Title

An Investigation into Validated Task-Based Evaluation Methods for Upper-Limb Prostheses

Author

Siegel, Joshua R.

Publication Date

2024

Peer reviewed|Thesis/dissertation

An Investigation into Validated Task-Based Evaluation Methods for Upper-Limb Prostheses

By

JOSHUA R. SIEGEL

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Mechanical and Aerospace Engineering

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Jonathon Schofield, Chair

Erkin Şeker

Wilsaan Joiner

Committee in Charge

2024

Table of Contents

Abstract

Despite significant mechatronic advancements in prosthetic hands, achieving the functionality of a biological extremity remains far from realized. For the development of next-generation prosthetics and improved patient outcomes, standardized, reliable, and validated task-based evaluation measures are essential. A task-based approach, entailing the manipulation of physical objects with prostheses, directly assesses patient performance in real-time and offers numerous benefits. Task-based assessments aid clinical decision-making, allowing clinicians to choose the best prosthetic device or strategy for individual patients and enables precise tracking of patient progress, highlighting treatment effectiveness or the need for adjustments. These measures also provide objective data for cost justification in insurance and public health systems, enhancing transparency among stakeholders, including researchers, clinicians, patients, and regulatory bodies. Finally, the standardization of task-based measures facilitates consistent comparisons across different prosthetic devices and control systems, promoting iterative improvement and innovation. Thus, properly implemented standardized evaluation measures are fundamental to the advancement of upper-limb prosthetics.

This thesis critically examines currently available task-based evaluation methods for upper-limb prosthetic technologies, highlighting the gap between the rapid advancement of prosthetic devices and the development of standardized assessment protocols. First, a comprehensive literature review, published in Frontiers in Robotics and AI, revealed that only 25 assessments for upper-limb prostheses have been validated since 1948, highlighting many researchers' reliance on non-standardized tests that may not have rigorously established validity and not fully address the diverse interests of clinical and research communities. This research then applies theory to practice by using one of the highest rated currently available methods, the

Anthropomorphic Hand Assessment Protocol (AHAP), to compare three open-source 3D-printed prosthetic hands with three commercially produced ones. This analysis, currently under review in BMC Biomedical Engineering, illustrates the practical challenges and advancements in prosthetic evaluation and paves the way for a more in-depth discussion on enhancing assessment methodologies. Finally, an online survey about task-based functional measures was conducted to understand and gather insights from diverse practitioners who interact with individuals prescribed upper-limb prostheses. The findings from this survey will influence future task-based evaluation methods and will be presented at the MyoElectric Control Conference in August 2024. This thesis not only provides insights into existing evaluation methods but also pinpoints areas for enhancement, significantly contributing to the development of effective and universally accepted evaluation techniques for upper-limb prostheses.

Dedication

I dedicate this thesis to my beloved grandparents: Trudy Rosenzweig Honigman, Frederic Herbert Honigman, Edie Fox Siegel, and Leonard Gordon Siegel.

Acknowledgments

First and foremost, I extend my heartfelt thanks to my parents, Erica Honigman Siegel and Matthew Jay Siegel; my brothers, Zachary Ryan Siegel and Jordan Russell Siegel; my turtle, Ronnie Coleman Siegel; and my extended family for their unwavering support throughout my master's studies. I also wish to express my profound gratitude to my mentor, Dr. Jonathon S. Schofield, for his invaluable knowledge, patience, advice, and guidance. Additionally, I am grateful to my peers in the Bionic Engineering and Assistive Robotics Laboratory for their support and knowledge. Finally, I would like to extend my gratitude to all the co-authors of my publications for their help and expertise: Marcus A. Battraw, Dylan J. A. Brenneis, Michael R. Dawson, Jedidiah K. Harwood, Michelle A. James, Wilsaan M. Joiner, Annette C. Lau, Patrick M. Pilarski, and Eden J. Winslow.

Copyright Notice

Chapter 1: Introduction

1.1: Background

Standardized, reliable, and validated task-based evaluation measures for upper-limb prostheses are crucial for advancing research and, most importantly, enhancing patient care. Using a task-based approach, which entails manipulating physical objects with prostheses, presents a distinct advantage as it directly assesses a patient's performance in real-time. This method is pivotal in evaluating functionality but also extends its influence to clinical decision-making, aiding clinicians in selecting the most suitable prosthetic device or strategy for individual patients. By integrating optimal evaluation measures, clinicians can also precisely track patient progress, thereby highlighting the efficacy of treatments and identifying necessary adjustments. Furthermore, task-based measures provide standardized, objective data to substantiate cost justifications for insurance and public health systems, facilitating transparency among stakeholders, including researchers, clinicians, patients, insurance agencies, regulatory bodies, and the general public. This standardization also ensures consistent comparisons across different prosthetic devices and control systems. Additionally, although self-reported surveys are invaluable for detailing patient functional outcomes, task-based methods offer unique and complementary information, helping mitigate the biases often encountered in surveys, such as: recall bias (participants might not accurately remember their experiences), social desirability bias (participants might answer in a way to be viewed favorably by others), and extreme response bias (participants might tend to choose the highest or lowest score on a rating scale) [1]. This objective data is crucial for iterative improvement and innovation, reducing uncertainties introduced by study-specific measurement techniques. In essence, standardized, validated task-based evaluation measures, when adeptly implemented, are the cornerstone for advancements in upper-limb prosthetics.

However, despite the clear significance of these measures, there is a notable discrepancy in the focus of upper-limb prosthetic research. While upper-limb prostheses have evolved from simple hook hands to sophisticated, dexterous, multi-articulating devices controlled through the nervous system, a standardized and well adopted assessment framework still remains absent [2]. This has caused researchers to resort to creating boutique tests for new features. To illustrate this point, Figure 1 shows a search conducted on PubMed for articles discussing new upper-limb prosthetic technology versus those discussing testing methodologies. For new prosthetic technologies, it included keywords: 'Prosthetic Hand', 'New Upper-Limb Prosthesis', and 'Hand Neuroprosthesis'. Keywords: 'prosthetic hand test', 'prosthetic hand dexterity assessment', 'prosthetic control system test', and 'prosthesis control system assessment' were used for evaluation methods. As depicted in Figure 1, the results of this search yield a stark discrepancy between the substantial volume of scientific literature reporting on upper-limb prosthetic technologies and the limited number of articles reporting on techniques to evaluate these same devices.

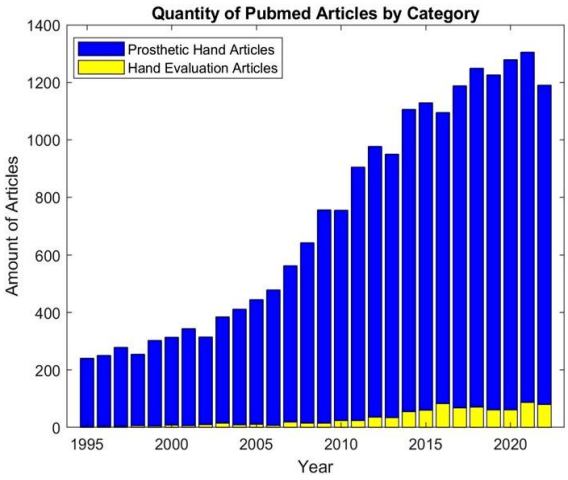


Figure 1. Quantity of PubMed Articles by Category

1.2: Thesis Objectives

This thesis aims to (1) critically examine currently available and validated task-based evaluation methods for upper-limb prosthetic technologies, highlighting the gap between the rapid advancement of prosthetic devices and the development of standardized assessment protocols; and (2) to provide directions for future research and evaluation methods.

This thesis consists of five chapters. **Chapter 1** outlines the significance along with current shortcomings of standardized, validated, task-based evaluation measures for upper-limb prosthetics. **Chapter 2** presents a comprehensive literature review on currently available options to assess prostheses, focusing on multi-grasp dexterity and the functional impact of varying control systems. **Chapter 3** details a study employing a common evaluation method, the Anthropomorphic Hand Assessment Protocol, to gain practical insights and to assess modern prosthetic hands in realistic grasping tasks. **Chapter 4** proposes and shows initial validation of an online survey among professionals to identify strategies for addressing the diverse needs in prosthetic evaluation. **Chapter 5** concludes the thesis with a discussion and future research directions.

Chapter 2: Literature Review

Much of this chapter was published in Frontiers in Robotics and AI [3]: JR Siegel, MA Battraw, EJ Winslow, MA James, WM Joiner, JS Schofield. Review and critique of current testing protocols for upper-limb prostheses: a call for standardization amidst rapid technological advancements. Frontiers in Robotics and AI, 1292632

2.1: Background

To gain a deeper insight into the existing and validated evaluation metrics for upper-limb prosthetic devices, a literature review was conducted to explore current offerings. The scope of this review was to analyze validated task-based assessments that evaluate both dexterity (the precise, voluntary movements required when handling objects), and the functional impact of varying control systems (the technology {electrical, mechanical, or other} that interfaces with the user to actuate a prosthetic device) [4]. This novel approach contrasted prior upper-limb prosthetic assessment reviews by Yancosek et al., in 2009, Resnik et al., in 2017, and Wang et al., in 2018 which utilized different evaluation criteria, focused solely on dexterity evaluations, and included surveys along with assessments for stroke-patients [5], [6], [7]. Conversely, this examination focused exclusively on task-based assessments for the multi-grasp dexterity and functional impact of varying control systems in pediatric and adult upper-limb prostheses.

The evaluation employed a diverse set of criteria, emphasizing accuracy, performance, reliability, and validity [8], [9]. Additionally, it is essential to acknowledge the importance of assessing patient performance, both with and without a prosthesis, especially in bimanual tasks. However, to maintain focus on evaluating the functionality of prostheses or the ability to use them, this metric was omitted from the evaluation criteria but will be noted in the descriptions of relevant tests, if applicable. Furthermore, the growing body of literature reporting on prosthesis interfaces that restore sensory feedback to users must also be recognized. Sensory feedback is

poised to be an integral component of future prostheses, leading to a growing body of literature that describes novel assessments, warranting its own review [10], [11], [12], [13], [14]. This review exclusively analyzed functional tasks sensitive to changes in patient motor-function without the requisite inclusion of a sensory feedback system. Thus, tests that are designed specifically to evaluate sensory enabled upper-limb prosthetic systems were excluded.

2.2: Methods

This review included 1,423 journal articles sourced from online databases: PubMed, Medline, Shirley-Ryan Rehabilitation Measures, Google Scholar, and ScienceDirect. Search terms for each database included: journal article [Publication Type] AND ‘prosthetic hand test’ OR ‘prosthetic hand dexterity assessment’ OR ‘prosthetic control system test' OR ‘prosthesis control system assessment’) and 19 articles from previous knowledge. After duplicates were removed, 1,434 articles remained. After reading titles and abstracts, the scope narrowed to 250 papers. From these, articles were selected that presented a validated task for measuring either upper-limb prosthetic dexterity, control systems, or both. The task could include a questionnaire but articles that solely consisted of questionnaires were excluded. After examination, 25 tests were identified for inclusion in our study. This selection process has been shown in Figure 2.

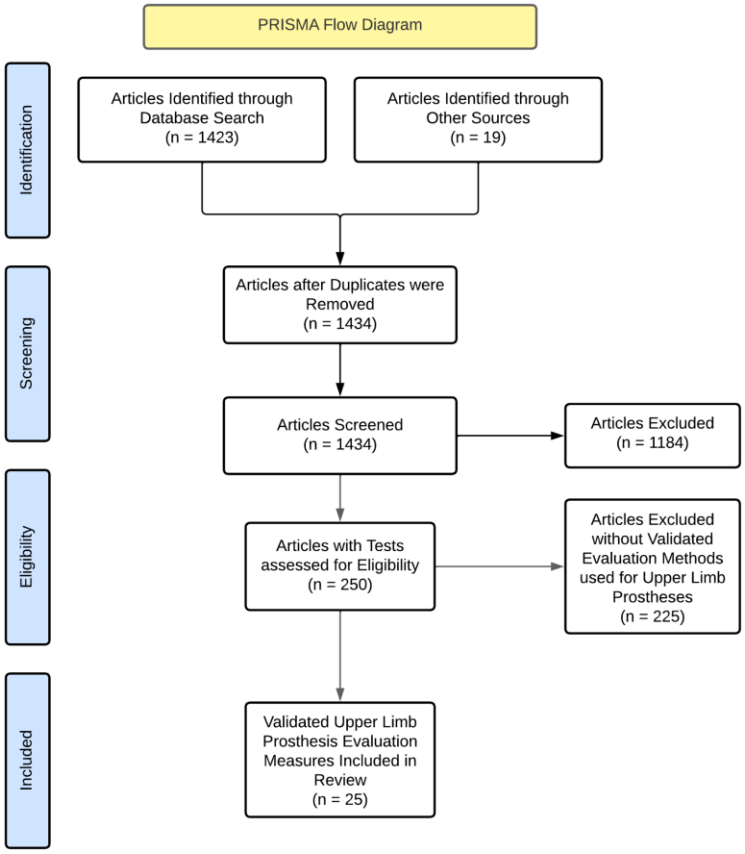


Figure 2. PRISMA Flow Diagram Showing Selection Process

2.2.1: Dexterity and Control Systems Criteria

It was first assessed whether the test offered a holistic assessment by analyzing both dexterity (defined as inclusion of at least five common hand grasps) and control systems (defined as the test's ability to be sensitive in detecting performance variations due to different control systems) [15]. Next, it was evaluated whether the manipulated test objects require the prosthesis to perform a range of grasping movements with varying degrees of hand closure and force, mirroring real-world applications, specifically considering object size and object compliance. This evaluation process also included in-hand manipulation capabilities, as the prosthesis's ability to securely hold and manipulate objects signifies its capacity for complex movements

beyond a simple grasp. Tool usage was the final dexterity and control systems criterion since manipulating tools, such as screwdrivers or toothbrushes, are integral to everyday life.

2.2.2: Additional Considerations

Monitoring progress was also a central aspect of the review, as gauging the test's ability to effectively track changes in dexterity or control systems is essential for assessing efficacy or indicating the need for adjustments. To evaluate this, it was identified whether the test or associated research reported its capability to effectively monitor progress. Furthermore, patient feedback, captured through questionnaires or other methods, was accounted for as it is invaluable as it provides a user-centric perspective on the prosthesis's performance. The criteria also accounted for the evaluators' expertise requirements; some examples include backgrounds in Physical Therapy, Occupational Therapy, or Engineering. Efficiency of test administration and accessibility also factored into the evaluation process since balancing time-constraints, availability, and affordability with quality insights is essential to widespread use of comprehensive evaluations.

Table 1. Custom Evaluation Criteria

Criteria	Assessment Scoring
Holistic Assessment	Did the evaluation test both dexterity and control systems, including multiple hand grasps?
Object Size	Did the test use varying sized objects?
Object Compliance	Did the test use objects that had varying compliances/densities?
In-Hand Manipulation	Did the test require the participant to move objects within the prosthetic hand?
Tool Usage	Did the test include utilizing tools?

Monitoring Progress	Does the test or associated research report its capability to effectively monitor patient progress?
Patient Feedback	Did the test provide a method to include patient feedback, such as a questionnaire?
Evaluator Expertise	Did the test require the evaluator to have prior training or undergo test-specific training?
Efficient Administration	How long does the test take to set-up, administer, and clean up?
Accessibility Considerations	How much did the test cost? Do the objects need custom manufacturing or can they be commonly bought?

2.2.3: Reliability metrics

In addition to the custom review criteria, the reliability of each measure was also assessed. Reliability is a fundamental psychometric property that pertains to the consistency of a measure or test over time [4], [9]. When the same test is administered to the same individual or group under identical conditions, a reliable test should yield the same or very similar results [4], [9]. This review included three types of reliability: test-retest, inter-rater, and internal consistency. Test-retest reliability refers to the stability of a test over time, meaning if the same test is given to the same participants multiple times, the results should be very similar or identical [9]. Inter-rater reliability evaluates the extent of agreement among multiple raters or observers [9]. This is crucial when human observers are involved in data collection to mitigate the risk of subjectivity or bias [9]. Internal consistency gauges the stability of results across items within a test [9]. High internal consistency suggests that the items of the test are likely measuring the same underlying construct [9]. Examining these three dimensions of reliability achieves a comprehensive and robust assessment of each measure's dependability.

To evaluate the test-retest and inter-rater reliability of each assessment, reported Intraclass Correlation Coefficient (ICC) values were used [16]. ICC values range from 0 to 1,

with values near 1 indicating high reliability, while those close to 0 suggest low reliability [16]. If internal consistency was applicable, reported Cronbach's alpha (α) values were referenced [16]. Cronbach's alpha values also range from 0 to 1; values near 1 indicate high internal consistency, and those close to 0 indicate low internal consistency [16]. The scoring range for ICC and α are shown in Table 2 [16].

Table 2. Reliability Scoring

Reliability Measure	>=0.90	0.90>ICC>=0.7	<0.7	Not Reported
ICC	Excellent	Good	Poor	✕
α	Excellent	Good	Poor	✕

2.2.4: Validation metrics

Another essential psychometric property for evaluating tests of prosthetic hand dexterity and control systems is validation. Validity refers to the degree to which the test accurately and reliably measures what it intends to, ensuring the inferences and conclusions drawn from the test results are appropriate and meaningful [4], [8], [9]. Furthermore, it is critical to incorporate various types of validity; this review included face, content, construct, external, concurrent, and predictive [4], [8], [9]. Face validity ensures that the test appears to be measuring what is intended [8], [9]. Content validity guarantees a comprehensive measurement of all facets of the subject [8], [9]. Construct validity, on the other hand, ensures that the designed measurement tool accurately assesses what it purports to measure [8], [9]. External validity confirms the generalizability of the test results to real-world scenarios or contexts beyond the experimental environment [8], [9]. Concurrent validity involves a comparison with an existing test or established criterion to ascertain the validity of the new test [8], [9]. Lastly, predictive validity establishes the link between test scores and performance in a specific domain, aiding in the

prediction of future performance based on current assessments [8], [9]. Each layer of validity was crucial in providing a comprehensive and effective evaluation of prosthetic hand dexterity and control system tests. The validity scoring criteria are shown in Table 3 and were adapted from Resnik et al. [6].

Table 3. Validity Criteria

Validity Type	Excellent	Adequate	Poor	No Evidence
Face	Test clearly appears to measure what is intended	Test appears to measure what is intended	Test does not appear to measure what is intended	No Evidence Available
Content	Test evaluates both dexterity with multiple hand grasps and control system	Test evaluates either dexterity with multiple hand grasps or control system	Test only evaluates a single hand grasp or control system	No Evidence Available
Construct	Results of the test match all the intended measures	Results of the test match some of the intended measures	Results of the test do not match the intended measures	No Evidence Available
External	The test uses objects and tasks that would be experienced or applicable to daily life	The test uses some objects and tasks that would be experienced or applicable to daily life	The test does not use objects and tasks that would be experienced or applicable to daily life	No Evidence Available
Concurrent	The test has been compared with multiple other validated tests	The test has been compared with one other validated test	The test has not been compared with other validated tests	No Evidence Available
Predictive	The test accounts for future improvements in prostheses	The test is capable of measuring current state-of-the-art prostheses	The test is incapable of measuring current state-of-the-art prostheses	No Evidence Available

2.3: Results

A brief description for each of the 25 tests is provided below, discussing them within the context of the review criteria and their psychometric properties. For context and comparison, the median administration time across all tests is ~25 minutes and the median cost per test is ~$200 (for supplies, expendables, etc.). The results of these evaluations have been summarized in tables in the Appendix.

1. Assessment of Capacity for Myoelectric Control [17], [18], [19], [20]:

The Assessment of Capacity for Myoelectric Control (ACMC) is a versatile evaluation tool consisting of 30 items, measured by a rater who observes the user perform self-selected functional tasks. Designed for all ages and different levels of prosthetic limb use, each ACMC item is rated on a four-point scale from zero (incapable) to three (spontaneously capable). The ACMC tests multiple hand grasps using varied objects, in-hand manipulation, tool usage, and monitors progress without requiring the evaluator to have prior training. Although relatively comprehensive, the ACMC limitations include: no active mechanism to capture patient feedback, lengthy administration time (>30 minutes), and significant acquisition costs (~$1,000 USD). The ACMC does demonstrate excellent test-retest and inter-rater reliability, with ICCs of 0.94 and 0.92-0.95 respectively, and performed at least adequately in all our validation scoring. However, its use with future multi-grasp prostheses may have limitations due to the score's upper limit, and its specificity to myoelectric devices prevents the assessment of body-powered prostheses.

2. Anthropomorphic Hand Assessment Protocol [21]:

The Anthropomorphic Hand Assessment Protocol (AHAP) is a specialized protocol designed to evaluate the grasping and retaining abilities of upper limb prostheses. The AHAP comprises 26 tasks involving 25 common household items (spatula, chips can, key, etc.). The AHAP tests

many grasp patterns and specifies which to use for each object. The grip patterns are: pulp pinch, lateral pinch, diagonal volar grip, cylindrical grip, extension grip, tripod pinch, spherical grip, and hook grip. Each task is individually scored for grasping and retaining by a rater using a 0, 0.5, or 1 rating, as dictated by the specific criteria provided in the AHAP paper appendix [21]. The AHAP performs strongly in many of our evaluation criteria due to its diverse range of grasp patterns, varying objects, and lack of prerequisite training. Nonetheless, the AHAP exhibits noteworthy limitations. Notably, scoring can be skewed due to the stringent guidelines for determining correct grasping. Additionally, it lacks an assessment of control systems, omits tasks involving tool usage, offers limited capability for tracking progress, lacks patient feedback, and necessitates a minimum of 30 minutes to complete. Despite these shortcomings, the AHAP exhibits excellent inter-rater reliability with an ICC of 0.969. Furthermore, the AHAP exhibits good test-retest reliability and internal consistency, as evidenced by an ICC of 0.839 and α of 0.846 respectively. The AHAP demonstrates adequate validity for evaluating the performance of current prostheses. Essentially, the AHAP is designed to assess the grasping capabilities of a prosthesis without factoring in control systems. While this serves as a useful measure in technical robotic contexts, its correlation with real-world prosthetic outcomes for patients remains uncertain.

3. Activities Measure for Upper Limb Amputees [22]:

The Activities Measure for Upper Limb Amputees (AM-ULA) is a comprehensive 18-item measure designed for adults with upper limb amputation, which evaluates task completion, speed, movement quality, skillfulness of prosthetic use, and independence. This assessment is designed for adults and is compatible with all types of prosthetic devices. Scoring for the AM-ULA ranges from zero to 40, with higher scores denoting better functional performance. While

the AM-ULA scored highly across most of our criteria with the only gap being patient feedback, it does pose significant challenges. The requirement for a certified hand therapist as a rater, coupled with lengthy administration (>30 minutes), makes the AM-ULA far less accessible. Consequently, inter-rater reliability has shown some variability, with ICCs ranging from 0.69 to 0.95. However, the test exhibits excellent test-retest reliability, with an ICC of 0.91, and strong internal consistency (α: 0.89-0.91). In terms of validity, the AM-ULA performs exceptionally well, earning 'excellent' ratings in nearly all types. Overall, despite its strengths, the limited accessibility of the AM-ULA may restrict its widespread use.

4. Accessible, Open-Source Dexterity Test [23]:

The Accessible, Open-Source Dexterity Test (AOSDT) is a novel assessment method designed to evaluate the performance of robots and adults with upper limb amputation. The methodology includes 24 tasks divided into five distinct manipulation categories, conducted on a rotating platform. The performance metrics of the AOSDT are derived from two parameters: the success rate and speed of task completion. Task success is measured on a 0-4 scale per specific criteria and the overall score is a weighted average of the scoring from each parameter [23]. While the AOSDT scored well on most of our criteria, despite the lack of patient feedback, its recent development in 2022 caused it to lack extensive reliability testing. Although it appears valid for measuring the dexterity of different hand grasps, it does not evaluate control systems. Furthermore, a significant obstacle is its requirement for custom 3D printed parts and rotational test board. This time-consuming construction process and the current absence of comprehensive reliability testing may significantly hinder its adoption.

5. Action Research Arm Test [24], [25]:

The Action Research Arm Test (ARAT) is a 19-item measurement tool designed for use by certified hand therapists to evaluate upper extremity and upper-limb prosthesis performance. The ARAT involves the completion of tasks grouped into four subscales: grasp, grip, pinch, and gross movement, with the performance of these tasks forming the individual's score. Tasks are arranged in a descending order of difficulty, with the most complex task attempted first, based on the hierarchy suggested by Lyle et. al, to enhance test efficiency [25]. Performance is rated on a four-point scale, with zero signifying no movement and three representing normal movement. While the ARAT scored well across most criteria, it does require a certified hand therapist to administer, lacks objects with varied compliances, excludes tool usage, has relatively high acquisition costs (~$650 USD), does not assess control systems, and lacks a method for patient feedback. However, the ARAT demonstrates excellent reliability, with ICCs of 0.965 and 0.998 for test-retest and inter-rater reliability, respectively. The ARAT also exhibits excellent internal consistency, with a α of 0.985, though this analysis was notably conducted with data from stroke patients rather than those with limb deficiencies. Overall, despite its strengths, the ARAT might be too time-consuming for a clinical setting, and could struggle supplying enough information for a research laboratory.

6. Brief Activities Measure for Upper Limb Amputees [26]:

The Brief Activities Measure for Upper Limb Amputees (BAM-ULA) was developed as an alternative to the more comprehensive AM-ULA, to address issues such as the lengthy completion time of approximately 30-35 minutes and a complicated scoring system that requires a trained clinician. The BAM-ULA streamlined this with a ten-item observational measure of activity performance, where each item is scored as either zero for 'unable to complete' or one for

'did complete'. The total score is derived from the sum of these individual item scores. The BAM-ULA demonstrated commendable reliability with a test-retest ICC of 0.91 and an internal consistency α of 0.83. Regarding validity, the BAM-ULA achieved at least an 'adequate' rating in all categories. However, there are concerns that the simplicity and binary scoring system of the BAM-ULA may not adequately reflect the functional capabilities of the prosthesis and might fail to distinguish between the performances of more advanced prosthetic hands.

7. Box and Block Test [27]:

The Box and Block Test (BBT) is a straightforward measure of dexterity and upper extremity function, involving 150 wooden cubes, each 2.5 centimeters per side. The score is based on how many blocks a participant can individually transfer from one compartment, over a partition, to another within 60 seconds. Each successfully moved block earns a point. In terms of our criteria, the BBT scored relatively low due to its inability to measure different hand grasps or control systems, and its exclusive use of identical cubes as the objects with no variation or inclusion of tools. However, the BBT is straightforward to administer, quick, and its reliability has been thoroughly evaluated, scoring highly with a test-retest ICC of 0.96 and inter-rater ICC of 0.99. While the BBT has been validated in various contexts and is commonly chosen in clinical settings due to its time efficiency, it does not comprehensively capture the capabilities of current prostheses and will likely become increasingly outdated.

8. Capacity Assessment of Prosthetic Performance for the Upper Limb [28], [29]:

The Capacity Assessment of Prosthetic Performance for the Upper Limb (CAPPFUL) is an outcome measure tailored for adults with upper-limb deficiencies. This measure evaluates a user's ability to perform 11 tasks (which require diverse hand grasp patterns to complete). It assesses across five distinct functional domains: control skills, component utilization,

maladaptive/adaptive compensatory movements, and task completion. The CAPPFUL is also the only currently validated test that features a complementary patient feedback mechanism, developed to integrate with the task-based evaluation. However, there are challenges associated with its use. Specifically, it necessitates the involvement of an occupational therapist, specialized in upper-limb prostheses, who must also undergo additional training. Furthermore, its duration can be considered lengthy for clinical environments, averaging between 25-35 minutes. In terms of its psychometric attributes, CAPPFUL exhibits commendable reliability, with inter-rater reliability ICCs ranging from 0.88-0.99 and an internal consistency α of 0.79 to 0.82. The measure has also achieved 'excellent' scores in the majority of validity categories. In summary, while the CAPPFUL stands out as a comprehensive assessment tool that includes patient feedback, its utility may be constrained by the specialized expertise required for its administration.

9. Carroll Quantitative Test of Upper Extremity Function [30], [31]:

The Carroll Quantitative Test of Upper Extremity Function (CQT) was originally developed to assess hand function in post-traumatic injury but has been adapted for evaluating upper-limb prostheses. It involves 32 tasks using 18 objects, testing diverse actions to assess dexterity, arm motion, and, to some degree, strength. While some tasks necessitate a power grip, the test predominantly focuses on pinch positions, dedicating 16 tasks to assess the ability to pinch using the thumb in conjunction with each of the four fingers. Performance is scored from 0-3, based on observed task completion quality and is supplemented by a reading from a Smedley dynamometer. However, the CQT has limitations: mandates administration on a custom table, lacks a patient feedback component, has not been compared with other tests, and its reliability has only been verified in pediatric patients with spastic hemiplegia [31]. Despite these

challenges, the CQT excelled in our validation criteria, receiving scores of ‘excellent’ in the majority of validation types. While the CQT has overall good validity, its focus on pinch grips and lack of reliability testing may limit its adoption.

10. Elliott and Connolly Benchmark [32]:

The Elliott and Connolly Benchmark (ECB) is a dexterity evaluation tool designed for use with robotic hands along with upper-limb prostheses. It involves eight objects and employs 13 manipulation patterns including pinch, dynamic tripod, squeeze, twiddle, rock, rock II, radial roll, index roll, full roll, rotary step, interdigital step, linear step, and palmar slide. The ECB initially scores performance on a binary basis—success or failure—based on the specific criteria for each pattern. Following this, it uses custom quantitative metrics to track the translations and rotations of each object along a specified hand coordinate axis [32]. The ECB performed well against our criteria, falling short only in tool usage, capability to monitor patient improvement, and providing a method for patient feedback. Furthermore, while the ECB does not explicitly require prerequisite qualifications, the calculations require extensive mathematical knowledge that likely necessitates training or a researcher to perform the test. Furthermore, its creation in 2020 has limited extensive reliability testing. Although the ECB performed satisfactorily according to our validity criteria, its lack of reliability testing suggests that more comprehensive testing may be needed before it can be considered for widespread use.

11. Functional Dexterity Test [33]:

The Functional Dexterity Test (FDT) is primarily intended to evaluate the functionality of the three-jaw chuck grasp pattern in individuals with hand injuries, but it has also been applied in the assessment of upper-limb prostheses. The test setup features a 16-peg board, placed ten centimeters from the edge of a table. Participants must pick up each peg, flip it over, and reinsert

it following a zig-zag trajectory. The FDT's scoring system considers both net time (the actual time taken to complete the test) and total time (net time plus any penalty seconds). Most errors incur a five second penalty, while dropping a peg results in a ten second penalty. A total time exceeding 55 seconds suggests a non-functional hand. The FDT scored relatively low against our criteria, as it only assesses one hand grasp, lacks a control system evaluation, does not evaluate tool usage, and lacks patient feedback. However, it does assess in-hand manipulation, and is quick and straightforward to conduct. Extensive reliability testing shows an excellent test-retest ICC of 0.92 and inter-rater ICC of 0.94:0.99. Despite this, our validity assessment found the FDT to be moderately poor as it fails to comprehensively evaluate current prostheses. While the FDT is a common choice clinically due to its simplicity and quick administration, it likely lacks the capability for a comprehensive assessment of current and future prostheses.

12. Gaze and Movement Assessment [12], [34], [35]:

The Gaze and Movement Assessment (GaMA) is an innovative tool for evaluating upper-limb prostheses, utilizing motion capture and eye tracking for functional tasks. It provides insights into hand-eye coordination and overall movement quality. While the GaMA does not directly assess multi-grasp dexterity, it reveals the influence of different grasps on movement kinematics. Furthermore, it is highly sensitive in detecting functional changes across control systems and prosthetic components. One significant limitation is its demanding setup, with equipment costs beginning at $2,000 USD, coupled with the need for specialized expertise for optimal use. The GaMA does feature good reliability, with a reported test-retest reliability ICC of 0.75 along with a RM-ANOVA determining its inter-rater reliability to be strong at a 95% confidence level. In terms of validity, the GaMA's detailed analysis of movement kinematics and compensatory actions suggests robust predictive validity, positioning it as an invaluable tool for studying

advanced prostheses in the future. However, its setup, duration, absence of patient feedback, intricate analysis, and cost might make it more suitable for research rather than routine clinical use.

13. Jebsen Hand Function Test [36], [37], [38]:

The Jebsen Hand Function Test (JHFT) is primarily used to evaluate the speed of upper-limb function. It includes seven subtests: writing, card-turning, moving small objects, stacking checkers, simulated feeding, moving light objects, and moving heavy objects. Each subtest measures the time taken in seconds to complete each task, beginning with the non-dominant hand, followed by the dominant hand. The JHFT scored well in our criteria, lacking only in the evaluation of objects with varying compliances and a method for patient feedback. Extensively tested for reliability, the JHFT is very reliable, demonstrating a test-retest ICC ranging from 0.84 to 0.97, an inter-rater ICC of 0.82 to 1.00, and an internal consistency α of 0.95. While the JHFT scored fairly well in our validity testing, it is limited by high acquisition costs (~$400 USD) and exclusively measuring speed. Consequently, the JHFT might be insufficient in delivering comprehensive data for research laboratories.

14. Minnesota Manual Dexterity Test [39]:

The Minnesota Manual Dexterity Test (MMDT) is an evaluation tool designed to measure hand-eye coordination and arm-hand dexterity, primarily focusing on gross motor skills. The MMDT consists of two-timed subtests: placing and turning. In the placing test, a participant is required to move 60 disks located above the testing board into the board's 60 corresponding cutouts. The turning test begins with the disks already placed in the board's cutouts. Participants must pick up each disk with their left hands, pass it to their right, flip it, and reinsert it into its original hole, following a zig-zag pattern. This process is then repeated in reverse, with the right hand passing

the disk to the left hand. The MMDT's scores are based on the speed of completion, not the quality of task performance. The MMDT also does not vary hand grasps, measure control systems, use tools, and is relatively expensive (~$350 USD). However, the MMDT does assess in-hand manipulation and shows good test-retest reliability with an ICC range of 0.79 to 0.87. Despite its quick evaluation time and good reliability, the MMDT might be limited in comprehensively evaluating all aspects of current or future prostheses.

15. Nine-Hole Peg Test [40]:

The Nine-Hole Peg Test (NHPT) is a functional assessment tool initially created to measure an impaired hand's speed of motor function. However, it has been adapted for use with upper-limb prostheses. The test requires participants to individually pick up nine pegs from a container and place them into corresponding holes on a board, with scoring dependent on completion speed (it does not consider quality of task execution). An alternative scoring method is provided which consists of the amount of pegs a participant can place within a designated time limit, typically 50 or 100 seconds. Despite its simplicity, speed, and high reliability—evidenced by a test-retest ICC of 0.92 to 0.95, and an inter-rater ICC of 0.93 to 0.98—the NHPT underperforms against our evaluation criteria. It only assesses in-hand manipulation over time, neglecting multiple grasp patterns, control system assessment, manipulation of diverse objects, tool usage, or patient feedback. In terms of validity, the NHPT performed poorly, given its inability to comprehensively evaluate the functionalities of hand prostheses. While the NHPT is time efficient and therefore a common choice clinically, it is likely unable to comprehensively evaluate the dexterity and control system capabilities of modern prostheses.

16. Prosthetic Hand Assessment Measure [41]:

The Prosthetic Hand Assessment Measure (PHAM) quantifies traditionally qualitative performance metrics for upper limb prostheses. Using a custom PVC frame with four LED-marked sections, it assesses prosthetic functionality at various arm angles. Participants move one of four basic objects (e.g., a cylinder representing a glass) between sections, depending on activated LEDs, using grip patterns like power, tripod, pinch, or key. The PHAM employs five inertial measurement units and a custom piezoresistive mat to record movement details, which are then analyzed using custom equations to assess metrics like 3D deviations in the chest and shoulder, 2D translational displacement, and completion rate [41]. As for its psychometric properties, the PHAM has yet to undergo reliability testing, though it achieved mostly 'excellent' scores in our validation criteria. Though comprehensive, the PHAM's reliance on specialized equipment, high cost (~$500 USD), intricate administration, and scoring, coupled with its untested reliability, might limit its broader adoption.

17. Purdue Pegboard [42], [43], [44]:

The Purdue Pegboard Test (PPT) is an evaluation measure designed to measure gross movements of fingers and finger dexterity. The board used in the test features four cups at the top and two vertical rows of 25 small holes down the center. The two outer cups hold 25 pins each, the cup to the immediate left contains 40 washers, and the one to the right of the center holds 20 collars. The test starts with the participant using their right hand to insert as many pins as they can into the right row within 30 seconds, followed by the left hand placing pins into the left row for 30 seconds. Subsequently, both hands insert pins into both rows within another 30 second period. The final task requires the participant to assemble as many pins with washers and collars as they can using both hands within 60 seconds. The test produces five scores: the

number of pins inserted by the right hand, the left hand, both hands, the sum of these three, and the number of assembled pins. In our evaluation, the PPT received a moderate score. While it is exceptional at assessing finger dexterity and indirectly measures multiple grasp patterns, the PPT does not directly evaluate control systems, tool usage, or provide a method for patient feedback. It does show excellent test-retest and inter-rater reliability, with ICCs of 0.91 for both. In terms of validity, while the PPT effectively evaluates finger dexterity and in-hand manipulation, it falls short in providing a comprehensive assessment of prostheses dexterity and control. However, the PPT is likely a good evaluation tool for clinicians who wish to effectively evaluate finger dexterity.

18. Refined Clothespin Relocation Test [45], [46]:

The Refined Clothespin Relocation Test (RCRT) was developed to evaluate individuals' proficiencies in using a prosthesis, specifically their compensatory movements and the time taken to perform a grasping and repositioning task. Initially researched in a motion capture laboratory, the RCRT has been adapted for clinical application [46]. The test requires patients to relocate three clothespins from a horizontal plane to three different locations on a vertical pole set at low, medium, and high levels. Unlike traditional tests, the RCRT does not employ a conventional scoring system. Instead, it compares the performance of prosthesis users to a control group of able-bodied individuals. Despite its simplicity and its ability to assess in-hand manipulation and the path of motion for clothespin relocation, the RCRT performed poorly against our evaluation criteria. It does not evaluate multiple grasp patterns, control systems, varying objects, tool usage, or offer a method for patient feedback. Furthermore, despite being established in 2016, the RCRT has no published reliability testing. For our validity assessment, the RCRT falls short in providing a comprehensive evaluation of current or future prostheses.

Due to its limitations and the lack of robust reliability data, the RCRT may face limited use in clinical and research applications.

19. Southampton Hand Assessment Procedure [47], [48]:

The Southampton Hand Assessment Procedure (SHAP) is specifically designed to evaluate the functionality and efficiency of upper limb prostheses, comprising six abstract objects and 14 Activities of Daily Living (ADL) tasks. All tasks are timed by the individual taking the test in an attempt to reduce the reliance on the observer or clinician's reaction times. The objects, placed on a dual-sided board with a blue felt side for abstract tasks and a red plastic side for ADL tasks, are timed and recorded by the assessor. These timings are then normalized to a score of 100 using a method devised by Light, Chappell, & Kyberd [47]. The scoring software, available for purchase through the SHAP website, associates each of the 26 tasks with one of six prehensile patterns, enabling the creation of a SHAP Functionality Profile—a numerical assessment of hand function highlighting areas of extraordinary skill or potential impairment. Notably, SHAP scores can exceed 100 for exceptionally quick task completion, while scores under 100 may indicate functional impairment. While the SHAP performed well in our evaluation, it lacks a method for patient feedback, and has two significant drawbacks: a trial takes at least 45 minutes to complete, and it costs over $2500USD for the equipment and license to the proprietary scoring software. However, with ICC values of 0.93 and 0.89 for test-retest and inter-rater reliability, respectively, the SHAP is considered very reliable. It also performed well in our validity assessment, indicating that it can accurately measure the dexterity of current and future prostheses. However, due to its lengthy administration time, absence of control system assessment, and high cost, its clinical and research use will likely be limited.

20. Sollerman Hand Function Test [49], [50], [51]:

The Sollerman Hand Function Test (SHFT) is an evaluative tool designed to measure the functionality of adult hands impaired due to injury or disease, but has been adapted for upper-limb prosthesis evaluations. It assesses the performance of seven distinct hand grips: pulp pinch, lateral pinch, tripod pinch, five-finger pinch, diagonal volar grip, transverse volar grip, spherical volar grip, and extension grip. The SHFT consists of 20 items, each encompassing 20 subtasks, each of which is rated on a scale of zero to four. Scores are determined based on the time taken to complete the task, the quality of task execution, and the assessor's perception of the task's difficulty. The SHFT scored highly in our evaluation criteria, only lacking a control system assessment and a method for patient feedback. The SHFT also boasts excellent reliability, with a test-retest ICC of 0.96:0.98 and an inter-rater ICC of 0.98. However, it is important to note that the test-retest reliability was predominantly evaluated on stroke patients [51]. In our validity assessment, we ascertain that the SHFT can accurately evaluate the dexterity of various hand grasps with current and future prostheses. The SHFT may be a good test for dexterity in a research environment, but would benefit from a reliability evaluation specifically with prostheses, the inclusion of a patient questionnaire, and an incorporated assessment of control systems.

21. Sequential Occupational Dexterity Assessment [52], [53]:

The Sequential Occupational Therapy Dexterity Assessment (SODA) is an assessment tool originally designed to evaluate hand function in individuals with rheumatoid arthritis. However, it has been adapted for use with upper-limb prosthetics. The SODA consists of 12 tasks, each rated from zero to four based on performance and zero to two based on perceived difficulty, contributing to a total evaluation score. A significant challenge in adapting the SODA for

prosthetic users is that six tasks necessitate the use of both hands, yet only one hand is scored. This approach is predicated on the presumption that rheumatoid arthritis would symmetrically affect both hands, thereby yielding identical scores. However, this may not be applicable to the vast majority of people with limb deficiencies, who have a single prosthetic limb. The SODA met most of our evaluation criteria, except for control system assessment, tool usage, needing at least 30 minutes to administer, and a method for patient feedback. The reliability of the SODA remains a point of contention in our view. It has an internal consistency α of 0.91, but the test-retest reliability was assessed using the Pearson correlation coefficient (r) instead of the ICC and lacks any testing for inter-rater reliability. While the SODA did receive an r value of 0.93, the ICC value would offer a more accurate representation of reliability because it accounts for the difference of the means of measures [54]. As for validity, the SODA shows promise but requires further adaptation for prosthetics before it can be used clinically.

22. Timed Measure of Activity Performance in Persons with Upper Limb Amputation [55]: The Timed Activity Performance in Persons with Upper Limb Amputation (T-MAP) was developed to provide a timed measure for the functional outcomes of persons with upper limb amputation. Although the T-MAP can also assess performance without a prosthesis, this aspect was not considered in our review. The T-MAP incorporates five tasks: drinking water, face washing, food preparation, eating, and dressing. Therapists assess both the time taken and the level of independence displayed during each activity. The independence metric uses a 3-point scale: 1 indicating dependency, 2 for verbal assistance required, and 3 for independent action, with or without aid. By aggregating the independence ratings and times, overall scores for both parameters are derived. In our evaluation, the T-MAP performed exceptionally well, only lacking a patient feedback mechanism. It boasts a test-retest reliability ICC of 0.93, but lacks

inter-rater reliability data. On most validity aspects, the T-MAP achieved an 'Excellent' rating. Overall, while the T-MAP provides insights into the time it takes for someone with a limb deficiency to perform daily activities, its sole focus on timing limits its depth of analysis.

23. Unilateral Below Elbow Test [56]:

The Unilateral Below Elbow Test (UBET) evaluates bimanual activities in both prosthesis wearers and non-wearers (for this review, we will be excluding the non-wearers portion). The UBET employs four age-specific categories, reflecting the developmental stages of hand function (2–4, 5–7, 8–10, and 11–21). Each category contains nine tasks, utilizing everyday household objects relevant to that age's hand development level. UBET's dual rating system, 'Completion of Task' and 'Method of Use', allows for assessing the overall task completion quality and recognizing functional disparities among different control systems. However, UBET's limitations include the need for an OT during administration, its lengthy procedure, an age ceiling of 21, and the absence of integrated patient feedback. In terms of reliability, the 'Completion of Task' has an inter-rater ICC ranging from 0.77 to 0.87, while 'Method of Use' shows a test-retest reliability ICC between 0.70 and 0.85, and a good Cohen’s kappa value (equal relevance to an ICC) of 0.68 to 0.82 for inter-rater reliability. In validity metrics, the UBET predominantly scores as 'excellent'. Overall, the UBET serves as a useful tool for pediatric research applications.

24. University of New Brunswick Test of Prosthetics Function [20], [57]:

The University of New Brunswick Test (UNBT) is an evaluative methodology created to measure upper-limb prosthetic function in those with limb loss aged two to 21. The UNBT, capable of evaluating both body-powered and myoelectric prostheses, categorizes participants into four age groups: 2-4, 5-7, 8-12, and 13-21. All age groups include three subtests featuring

ten tasks each. These tasks are assessed on two aspects: spontaneity and skill of prosthetic function. Both elements are scored on a 0-4 scale, and the scores for each category are tallied at the end of each subtest. The UNBT also provides a therapy recommendation chart based on the scores obtained in each category. In our evaluation criteria, the UNBT performs exceptionally well, falling short only in the relatively high cost of $500, absence of a questionnaire—likely due to patient age—, and a rather lengthy administration time of at least 30 minutes. The UNBT also demonstrates good reliability, with a test-retest ICC of 0.74:0.79, an inter-rater ICC of 0.72:0.73, and an internal consistency α of 0.74. In terms of validity, the UNBT is among the most comprehensive tests we reviewed for evaluating the dexterity and control systems of modern and future prostheses. The UNBT may be a great option for use in a research laboratory but is likely too long to feasibly administer in a clinical setting.

25. Wolf Motor Function Test [58]:

The Wolf Motor Function Test (WMFT) is a diagnostic tool initially designed for assessing upper limb dexterity and strength in stroke recovery patients, but it has been adapted for use with prostheses. The WMFT, originally consisting of 21 items and tasks, is now typically used with 17 items and tasks, with each task capped at a maximum of 120 seconds. The first six objects are used for timed functional tasks; items seven through fourteen are used to measure strength, and the remaining objects assess movement quality. Each item is scored on a scale of zero to five, and the final score is the aggregate of all item scores. In our evaluation criteria, the WMFT scored well, falling short only in control system assessment, the absence of a patient feedback questionnaire, and a relatively long administration time of at least 30 minutes. Although the WMFT demonstrates excellent reliability results, primarily with stroke patients, with a test-retest ICC of 0.97, an inter-rater ICC of 0.92:0.99, and an internal consistency α of 0.91, it would

benefit from further reliability testing with prostheses. Our validity assessment also gave the WMFT high marks. However, its administration time may restrict clinical use and it should likely incorporate a control systems assessment, along with updated reliability testing, before widespread implementation could occur in a research setting.

2.4: Discussion

This review revealed that in the current state of upper-limb prosthetic research, a significant discrepancy exists between the pace of mechatronic development and the available evaluation methodologies. Among the reviewed tests, there is a limited ability to comprehensively measure the performance of upper-limb prostheses in terms of directly quantifying both multi-grasp dexterity along with the impact of varying control systems. This highlights the need for urgent action to establish standardized, comprehensive evaluation methodologies suitable for both clinical and research settings. However, addressing this issue is a nuanced challenge since the success of a test can often be influenced by competing interests. Clinical environments, typically overseen by physical therapists, occupational therapists, and certified hand therapists, often face tight schedules that blend evaluation with treatment. Here, quick tests like the BBT, MMDT, and NHPT are likely preferred. Yet, their current forms inadequately capture the capabilities of modern prostheses, particularly in multi-grasp dexterity. Conversely, research laboratories tend to lean toward exhaustive tests like the AHAP, AM-ULA, SHAP, and UNBT. However, these tests' extensive setups, niche objects, significant costs, and intricate procedures have hampered widespread implementation. Therefore, while validated tests are available, their limitations have resulted in inconsistent adoption that has also hindered a unified and standardized evaluation framework. This lack of standardization and validated evaluation measures has caused multiple problematic consequences. Notably, new prosthetic

devices tend to be assessed using methods devised by their own developers, introducing potential bias and undermining validity. Furthermore, the lack of a unified, standard assessment process has resulted in redundancy among current validated tests, further complicating the evaluation process. One example of this was demonstrated by Burger et al., who found that the ACMC and UNBT are equally capable of evaluating myoelectric prostheses [20].

An essential component lacking in 24 of the 25 previously mentioned validated tests is an integrated mechanism to gather patient feedback. We maintain that patient feedback, often collected via questionnaires and surveys, forms a crucial part of a comprehensive upper-limb prosthesis evaluation [59], [60]. Currently, self-reported surveys and task-based assessments are designed independently, leaving assessors to combine and interpret data from separate, potentially incompatible sources. Research, such as the study by Burger et al., has underscored the importance of questionnaires for yielding invaluable insights and revealed the limitations of exclusively depending on clinical tests [20]. There are deeper underlying predictors of performance, often first identified by patients, that extend beyond just the quality of the prosthesis or its control system. Factors such as socket fit, heat or sweat management, skin irritation, suspension or harnessing, and overall discomfort not only impact an individual's ability to use and effectively operate their prosthesis but also determine its consistent use (if it is not comfortable, the prosthesis will not be used) [61], [62], [63]. Furthermore, the psychosocial impact of a prosthetic device on a person's self-image, confidence, and social interactions can only be assessed through patient feedback, informing design improvements and support services [62], [64], [65]. Patient feedback is indispensable in evaluating the prosthesis's overall effectiveness, and should be included to provide a holistic view that not only considers the physical and mechanical aspects but also addresses the psychological and social implications.

2.5: Conclusion

This narrative review assessed upper-limb prosthetic dexterity and control system assessment techniques. The primary objective was to analyze the essential characteristics and psychometric attributes of these evaluations and identify any existing gaps in the field. The analysis revealed a dichotomy in current clinical assessments for upper-limb prosthetics. Many commonly used tests are quick and easy to administer but offer limited insights, often failing to capture the full range of dexterity and control system capabilities of advanced prosthetics. Conversely, comprehensive tests provide extensive data but are lengthy, costly, and complex, limiting their widespread use. Furthermore, an integrated method for patient feedback is currently absent in nearly all testing methods. Thus, both clinical and research environments urgently require the creation or refinement of tests to more accurately evaluate the dexterity and control systems of modern upper-limb prosthetics.

Chapter 3: Using the Anthropomorphic Hand Assessment Protocol

A majority of this chapter is submitted to be published in BMC Biomedical Engineering: Siegel JR, Harwood JK, Lau AC, Brenneis DJA, Dawson MR, Pilarski PM, Schofield JS. A Performance Evaluation of Commercially Available and 3D-Printable Prosthetic Hands using the Anthropomorphic Hand Assessment Protocol, BMC Biomedical Engineering. 4640dab0-08cb-4033-9beb-f3d94ddd7ffa

3.1: Background

To gain a practical and comprehensive understanding of modern prosthetic devices and available evaluation methods, this study employed the Anthropomorphic Hand Assessment Protocol (AHAP). Posing a research question about the impact of 3D printing on prosthetic development facilitated the investigation of the AHAP in a research setting. This question necessitated a standardized, validated, and informative task-based evaluation method, as it is essential for accurately and fairly assessing 3D-printed prostheses against clinically prescribable and commercially available (CPCA) devices. Thus, this study not only provides valuable insights into the efficacy of 3D-printed prostheses compared to CPCA devices but also underscores the importance of robust evaluation protocols and sheds light on their potential shortcomings for advancing prosthetic technology.

Additive manufacturing, or 3D printing, began to influence prosthetics in 2012, when the first 3D-printed prosthesis, “Robohand”, was introduced [66]. Since 2012, the field has continuously evolved, bringing to market a variety of open-source 3D-printed prosthetic models. This method significantly reduces production costs, with 3D-printed prostheses starting as low as $19USD in raw materials and parts, in stark contrast to clinically prescribable and commercially available devices that can often cost upwards of $20,000 USD [67], [68], [69]. The affordability and accessibility of 3D printing have also made prosthetics research more inclusive, enabling

studies across labs with varying funding levels and promoting a diverse research ecosystem. This technology not only facilitates rapid prototyping, allowing for the swift design, production, and testing of prosthetic components but also supports the customization of designs to meet individual anatomical needs and preferences. Such flexibility is key from a research and design perspective when exploring prosthetic functionality and enhancing user comfort. Despite these advancements, 3D printing technology in prosthetics is still maturing, and comprehensive research into the dexterity and functionality of 3D-printed prosthetic hands, as juxtaposed with CPCA prostheses, remains in the nascent stages.

Previous research in prosthetics has often been compartmentalized, with studies typically focusing on either 3D-printed or CPCA prosthetic hands in isolation. For example, Llop-Harillo et al. and Cabibihan et al. performed work to evaluate the performance of a variety of 3D-printed hands [69], [70]. In contrast, Belter et al. and Kannenberg et al. have independently explored the efficacy of CPCA prosthetic hands [71], [72]. The varied directions taken in prosthetics research have created a significant knowledge gap concerning the systematic evaluation and benchmarking of both 3D-printed and CPCA prostheses within the same standardized testing framework. This study aims to fill this void and contribute to this evolving landscape by conducting an extensive evaluation that simultaneously assesses both types of prosthetic hands during grasping tasks with physical objects that are representative of those likely encountered by users in their daily activities.

3.2: Methods

3.2.1: Testing Methods:

The Anthropomorphic Hand Assessment Protocol (AHAP) was the employed evaluation metric. As described in Chapter 2, this specialized and validated procedure was developed to

assess the multi-grasp dexterity of hand prostheses through grasping and holding common items [13]. The AHAP involves 26 specific tasks using 25 household objects and encompasses 10 different grip patterns: hook grip, spherical grip, tripod pinch, extension grip, cylindrical grip, diagonal volar grip, lateral pinch, pulp pinch, index pointing/pressing, and platform [13]. The AHAP tasks and grips are shown in Figure 3. The AHAP's reliability is notable, scoring a test-retest reliability intraclass correlation coefficient (ICC) of 0.839, an inter-rater reliability ICC of 0.969, and an internal consistency Cronbach's alpha of 0.846 [13], [14]. This study included three open-source 3D-printed prosthetic hands—HACKberry Hand, HANDi Hand, and BEAR PAW—and three frequently prescribed, CPCA prosthetic hands—Össur i-Limb Quantum, RSL Steeper BeBionic Hand V3, and Psyonic Ability Hand. Additionally, the AHAP results from these six hands were then combined with previously published AHAP scores from four additional 3D-printed hands: Dextrus v2.0, IMMA, InMoov, and Limbitless [13].

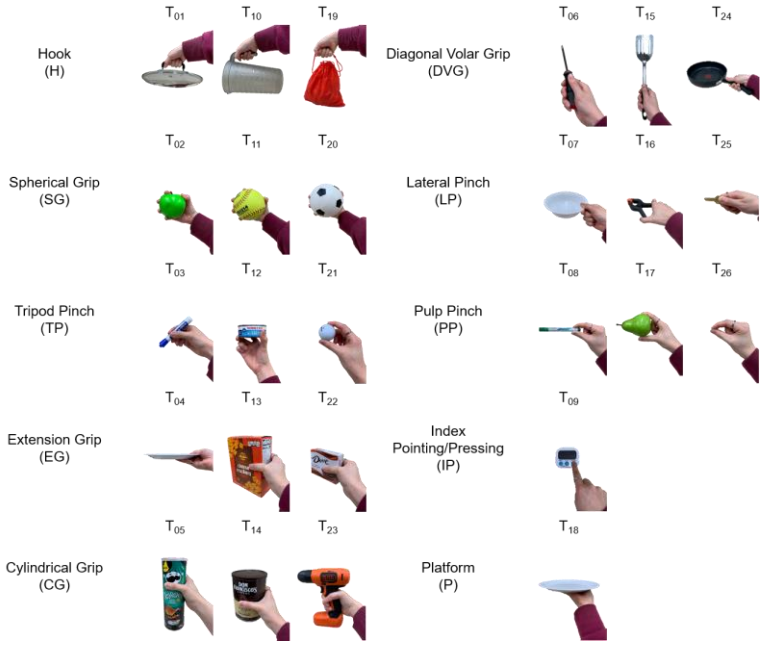


Figure 3. AHAP Tasks and Grips

Execution of the AHAP followed the established protocol described in [21]. This involved a lead investigator who conducts the testing and scoring, and three test investigators who are responsible for operating the prosthesis. Each test investigator performed three trials, for a total dataset of nine trials per prosthesis (3 test investigators x 3 trials each = 9 total trials). The standard AHAP procedure required replicating the test with multiple trials and three separate test investigators to account for potential variability in the way objects might have been manipulated [21]. Before starting the testing protocol, the lead investigator briefed the test investigators on the proper grip type for each object and allowed a one-minute familiarization period. Each AHAP trial commenced with the lead investigator presenting one of the 26 objects to a test investigator in a specific orientation. For each grip type (except index pointing/pressing), the prosthesis was initially positioned with the palm facing upwards. Upon securing the object, the prosthesis was required to sustain its grip on the object for a duration of three seconds (the grasping phase). This was followed by a 180° pronation to a palm-down position (clockwise for left-handed prostheses and anti-clockwise for right-handed prostheses), again trying to maintain its grip for an additional three seconds (the maintaining phase). Further descriptions of the grasping and maintaining phases for each grip type and posture can be found in the original AHAP instructions by Llop-Harillo et al. [21].

Following the AHAP protocol [21], during the grasping and maintaining phases for each object the lead investigator scored the prosthesis's performance. Accordingly, a score of 1 was received if the object was held with the specified grip for the allotted time. A score of 0.5 was given if the prosthesis held the object for the designated time but did not follow the specific grip requirements described by the AHAP. Finally, a score of 0 was received if the prosthesis was unable to hold the object at all. Then, if there was no movement of the object within the hand

during the maintaining phase, a score of 1 was awarded. If the object moved but did not drop, then a score of 0.5 was received, and a score of 0 was given if it was not able to maintain the object. A score of 0 may also be assigned at the lead investigator's discretion, without attempting the grip, when it was deemed likely that an attempt would cause functional damage to the hand.

Scores were separated by phase: grasping or maintaining. The scores for each prosthetic hand were further separated into 10 categories for grasping and nine categories for maintaining classified by grip type/posture. These scores were averaged across the three test investigators such that individual grasping and maintaining comparisons could be made between hands. Finally, an overall grasping ability score (GAS) was given for each hand by averaging all scores.

3.2.2: Hands Tested

A brief description for each of the six tested hands is provided below. Technical data has also been summarized in Table 4. Data in this table were tested and published by the respective manufacturers of each hand.

Table 4. Technical Data of the Six Tested Prosthetic Hands

Hand	Mass (g)	Active Digit Flexion	Active Thumb Rotation	Closing Speed (s)	Pinch Grip Force (N)	Power Grip Force (N)	Max Static Load (kg)
i-Limb Quantum	528	✓	✓	0.8	35*	136*	90
BeBionic Hand V3	614	✓	✕	1.0	12.5	140	45
Ability Hand	500	✓	✓	0.2	9.3	66.0	35.8
HACKberry Hand	223	✓**	✓	0.3	3.0	6.0	2
HANDi Hand * **	256	✓	✓	0.8	0.3	1.5	0.6
BEAR PAW	177	✓	✓	0.7	2.0	7.2	-

*Grip forces are no longer stated in spec sheets for the i-Limb Quantum so these values are estimated from the similar i-Limb Ultra Revolution model [73]

** For HACKberry hand: 1st motor flexes index, 2nd motor flexes D3-D5, and Thumb flexion is passive
*** For HANDi Hand: used Dymond D47 servo motors instead of the original Hitec HS35-HD motors

Össur i-Limb Quantum

Figure 4. Össur i-Limb Quantum

The Össur i-Limb Quantum [73], [74], [75] is a CPCA prosthetic hand. Built with titanium digits, the i-Limb Quantum weighs 658 g, has a static limit finger carry load of 48 kg and a hand load static limit of 90 kg. It is equipped with five independently motorized fingers and a powered thumb rotation with manual override. Using the My i-Limb™ iOS apps, it has up to 36 selectable grips, both pre-programmed and customizable.

RSL Steeper BeBionic V3

Figure 5. RSL Steeper BeBionic V3

The RSL Steeper BeBionic Hand V3 [76], [77], [78] has been commonly prescribed as a hand prosthesis since 2010 [76]. Built with carbon fiber digits, the BeBionic V3 weighs 588 g, has a static limit finger carry load of 25 kg and a hand load static limit of 45 kg. The hand's individual motors located in each finger and at the thumb base enable the user to have five degrees of actuation (passive thumb rotation) and 14 different grips and hand positions [76].

Psyonic Ability Hand

Figure 6. Psyonic Ability Hand

The Psyonic Ability Hand [79], [80] is a CPCA prosthetic hand. It is equipped with fingertip sensors that detect pressure during gripping and send vibrations to the user's arm, offering some tactile feedback. With its carbon fiber shell, the Ability Hand weighs 520 g and has a maximum grip force of 66 N. The Ability Hand has five independently motorized fingers and a powered thumb rotation with manual override. The hand comes pre-programmed with 32 grip patterns, including 19 predefined options. The Ability Hand is compatible with third-party EMG pattern recognition systems, EMG direct control systems, linear transducers, and force-sensitive resistors along with integration with iOS and Android mobile apps for adjustment of settings and updates.

HACKberry Hand

Figure 7. HACKberry Hand

The HACKberry Hand [81], developed by Japanese startup Exiii, is an open-source 3D-printable bionic prosthetic hand. The HACKberry hand is made of polylactic acid (PLA), weighs 475 g, and can support loads up to 2 kg. Equipped with three motors, it has partially-motorized long fingers and powered thumb rotation. The third, fourth, and fifth fingers are coupled, allowing them to flex and extend as a group. Additionally, it has passive thumb flexion, allowing it to meet the index finger for a pinch grip. When attached to its arm component, it has passive wrist flexion and rotation. The HACKberry hand does not come with pre-programmed grips, however they can be defined at the user's discretion so long as it remains within the range of motion of the finger joints.

HANDi Hand

Figure 8. HANDi Hand

The Humanoid, Anthropometric, Naturally Dextrous Intelligent (HANDi) Hand [82] is a 3D-printed multi-articulating hand. It was developed at the Bionic Limbs for Improved Natural Control Laboratory (BLINC Lab). The hand can be used in conjunction with the Bento Arm, a five-degree-of-freedom robotic arm designed for myoelectric training and research applications [83]. Made of PLA, the HANDi Hand weighs 256 g and has a maximum grip force of 4.2 N. Six integrated Dymond D47 servo motors allow for individual finger articulation, with separate thumb rotation and flexion. Rotary potentiometers in the joints and force-sensitive resistors in the fingertips can provide finger position and force information to machine learning algorithms. Further, a USB webcam is integrated into the palm, providing visual information about the hand's workspace. The HANDi hand must be programmed by the user as it does not come with pre-programmed grips.

BEAR PAW

Figure 9. BEAR PAW

The Bionic Engineering and Assistive Robotics Pediatric Assistive Ware (BEAR PAW) is a 3D-printed pediatric prosthetic hand developed at the UC Davis Bionic Engineering and Assistive Robotics Laboratory (BEAR Lab) [84]. Modeled after the anatomical proportions of an 8-year-old child and printed using PLA, it has a weight of 177 g with a maximum grip force of 7.216 N. The BEAR PAW has five independently motorized fingers and a powered thumb rotation. The hand comes pre-programmed with 10 grip patterns, and additional grips can be created by the user.

Dextrus v2.0, IMMA, InMoov, and Limbitless

Llop-Harillo et al. published AHAP data for four open-source adult 3D-printed hands (Dextrus v2.0, IMMA, InMoov, and Limbitless) [70]. Printed either using PLA or Ninjaflex®, they ranged in weight from 131 g to 201.5 g, but did not publish grip force or static load values. These four adult hands were all underactuated systems with a range from 14 to 17 degrees of

freedom and 1–6 degrees of actuation. The testing for these hands was done using the custom-made Able-Bodied Adapter presented by Llop-Harillo and Pérez-González in 2017 [85].

3.2.3: Data Analysis:

The data analysis was designed to allow for the group-based comparison of overall grasping scores between the 3D-printed and CPCA hands, followed by an assessment of grasping versus maintaining scores within these specific groups. Subsequently, pairwise grip comparisons were conducted within each subgroup. This procedure is depicted in Figure 10, with further descriptions of each comparison provided below.

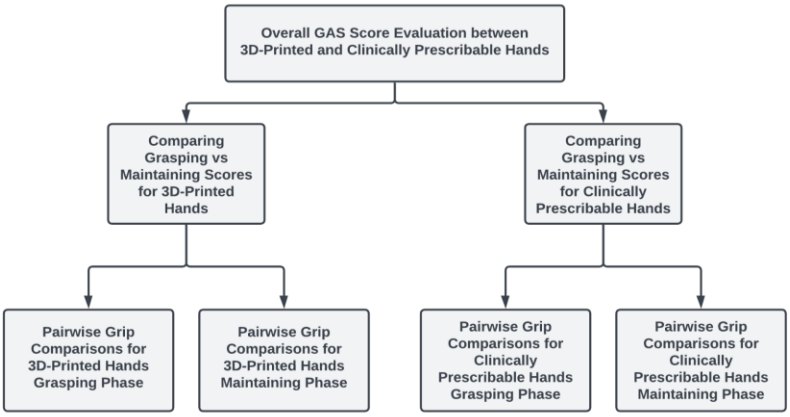


Figure 10. Statistics Logic Flow Diagram

3.2.3.1: Overall GAS Score Evaluation

The Mann-Whitney U-Test was utilized to compare the overall performance (GAS scores) between CPCA and 3D-printed prosthetic hands. The Mann-Whitney U-Test was chosen to accommodate the small sample size, non-parametric data, and to fit the assumption of a bimodal distribution, as a normal distribution could not be assumed for the GAS scores [25]. The Mann-Whitney U-Test operates by converting actual data points into ordered ranks to form a permutation distribution, from which it calculates the p-value [28]. It was also confirmed, using

our data characteristics, that the prerequisites for the Mann-Whitney U-Test were met [28]. This included ensuring that the observations within and across groups were independent and that the response variable was ordinal or continuous [28].

3.2.3.2: Comparing Grasping vs Maintaining Scores

It was also investigated whether there were differences between grasping and maintaining scores for both the 3D-printed and CPCA hands as individual groups. To do this, two additional Mann-Whitney U-Tests were conducted. In consideration of conducting a series of related tests, the Bonferroni Correction was used to keep the family-wise error rate below 5%. Consequently, the inclusion of these two Mann-Whitney U-Tests, alongside the overall GAS score comparison, established a test-wise significance threshold of 1.67% [29].

3.2.3.3: Pairwise Grip Comparisons

Finally, Friedman's Test (a non-parametric alternative to a Two-Way ANOVA model without an interaction term), followed by the Nemenyi Test (if applicable) were used to determine which specific grips the hands may have struggled with [30], [31]. Similar to the Mann-Whitney U-Test, Friedman's Test works by converting the data into ordered ranks and using a permutation distribution to calculate the p-value (indicating if there were differences among grips at a significance threshold of 5%). The Nemenyi Test, designed as a non-parametric pairwise comparison method, enabled the identification of specific grip patterns that exhibited lower performance, provided that Friedman's Test indicated significant differences [30].

3.3: Results

3.3.1: AHAP Results

Table 5 presents the average overall GAS, grasping, and maintaining scores along with their standard deviations for each prosthetic hand, as measured using the AHAP. The scores in Table 5 are expressed as a percentage of the maximum achievable score [10].

Table 5. AHAP Scoring Results

Type	Hand	GAS (%)	Grasping (%)	Maintaining(%)
CPCA	Össur i-Limb Quantum	90 ± 2	85 ± 1	96 ± 3
	RSL Steeper BeBionic V3	77 ± 0	77 ± 0	79 ± 1
	Psyonic Ability Hand	95 ± 0	97 ± 1	94 ± 1
3D-Printed	HACKberry Hand	53 ± 2	64 ± 0	36 ± 5
	HANDi Hand	66 ± 2	77 ± 1	53 ± 3
	BEAR PAW	68 ± 1	77 ± 3	57 ± 0
	Dextrus v2.0	48 ± 4	61 ± 3	34 ± 6
	IMMA	57 ± 2	77 ± 1	37 ± 4
	InMoov	49 ± 1	57 ± 1	40 ± 2
	Limbitless	50 ± 3	63 ± 2	37 ± 4
Color Scale:	25%			100%

3.3.2: Data Analysis Results

3.3.2.1: Overall GAS Score Evaluation:

The analysis indicated a statistically significant difference in overall GAS scores between the CPCA and 3D-printed prostheses, with data inspection indicating lower performance by the

3D-printed prostheses. The initial assumption under the null hypothesis was that both groups of prostheses (3D-printed and CPCA) showed identical underlying distributions. Figure 11 displays the Mann-Whitney U-Test results, including a P-value of 0.008333. This permitted rejecting the null hypothesis using a test-wise significance level of 1.67% (post Bonferroni Correction).

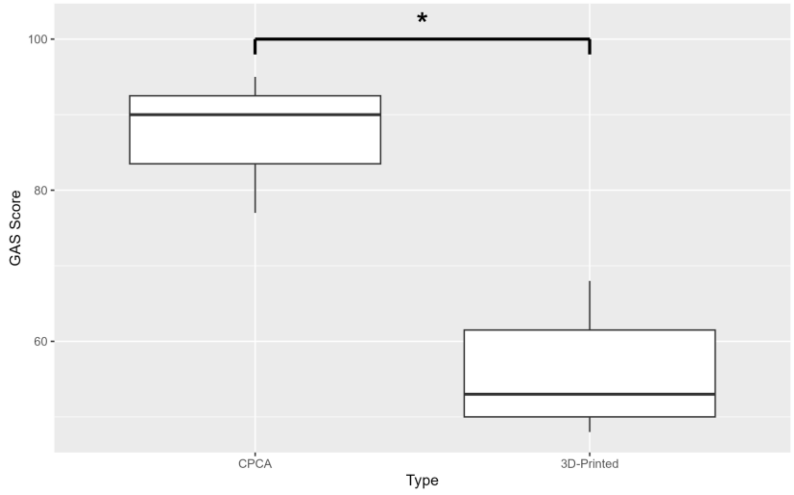
GAS Score Comparison Between CPCA and 3D-Printed Prostheses		
	Range	Clinical: 18
		3D-printed: 20
	Median	Clinical: 90
		3D-printed: 53
	Test Statistic	21
	P-Value	0.0083

Figure 11. Overall GAS Score Mann-Whitney U-Test Results: CPCA hands significantly outperformed 3D-Printed

3.3.2.2: Comparing Grasping vs Maintaining Scores:

The outcomes of these analyses demonstrated statistically significant disparities within the 3D-printed hands, also at a test-wise significance level of 1.67%, indicated by a P-value of 0.0011 and a test statistic of 49. Conversely, the comparisons within CPCA prosthetic hands did not reveal statistical significance, as shown by a P-value of 0.50 and a test statistic of 4. The findings of these tests are detailed in Figure 12. For these tests, the null hypothesis assumed that the scores for grasping and maintaining were identical within each type of prosthetic hand—first examining this assumption for 3D-printed hands, and then separately for CPCA hands.

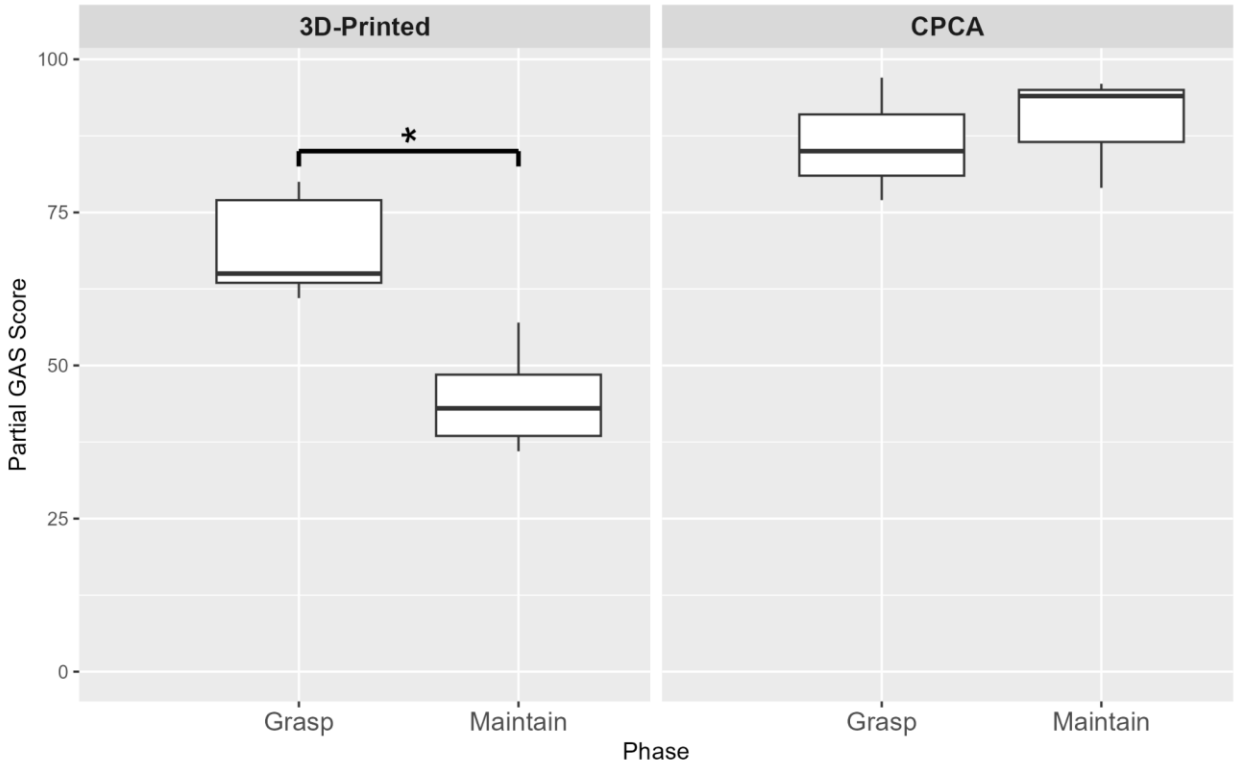


Figure 12. Comparing Grasping vs Maintaining Scores: 3D-Printed hands particularly struggled during the maintaining phase of the AHAP. CPCA hands had consistent results between grasping and maintaining.

3.3.3.3: Pairwise Grip Comparisons

Differences Among Specific Grip Patterns

Figure 13 presents the outcomes of the Friedman's Tests, highlighting statistically significant differences in the grip patterns within both the grasping and maintaining scores for the 3D-printed prosthetic hands, with P-values of 4.11e-05 and 1.14e-05, and test statistics of 35.91 and 37.03, respectively. For the CPCA prosthetic hands, the results demonstrated statistically significant differences within the grasping scores across different grips, as indicated by a P-value of 0.033 and a test statistic of 18.16. However, the maintaining scores of CPCA hands did not exhibit significant variations across grip types, evidenced by a P-value of 0.24 and a test statistic of 10.45.

Identifying Specific Difficult Grips for 3D-Printed Hands

The Nemenyi Test results highlighted significant disparities in the grasping scores of the 3D-printed hands, and can be found in Figures 14. Specifically, the Diagonal Volar grip demonstrated considerable differences compared to the Hook, Index Pointing/Pressing, Platform, and Tripod Pinch grips. Additionally, for maintaining scores, the Index Pointing/Pressing grip was significantly different from the Cylindrical Grip, Extension Grip, and Spherical Grip. For the other grip types evaluated within the 3D-printed hands, there were no statistically significant differences.

Identifying Specific Difficult Grips for CPCA Hands

Unfortunately, while Friedman's Test indicated statistically significant differences among the grasping scores for the CPCA hands, the subsequent Nemenyi Test P-values were statistically insignificant. Therefore, it could not specify which grip pairs exhibited these differences. However, inspection of Figure 13 suggests that the Extension and Platform grips underperformed compared to other grips. Finally, because Friedman's Test was unable to show differences among the maintaining scores of the CPCA hands, it excluded them from the Nemenyi Test analysis.

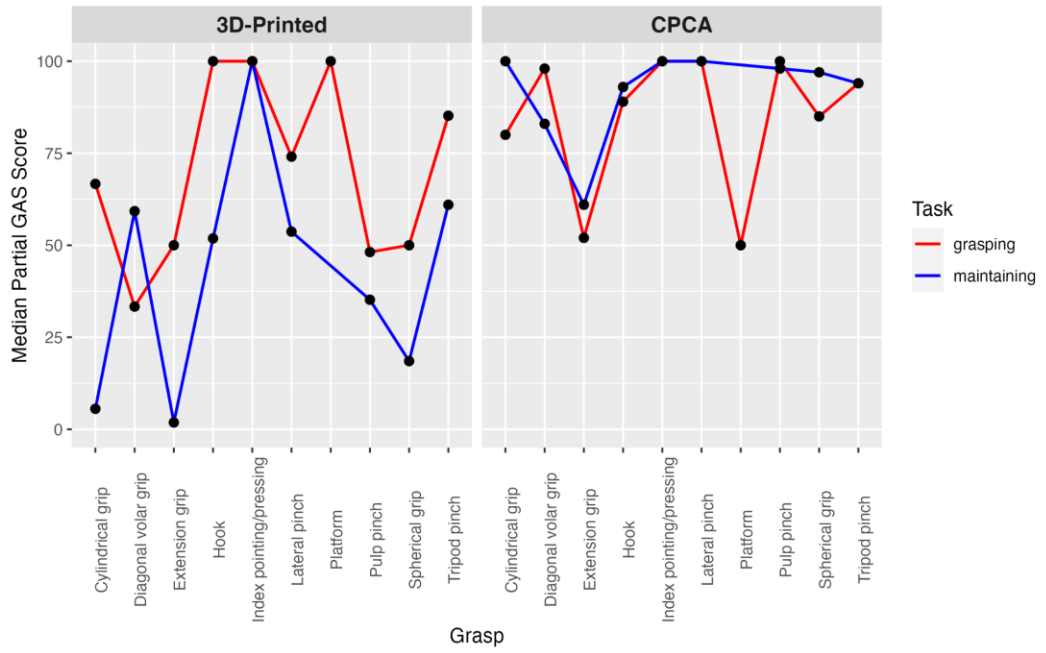


Figure 13. Grip-wise Comparison using Friedman's Test:
3D-Printed Hands struggled with certain grips during both the grasping and maintaining phase. CPCA hands only struggled with certain grips during the grasping phase.

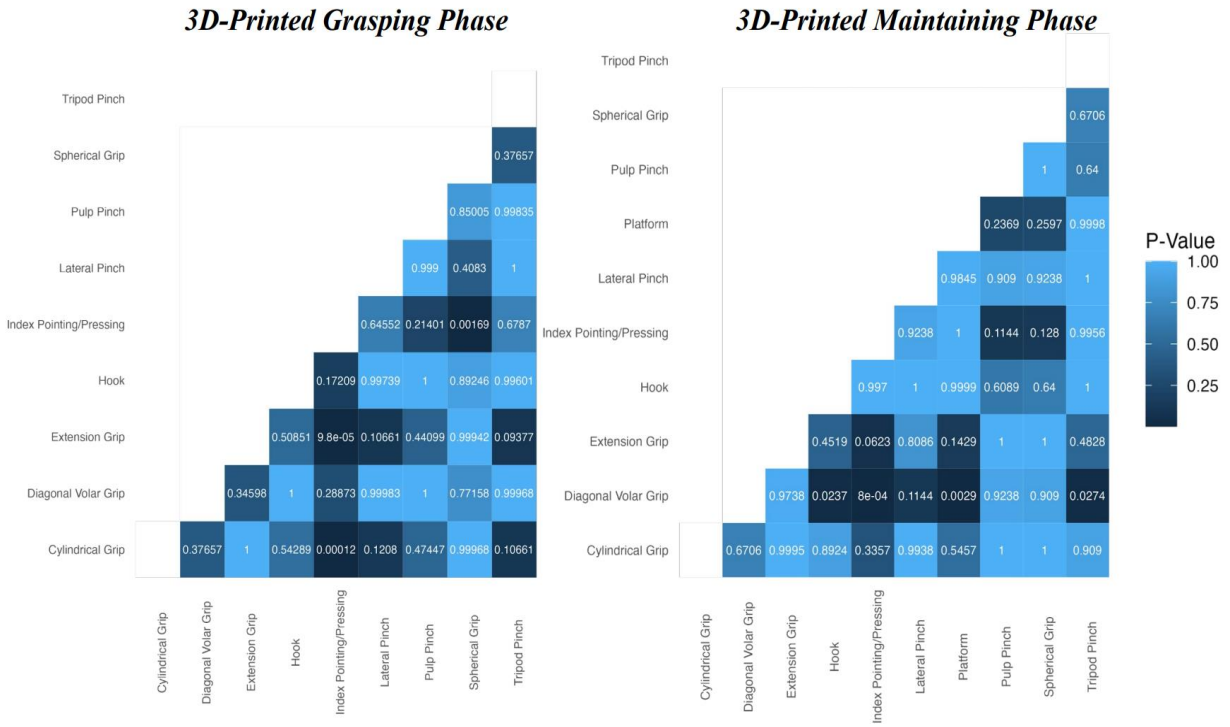


Figure 14. Specific Pairwise Comparison using the Nemenyi Test showed the 3D-printed hands struggled with Diagonal Volar grip during the grasping phase along with the Cylindrical Grip, Extension Grip, and Spherical Grip during the maintaining phase.

3.4: Discussion:

The analysis of CPCA and 3D-printed prosthetic hands, conducted using the AHAP, highlighted significant differences in their multi-grasp dexterity and performance. Specifically, it was observed that the 3D-printed prosthetic hands had lower scores compared to their CPCA alternatives. Furthermore, the performance gap observed during the maintaining phase is particularly significant because, during this phase, the AHAP only requires the hand to securely manipulate objects without imposing specific grip requirements.

To ensure the secure handling of an object, the sum of all contact forces and the resulting friction forces must balance to zero. Given this principle, the inability of the 3D-printed hands to succeed in the maintaining phase suggests potential issues such as insufficient friction, inadequate grip force, or geometric limitations preventing the effective reorientation of contact forces. This suggests that the factors contributing to the lower performance of the 3D-printed prosthetic hands are rooted in their design or construction. This is likely related to the primary challenge in the development of 3D-printed prostheses: balancing cost-effectiveness with the quality of materials and components. This often necessitates making strategic choices to maintain affordability but can compromise functionality. For instance, opting for cost-effective motors often compromises their achievable torque output, which directly impacts the prosthetic's ability to securely hold and manipulate objects, thereby contributing to lower maintaining scores. Additionally, it is essential to acknowledge that the level of dimensional precision achievable by a 3D printer may not always match that of mass-produced commercial prostheses. These dimensional differences can have implications for the fit and functionality of the prosthesis, underscoring the need for ongoing improvements in 3D printing technology. Durability and strength are also crucial, as commercial prostheses are designed for more rigorous use and are constructed from tested robust materials. It was observed that the 3D-printed hands were limited

in strength, being unable to support the weight of certain objects (skillet lid, wooden blocks with rope, and skillet). However, there are also clear opportunities for impactful enhancements through the cost-effective addition of certain features. For example, the observed slipping of certain objects (large marker, small marker, and golf ball) from the HACKberry hand was not due to a lack of grip strength but rather to insufficient friction. Consequently, adding rubberized grips or gloves to the fingers and palm could be a relatively simple yet effective method to significantly enhance friction and, by extension, the usability of the prosthesis. Another notable example is the difference in transmission mechanisms of the CPCA compared to the 3D-printed hands. The CPCA hands all typically have locking mechanisms with low backlash, so that when power is turned off the hands still hold their position, a useful feature for conserving battery as well as providing additional mechanical stability in the joint. None of the examined 3D-printed options currently employ these mechanisms and this may be another key feature that could be explored in future versions of 3D-printed hands to improve their maintaining scores.

Given that these findings demonstrate CPCA prostheses achieved significantly higher GAS scores than the 3D-printed hands, it suggests that these two types of devices may currently be best suited for distinct end uses. CPCA prostheses are evidently more practical for everyday use, given their ability to achieve a wider range of grip positions and superior performance in securely manipulating objects. Conversely, 3D-printed prosthetic devices, despite their limitations in grip capabilities, present a significantly more affordable option. With production costs as low as $19 USD [69], they stand in stark contrast to the often prohibitive expenses associated with CPCA devices, which often exceed $20,000 USD [67], [68]. Furthermore, the accessibility of 3D-printed prosthetic technology supports research to be conducted in labs of most funding levels, fostering a more inclusive research environment. This affordability is

complemented by the technology's capability for rapid prototyping, allowing researchers to quickly design, print, and test different prosthetic components, accelerating the pace of innovation and development. Open-source 3D-printed devices also benefit from non-proprietary firmware, allowing for more control over their devices. Additionally, researchers can tailor designs to fit various anatomical needs or specific user preferences, a crucial aspect in studying prosthetic functionality and comfort. This study underlines the need for ongoing research and development in prosthetic technology, aiming to bridge the gap between affordability and functionality. Nevertheless, the potential of 3D printing in this field is immense, and recognizing these limitations through this and subsequent research will contribute to refining the design and functionality of both 3D-printed and CPCA prosthetic devices.

While this study offers valuable insights into the performance of modern prosthetic hands, it does have limitations. The sample size for our study was limited, and the investigation concentrated on a specific subset of CPCA and 3D-printed prosthetic hands. For CPCA hands, this focus was primarily dictated by the challenges associated with acquiring these devices, often related to their high costs. Furthermore, there is a vast array of 3D-printed hands available, and the ones we tested do not encompass all existing models. The selection was based on a limited subset chosen for the availability of documentation and open-access print files. It would be beneficial for future research to include a broader range of prosthetic hand technologies, especially those utilizing emerging designs and materials, particularly highlighting the potential of innovative devices from companies such as Unlimited Tomorrow and Open Bionics, known for their CPCA, 3D-printed prostheses [86], [87]. Unfortunately, their cost, programming interfaces, and limited availability excluded them from this study.

Additionally, certain limitations in the AHAP must be acknowledged [21]. Specifically, the grasping phase criteria may under-represent actual dexterity, as the strict guidelines for a correct grip do not always align with practical, clinical scenarios where various secure grasping methods could be employed to manipulate an object. The AHAP also does not have a measure of the ease of attaining a prescribed grip, an essential component of daily use. Additionally, the AHAP does not distinguish between grasping pressures, allowing hands with superior grip strength to achieve a higher maintaining score by simply applying excessively high gripping forces, a result that could be undesirable during many activities of daily living.

The impressive dexterity demonstrated by both the CPCA and 3D-printed prostheses was evaluated without adequate consideration for control systems. As dexterity improves, there is a consequent need for more advanced control systems, which underscores the necessity for accurate and precise assessment methods. Technologies such as electromyography (EMG), force myography (FMG), electrical impedance tomography (EIT), electroencephalography (EEG), mechanomyography (MMG), sonomyography (SMG), and functional magnetic resonance imaging should be considered significant for body information monitoring [88], [89]. These, along with future technologies, are crucial for ensuring the effective integration and control of prosthetic limbs, necessitating their thorough evaluation.

Finally, the AHAP scores hand performance on an ordinal scale, which constrains the range of statistical inference methods that can be applied and necessitates the use of non-parametric methods. These methods, while appropriate, typically offer less statistical power compared to a parametric alternative. As a result, some grasps that may appear to differ greatly in their respective scores may not be considered significantly different by the Nemenyi Test. For example, during testing there appeared to be variances in the performance of the pulp pinch grip;

however, these differences could not be statistically confirmed as significant through the Nemenyi Test. This suggests a need for further investigations, encompassing the need for additional standardized assessments to accurately identify performance shortcomings and devise effective strategies for their improvement, as it could reveal specific design enhancements necessary to improve the functionality of 3D-printed and CPCA prosthetic hands.

Chapter 4: Online Survey to Refine Task-Based Evaluation Methods

A significant portion of this section will be presented/published at the Myoelectric Control Conference in August 2024.

4.1: Background

Chapters 1-3 established that assessments for upper-limb prosthetic devices are critically important while highlighting a significant research gap in this area. Despite the rapid advancements in mechatronic technologies for upper-limb prosthetics, the literature review in Chapter 2 revealed that, since 1948, only 25 task-based evaluation measures have been reported and validated [3]. This gap between technological progress and the slow evolution of standardized assessment measures underscores the urgent need for universally accepted and continuously updated evaluation frameworks. Chapter 3 specifically underlines that as prosthetic hands become more sophisticated, the necessity for improved and more accurate assessments intensifies. Furthermore, with dexterity capabilities continually increasing, there arises a need for more advanced control systems, necessitating precise and accurate assessment methods. Technologies such as EMG, FMG, EIT, EEG, MMG, SMG, and functional magnetic resonance imaging, among others, are considerable options for body information monitoring [88], [89]. These technologies, along with other emerging and future innovations, demand comprehensive and meticulous evaluation to verify their effectiveness and suitability for prosthetic control.

The challenge of bridging this gap is further compounded by the varying priorities across professional settings. Clinical settings, which are often under time constraints, may prefer more rapid tests for their efficiency in assessing patient outcomes. However, this may come at the expense of the depth of data collected. On the other hand, research settings may opt for more comprehensive tests which, despite their thoroughness, face challenges in wider spread adoption

due to their extensive setups, accessibility of testing materials, significant costs, and more complex protocols. The goal of this study was to gather insights on current task-based evaluation methods in the context of the unique needs and expectations across the diversity of practitioners that may interact with individuals prescribed upper limb prostheses. To do so, an online survey was conducted with a wide array of individuals across the professional spectrum, including physical and occupational therapists, prosthetists, medical practitioners, and academic researchers.

4.2: Methods

4.2.1: Survey Design

This online survey was strategically designed to gather data on professionals' experiences, preferences, and practices related to upper-limb prostheses and task-based functional measures. The study was approved by the University of California, Davis Institutional Review Board. Recruitment was performed through email via the Bionic Engineering and Assistive Robotics Laboratory's professional networks. Once participants agreed to take part in the study, they were provided with a link to an anonymous survey hosted on Qualtrics. This began with an introduction outlining the study's objectives, confidentiality assurances, detailed instructions, and contact information for any follow-up questions. Consent to proceed led participants through a questionnaire that required no more than 15 minutes to complete. The survey incorporated a variety of question types, including multiple-choice, checkboxes, and questions that allowed respondents to order their preferences. This design facilitated the easy and efficient capture of detailed responses across a range of topics. The questionnaire was structured to progress through a series of questions aimed at anonymously characterizing each participant's profession, experience, training, and exposure to individuals with upper-limb prostheses.

Following this initial characterization, the survey focused on identifying which validated task-based measures participants were aware of and actively used. Finally, participants were asked to prioritize a list of factors they deemed most important in a task-based measure for upper-limb prosthetic assessment.

4.2.2: Data Analysis

Binning and response counts were employed as the primary analytical methods. Data collected from the survey were first separated (binned) by profession, allowing for a detailed analysis of the perspectives of different professionals. Response counts were utilized to quantify the prevalence of specific views and practices among the participants, providing a straightforward method to identify the most used task-based measures and the factors considered most important for evaluating upper-limb prosthetic devices.

4.3: Results

In this thesis, initial validation data is presented from N=30 participants, whose professional backgrounds are outlined in Table 6. Further data will be gathered during the MyoElectric Control Conference in August 2024. The distribution of participants by profession was as follows: 5 physical/occupational therapists (PT/OTs), 4 certified prosthetist/orthotists (CPOs), 14 medical doctors (MD/DOs), and 7 who are primarily researchers (PRs). Additionally documented was the median duration of practice in their respective fields by having them select from a list of time ranges: PT/OTs and CPOs professionals had a median range of experience between 10 to 15 years; MD/DOs participants reported a median range of 12.5 to 17.5 years; and for those primarily involved in research (PRs), the median experience ranged from 15 to 20 years. The survey also required participants to select from a list highlighting the frequency range of interaction with upper limb prosthesis users. The median rate of patient interactions revealed a

spectrum of engagement frequencies: PT/OTs and PRs typically interacted with patients once every 2 to 5 months; CPOs reported at least one patient interaction per week; and MD/DOs professionals engage with patients at least once per month.

Table 6: Respondent Background

Profession	Respondents	Median Time Practicing	Median Patient Interaction Rate
Physical/Occupational therapist	5	10 – 15 years	Once every 2-5 months
Certified Prosthetist/Orthotist	4	10 – 15 years	At least once per week
Medical Doctor	14	12.5 – 17.5 years	At least once per month
Primarily Researcher	7	15 – 20 years	Once every 2-5 months

Table 7 highlights the results from a survey question that prompted participants to select task-based measures, from a list of 25 (identified in Chapter 2 [3]), that they were familiar with and would likely use with patients in their professional practice. The Box and Block Test (BBT) was identified as the most favored test across all professions for patient use. This finding is particularly significant considering the test's brevity and limited scope in assessing functional capabilities. Despite these constraints, the Box and Block Test is valued for its comprehensive validation with numerous patient populations, endorsement through peer review, straightforward administration, affordability, and ease of learning. Conversely, more involved evaluations such as the Southampton Hand Assessment Procedure (SHAP), Activities Measure for Upper-Limb Amputees (AM-ULA), and Gaze and Movement Assessment (GaMA) were primarily chosen for research purposes. It is important to note that a significant portion of the MD/DOs reported a lack of familiarity with many of the tests listed. Several doctors indicated in their responses that they would prefer to delegate the responsibility of administering these tests to PT/OTs.

Table 7: Perspectives on Currently Available Tests

Profession	Top Rated Tests to be used with a Patient – Percentage of Respondents * Indicates tie
Physical/ Occupational therapist	1: Box and Block Test (BBT) – 71.43%
	2*: Action Research Arm Test (ARAT) – 50.00% 2*: Jebsen Hand Function Test (JHFT) – 50.00% 2*: Nine-Hole Peg Test – 50.00%
Certified Prosthetist/ Orthotist	1: Box and Block Test (BBT) – 50.00%
	2*: Assessment of Capacity for Myoelectric Control (ACMC) – 33.33% 2*: University of New Brunswick Test of Prosthetic Function (UNBT) – 33.33%
Medical Doctor	1*: Box and Block Test (BBT) – 21.42% 1*: Jebsen Hand Function Test (JHFT) – 21.42% 1*: Nine-Hole Peg Test (NHPT) – 21.42%
	2*: Purdue Pegboard Test (PPT) – 14.29% 2*: Unilateral Below Elbow Test (UBET) – 14.29%
Primarily Researcher	1*: Southampton Hand Assessment Procedure (SHAP) – 44.44% 1*: Box and Block Test (BBT) – 44.44%
	2*: Activities Measure for Upper-Limb Amputees (AM-ULA) – 42.86% 2*: Gaze and Movement Assessment (GaMA) – 42.86%

Table 8 shows the results when participants selected from a list of maximum time ranges they felt was acceptable to administer a task-based measure in their practice. Additionally, Table 8 highlights the top three criteria they viewed as important when selecting a task-based measure, underscoring a universal preference for validated and peer-reviewed tools. Clinical practitioners reported a significantly shorter maximum testing time compared to their research-focused peers, highlighting a prioritization of efficiency in clinical settings. This emphasis on time efficiency is reflected in the ranking of the total administration time as a key factor for its selection among clinical professionals. Despite these differences, there's a unanimous agreement on the

importance of using tests that effectively monitor patient progress, illustrating a common objective to employ assessments that are both practical and beneficial for patient care across diverse professional landscapes.

Table 8: Desired Characteristics for Evaluation Methods

Profession	Median Max Time for Test	Ranking of Most Important Factors
Physical/ Occupational therapist	Between 10-20 minutes	1: The test has been validated and peer-reviewed
		2: Efficacy of monitoring patient progress
		3: Total administration time
Certified Prosthetist/ Orthotist	Between 15-25 minutes	1: The test has been validated and peer-reviewed
		2: Total administration time
		3: Comprehensive analysis of multi-grasp dexterity and impact of varying control systems
Medical Doctor	Between 5-10 minutes	1: The test has been validated and peer-reviewed
		2: Total administration time
		3: Efficacy of monitoring patient progress
Primarily Researcher	Between 30-60 minutes	1: The test has been validated and peer-reviewed
		2: Efficacy of monitoring patient progress
		3: Comprehensive analysis of multi-grasp dexterity and impact of varying control systems

4.4: Discussion

This study unveiled insightful findings regarding the prevailing views on task-based evaluation methods for upper-limb prostheses. Notable variations were observed in how frequently different professional groups engage with patients equipped with upper-limb

prosthetic devices. It is essential to mention that these interactions ranged from weeks to months, highlighting a considerable variance among professionals. However, a potential limitation of this study was the methodology used to contact respondents—email outreach within the Bionic Engineering and Assistive Robotics Laboratory's network of researchers and clinicians specializing in upper-limb absence. This approach might have led to an overestimation of interaction frequency, as it may not accurately represent the engagement levels of the average practitioner. Despite this limitation, the importance of addressing the prosthetic needs and managing patient expectations cannot be overstated, especially considering the challenges posed by the advancing technology in upper-limb prosthetics. These challenges are compounded by the mobility requirements of the upper-limb and the vital role that hands and arms play in our daily activities. These findings also shed light on the "upper extremity dilemma [90]," where prosthetics are becoming more technologically advanced and specialized. However, the relatively infrequent encounters with upper-limb prosthetic users make it difficult for many clinicians to expand their knowledge and expertise [91]. This gap necessitates a high level of specialized care for a group of patients seen less frequently by practitioners, leading to potential challenges in meeting their specific needs [92]. To bridge this gap, validated task-based measures and a more universally applicable analysis framework could play a crucial role. Such tools would provide practitioners with objective data, facilitating more informed decision making and ultimately enhancing care for patients using upper-limb prosthetic devices.

Categorizing the data from 30 participants by profession revealed distinct preferences in testing goals and methods. Although all groups emphasized the necessity of validated and peer-reviewed tests, notable differences emerged: clinical settings prioritize quick evaluations next, with the box and block test—likely favored for its sub-5-minute completion time—ranking high

among physical and occupational therapists, prosthetists/orthotists, and medical doctors. Notably, the maximum time reported for testing in these clinical groups was significantly shorter than that for research focused professions. Furthermore, professionals across these fields consistently rank the total time required to administer a test as one of the top three criteria for determining its effectiveness. In contrast, research environments valued the comprehensive analysis which likely explains the preference for the more intensive Southampton Hand Assessment Procedure, though the box and block test does remain in use for this group. Nevertheless, professionals unanimously agree on the importance of tests that effectively monitor patient progress. These findings highlight the shared and unique priorities across professions, underscoring the need for a balanced approach in developing and selecting upper-limb prosthetic evaluation methods to accommodate the quick assessment preferences of clinical practitioners and the detailed analytical needs of researchers.

Chapter 5: Discussion and Future Directions

5.1: Discussion

This thesis underscores the necessity of standardized, reliable, and validated task-based evaluation measures for the advancement of next-generation prosthetics and improved patient outcomes. It highlights a significant gap between the rapid technological advancements in upper-limb prosthetics and the slow evolution of assessment measures. Despite the rapid progress in prosthetic technology, only 25 task-based evaluation measures have been validated since 1948, as detailed in Chapter 2. A critical examination reveals a dependence of many researchers on non-standardized tests, which may compromise validity and fail to meet the comprehensive needs of the clinical and research communities.

Chapter 3 of the thesis illustrates the lag in assessment measures compared to the rapid advancements in prosthetic technology. It details applying the Anthropomorphic Hand Assessment Protocol (AHAP) to evaluate both open-source 3D-printed and commercially produced prosthetic hands, effectively bridging the gap between theoretical concepts and practical applications in prosthetic assessment. This analysis revealed that as prosthetics become more sophisticated, there is a pressing need for more refined and accurate assessment methods that can adequately analyze the advancements in control systems and the enhanced dexterity of multi-grasp functions.

Chapter 4 introduces preliminary findings from an online survey of diverse practitioners, offering insights into the specific needs for upper-limb prostheses and guiding future enhancements in evaluation methods. This chapter illuminates the current state of evaluation practices and pinpoints improvement areas. It makes a significant contribution to the

development of universally accepted evaluation standards for upper-limb prostheses, reflecting a comprehensive approach to understanding and advancing prosthetic assessment.

5.2: Future Directions

The primary challenge in current testing methodologies for upper-limb prosthetics is the lack of standardization, compounded by the varying testing requirements of professionals. Rather than creating adaptable tests that can be refined or expanded, many research groups produce new, study-specific tests that lack universal applicability. This has led to the development of tests that may be limited in the information they can provide, like the NHPT or BBT, failing to encapsulate the complexity of daily tasks, or data rich, but at the expense of complexity of implementation, such as the SHAP or AM-ULA. To overcome these issues, a dual or modular testing approach is suggested. In clinical settings, a straightforward, user-friendly test should be used to prioritize patient progress, multi-grasp dexterity, and the impact of different control systems. In contrast, research settings require a more comprehensive test to thoroughly evaluate a prosthesis's functionality. Both testing approaches must share key elements to allow for the alignment and comparison of clinical and research data, which is vital for fostering communication between clinicians and researchers and ensuring access to consistent, relevant, and comprehensive information. Furthermore, the standardized and modular test proposed must fulfill the criteria set out in Chapter 2's evaluation guidelines and the needs identified in Chapter 4. Implementing standardized, validated task-based evaluation measures is essential and urgently needed to drive progress and underpin advancements in upper-limb prosthetics.

References

[1] B. C. K. Choi and A. W. P. Pak, "A Catalog of Biases in Questionnaires," *Prev. Chronic. Dis.*, vol. 2, no. 1, p. A13, Dec. 2004.

[2] I. Vujaklija, D. Farina, and O. C. Aszmann, "New developments in prosthetic arm systems," *Orthop. Res. Rev.*, vol. 8, pp. 31–39, Jul. 2016, doi: 10.2147/ORR.S71468.

[3] J. R. Siegel, M. A. Battraw, E. J. Winslow, M. A. James, W. M. Joiner, and J. S. Schofield, "Review and critique of current testing protocols for upper-limb prostheses: a call for standardization amidst rapid technological advancements," *Front. Robot. AI*, vol. 10, 2023, Accessed: Nov. 14, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frobt.2023.1292632

[4] "Assessment of Hand Function: The Relationship between Pegboard Dexterity and Applied Dexterity." Accessed: Aug. 15, 2023. [Online]. Available: https://journals.sagepub.com/doi/epdf/10.1177/000841749205900406

[5] K. E. Yancosek and D. Howell, "A Narrative Review of Dexterity Assessments," *J. Hand Ther.*, vol. 22, no. 3, pp. 258–270, Jul. 2009, doi: 10.1016/j.jht.2008.11.004.

[6] L. Resnik, M. Borgia, B. Silver, and J. Cancio, "Systematic Review of Measures of Impairment and Activity Limitation for Persons With Upper Limb Trauma and Amputation," *Arch. Phys. Med. Rehabil.*, vol. 98, no. 9, pp. 1863-1892.e14, Sep. 2017, doi: 10.1016/j.apmr.2017.01.015.

[7] S. Wang *et al.*, "Evaluation of Performance-Based Outcome Measures for the Upper Limb: A Comprehensive Narrative Review," *PM R*, vol. 10, no. 9, pp. 951-962.e3, Sep. 2018, doi: 10.1016/j.pmrj.2018.02.008.

[8] D. Cornell (PhD) and C. Drew (PhD), "9 Types of Validity in Research (2023)." Accessed: Aug. 15, 2023. [Online]. Available: https://helpfulprofessor.com/types-of-validity/

[9] "4.2 Reliability and Validity of Measurement – Research Methods in Psychology." Accessed: Aug. 15, 2023. [Online]. Available: https://opentext.wsu.edu/carriecuttler/chapter/reliability-and-validity-of-measurement/

[10] M. Markovic *et al.*, "The clinical relevance of advanced artificial feedback in the control of a multi-functional myoelectric prosthesis," *J. NeuroEngineering Rehabil.*, vol. 15, p. 28, Mar. 2018, doi: 10.1186/s12984-018-0371-1.

[11] D. T. Beckler, Z. C. Thumser, J. S. Schofield, and P. D. Marasco, "Using sensory discrimination in a foraging-style task to evaluate human upper-limb sensorimotor performance," *Sci. Rep.*, vol. 9, p. 5806, Apr. 2019, doi: 10.1038/s41598-019-42086-0.

[12] H. E. Williams, C. S. Chapman, P. M. Pilarski, A. H. Vette, and J. S. Hebert, "Gaze and Movement Assessment (GaMA): Inter-site validation of a visuomotor upper limb functional protocol," *PLoS ONE*, vol. 14, no. 12, p. e0219333, Dec. 2019, doi: 10.1371/journal.pone.0219333.

[13] M. A. Battraw, P. R. Young, M. E. Welner, W. M. Joiner, and J. S. Schofield, "Characterizing Pediatric Hand Grasps During Activities of Daily Living to Inform Robotic Rehabilitation and Assistive Technologies," in *2022 International Conference on Rehabilitation Robotics (ICORR)*, Jul. 2022, pp. 1–6. doi: 10.1109/ICORR55369.2022.9896512.

[14] K. Y. Cheng, M. Rehani, and J. S. Hebert, "A scoping review of eye tracking metrics used to assess visuomotor behaviours of upper limb prosthesis users," *J. NeuroEngineering Rehabil.*, vol. 20, p. 49, Apr. 2023, doi: 10.1186/s12984-023-01180-1.

[15] M. Vergara, J. L. Sancho-Bru, V. Gracia-Ibáñez, and A. Pérez-González, "An introductory

study of common grasps used by adults during performance of activities of daily living," *J. Hand Ther.*, vol. 27, no. 3, pp. 225–234, Jul. 2014, doi: 10.1016/j.jht.2014.04.002.
[16] G. Bravo and L. Potvin, "Estimating the reliability of continuous measures with Cronbach's alpha or the intraclass correlation coefficient: toward the integration of two traditions," *J. Clin. Epidemiol.*, vol. 44, no. 4–5, pp. 381–390, 1991, doi: 10.1016/0895-4356(91)90076-l.
[17] H. Lindner and L. Hermansson, "Assessment of Capacity For Myoelectric Control - Evluation of Task Difficulty, Newly Merged Items and Redefined Rating Scale," 2011. Accessed: Aug. 15, 2023. [Online]. Available: https://www.semanticscholar.org/paper/ASSESSMENT-OF-CAPACITY-FOR-MYOELECTRIC-CONTROL-OF-Lindner-Hermansson/ffb4aa1de734ab3fe1b2451ca8d7267469ad8ba4
[18] H. Y. N. Lindner and J. M. Linacre, "Assessment of Capacity For Myoelectric Control: Construct Validity and Rating Scale Structure," 2008.
[19] H. Y. N. Lindner, A. Langius-Eklöf, and L. M. N. Hermansson, "Test-retest reliability and rater agreements of assessment of capacity for myoelectric control version 2.0," *J. Rehabil. Res. Dev.*, vol. 51, no. 4, pp. 635–644, 2014, doi: 10.1682/JRRD.2013.09.0197.
[20] H. Burger, D. Brezovar, and G. Vidmar, "A comparison of the University of New Brunswick Test of Prosthetic Function and the Assessment of Capacity for Myoelectric Control," *Eur. J. Phys. Rehabil. Med.*, vol. 50, Jan. 2014.
[21] I. Llop-Harillo, A. Pérez-González, J. Starke, and T. Asfour, "The Anthropomorphic Hand Assessment Protocol (AHAP)," *Robot. Auton. Syst.*, vol. 121, p. 103259, Nov. 2019, doi: 10.1016/j.robot.2019.103259.
[22] L. Resnik *et al.*, "Development and Evaluation of the Activities Measure for Upper Limb Amputees," *Arch. Phys. Med. Rehabil.*, vol. 94, no. 3, pp. 488-494.e4, Mar. 2013, doi: 10.1016/j.apmr.2012.10.004.
[23] N. Elangovan, C.-M. Chang, G. Gao, and M. Liarokapis, "An Accessible, Open-Source Dexterity Test: Evaluating the Grasping and Dexterous Manipulation Capabilities of Humans and Robots," *Front. Robot. AI*, vol. 9, p. 808154, Apr. 2022, doi: 10.3389/frobt.2022.808154.
[24] "Action Research Arm Test," Shirley Ryan AbilityLab. Accessed: Aug. 15, 2023. [Online]. Available: https://www.sralab.org/rehabilitation-measures/action-research-arm-test
[25] "Action Research Arm Test (ARAT)," Physiopedia. Accessed: Aug. 15, 2023. [Online]. Available: https://www.physio-pedia.com/Action_Research_Arm_Test_(ARAT)
[26] L. Resnik, M. Borgia, and F. Acluche, "Brief activity performance measure for upper limb amputees: BAM-ULA," *Prosthet. Orthot. Int.*, vol. 42, no. 1, pp. 75–83, Feb. 2018, doi: 10.1177/0309364616684196.
[27] "Box and Block Test," Shirley Ryan AbilityLab. Accessed: Aug. 15, 2023. [Online]. Available: https://www.sralab.org/rehabilitation-measures/box-and-block-test
[28] N. T. Kearns, J. K. Peterson, L. Smurr Walters, W. T. Jackson, J. M. Miguelez, and T. Ryan, "Development and Psychometric Validation of Capacity Assessment of Prosthetic Performance for the Upper Limb (CAPPFUL)," *Arch. Phys. Med. Rehabil.*, vol. 99, no. 9, pp. 1789–1797, Sep. 2018, doi: 10.1016/j.apmr.2018.04.021.
[29] A. Dynamics, "The Comprehensive Assessment of Prosthetic Performance for Upper Limb | Arm Dynamics." Accessed: Aug. 24, 2023. [Online]. Available: https://www.armdynamics.com/the-comprehensive-assessment-of-prosthetic-performance-for-upper-limb

[30] D. Carroll, “A quantitative test of upper extremity function,” *J. Chronic Dis.*, vol. 18, no. 5, pp. 479–491, May 1965, doi: 10.1016/0021-9681(65)90030-5.
[31] H. Lu, K. Xu, and S. Qiu, “Study on reliability of Carroll upper extremities functional test in children with spastic hemiplegia,” *Chin. J. Rehabil. Med.*, vol. 26, pp. 822–825, Sep. 2011, doi: 10.3969/j.issn.1001-1242.2011.09.007.
[32] R. Coulson, C. Li, C. Majidi, and N. S. Pollard, “The Elliott and Connolly Benchmark: A Test for Evaluating the In-Hand Dexterity of Robot Hands,” in *2020 IEEE-RAS 20th International Conference on Humanoid Robots (Humanoids)*, Munich, Germany: IEEE, Jul. 2021, pp. 238–245. doi: 10.1109/HUMANOIDS47582.2021.9555798.
[33] “Functional Dexterity Test,” Shirley Ryan AbilityLab. Accessed: Aug. 15, 2023. [Online]. Available: https://www.sralab.org/rehabilitation-measures/functional-dexterity-test
[34] P. D. Marasco *et al.*, “Neurorobotic fusion of prosthetic touch, kinesthesia, and movement in bionic upper limbs promotes intrinsic brain behaviors,” *Sci. Robot.*, vol. 6, no. 58, p. eabf3368, Sep. 2021, doi: 10.1126/scirobotics.abf3368.
[35] “Gaze and Movement Assessment (GaMA),” BLINC Lab. Accessed: Aug. 18, 2023. [Online]. Available: https://blinclab.ca/research/gama/
[36] F. Panuccio *et al.*, “Internal consistency and validity of the Italian version of the Jebsen–Taylor hand function test (JTHFT-IT) in people with tetraplegia,” *Spinal Cord*, vol. 59, no. 3, Art. no. 3, Mar. 2021, doi: 10.1038/s41393-020-00602-4.
[37] İ. C. Sığırtmaç and Ç. Öksüz, “Investigation of reliability, validity, and cutoff value of the Jebsen-Taylor Hand Function Test,” *J. Hand Ther. Off. J. Am. Soc. Hand Ther.*, vol. 34, no. 3, pp. 396–403, 2021, doi: 10.1016/j.jht.2020.01.004.
[38] “Jebsen Hand Function Test,” Shirley Ryan AbilityLab. Accessed: Aug. 15, 2023. [Online]. Available: https://www.sralab.org/rehabilitation-measures/jebsen-hand-function-test
[39] J. Desrosiers, A. Rochette, R. Hébert, and G. Bravo, “The Minnesota Manual Dexterity Test: Reliability, Validity and Reference Values Studies with Healthy Elderly People,” *Can. J. Occup. Ther.*, vol. 64, no. 5, pp. 270–276, Dec. 1997, doi: 10.1177/000841749706400504.
[40] “Nine-Hole Peg Test,” Shirley Ryan AbilityLab. Accessed: Aug. 15, 2023. [Online]. Available: https://www.sralab.org/rehabilitation-measures/nine-hole-peg-test
[41] C. Hunt, R. Yerrabelli, C. Clancy, L. Osborn, R. Kaliki, and N. Thakor, “PHAM: Prosthetic hand assessment measure,” Aug. 2017.
[42] D. K. Lindstrom-Hazel and N. VanderVlies Veenstra, “Examining the Purdue Pegboard Test for Occupational Therapy Practice,” *Open J. Occup. Ther.*, vol. 3, no. 3, Jul. 2015, doi: 10.15453/2168-6408.1178.
[43] J. Tiffin and E. J. Asher, “The Purdue Pegboard: norms and studies of reliability and validity,” *J. Appl. Psychol.*, vol. 32, no. 3, pp. 234–247, 1948, doi: 10.1037/h0061266.
[44] “Purdue Pegboard Test,” Shirley Ryan AbilityLab. Accessed: Aug. 15, 2023. [Online]. Available: https://www.sralab.org/rehabilitation-measures/purdue-pegboard-test
[45] A. Hussaini and P. Kyberd, “Refined clothespin relocation test and assessment of motion,” *Prosthet. Orthot. Int.*, vol. 41, no. 3, pp. 294–302, Jun. 2017, doi: 10.1177/0309364616660250.
[46] A. Hussaini, W. Hill, and P. Kyberd, “Clinical evaluation of the refined clothespin relocation test: A pilot study,” *Prosthet. Orthot. Int.*, vol. 43, no. 5, pp. 485–491, Oct. 2019, doi: 10.1177/0309364619843779.
[47] “SHAP: Southampton Hand Assessment Procedure.” Accessed: Aug. 15, 2023. [Online].

Available: http://www.shap.ecs.soton.ac.uk/
[48] J. Adams, K. Hodges, J. Kujawa, and C. Metcalf, “Test-retest Reliability of the Southampton Hand Assessment Procedure,” *Int. J. Rehabil. Res.*, vol. 32, p. S18, Aug. 2009.
[49] C. Sollerman and A. Ejeskär, “Sollerman Hand Function Test: A Standardised Method and its Use in Tetraplegic Patients,” *Scand. J. Plast. Reconstr. Surg. Hand Surg.*, vol. 29, no. 2, pp. 167–176, Jan. 1995, doi: 10.3109/02844319509034334.
[50] “Sollerman Hand Function Test,” Shirley Ryan AbilityLab. Accessed: Aug. 15, 2023. [Online]. Available: https://www.sralab.org/rehabilitation-measures/sollerman-hand-function-test
[51] E. Ekstrand, J. Lexell, and C. Brogårdh, “Test−Retest Reliability and Convergent Validity of Three Manual Dexterity Measures in Persons With Chronic Stroke,” *PM&R*, vol. 8, no. 10, pp. 935–943, Oct. 2016, doi: 10.1016/j.pmrj.2016.02.014.
[52] W. Van Lankveld, P. van’t Pad Bosch, J. Bakker, S. Terwindt, M. Franssen, and P. van Kiel, “Sequential Occupational Dexterity Assessment (SODA): A new test to measure hand disability,” *J. Hand Ther.*, vol. 9, no. 1, pp. 27–32, Jan. 1996, doi: 10.1016/S0894-1130(96)80008-1.
[53] W. Van Lankveld, P. Van’T Pad Bosch, J. Bakker, S. Terwindt, M. Franssen, and P. Van Kiel, “Sequential Occupational Dexterity Assessment (SODA),” *J. Hand Ther.*, vol. 9, no. 1, pp. 27–32, Jan. 1996, doi: 10.1016/S0894-1130(96)80008-1.
[54] J. Liu, W. Tang, G. Chen, Y. Lu, C. Feng, and X. M. Tu, “Correlation and agreement: overview and clarification of competing concepts and measures,” *Shanghai Arch. Psychiatry*, vol. 28, no. 2, pp. 115–120, doi: 10.11919/j.issn.1002-0829.216045.
[55] L. Resnik, M. Borgia, and F. Acluche, “Timed activity performance in persons with upper limb amputation: A preliminary study,” *J. Hand Ther.*, vol. 30, no. 4, pp. 468–476, Oct. 2017, doi: 10.1016/j.jht.2017.03.008.
[56] A. M. Bagley, F. Molitor, L. V. Wagner, W. TomhaveO, and M. A. James, “The Unilateral Below Elbow Test: a function test for children with unilateral congenital below elbow deficiency,” *Dev. Med. Child Neurol.*, vol. 48, no. 7, pp. 569–575, 2006, doi: 10.1111/j.1469-8749.2006.tb01317.x.
[57] L. Resnik, K. Baxter, M. Borgia, and K. Mathewson, “Is the UNB test reliable and valid for use with adults with upper limb amputation?,” *J. Hand Ther.*, vol. 26, no. 4, pp. 353–359, Oct. 2013, doi: 10.1016/j.jht.2013.06.004.
[58] “Wolf Motor Function Test,” Shirley Ryan AbilityLab. Accessed: Aug. 15, 2023. [Online]. Available: https://www.sralab.org/rehabilitation-measures/wolf-motor-function-test
[59] F. Virginia Wright, S. Hubbard, J. Jutai, and S. Naumann, “The prosthetic upper extremity functional index: Development and reliability testing of a new functional status questionnaire for children who use upper extremity prostheses,” *J. Hand Ther.*, vol. 14, no. 2, pp. 91–104, Apr. 2001, doi: 10.1016/S0894-1130(01)80039-9.
[60] F. V. Wright, S. Hubbard, S. Naumann, and J. Jutai, “Evaluation of the validity of the prosthetic upper extremity functional index for children,” *Arch. Phys. Med. Rehabil.*, vol. 84, no. 4, pp. 518–527, Apr. 2003, doi: 10.1053/apmr.2003.50127.
[61] “Comparison of clinical test and questionnaires for the evaluation of upper limb prosthetic use in children.” Accessed: Aug. 15, 2023. [Online]. Available: https://www.tandfonline.com/doi/epdf/10.1080/09638280410001708931?needAccess=true
[62] L. C. Smail, C. Neal, C. Wilkins, and T. L. Packham, “Comfort and function remain key

factors in upper limb prosthetic abandonment: findings of a scoping review," *Disabil. Rehabil. Assist. Technol.*, vol. 16, no. 8, pp. 821–830, Nov. 2021, doi: 10.1080/17483107.2020.1738567.
[63] S. Salminger *et al.*, "Current rates of prosthetic usage in upper-limb amputees – have innovations had an impact on device acceptance?," *Disabil. Rehabil.*, vol. 44, no. 14, pp. 3708–3713, Jul. 2022, doi: 10.1080/09638288.2020.1866684.
[64] T. W. Armstrong, M. L. C. Williamson, T. R. Elliott, W. T. Jackson, N. T. Kearns, and T. Ryan, "Psychological distress among persons with upper extremity limb loss," *Br. J. Health Psychol.*, vol. 24, no. 4, pp. 746–763, Nov. 2019, doi: 10.1111/bjhp.12360.
[65] A. C. Roșca, C. C. Baciu, V. Burtăverde, and A. Mateizer, "Psychological Consequences in Patients With Amputation of a Limb. An Interpretative-Phenomenological Analysis," *Front. Psychol.*, vol. 12, p. 537493, May 2021, doi: 10.3389/fpsyg.2021.537493.
[66] K. Wendo *et al.*, "Open-Source 3D Printing in the Prosthetic Field—The Case of Upper Limb Prostheses: A Review," *Machines*, vol. 10, no. 6, Art. no. 6, Jun. 2022, doi: 10.3390/machines10060413.
[67] W. Williams, "Bionic Hand Price List," Bionics For Everyone. Accessed: Aug. 23, 2023. [Online]. Available: https://bionicsforeveryone.com/bionic-hand-price-list/
[68] "Getting to grips with bionic costs," Eureka. Accessed: Jan. 09, 2024. [Online]. Available: https://www.eurekamagazine.co.uk/content/technology/getting-to-grips-with-bionic-costs/
[69] J.-J. Cabibihan *et al.*, "Suitability of the Openly Accessible 3D Printed Prosthetic Hands for War-Wounded Children," *Front. Robot. AI*, vol. 7, 2021, Accessed: Jan. 08, 2024. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frobt.2020.594196
[70] I. Llop-Harillo, A. Pérez-González, and J. Andrés-Esperanza, "Grasping Ability and Motion Synergies in Affordable Tendon-Driven Prosthetic Hands Controlled by Able-Bodied Subjects," *Front. Neurorobotics*, vol. 14, p. 57, Aug. 2020, doi: 10.3389/fnbot.2020.00057.
[71] J. T. Belter, J. L. Segil, A. M. Dollar, and R. F. Weir, "Mechanical design and performance specifications of anthropomorphic prosthetic hands: A review," *J. Rehabil. Res. Dev.*, vol. 50, no. 5, p. 599, 2013, doi: 10.1682/JRRD.2011.10.0188.
[72] A. Kannenberg, R. Lundstrom, K. D. Hibler, and S. S. Johnson, "Differences in Two Multiarticulating Myoelectric Hands for Facilitating Activities of Daily Living in Individuals with Transradial Amputation: A Cross-Sectional Study," *JPO J. Prosthet. Orthot.*, vol. 35, no. 1, p. 38, Jan. 2023, doi: 10.1097/JPO.0000000000000411.
[73] "ma00001a_i-limb_ultra_datasheet.pdf." Accessed: Feb. 27, 2024. [Online]. Available: https://training.touchbionics.com/pdfs/ma00001a_i-limb_ultra_datasheet.pdf
[74] "i-Limb® Quantum Bionic Hand. Ossur.com." Accessed: Jan. 05, 2024. [Online]. Available: https://www.ossur.com/en-us/prosthetics/arms/i-limb-quantum
[75] "i-Limb_Quantum_Catalog_en-us_Q4-22.pdf." Accessed: Feb. 27, 2024. [Online]. Available: https://media.ossur.com/ossur-dam/image/upload/product-documents-global/i-Limb_Quantum_Catalog_en-us_Q4-22.pdf
[76] "bebionic Hand EQD | The most lifelike prosthetic hand." Accessed: Jan. 05, 2024. [Online]. Available: https://www.ottobock.com/en-us/product/8E70
[77] "14112_bebionic_user_guide_lo.pdf." Accessed: Jan. 31, 2024. [Online]. Available: https://www.ottobockus.com/media/local-media/prosthetics/upper-limb/files/14112_bebionic_user_guide_lo.pdf
[78] "127579-1-en_master-01-1704w_2508822.pdf." Accessed: Jan. 05, 2024. [Online].

Available: https://media.ottobock.com/_web-site/prosthetics/upper-limb/bebionic-hand/files/127579-1-en_master-01-1704w_2508822.pdf
[79] "Ability Hand™," PSYONIC. Accessed: Jan. 12, 2024. [Online]. Available: https://www.psyonic.io/ability-hand
[80] "PSYONIC+User+Manual+2.24+1.9.2024.pdf." Accessed: Feb. 27, 2024. [Online]. Available: https://static1.squarespace.com/static/61229f1b98e6c829c570bace/t/659ec6d3caa57e12edfa3b46/1704904420631/PSYONIC+User+Manual+2.24+1.9.2024.pdf
[81] "HACKberry Hand," HACKberry. Accessed: Jan. 12, 2024. [Online]. Available: https://www.exiii-hackberry.com
[82] D. J. A. Brenneis, M. R. Dawson, and P. Pilarski, "DEVELOPMENT OF THE HANDI HAND: AN INEXPENSIVE, MULTI-ARTICULATING, SENSORIZED HAND FOR MACHINE LEARNING RESEARCH IN MYOELECTRIC CONTROL," 2017. Accessed: Jan. 31, 2024. [Online].
[83] M. Dawson, C. Sherstan, J. Carey, J. Hebert, and P. Pilarski, "Development of the Bento Arm: An Improved Robotic Arm for Myoelectric Training and Research," Aug. 2014. doi: 10.13140/2.1.3118.4640.
[84] M. A. Battraw, P. R. Young, W. M. Joiner, and J. S. Schofield, "A multiarticulate pediatric prosthetic hand for clinical and research applications," *Front. Robot. AI*, vol. 9, 2022, Accessed: Jan. 31, 2024. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frobt.2022.1000159
[85] I. Llop-Harillo and A. Pérez-González, "System for the experimental evaluation of anthropomorphic hands. Application to a new 3D-printed prosthetic hand prototype," *Int. Biomech.*, vol. 4, no. 2, pp. 50–59, Nov. 2017, doi: 10.1080/23335432.2017.1364666.
[86] "TrueLimb® Prosthetic Arm Manufacturer," Unlimited Tomorrow. Accessed: Jan. 09, 2024. [Online]. Available: https://www.unlimitedtomorrow.com/
[87] "Open Bionics - Turning Disabilities into Superpowers." Accessed: Jan. 09, 2024. [Online]. Available: https://openbionics.com/
[88] Z. Zheng, Z. Wu, R. Zhao, Y. Ni, X. Jing, and S. Gao, "A Review of EMG-, FMG-, and EIT-Based Biosensors and Relevant Human–Machine Interactivities and Biomedical Applications," *Biosensors*, vol. 12, no. 7, p. 516, Jul. 2022, doi: 10.3390/bios12070516.
[89] M. Connan, E. Ruiz Ramírez, B. Vodermayer, and C. Castellini, "Assessment of a Wearable Force- and Electromyography Device and Comparison of the Related Signals for Myocontrol," *Front. Neurorobotics*, vol. 10, Nov. 2016, doi: 10.3389/fnbot.2016.00017.
[90] C. Lake and R. Dodson, "Progressive upper limb prosthetics," *Phys. Med. Rehabil. Clin. N. Am.*, vol. 17, no. 1, pp. 49–72, Feb. 2006, doi: 10.1016/j.pmr.2005.10.004.
[91] C. Brenner, "Wrist disarticulation and transradial amputation: prosthetic management.," in *Atlas of amputations and limb deficienciesd surgical, prosthetic and rehabilitation principles*, 3rd ed., Rosemont: American Academy of Orthopaedic Surgeons, 2004, pp. 223–230.
[92] J. S. Schofield, K. R. Schoepp, H. E. Williams, J. P. Carey, P. D. Marasco, and J. S. Hebert, "Characterization of interfacial socket pressure in transhumeral prostheses: A case series," *PLOS ONE*, vol. 12, no. 6, p. e0178517, Jun. 2017, doi: 10.1371/journal.pone.0178517.

Appendix

Dexterity and Control System Scoring

Test	Holistic assessment	Object size	Object compliance	In-hand manipulation	Tool usage
ACMC	DM	SML	MC	✔	✔
AHAP	DM	SML	MC	✔	✖
AM-ULA	DCM	SML	MC	✔	✔
AOSDT	DM	SM	SC	✔	✔
ARAT	DM	SM	CO	✔	✖
BAM-ULA	DM	SML	SC	✔	✔
BBT	DO	S	CO	✖	✖
CAPPFUL	DCM	SML	SC	✔	✔
CQT	DCM	SM	CO	✔	✔
ECB	DM	SML	SC	✔	✖
FDT	DO	S	CO	✔	✖
GaMA	DCM	M	SC	✖	✖
JHFT	DM	SM	SC	✔	✔
MMDT	DO	S	CO	✔	✖
NHPT	DO	S	CO	✔	✖
PHAM	DCM	SM	CO	✔	✖
PPT	DO	S	CO	✔	✖
RCRT	DO	S	CO	✔	✖
SHAP	DCM	SM	SC	✔	✔
SHFT	DM	SML	SC	✔	✔
SODA	DM	SM	SC	✔	✖
T-MAP	DCM	SML	MC	✔	✔
UBET	DCM	SML	SC	✔	✔
UNBT	DCM	SML	SC-MC****	✔	✔
WMFT	DM	SM	SC	✔	✖

Key: DO, dexterity only; DM, dexterity only but uses multiple hand grasps; DC, dexterity and control system; DCM, dexterity and control system and uses multiple hand grasps; MC, multiple custom objects; MH, multiple household objects; SO, single object, S = Small* Objects Only, M = medium objects only, SM, Small and Medium** Objects, SML, small, Medium, and Large*** Objects, MC, Multiple Compliances (≥ 5), SC, Some Compliances (5>SC ≥ 2), CO, single compliance only.

[a]A small object can be moveable using only fingers on an average size able-body participant.

[b]A medium object can be held mostly within the hand of an average size able-body participant.

[c]A large object will have an external component outside of the hand (such as a frying pan) or requires both hands (such as a tire).

[d]Depends on the subset used.

Additional Considerations Scoring

Test	Monitoring progress	Patient feedback	Evaluator Expertise	Efficient administration	Accessibility considerations	Year
ACMC	✔	✖	NQ	>30 min	~$1,000, P	2008
AHAP	✖	✖	NQ	>30 min	~$200, P	2019
AM-ULA	✔	✖	CHT	>30 min	~$200, P	2012
AOSDT	✖	✖	NQ	>30 min	~$250, C	2022
ARAT	✔	✖	CHT	<30 min	~$650, P	1981
BAM-ULA	✔	✖	NQ	<15 min	~$100, P	2017
BBT	✔	✖	NQ	<5 min	~$200, P	1985
CAPPFUL	✔	✔	OT	>25 min	~$300, P	2018
CQT	✔	✖	NQ	>15 min	~$200, P	1965
ECB	✖	✖	NQ	>30 min	~$200, P	2020
FDT	✔	✖	NQ	<5 min	~$90, P	2003
GaMA	✔	✖	R	>1 h	>$2000, P	2018
JHFT	✔	✖	NQ	<25 min	~$400, P	1969
MMDT	✔	✖	NQ	<10 min	~$350, P	1997
NHPT	✔	✖	NQ	<5 min	~$100, P	1985
PHAM	✔	✖	R	>1 h	>$500, P/C	2017
PPT	✔	✖	NQ	<5 min	~$150, P	1948
RCRT	✔	✖	NQ	<5 min	~120, P	2016
SHAP	✔	✖	NQ	>45 min	>$2,500, P	2002
SHFT	✔	✖	NQ	<25 min	~$100, P	1995
SODA	✔	✖	NQ	>30 min	~$100, P	1996
T-MAP	✔	✖	NQ	>10 min	~$50, P	2017
UBET	✔	✖	OT	>30 min	~$400, P	2006
UNBT	✔	✖	NQ	>30 min	~$500, P	1985
WMFT	✔	✖	NQ	>30 min	~$100, P	2001

Key: CHT, certified hand therapist; OT, occupational therapist, R = researcher, NQ, no required qualifications, C = custom made, P = Purchasable/commonly available.

Reliability Scoring

Test	Test-retest	Inter-rater	Internal consistency
ACMC	ICC: 0.94	ICC: 0.92 to 0.95	✖
AHAP	ICC: 0.839	ICC: 0.969	α: 0.846
AM-ULA	ICC: 0.91	ICC: 0.69 to 0.95	α: 0.89 to 0.91
AOSDT	✖	✖	✖
ARAT	ICC: 0.965	ICC: 0.998	α: 0.985[a]
BAM-ULA	ICC: 0.91	✖	α: 0.83
BBT	ICC: 0.96	ICC: 0.99	✖
CAPPFUL	✖	ICC: 0.88 to 0.99	α: 0.79 to 0.82
CQT	✖	ICC: 0.92 to 1[a]	✖
ECB	✖	✖	✖
FDT	ICC: 0.92	ICC: 0.94 to 0.99	✖
GaMA	ICC: 0.75	$p^b<0.05$	✖
JHFT	ICC: 0.84 to 0.97	ICC: 0.82 to 1.00	α: 0.95
MMDT	ICC: 0.79 to 0.87	✖	✖
NHPT	ICC: 0.92 to 0.95	ICC: 0.93 to 0.98	✖
PHAM	✖	✖	✖
PPT	ICC: 0.91	ICC: 0.91	✖
RCRT	✖	✖	✖
SHAP	ICC: 0.93	ICC: 0.89	✖
SHFT	ICC: 0.96 to 0.98[a]	ICC: 0.98	✖
SODA	r: 0.93	✖	α: 0.91
T-MAP	ICC: 0.93	✖	✖
UBET	$ICC_{[COT]}$: ✖	$ICC_{[COT]}$: 0.77 to 0.87	✖
	$ICC_{[MOU]}$: 0.70 to 0.85	$\kappa_{[MOU]}$: 0.68 to 0.82	
UNBT	ICC: 0.74 to 0.79	ICC: 0.72 to 0.73	α: 0.74
WMFT	ICC: 0.97[c]	ICC: 0.93 to 0.99[c]	α: 0.91[c]

Key: ICC, intraclass correlation coefficient, α = Cronbach's alpha, r = Pearson Correlation Coefficient, κ = Cohen's kappa, COT, completion of task; MOU, method of use.
[a]Tested with pediatric patients with spastic hemiplegia.
[b]*p*-value was used for a two and three factor RMANOVA.
[c]Primarily tested with stroke patients.

Validity Scoring

Test	Face	Content	Construct	External	Concurrent	Predictive
ACMC	Excellent	Adequate	Adequate	Excellent	Excellent	Adequate
AHAP	Adequate	Adequate	Adequate	Excellent	Excellent	Adequate
AM-ULA	Excellent	Excellent	Adequate	Excellent	Excellent	Excellent
AOSDT	Adequate	Adequate	No Evidence	Adequate	Excellent	Adequate
ARAT	Excellent	Adequate	Adequate	Excellent	Excellent	Adequate
BAM-ULA	Excellent	Adequate	Adequate	Excellent	Excellent	Adequate
BBT	Adequate	Poor	Adequate	Poor	Excellent	Poor
CAPPFUL	Excellent	Excellent	Adequate	Excellent	Excellent	Excellent
CQT	Excellent	Excellent	Excellent	Adequate	No Evidence	Excellent
ECB	Adequate	Adequate	No Evidence	Adequate	Excellent	Adequate
FDT	Adequate	Poor	Adequate	Poor	Excellent	Poor
GaMA	Adequate	Adequate	Adequate	Excellent	Adequate	Excellent
JHFT	Adequate	Adequate	Adequate	Adequate	Excellent	Poor
MMDT	Adequate	Poor	Poor	Poor	Excellent	Poor
NHPT	Adequate	Poor	Poor	Poor	Excellent	Poor
PHAM	Excellent	Excellent	No Evidence	Excellent	Excellent	Excellent
PPT	Adequate	Adequate	Adequate	Adequate	Excellent	Poor
RCRT	Adequate	Adequate	No Evidence	Poor	No Evidence	Poor
SHAP	Adequate	Adequate	Adequate	Excellent	Excellent	Excellent
SHFT	Adequate	Adequate	Adequate	Excellent	Excellent	Adequate
SODA	Adequate	Adequate	Adequate	Adequate	No Evidence	Adequate
T-MAP	Excellent	Excellent	Adequate	Excellent	Excellent	Excellent
UBET	Excellent	Excellent	Adequate	Excellent	Adequate	Excellent
UNBT	Excellent	Excellent	Adequate	Excellent	Excellent	Excellent
WMFT	Adequate	Adequate	Adequate	Adequate	Excellent	Adequate