific assessments within a telehealth service delivery model (Boone et al., 2022; Davies et al., 2025; Yang et al., 2023).

Due to the breadth of occupational therapy practice, occupational therapy evaluations can use varied assessments depending on the diagnosis of the client. Many clients live in rural areas or have conditions that prevent them from attending visits in person. Because of this, it is important that assessments are validated for a telehealth delivery model (Yang et al., 2023)

These studies demonstrated that the eACS, AIMS, and ArmCAM tool have advantages over the traditional in-clinic version of the assessments. The eACS was found to be more visually appealing, user friendly, and is inclusive for people at home (Boone et al., 2022). The

AIMS, when performed via telehealth, affords the ability to pause and rewatch sections of the video which results in more accurate and reliable results when observing infant motor behavior (Davis et al., 2025). The ARMCAM tool demonstrated that in a remote setting it was easy to complete the assessment because household items could be used for the evaluation process and there was strong interrater reliability when administered through telehealth (Yang et al.,2023). The Frailty toolkit (FT) and inertia devices were easy to use but require additional research to determine their effectiveness in assessing frailty and fall risk in older adults as they were tested on a different population, healthy young adults (Bian et al., 2022; García-Villamil & Neira-Álvarez, 2021). While the emerging evidence is promising, more studies are needed to evaluate these and other OT assessments when used with the intended populations in home and community settings through a telehealth service delivery model.

Strengths and Limitations

A strength of this systematic review was the use of PRISMA guidelines and the flow diagram that ensured a replicable search and selection process. This process made the review stronger and minimized bias in article inclusion. Furthermore, having a group of researchers collaboratively evaluate the literature and select appropriate studies based on inclusion criteria greatly reduced the risk of bias. Another factor that contributed and greatly reduced the risk of bias analysis using the Cochrane risk of bias tools.

Limitations included only a few studies met the inclusion criteria and limitations within the studies included small sample size, lack of diversity, and limited external validity. In conclusion, the limitations indicate that more research is needed to validate the use of existing OT assessments for administration through telehealth.

Implications for Occupational Therapy Practice

This systematic review found that telehealth, assessments completed at home without the therapist present, and ambient technologies and wearable sensors can be effective and useful alternatives to traditional, in-person OT assessments. As advances in technologies support evidence-based practice for clients in rural or underserved areas, valid and reliable tools are essential to provide and maintain high standards of care. This review highlights traditional assessments adapted for virtual delivery and examined research related to technology using sensor-based tools and devices that can be worn to provide assessment data. More research is needed for occupational therapists to begin to integrate the tools into practice with confidence with diverse populations.

Key takeaways:

- Some occupational therapy assessments have demonstrated good reliability and validity using telehealth including the eACS, AIMS, and ArmCAM (Boone et al., 2022; Davies et al., 2025; Yang et al., 2023).
- Telehealth assessments are essential for populations with barriers to in-person care (Yang et al., 2023).
- Technology used at home like Frailty Toolkit and wearable sensors allow for ease of use and equally significant data as in-clinic assessments (Bian et al., 2021; García-Villamil et al., 2021).
- Training needs to be implemented for occupational therapists practicing via telehealth to have proper placement and device set up to ensure that assessment results are accurate (Davies et al., 2025).

- More research is needed to establish psychometrics (reliability and validity) of OT assessments administered through technologies in different settings and with various populations.

Conclusion

Studies included within this systematic review provide evidence on the validity of several assessments used in occupational therapy evaluations administered through telehealth (i.e., synchronously through videoconferencing, using sensor technologies, and asynchronously completed without the therapist present). More research is necessary to validate additional assessments across the diverse populations served by occupational therapists. Telehealth and remote assessment methods show potential to facilitate and promote access, efficiency, and client centered care in occupational therapy.

References

Bian, C., Ye, B., & Mihailidis, A. (2022). The development and concurrent validity of a multi-sensor-based frailty toolkit for in-home frailty assessment. *Sensors, 22*(9), 3532. https://doi.org/10.3390/s22093532

Boone, A. E., Wolf, T. J., & Baum, C. M. (2022). Brief report—development and initial testing of the electronic Activity Card Sort (ACS3) among community-dwelling adults. *American Journal of Occupational Therapy, 76*(3), 7603345040. https://doi.org/10.5014/ajot.2022.047522

Davies, S., Lucas, B. R., & Dwyer, G. M. (2025). Reliability of the Alberta Infant Motor Scale (AIMS) when used via telehealth for neurodevelopmentally high-risk infants. *Physical & Occupational Therapy in Pediatrics*, *45*(1),1-14. https://doi.org/10.1080/01942638.2025.2451406

García-Villamil, G., Neira-Álvarez, M., Huertas-Hoyas, E., Ramón-Jiménez, A., & Rodríguez-Sánchez, C. (2021). A pilot study to validate a wearable inertial sensor for gait assessment in older adults with falls. *Sensors, 21,* 4334. https://doi.org/10.3390/s21134334

Yang, C.-L., Simpson, L. A., & Eng, J. J. (2023). A pilot study for remote evaluation of upper extremity motor function after stroke: The Arm Capacity and Movement Test (ArmCAM). *American Journal of Occupational Therapy, 77*(1), 7701205020. https://doi-org.hpu.idm.oclc.org/10.5014/ajot.2023.050020

Appendix A: Search Terms

occupational therapy OR occupational therapist OR occupational therapists OR OT AND assessment OR assessment tool OR assessment tools OR assessing OR assessment methods AND psychometrics OR validity OR validation OR reliability

Appendix B

Evidence Table

Which OT assessment tools have been validated for telehealth use?					
Author/Year	Level of Evidence Study Design Risk of Bias	Participants Inclusion Criteria Study Setting	Intervention and Control Groups	Outcome Measures	Results
Bian et al., (2022) https://doi.org/10.3390/s22093532	Level IIIb One-group design *Risk of bias:* Moderate	9 healthy adults age 18+; English-speaking; Able to provide consent and attend session; Were excluded if they were using a wheelchair, or with hearing or speech impairment	Intervention: Use of the Frailty Toolkit (FT) system with ambient sensors (mat, motion, weight scale, door sensor, smart speaker); No control group	Frailty Toolkit components: motion sensors, smart speaker responses, weight scale, mat sensors, and distance sensors	Excellent agreement for motion sensors (x= 0.938) - smart speaker (100% accuracy), - mat sensors (95.2% within LoA). Strong correlation for weight scale (r = 0.942). The distance sensors had only a 50% success rate for stair detection.
Boone et al. (2022) https://doi.org/10.5014/ajot.2022.047522	Level IV Cross-sectional single group *Risk of bias:* Moderate	N = 20 adults, M age 57.5 yr., 60% female Inclusion Criteria: > 45 years old Exclusion Criteria: None	Intervention 1: Activity Card Sort (ACS) Intervention 2: An electronic version of the Activity Card Sort (ACS3) All participants completed both interventions where they see a picture of an activity on the card and sort it into	Tests IADLs, social activities, low-demand leisure activities, and high demand leisure activities. Results from both tests were compared for similarity	The two tests were found to be very similar in outcomes, having r_s = .863, meaning the electronic version has validity. The ACS3 was found by participants to be easier to use, visually more appealing, more representative, and better at helping people realize

			categories of how often they engage in it. Even numbered participants conducted ACS first, then ACS3, odd numbered participants completed ACS3 first, then ACS		occupations they wanted to do.
Davies et al. (2025) https://doi.org/10.1080/01942638.2025.2451406	Level IIIb Two group non-randomized study *Risk of bias:* moderate	N = 23 Infants, age less than 27 weeks, 65% female N = 12 raters, 4 experts with > 5 years experience, 8 novices (final year physiotherapy students) Inclusion Criteria: infant admitted to NICU with at least one of five neurodevelopmental risk factor Exclusion Criteria: Parents did not speak English, Infants with a diagnosed syndrome, cortical visual impairment or retinopathy of prematurity grade 3,	1: Each rater rated a video recording of an infant receiving the assessment over telehealth by a skilled therapist. 2: Each participant rated the videos again after a minimum of two weeks passed. The two assessments were compared to identify inter- and intra-rater reliability.	Assessment: The Alberta Infant Motor Scale (AIMS), a norm-referenced test to evaluate spontaneous gross motor skills in infants 0-18 months old, in prone, supine, sitting, and standing. The assessor only credits observed skills.	AIMS assessments via recorded tele-health assessments are reliable. Inter- and intra-rater reliability was excellent in total score, only the standing position had less than good reliability due to inadequate camera angles being used.

		significant auditory impairment, received major surgery			
García-Villamil, & Neira-Álvarez (2021) https://doi.org/10.3390/s21134334	Level IIIb Cohort Study *Risk of bias:* Low	21 participants, M age 81 YO, 57.1% female Inclusion Criteria: >70 YO, walk without assistance from another person and met one of the three criteria: (a) One or more falls with consequences in the last year; (b) gait and balance disorder or fear of falling; (c) post-fall syndrome. Exclusion Criteria: terminal illness with a life expectancy of fewer than six months or not providing informed consent	21 participants wore a foot based inertial measurement unit. Results were divided into groups of fallers and non-fallers and analyzed, and were also divided into three levels of frailty and analyzed. A falls assessment and frailty assessment was conducted on each participant prior to the intervention. The intervention consisted of wearing the device for 30 minutes during normal walking activities.	Mean Stride speed Mean Stride length SD Stride length Mean Swing time SD Swing time Cadence Steps Total Distance Total Walking Time (SD = standard deviation)	The device accurately measures gait parameters in older adults, it provides assessment results in less time than conventional tests, and can analyze walking behavior on different surfaces. In the observed gait assessments they are of value to identify fall risks and frailty level in which occupational therapy is equipped with fall prevention strategies.
Yang et al. (2023) https://doi-org.hpu.idm.oclc.org/	Level IIIb One-group design *Risk of bias*: Low	31 adults that were 19+ years old, 6 months+ post-stroke. They are able to follow instructions and have	Arm Capacity and Movement (ArmCAM) remote assessment w/no control group.	Test–retest reliability (ICC), interrater reliability, standard error of measurement (SEM), minimal detectable change	ArmCAM showed excellent reliability (ICC > .99) - low SEM (0.74), - high validity with established tools

10.5014/ajot.2023.050020		access to the internet/webcam. Excluded if other neurological conditions or severe UE pain.	Participants were tested over video on 10 items that required use of the UE and items commonly found in a home. They were retested within one week to determine test/retest validity.	(MDC), construct validity via correlations with FM–UE, ARAT, SIS-Hand, REACH	(e.g., r = .944 with FM–UE). Tool took about 15 minutes to administer and was well tolerated by participants.
Note. [Define any acronyms used] ROM: Range of motion, UE: upper extremity, ICC: INtraclass correlation coefficient, SEM:Standard error of measure, MDC: minimal detectable change, FM-UE: Fugl myer assessment for upper extremity, ARAT:Action research arm test, SIS-hand: stroke impact scale, REACH: rating of everyday arm in use, ACS: Activity card sort, ACS3: electronic activity card sort, IADLs: instrumental activities of daily living, rs: spearman's rank order correlation, SD:standard deviation,Alberta Infant Motor Scale (AIMS)					

Appendix C

Risk-of-Bias Tables

Risk-of-Bias Table for Randomized Controlled Trial (RCT) and Non-RCT (Two or More Group Design)										
	Selection Bias (Risk of bias arising from randomization process)			Performance Bias (effect of assignment to intervention)		Detection Bias		Attrition Bias	Reporting Bias	Overall risk-of-bias (low, moderate, high
Citation	Random Sequence Generation	Allocation Concealment (until participants enrolled and assigned)	Baseline difference between intervention groups	Blinding of Participants During the Trial	Blinding of Study Personnel During the Trial	Blinding of Outcome Assessment: Self-reported outcomes	Blinding of Outcome Assessment: Objective Outcomes (assessors aware of intervention received?)	Incomplete Outcome Data (data for all or nearly all participants	Selective Reporting (results being reported selected on basis of the results?)	
Davies et al. (2025)	-	-	+	+	-	-	-	+	+	moderate

Note. Categories for risk of bias are as follows: Low risk of bias (+), unclear risk of bias (?), high risk of bias (–). Scoring for overall risk of bias assessment is as follows: 0–3 minuses, low risk of bias (L); 4–6 minuses, moderate risk of bias (M); 7–9 minuses, high risk of bias (H).

Citation. Table format adapted from Higgins, J. P. T., Sterne, J. A. C., Savović, J., Page, M. J., Hróbjartsson, A., Boutron, I., . . . Eldridge, S. (2016). A revised tool for assessing risk of bias in randomized trials. Cochrane Database of Systematic Reviews 2016, Issue 10 (Suppl. 1), 29–31. https://doi.org//10.1002/14651858.CD201601

Risk of Bias for Before-After (Pre-Post) Studies with No Control Group (One Group Design)												
Citation	Study question or objective clear	Eligibility or selection criteria clearly described	Participants representative of real-world patients	All eligible participants enrolled	Sample size appropriate for confidence in findings	Intervention clearly described and delivered consistently	Outcome measures pre-specified, defined, valid/reliable, and assessed consistently	Assessors blinded to participant exposure to intervention	Loss to follow-up after baseline 20% or less	Statistical methods examine changes in outcome measures from before to after intervention	Outcome measures were collected multiple times before and after intervention	Overall risk of bias assessment (low, moderate, high risk)
Bian et al. (2022)	Y	Y	N	Y	N	Y	Y	Y	Y	Y	N	M
Boone et al. (2022)	Y	Y	N	Y	N	Y	Y	N	Y	Y	N	M
Garcia et al., 2021	Y	Y	Y	Y	N	Y	Y	N	Y	Y	N	L
Yang et al. (2023)	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	L

Note. Y = yes; N = no; NR = not reported. Scoring for overall risk of bias assessment is as follows: 0–3 N, Low risk of bias (L); 4–8 N, Moderate risk of bias (M); 9–11 N, High risk of bias (H).

Citation. Table format adapted from National Heart Lung and Blood Institute. (2014). Quality assessment tool for before–after (pre–post) studies with no control group. Retrieved from https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools