2024

EXPLORING USABILITY IN EXERCISE INTERVENTIONS: FROM CONCEPTUALIZATION TO MEASUREMENT AND APPLICATION

Anne Inger Mørtvedt
Michigan Technological University, amrtvedt@mtu.edu

EXPLORING USABILITY IN EXERCISE INTERVENTIONS: FROM CONCEPTUALIZATION TO MEASUREMENT AND APPLICATION

By

Anne Inger Mørtvedt

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Applied Cognitive Science and Human Factors

MICHIGAN TECHNOLOGICAL UNIVERSITY

2024

This dissertation has been approved in partial fulfillment of the requirements for the Degree of DOCTOR OF PHILOSOPHY in the Applied Cognitive Science and Human Factors Program.

Department of Cognitive and Learning Sciences

Dissertation Advisor:	*Dr. Erich J. Petushek*
Committee Member:	*Dr. Steven J. Elmer*
Committee Member:	*Dr. Kevin Trewartha*
Committee Member:	*Dr. Eva Ageberg*
Department Chair:	*Dr. Kelly Steelman*

Table of Content

1 Introduction 1
1.2 Importance of Exercise 1
1.3 Adherence to Prescribed Exercise 1
1.4 Usability of Exercise Interventions 1
1.5 Dissertation Overview 2
2 Usability Testing and its Application in Exercise Medicine: A Literature Review 6
2.1 How to Measure Usability 6
2.2 Usability Scales and Questionnaires and Their Applicability to Exercise Interventions 10
2.3 Psychometric Properties in Scale Measurement 17
2.4 Concepts and Theories Related to Usability 20
2.5 Summary and implications 31
3 Usability Considerations for Enhancing Exercise-Based Injury Prevention Interventions 32
3.1 Abstract 32
3.2 Introduction 33
3.3 Material and Methods 34
3.4 Results 36
3.5 Discussion 38
3.6 Conclusion 40
3.7 Acknowledgements 40
4 Perspectives on Usability and Adoption of a New ACL Injury Prevention Program for Female Handball Players: A Mixed Methods Approach 41
4.1 Abstract 41
4.2 Introduction 43
4.3 Methods and Materials 44
4.4 Results 47
4.5 Discussion 55
4.6 Conclusion 57
4.7 Clinical implications 58
5 EVALUATING THE PULSE OF EXERCISE PROGRAMS: DEVELOPING AND VALIDATING THE INTERVENTION USABILITY SCALE FOR EXERCISE (IUSE) TO ENHANCE IMPLEMENTATION AND ADHERENCE 59
5.1 Abstract 59
5.2 Introduction 60
5.3 Methods 61
5.4 Results 67
5.5 Discussion 77
5.6 CONCLUSION 80
6 Conclusion and Future Research 81
6.1 Theoretical contributions 81
7 Reference List 86

8 Appendices 110
8.1 Strength training programs (Study 3) 110
8.2 Qualtrics survey 112
8.3 Study 2: Interview Guides 120
8.4 IRB Documents.....123

Author Contribution Statement

Study 1: Usability Considerations for Enhancing Exercise-Based Injury Prevention Interventions

Anne Inger Mørtvedt[1], Tron Krosshaug[2,3], Erich J. Petushek[1]

[1]Michigan Technological University, Department of Cognitive and Learning Sciences, Michigan, USA
[2]Oslo Sports Trauma Research Center, Department of Sports Medicine, Oslo, Norway
[3]Norwegian School of Sports Sciences, Oslo, Norway

TK was involved in designing the intervention study project and included participants. All authors designed the survey project. AIM collected and analyzed data and wrote the first draft of the paper. All authors contributed to the final manuscript. All authors read and approved the final manuscript.

Submitted to the HFES International Symposium on Human Factors and Ergonomics in Health Care.

Study 2: Perspectives on Usability and Adoption of a New ACL Injury Prevention Program for Female Handball Players: A Mixed Methods Approach

Anne Inger Mørtvedt[1], Tron Krosshaug[2,3], Erich J. Petushek[1]

[1]Michigan Technological University, Department of Cognitive and Learning Sciences, Michigan, USA
[2]Oslo Sports Trauma Research Center, Department of Sports Medicine, Oslo, Norway
[3]Norwegian School of Sports Sciences, Oslo, Norway

TK was involved in designing the intervention study project and included participants. All authors designed the survey project. AIM collected and analyzed data and wrote the first draft of the paper. All authors contributed to the final manuscript. All authors read and approved the final manuscript.

Submitted to the BMJ Open Sport & Exercise Medicine.

Study 3: The Development and Validation of the Intervention Usability Scale for Exercise (IUSE)

Anne Inger Mørtvedt[1] & Erich J. Petushek[1]

[1]Michigan Technological University, Department of Cognitive and Learning Sciences, Michigan, USA

Both authors designed the study. AIM collected and analyzed data and wrote the first draft of the paper. Both authors contributed to the final manuscript. Both authors read and approved the final manuscript.

Written up for submission to the British Journal of Sports Medicine.

Acknowledgements

I would like to express my sincere gratitude to all those who have played important roles in this journey towards completing my dissertation.

First and foremost, I extend my appreciation to my committee members, Dr. Eva Ageberg, Dr. Steven Elmer, and Dr. Kevin Trewartha, for their insightful feedback, constructive criticism, and expert advice, which have significantly enhanced the quality of my research.

I am thankful to the faculty and staff at the Department of Cognitive and Learning Sciences, the Graduate School, and the International Student Programs for their help and support throughout these years.

Special thanks go to my peers in the ACSHF program, particularly Katrina Carlson and Dr. Brittany Nelson, and to the wonderful friends I have been so lucky to make during my time in Michigan.

A very special thanks is extended to my partner, Brendan, for taking me on adventures, providing warmth, love, and continuous encouragement.

Great appreciation is also extended towards my motherland. I am especially grateful to my mentor, Tron Krosshaug, for collaborations, support, and guidance. I want to thank my beloved friends back home for their unconditional love and encouragement during these years.

Most importantly, I am forever grateful and in debt to my family for their unwavering love and support. You have always had my back and provided the safe ground I have needed to accomplish my goals and explore the world.

I would like to acknowledge the financial support provided by the Health Research Institute at MTU and a finishing fellowship, which made this research possible. Their generosity and investment in my academic pursuits are deeply appreciated.

I extend thanks to Muscle Animations AS, ExorLive AS, and my friend, Åsta Lindstad Isum, for help with research materials. I also want to express my heartfelt gratitude to all research participants, without whom this research would not have been possible.

Beyond all, I am grateful to my advisor, Dr. Erich Petushek, for his guidance, encouragement, engagement, and timely feedback. Thank you for believing in me, emphasizing work-life balance, being patient, and for lifting me up at times when things have felt extra challenging. I am proud to look back at our accomplishments.

Abstract

Exercise interventions hold promise for preventing and treating numerous conditions, diseases, and injuries. However, these interventions will only be effective if they are being used. Unfortunately, uptake and adherence to prescribed exercise and physical activity guidelines are insufficient. Some reasons for this include lack of knowledge, resources, flexibility, and enjoyment. Exercise program developers need to not only consider the effectiveness of the program during the development phase, but also involve end-users and receive feedback on program usability to determine likelihood of uptake and adoption. Usability testing can be used to detect barriers to use and implementation likelihood but has not yet been utilized within the domain of exercise-based interventions. The goal of this research was to better characterize and quantify exercise program usability to promote the design of more usable exercise programs. In the first study, a modified usability scale was used to assess and identify important program characteristics and their relationship to female handball players' intention to use a newly developed anterior cruciate ligament (ACL) injury prevention program (IPP). Study 2 involved a mixed methods approach to gain deeper insight into factors affecting use of an IPP and the relationship between perceived program characteristics and effectiveness of the program utilizing interviews with coaches and players, and surveys. From study 1 and 2, results indicated that perceived effectiveness, enjoyability, efficiency and flexibility affected players' and coaches' intention and willingness to continue using the IPP. Building on these findings, the Intervention Usability Scale for Exercise (IUSE) was developed and validated in study 3. Exercise intervention stakeholders and target users of an exercise program contributed to item generation and content validation. Subsequently, a large sample of target users used the full scale to assess the usability of an exercise program. Following an extensive data analysis process, the 8-item IUSE indicated good psychometrics properties. Collectively, this research sought to improve exercise program usability by developing a tool exercise intervention developers can utilize as part of their program development and assessment process. Future studies should evaluate the predictive utility of the scale on actual uptake and adherence to an exercise intervention.

1 Introduction

1.2 Importance of Exercise

Exercise interventions have been proven to prevent and treat numerous injuries and illnesses and improve quality of life (Pedersen & Saltin, 2015; World Health Organization, 2022). This includes prevention and management of diseases such as cancer, diabetes, cardiovascular diseases, Parkinson's disease, multiple sclerosis, obesity, chronic obstructive pulmonary disease, osteoarthritis, chronic back pain, muscle strains, and several others (Arnason et al., 2008; Pedersen & Saltin, 2015). Physical activity also improves brain health, including enhanced thinking and learning and judgment skills (World Health Organization, 2022). The World Health Organization also reports that insufficient physical activity levels increase risk of death by 20-30% compared to recommended activity levels.

1.3 Adherence to Prescribed Exercise

The recommended guidelines for physical activity are considered "prescribed exercise" that apply to the whole world's population. The WHO and the CDC state that about 1 in 4 Americans meet the recommended guidelines for physical activity. Globally, even fewer adolescents are sufficiently active (< 20%) (World Health Organization, 2022). This is an increasing trend and a country's level of inactivity tends to correspond with high or rising gross national product. The Physical Activity Guidelines for American adults aged 18-64 is defined as performing 150-300 minutes of moderate-intensity or 75-150 minutes of vigorous-intensity aerobic physical activity a week, or an equivalent combination, and muscle-strengthening activities at least two days per week (U.S. Department of Health and Human Services, 2018). The strength training component seems to be less adhered to compared to the aerobic component. Among all American adults (18+), 22.7% met only the guidelines for aerobic activity and 6.8% met only the guidelines for muscle-strengthening activity. Only 30.2% of adults in America meet the recommended two times per week of muscle-strengthening activities (Bennie et al., 2018), compared to 46.9% meeting the guidelines for aerobic activity. There is a need to increase adherence to prescribed/recommended exercise interventions to improve health and prevent injuries and diseases. Targeting the development design and usability assessment of exercise interventions might be one way to increase adherence.

1.4 Usability of Exercise Interventions

To increase and predict use of exercise programs, usability assessment can be conducted to identify the program's effectiveness, efficiency and user satisfaction (International Organization of Standardization, 2018). There are currently no studies on usability testing of exercise programs when performing literature searches, beyond assessing the use of some kind of technical device used in the exercise intervention and not the program/intervention itself. For example, Batsis et al. (2019) performed a usability assessment of a Bluetooth-enabled resistance exercise band in a population of older adults with obesity. Other examples are studies that have assessed the usability of different tablet, mobile and computer-tailored exercise interventions (Evans et al., 2021;

Hawley-Hague et al., 2020; Mehra et al., 2019). However, there are no known studies that have assessed the usability of the specific program characteristics (e.g., the order and choice of exercises, variation, difficulty).

Previous research focusing on exercise interventions and increasing adherence of evidence-based interventions has been conducted. Typically, these studies focus on identifying barriers and facilitators, implementation strategies and feasibility studies (Ageberg et al., 2019; Donaldson et al., 2019; Heywood et al., 2017; Joy et al., 2013; Richmond et al., 2020). These concepts seem to have a strong connection with usability based on their definitions, methodologies, and overall goal/purpose. Usability testing typically focuses on identifying barriers to use throughout the development process. In contrast, implementation science typically focuses on the implementation process of an intervention that has already been designed, developed, and tested in a controlled research setting. Feasibility, Appropriateness and Acceptability are three outcome measures in implementation science (Weiner et al., 2017). The Feasibility Intervention Measure, which assesses perceived ease of use (e.g. "This EBP seems easy to use") has been suggested as a direct measure of adherence and/or completion (Heywood et al., 2017; Weiner et al., 2017).

Evidence-based exercise interventions are typically static tools that should be used exactly as prescribed by the researchers that have documented the desired effect of the program (e.g., injury prevention/rehabilitation, weight loss, hypertension). We can only assume that the program will have similar effects if being slightly modified, as the modified version has not had its effectiveness documented. The purpose and documented effect of the program are naturally what drives people to use it. However, assessing and documenting the effect of a new intervention is a time-consuming and intricate process and can take months to years to document. Furthermore, the translation from research to practical implementation can take up to 17 years (Arundale et al., 2022; Green et al., 2009; Morris et al., 2011; Rubin, 2023; Trochim, 2010).

Lyon & Koerner (2016) have proposed a user-centered approach to the design, development, and implementation process of health-related interventions with the goal of addressing existing divisions between intervention design and intervention implementation processes, and to target the key design issues in evidence-based interventions, such as flexibility, complexity, and effectiveness. Within injury prevention in sports, where adherence to evidence-based prevention programs are typically low, Petushek & Donaldson (2020) has proposed the use of human factors methods, such as usability testing, to help develop checklists for what the program need to include and focusing on a user-centered approach when designing exercise programs. This approach might help shorten the lag between research and implementation.

1.5 Dissertation Overview

An overview of this dissertation is presented in Figure 1.1. Funding and support for this research was provided by a graduate student finishing fellowship and the Health Research Institute of Michigan Technological University.

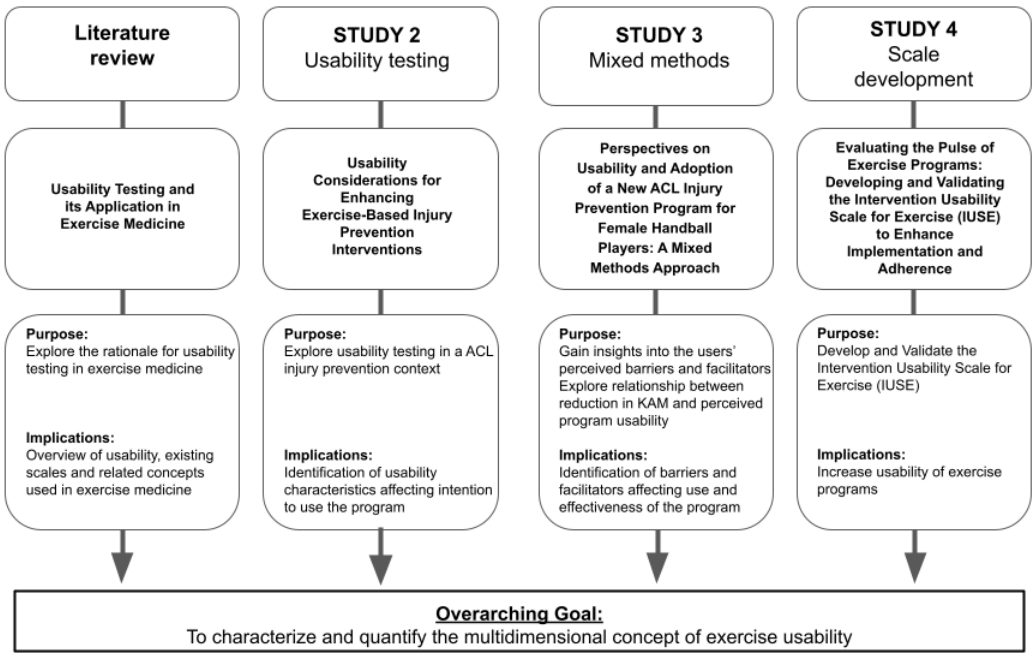


Figure 1.1. Overview of dissertation

The following sections are written up for a submission to the BJSM PhD Academy Award.

1.5.1 What did I do?

A literature review and three consecutive studies were developed to offer insights into usability testing and construct clarity within the realm of exercise interventions. The overarching goal was to accurately characterize and quantify the multidimensional concept of exercise program usability, laying a robust foundation for measurement and subsequent improvement. Through the development and rigorous psychometric assessment of the Intervention Usability Scale for Exercise (IUSE), subconstructs contributing to intervention usability and their impact on the intention to use an exercise intervention were proposed.

1.5.2 Why did I do it?

Adherence to prescribed exercise interventions are typically low (World Health Organization, 2022), and some proposed barriers for use involve program design issues, lack of enjoyment, knowledge gaps, and resource constraints (Collado-Mateo et al., 2021; Donaldson et al., 2019; Minnig et al., 2022). Drawing inspiration from successful methodologies in product development, particularly in technological domains, this study explored the potential of usability testing as a tool for enhancing exercise intervention

efficacy. However, first, a foundational understanding of the concept of "exercise intervention usability" was deemed necessary. For instance, it was unclear what factors seem to determine exercise usability and further affect uptake of and adherence to exercise interventions.

1.5.3 How did I do it?

A critical literature review on usability and its application in exercise medicine laid the groundwork. This review justified further investigations, encompassing quantitative, qualitative, and psychometric research methodologies. The purpose of the first two studies was to explore the use of usability testing during the assessment of a newly developed anterior cruciate ligament (ACL) injury prevention program (IPP). Data on factors influence program effectiveness (e.g. reduction in knee abduction moment) and players' and coaches' intention to use the program were collected through surveys, employing a modified usability scale, and through interviews. These insights guided the development of a usability scale for exercise interventions. The scale development and validation study included a content validation phase with exercise intervention developers and stakeholders (e.g., strength and conditioning coaches, physical therapists, athletic trainers) and target users of an exercise program. The long version scale derived from the content validation phase was tested for assessing a strength training program based on recommendations for weekly physical activity by 526 users from the general population (U.S. Department of Health and Human Services, 2018). Following COSMIN guidelines, Standards for Educational and Psychological testing, and modern psychometric techniques like item response theory, the efficient 8-item IUSE scale was developed and the Initial validity evidence established (American Educational Research Association, 2018; Mokkink, Terwee, Knol, et al., 2010).

1.5.4 What did I find?

The psychometric properties and construct clarity in scales identified through the literature review, and their appropriateness for assessing exercise intervention usability, were considered poor. The ACL IPP studies indicated that perceived effectiveness, enjoyability, efficiency, and flexibility affected players' and coaches' implementation likelihood post-intervention. The key factor for further use in this sample seemed to be perceived effectiveness of the program (e.g., injury risk reduction and/or enhanced performance), which at the time was unknown. Additionally, barriers such as the rigid design, time commitment, and need for equipment were identified. Through rigorous testing of the IUSE scale, three factors seem to make up the latent trait of usability: $IUSE_{Value}$, $IUSE_{Ease}$, and $IUSE_{Social}$. Notably, the $IUSE_{Value}$ subscale was most closely related to intention to use the program (r = 0.82 [95% CI 0.79, 0.85]). The 8-item IUSE demonstrated satisfactory psychometric properties. Figure 1.2 provides a summary of the collective findings.

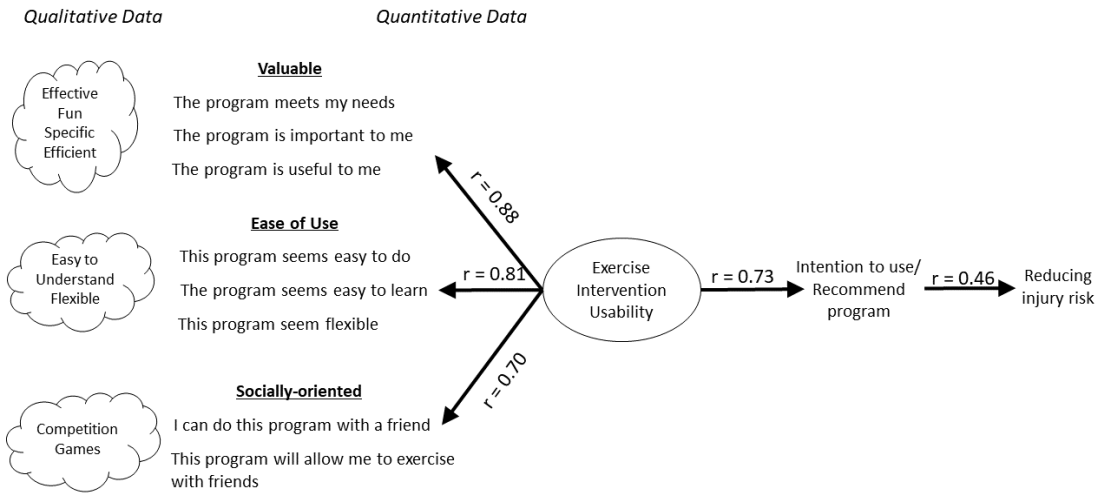


Figure 1.2: Visual summary of collective findings

1.5.5 What is the most important clinical impact/practical application?

Collectively, research conducted for this PhD is a significant stride conceptualizing exercise intervention usability. All studies indicate that as long as users perceive the program as valuable to them (e.g., meet their needs, fulfill its purpose), they seem more inclined to use it. Similarly, results across these studies imply that ease of use is less important for intention to use. However, this factor seems to add important information to the concept of exercise intervention usability and might play a role for initial uptake. Additionally, a unique contribution of IUSE compared to similar scales is the inclusion of the socially oriented factor. As social support has been proposed as important for adherence to exercise interventions (Smith et al., 2023), this subscale might capture valuable information in specific populations. Of note, this research assessed intention to use and not actual use/behavior, and further research is needed to add predictive validity evidence to IUSE. Ultimately, the long-term aim is to furnish exercise intervention designers with an efficient tool for assessing and enhancing exercise program usability.

2 Usability Testing and its Application in Exercise Medicine: A Literature Review

This review discusses critical issues within the domains of usability, user experience, and implementation science. It provides the rationale for developing a valid and reliable scale. Such a scale would assist program designers in creating exercise programs that the target population is more likely to implement and use.

2.1 How to Measure Usability

2.1.1 Usability and usability testing

The International Organization of Standardization (ISO) defines usability as the "extent to which a system, product, or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" (International Organization of Standardization, 2018). Furthermore, ISO defines usability testing as an "evaluation that involves representative users performing specific tasks with the system to enable the measurement of efficiency, effectiveness, and/or user satisfaction." This definition has been challenged and modified over the years but it is still the definition most frequently used, which includes more recent usability studies (Weichbroth, 2020). Based on these definitions, usability assessments should aim at capturing the system's effectiveness, efficiency, and user satisfaction.

Usability testing is important as part of the development process of new products and systems. Within technology-based products and systems, usability and perceived usefulness influence our acceptance of the new system and predicts actual use (Keenan et al., 2022). Usability testing helps system developers to uncover problems in the design, discover ways to improve the design, and learn about the specific users' behaviors and preferences. Depending on the system and context in which it is used, some developers may emphasize effectiveness and speed-accuracy tradeoff (e.g. medical device), whereas others prioritize optimizing user satisfaction and enjoyability (e.g. video game).

Although not included in the ISO 9241-11 definition of usability, accessibility is considered a part of the ISO standard titled "Usability of consumer products and products for public use," and has been operationalized as "the extent to which a product can be used with effectiveness, efficiency, and satisfaction by people from a population with the widest range of characteristics and capabilities to achieve a specified goal in a specified context of use" (International Organization of Standardization, 2018). Like accessibility, the ISO 9241-11 also reports user experience as a concept or discipline related to usability. User experience is defined by the ISO as "a user's perceptions and responses that result from the use and/or anticipated use of a system, product or service."

The concepts of both accessibility and user experience seems to be important to consider when assessing usability of a system. The ISO 9241-11:2018 proposes that the concept of usability involves both the context of use and the outcomes of use. The context of use depends not only on the system, product, or service, but the specific users, environment, resources, and the specific goals and tasks. For the outcomes of use, they include the

three usability constructs and other outcomes, such as accessibility, user experience and avoidance of harm.

2.1.2 Nomenclature of usability measures

The nomenclature within research on usability scales and questionnaires can be confusing. Words such as constructs, components, concepts, criteria, metrics, attributes, aspects, factors, and features are used interchangeably to describe what is being assessed through a usability scale. A construct is defined by the American Psychological Association (2020) as "a complex idea or concept formed from a synthesis of simpler ideas." *Usability* can be one construct, but usability can also involve assessment of other constructs (e.g., learnability, efficiency and enjoyability). These sub-constructs are sometimes also referred to as attributes or aspects and can be important for a more comprehensive and in-depth assessment of the higher-order construct of usability. In addition, a more in-depth assessment of usability may be informative for suggesting changes to the various modifiable features, such as efficiency and satisfaction.

2.1.3 Usability models and metrics

Many usability questionnaires and scales have been developed and researched. There is a large range of items and a large variety of sub-constructs defined in the different scales. There is very little consistency in use of constructs, and literature reviews of existing usability scales/questionnaires reveal that the usability criteria defined by the ISO are rarely the predefined constructs being assessed in usability questionnaires (Assila et al., 2016; Madan & Kumar, 2012). This raises the question of whether the scales are not appropriate for usability testing or whether the definition of usability needs to be adjusted. The existence of a "usability construct" has been debated in previous research. Particularly, Tractinsky (2018) published a provocative article critiquing the use of umbrella concepts in the Human-Computer Interaction community, and particularly the usability construct. His argument has three main points: 1. Usability is an umbrella concept; 2. there is a mismatch between the construct of usability and how it is measured empirically; 3. the usability construct needs to be unbundled and replaced by more well-defined constructs in order to progress scientifically. As a response to this article, Lewis (2018) wrote an editorial arguing for why the usability construct is "alive and well," emphasizing evidence in structural analysis of both objective and subjective data from usability studies, indicating that there is a consistent underlying construct of usability.

Madan & Kumar (2012) describes the evolution of usability concepts and evaluation models and how they all use different sub-attributes (hypothetical constructs) for describing the success of a system. All five models included have different combinations of sub-attributes. The Eason Model from 1984 has frequency, openness, knowledge, motivation, discretion, ease of learning, ease of use and task match as its sub-attributes. The Shackel Model (1991) uses effectiveness, learnability, flexibility, and attitude. The Nielsen Model (1993) uses learnability, efficiency, memorability, errors, and satisfaction. Finally, the ISO 9241-11 (1998) has the three attributes from their definition (effectiveness, efficiency, and satisfaction), whereas the newer ISO 9126 (2001) uses understandability, learnability, operability, attractiveness, and usability compliance. With

this review, Madan & Kumar (2012) states that they have determined the sub attributes that form the basis for the usability evaluation of a software system. In other words, they believe that there is more to a usability assessment than the three original constructs effectiveness, efficiency, and satisfaction.

Weichbroth (2020) did a systematic literature review on the different usability definitions, attributes and measures used in usability testing of mobile applications. In total, 75 attributes were identified among 53 studies. The author found that the three most used attributes were efficiency (70%), satisfaction (66%) and effectiveness (58%). This aligns perfectly with the ISO (year) definition, which was also the definition used in 88 % of the studies included (n=66). Other commonly used attributes were learnability (45%), memorability (23%), cognitive load (19%) and errors (17%). For the studies included in the review, controlled observation and surveys were used to measure efficiency and effectiveness, and only surveys were used to measure satisfaction. Overall, survey/questionnaire was the most frequent method to assess usability of mobile applications. Weichbroth (2020) distinguishes between usability of the product and the performance of the user (e.g., attention, the ability to learn and memorize etc.). Two interesting remarks from this review were that 91% of the studies from their literature search were lacking a usability definition, and that combining both objective and subjective assessments produce an outcome which neither refers to the usability of the application nor the user experience (e.g. performance of the device vs performance of the user). The author states that this affects the validity of the results negatively due to an ignorance to methodological rigor.

Wronikowska et al. (2021) published a systematic review of the usability metrics and methods used in the assessment of electronic health record systems. They found 11 different usability methods among the 51 studies included, and 78% (n=38) of the studies used questionnaires/scales in their evaluation. Out of the 38 studies using questionnaires/scales, the System Usability Scale (SUS), NASA Task Load Index, Post-Study System Usability Questionnaire, and the User Interaction Satisfaction Questionnaire were the most frequently used questionnaires (42%, 16%, 13%, and 11%, respectively). Components or metrics they included were satisfaction, efficiency, effectiveness, learnability, memorability, and error components. These metrics correspond to the Nielsen Model and ISO 9241-11 presented above. Satisfaction was most frequently measured in the 38 studies (75%), followed by efficiency (63%), effectiveness (61%), errors (31%), learnability (24%), and memorability (2%). In addition, they identified and reported a new metric, usefulness, that was found in 39% of the studies. However, when only considering usability metrics within studies using questionnaires, effectiveness and efficiency were only measured in two of the studies, whereas satisfaction, learnability and usefulness were used in 31, 10 and 11 studies, respectively. The effectiveness and efficiency metric were typically measured objectively (e.g., time to complete task, number of successfully completed tasks). If these metrics are to be included in a scale/questionnaire, different measures must be created (e.g. perceived or self-report measures).

Assila et al. (2016) did an extensive review of 24 standardized usability questionnaires. Detailed information and full name of the questionnaires discussed below are presented in Table 2.1. Seventeen of them have been used in several kinds of software applications, whereas the others have more specific applicability (e.g., mobile applications and websites). Within the more universal questionnaires, system usefulness, usability, overall ease of task, completion and "overall system" are the most frequent criteria (sub attributes/constructs) considered by the questionnaires. The Usability Metric for User Experience (UMUX) is the only questionnaire from the review that explicitly reports to measure the three usability characteristics that define usability (effectiveness, efficiency, and satisfaction). However, when evaluating the meaning of each item in all of the universal questionnaires and their association to the three usability characteristics/criteria, the authors found that five other questionnaires capture the three characteristics as well (PSSUQ, CSUQ, T-CSUQ, SUMI and ASQ). For example, the item "I can effectively complete my work using this system" from the PSSUQ and CSUQ function as the effectiveness criteria; "This software responds too slowly to inputs" from the SUMI functions as the efficiency criteria; and "Overall, I am satisfied with the ease of completing the task in this scenario" from the ASQ functions as the satisfaction criteria. However, these questionnaires also entail usability criteria such as guidance, comprehensibility, consistency, and workload.

Hodrien et al. (2021) recently published a review of subjective questionnaires to guide assessment of system usability. Similarly to Assila et al. (2016), the authors reported whether the three usability characteristics from the ISO definition were measured in the different questionnaires even though they were not specifically reported as or validated as constructs. In addition, they included learnability as a fourth common characteristic of usability. Seven additional questionnaires were reported to measure both effectiveness, efficiency, and satisfaction: AttrakDiff2, meCUE, QUIS, TAM, UEQ, USE and EUCS. However, the QUIS, TAM and USE questionnaires were reported to lack one of the characteristics according to the review by Assila et al. (2016).

The SUS is a generic and widely used usability scale consisting of 10 items (Brooke, 2013). It has previously been proposed that the SUS had a 2-dimensional factor structure assessing Learnability and Usability (Sauro & Lewis, 2009). However, this finding has never been replicated and the authors later published an article suggesting it likely is a unidimensional scale assessing perceived usability (Lewis & Sauro, 2017). The Usability Metric for User Experience (UMUX) consists of 4 items that are responded to through a 7-point Likert scale. Through the development of the UMUX, the goal of Finstad (2010) was to address the issues identified with using the system usability scale (SUS): it does not map onto the concepts that comprise usability according to ISO standards, and it is not considered a diagnostic tool. The four items included are 1) [This system's] capabilities meet my requirements (*Effectiveness*), 2) Using [this system] is a frustrating experience (*Satisfaction*), 3) [This system] is easy to use (*Overall*), and 4) I have to spend too much time correcting things with [this system's] (*Efficiency)*. Although including items of three different concepts, principal components analysis revealed a unidimensional scale (eigenvalue 3.37, 84.37% of variance explained). The correlation between the SUS and UMUX was r = 0.96, which is above the 0.8 criterion for providing

evidence of validity. The scale was also sensitive in detecting differences in usability between two different systems rated as having poor and good usability using the SUS.

2.1.4 Correlations between usability characteristics

The relationship between effectiveness, efficiency and satisfaction has been debated in previous literature with mixed results. Hornbæk & Law (2007) published a meta-analysis on the correlations among usability measures looking at raw data from 73 studies on usability. As a result of their study, they raise an important question about the lack of predictive theories about the relations between usability aspects. In their meta-analysis they found that correlations between the three usability measures classified by the ISO are medium to low (r = 0.16-0.25). Similarly, Frøkjær et al. (2000) implied that effectiveness, efficiency, and satisfaction should be measured independently, and one should not expect correlations between them. They found no correlations between the variables (r = 0-0.05). However, their findings were different to a study by Sauro & Kindlund (2005) that found correlations between r = 0.3-0.5 and thus proposed using a single score for usability. Hornbæk & Law (2007), on the other hand, warranted the aggregation of the three metrics into one usability score given the weak correlations they found. Sauro & Lewis (2009) extended the work of Hornbæk & Law (2007) by using data from actual usability tests (in contrast to including studies that are not traditional scenario-based usability tests). In their study, they use time, errors, completion, task-level satisfaction and test level satisfaction as variables, whereas Hornbæk & Law (2007) used time-on-task, binary completion rates, error rates and user satisfaction. These variables are all supposed to represent the three usability measures from the ISO classification. Sauro & Lewis (2009) found high correlations between time and error, task satisfaction and test satisfaction ($r > 0.6$). In addition, completion and error, completion and satisfaction, completion and time, time and satisfaction, and error and satisfaction were all around r = 0.5. Based on this, they suggest that effectiveness and satisfaction have a correlation of r = 0.35, effectiveness and efficiency have a correlation of r = 0.53, and efficiency and satisfaction have a correlation of r = 0.37.

If effectiveness and efficiency are to be measured through usability scales and questionnaires, this can be done through subjectively reporting how efficient they find the system and how well they think it works. However, this might result in higher correlations between the usability metrics. For example, if a system is perceived as inefficient, it is very likely that the user will not find it very enjoyable/satisfactory to use the system. Similarly, if the user thinks the system works as intended and for its purpose, their satisfaction score will likely be remarkably higher than if they think the system is not doing what it should.

2.2 Usability Scales and Questionnaires and Their Applicability to Exercise Interventions

The table below (Table 2.1) displays different usability scales/questionnaires that were identified through the literature review on usability. The table is based on the ABC of test construction from Ziegler (2014) which aims to answer three questions; A) What is the

construct being measured?, B) What are the intended uses of the measure?, and C) What is the targeted population?. For scales that have items relevant for usability testing of exercise programs, their psychometric properties and validity evidence are displayed in Table 2.2.

Table 2.1 Usability Scales and Questionnaires

Scale	**Construct Measured**	**Intended Use of Measure**	**Target population**	**Items Applicable to Exercise Interventions**
System Usability Scale (SUS)	Usability (Learnability)	Wide variety of systems and products	Users of system	Yes
Intervention Usability Scale (IUS)	Usability Learnability	Wide variety of interventions	Health care professionals Users of intervention	Yes
Questionnaire for User Interface Satisfaction (QUIS)	Not specified	Human Computer Interaction Products and computer software	Users of computer systems	No
Software Usability Measurement Inventory (SUMI)	Learnability Efficiency Affect Helpfulness Control	Software applications	Software users	No
Computer System Usability Questionnaire (CSUQ)	System Usefulness Information Quality Interface Quality Overall Satisfaction	Computer systems	Users of computer systems	No
After-Scenario Questionnaire (ASQ)	Not reported	Scenario-based usability studies	Users of the system	Yes
The Usefulness, Satisfaction, and Ease of Use Questionnaire (USE)	4 factors, poor model fit	Wide variety of technology-based systems and products	Users of system	Most items
The Usability Metric of User Experience (UMUX)	Perceived Usability	User experience	Users of system	No

UMUX-LITE	Perceived Usability	User experience	Users of system	No
Purdue Usability Testing Questionnaire (PUTQ)	Compatibility Consistency Flexibility Learnability Minimal action Minimal memory load Perceptual limitation User guidance	Human-computer interactions	Users of system	No
End User Computing Satisfaction (EUCS)	Content Accuracy Format Ease of use Timeliness	End-user computing	Users of system	No
Post-Study System Usability Questionnaire (PSSUQ)	System Usefulness Information Quality Interface Quality Overall Satisfaction	Computer systems	Users of computer systems	No
DEsign-oriented Evaluation of Perceived usability (DEEP)	Content Structure and information architecture Navigation Layout consistency Visual guidance	Design aspects of web-based systems	Users of web-based system	No
Usability Magnitude Estimation (UME)	Usability (one question)	Software and information systems	Users of system	Yes
Technology Acceptance Model (TAM)	Perceived usefulness Perceived ease of use	Intended use of Information Technology	Users of system	No
Subjective Mental Effort Questionnaire (SMEQ)	Mental Effort (one question)	Software and information systems	Users of system	Yes
SEQ	Difficulty (one question)	Wide variety of technology-based systems and products	Users of system	Yes

Modular Evaluation of Key Components of User Experience (meCUE)	Usefulness Usability Visual aesthetics Status Commitment Positive emotions Negative emotions Intention to Use	Technical devices	Users of system	Most items
AttrakDiff2	Pragmatic Quality Perception Hedonic Quality Perception	Interactive products	Users of interactive product	No
User Experience Questionnaire (UEQ)	Attractiveness Perspicuity Efficiency Dependability Stimulation Novelty	Interactive products	Users of interactive product	Yes

Table 2.2. *Psychometric Properties of Relevant Usability Scales/Questionnaires following COSMIN guidelines*

Scales	**IC**	**GR**	**ME**	**ConV**	**SV**	**HT**	**CCV**	**CriV**	**Resp**	**ITEMS**
SUS	0.92	0.7-0.91		X	X		X	X	X	1. I think that I would like to use this system frequently. 2. I found the system unnecessarily complex. 3. I thought the system was easy to use. 4. I think that I would need the support of a technical person to be able to use this system. 5. I found the various functions in this system were well integrated. 6. I thought there was too much inconsistency in this system. 7. I would imagine that most people would learn to use this system very quickly. 8. I found the system very cumbersome to use. 9. I felt very confident using the system. 10. I needed to learn a lot of things before I could get going with this system.
IUS	.83			NR	X		NR			1. I like to use the intervention frequently 2. I find the intervention unnecessarily complex 3. I think the intervention is easy to use 4. I need the support of an expert consultant to be able to use the intervention 5. I find the various components of the intervention are well integrated 6. I think there is too much inconsistency in the intervention 7. I would imagine that most people would learn to use the intervention very quickly 8. I find the intervention very cumbersome to use 9. I feel very confident using the intervention 10. I needed to learn a lot of things before I could get going with the intervention
ASQ		0.90-96		NR	X	X		X		1. Overall, I am satisfied with the ease of completing the task in this scenario.

										2. Overall, I am satisfied with the amount of time it took to complete the task in this scenario. 3. Overall, I am satisfied with the support information (on-line help, messages, documentation) when completing the task.
USE		–		X	X			X		1. It helps me be more effective 2. It helps me be more productive 3. It is useful 7. It meets my needs 8. It does everything I would expect it to do 9. It is easy to use 12. It requires the fewest steps possible to accomplish what I want to do with it 14. Using it is effortless 15. I can use it without written instructions 24. I am satisfied with it 26. It is fun to use 27. It works the way I want it to work • 18 more items
UME		–						X		One question, dependent on task (ex. Give a value for task difficulty)
SMEQ		>0.94						X		“How hard was this task to execute”
SEQ		>0.94						X		Overall, how difficult or easy was the task to complete
meCUE				X	X		X	X		1. The product is creatively designed 6. It is quickly apparent how to use the product 7. By using the product, I would be perceived differently 11. The product is like a friend to me 14. I consider the product extremely useful 18. The product frustrates me 20. The product makes me tired 24. The product makes me happy

										31. I can hardly wait to use this product again 32. If i could, I would use the product daily 33. When using this product, I lose track of time • 22 more items
UEQ		.70-.90		X	X		X	X		Annoying/Enjoyable Unlikeable/Pleasing Inefficient/Efficient Difficult to learn/Easy to learn Boring/Exciting Dull/Creative Does not meet expectations/Does meet expectations • 19 other pairs
IC = Internal Consistency, GR = Global Reliability, ME = Measurement Error, ConV = Content Validity, SV = Structural Validity, HT = Hypothesis Testing, CCV = CrossCultural Validity, CriV = Criterion Validity, Resp = Responsiveness										

As outlined in Table 2.1. and 2.2, several usability scales exist and many of them include items that may be appropriate for an exercise intervention usability scale. However, psychometric properties are often not reported or seem insufficient. Future research is needed to determine reliability and validity of most scales in order to gain confidence that the scales measure what they are intended to.

2.3 Psychometric Properties in Scale Measurement

The foundation of scale development is assessing and revealing its psychometric properties (e.g., reliability and validity evidence). The American Psychological Association defines psychometrics as "the branch of psychology concerned with the quantification and measurement of mental attributes, behavior, performance, and the like, as well as with the design, analysis, and improvement of the tests, questionnaires, and other instruments used in such measurement". Many scales and questionnaires for usability assessments are available for use in clinical, industry, educational and research settings. However, literature warns that many of them do not have adequate validity evidence and should be used and interpreted with caution until their psychometric properties have been better evaluated (Salmond, 2008; Souza et al., 2017).

There are typically three phases of scale development (See Figure 2.1): 1) Identifying the domain and generating domain-relevant items, 2) Performing item reduction (e.g. Item Response Theory), assessing the reliability and dimensionality (internal structure) of the scale and items, and 3) adding validity evidence to the scale (e.g., convergent/discriminant and criterion validity) (Boateng et al., 2018). They each have several steps and analyses to be conducted to ensure adequate validity evidence which will be talked about more during our two proposed studies on exercise intervention usability scale development. The psychometric properties in scale development are further discussed below.

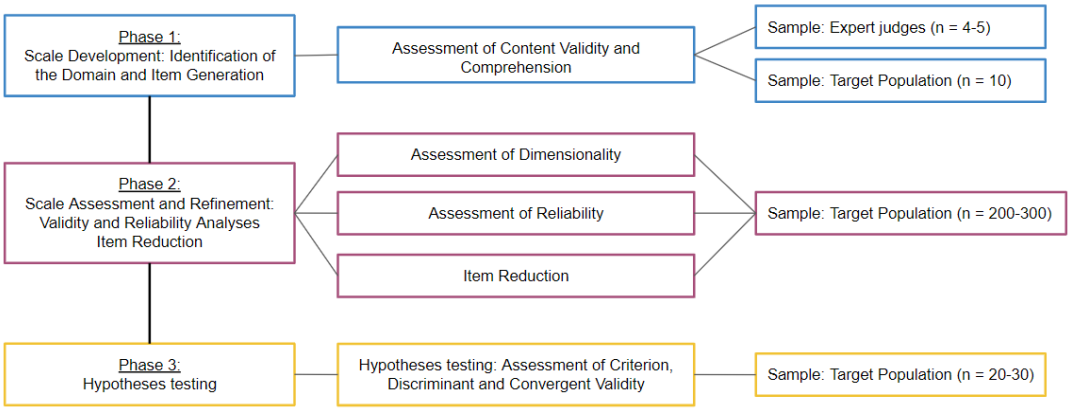


Figure 2.1. The scale development process

Several articles have been published discussing how to develop rigorous measurements (e.g. a scale), providing guidelines for item construction and selection, sample and sample sizes, and testing for validity and reliability (Boateng et al., 2018; Clark & Watson, 2019; Messick, 1995; Mokkink, Terwee, Knol, et al., 2010). Although some research reports reliability and validity as two single measures, that is not the case. Within test development and use in education and psychology, validity is defined as the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests (American Educational Research Association, 2018). Validity is a unitary concept, meaning that although different sources of evidence may illuminate different aspects of validity, they are not representing distinct types of validity.

Adequate validity evidence and reliability provide confidence about the quality and accuracy of the results of the scale. Both the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) taxonomy of measurement properties (Mokkink, Terwee, Patrick, et al., 2010) and the Standards for Educational and Psychological Testing (2014 edition) was made to provide guidelines for assessing the validity of interpretations of test scores for the intended test uses. The COSMIN guidelines describe seven measurement properties: internal consistency, reliability, measurement error, content validity, construct validity, criterion validity and responsiveness. On the other hand, the *Standards* describe five sources of validity evidence: 1) evidence based on test content, 2) evidence based on response processes, 3) evidence based on internal structure, 4) evidence based on relations to other variables, and 5) evidence for validity and consequences of testing. The COSMIN and the Standards are largely covering the same measurement properties. Responsiveness and cross-cultural validity will not be part of the initial scale development phases but will be important to assess at a later stage. Table 2.3 below describes the first four sources of validity evidence from the Standards that are relevant for scale development and how they can be tested statistically.

Table 2.3 Sources of Validity Evidence

Sources of Validity Evidence	**What is it?**	**Related Terms Used in Other Scientific Domains** ***Definition***	**Appropriate statistical tests**
Evidence Based on Test Content	The relationship between the content of a test and the construct it is intended to measure	Content Validity *The extent to which a test measures a representative sample of the subject matter or behavior under investigation.*	Logical or empirical analysis of adequacy Expert judgment

		(American Psychological Association, 2020)	
Evidence Based on Response Processes	The thought processes of users of the tested inventory as they respond to the questionnaire. This typically involves clarity of instructions and language used in the assessment tool, as well as the comprehension of instruction.	Face Validity *An informal review of a questionnaire by non-experts, assessing its clarity, comprehensibility, and appropriateness for the target-group* (Tanner, 2018)	Analyses of individual responses (e.g. concurrent verbal protocol or think aloud)
Evidence Based on Internal Structure	The degree to which the relationships among test items and test components conform to the construct on which the proposed test score interpretations are based.	Structural Validity *The degree to which scores of a scale are an adequate indication of the dimensionality of the construct, attribute or factor being measured (Mokkink et al., 2010)*	Factorial analysis Structural equation modeling Omega/Bifactor models
Evidence Based on Relations to Other Variables	The degree to which these relationships are consistent with the construct underlying the proposed test score interpretations. Includes discriminant and convergent validity (relationship between test scores and measures intended to assess different and/or similar constructs), test-criterion relationship (e.g. how accurately do test scores predict criterion performance), and generalizability validity (e.g. the degree to which validity evidence based on test-criterion relations can be generalized to a new situation without further study of validity in that new situation).	Criterion Validity *Evaluating the validity of a measure based on its relationship to a specific criterion.* Convergent Validity *The extent to which test scores or responses demonstrate a strong relationship with scores or responses on conceptually similar tests.* Discriminant Validity *A form of construct validity that defines how well the test or measure does not correlate with unrelated constructs.* Predictive Validity *The degree to which a test score predicts future behavior or performance on an accepted criterion measure.* (American Psychological Association, 2020)	Correlation Regression Structural Equation Modeling

2.4 Concepts and Theories Related to Usability

2.4.1 Usability versus implementation science

The goal of developing new systems, products and tools is usually to make something people want to use and implement. Within technology-based systems, there has been an increased focus on usability testing as part of the design and development process to assess the extent to which the system can be used by the specific users to accomplish their goals effectively, efficiently and satisfactory (International Organization of Standardization, 2018). Similarly, implementation science is a relatively new field with the goal involving methods to promote the systematic uptake of research findings and evidence-based practices (Bauer et al., 2015). Usability science and implementation science do therefore seem to be related disciplines. Both usability and implementation science are concerned with factors that affect adoption, acceptance, and sustained use of a "system".

Previous research does distinguish between the two disciplines. Although they may share a common goal or desired outcome, they might play important roles during different phases of system development and use. Usability testing can be part of all the different steps of system development, whereas implementation science typically focuses on ways to implement the already developed system (Bauer et al., 2015). Usability testing focuses on the system or product itself, whereas implementation science focuses more on the strategies to have the system being used after the system is developed. As part of the implementation phase, factors that do not necessarily seem like obvious barriers during usability testing might be identified, such as resources, personnel, and accessibility.

Within implementation research, *systems, products* and *tools* are typically swapped with terms like *interventions*, *evidence-based treatments/practices* or *"the thing"*. Lyon & Koerner (2016) describes how interventions are typically developed and tested before being disseminated more widely as a static intervention protocol. This disconnect between the intervention design context and intervention implementation context often leads to designs that are not applicable or appropriate in real-world contexts. To mitigate the risk of developing interventions that are not appropriate, feasible or accepted by the target population, Lyon & Koerner (2016) suggests that a user-centered approach to a greater extent becomes part of intervention development. Usability is the principal outcome of user-centered design, and usability assessment is typically an integral part of the user-centered design process. There is a need for more research on the connections between the aspects of intervention design (e.g. usability) and identified implementation outcomes.

2.4.2 Usability versus user experience versus accessibility

According to the ISO 9241 – 210 definition, user experience is "a person's perceptions and responses that result from the use and/or anticipated use of a product, system or service." (International Organization of Standardization, 2018) This indicates that user experience involves user satisfaction, perceived efficiency, and perceived effectiveness,

which are all characteristics included in the definition of the “classic” usability term. User experience has been considered an extension of classical usability, and focuses more extensively on emotional outcomes such as trust, satisfaction and perceived beauty (Lewis & Sauro, 2021). As with the definition of usability, the definition of user experience has been debated. There have particularly been three main types of user experience definitions; 1) the holistic view that refers to all actions, sensations, considerations, feelings and sense making of a person that interacts with the system/product, 2) the extension of usability, where affect and emotions (hedonic usability) are added in addition to effectiveness, efficiency and satisfaction (pragmatic usability), and 3) the primary focus on the emotion aspect. The first type of definitions is considered relatively broad, whereas the third type is more narrow.

There seems to have been a growing interest in the concept of user experience rather than usability the past two decades. For example, the Usability Professionals Association changed its name to the User Experience Professionals Association in 2012. Similarly, the Journal of Usability Studies changed its name to the Journal of User Experience in 2022. Before the 2000’s, product/system developers primarily focused on the more classical usability outcomes (e.g., successful task completion, perceived usability, and user satisfaction), however, in the early 2000’s, Marc Hassenzahl and colleagues started to publish research where they made a distinction between classical usability (pragmatic usability) and hedonic usability, where hedonic usability to a greater extent focused on the products’ appealingness and attractiveness. Another description of hedonic versus pragmatic is the focus on “be-goals” versus “do-goals", respectively (Hassenzahl, 2007). Hassenzahl argued that the classic definition of usability neglected the contribution of fun and enjoyment. He believes that user experience differs fundamentally from usability in three ways: 1) its emphasis on positive aspects of the user-product relationship, 2) its emphasis on the understanding and management of the subjective side of product use as opposed to objective performance criteria, and 3) its incorporation of hedonic aspects (Hassenzahl, 2007). Historically, usability was operationalized and measured primarily in terms of the user’s effectiveness and efficiency within industry work (Bevan et al., 2015). There has been a revision of the ISO 9241-11:1998 definition of usability in order to clarify the satisfaction component and address the importance of understanding the user experience (e.g., the user’s subjective reactions and emotional experience) (Bevan et al., 2015). In the 1998 definition, satisfaction meant “freedom from discomfort, and positive attitudes towards the use of the product”. The new edition includes personal responses that have been highlighted in research on user experience: “the extent to which attitudes related to the use of a system, product or service and the emotional and physiological effects arising from use are positive or negative” (Bevan et al., 2015). The revised version also provides an explanation of the relationship usability has to related concepts such as user experience, accessibility and human-centered quality. User experience and the satisfaction component of usability seem to overlap to some extent. According to the ISO standards, satisfaction does not include anticipated use. However, they acknowledge that anticipated use can affect satisfaction of actual use.

The concepts and use of usability, user experience and accessibility has been debated because of their relatedness and overlap (Sauer et al., 2020). Accessibility is considered even broader than usability and user experience. It involves both the design of websites and consumer products, but also the design of buildings and transportation. Accessibility particularly emphasizes the consideration of different user groups in designing and testing of systems/products. ISO FDIS 9241-210 defines accessibility as "the usability of a product, service, environment or facility by people with the widest range of capabilities".

There are several different methods that can be used for assessing usability, user experience and accessibility. According to Sauer et al. (2020), accessibility is usually not assessed through user testing and user involvement to the same extent. Sauer et al. (2020) reports that lack of resources is the main reason why accessibility is not as frequently measured through user testing. For accessibility evaluation, expert-based approaches are used to develop checklists and algorithms for automatic checking of quantifiable parameters (Sauer et al., 2020). However, there are examples of successful studies assessing accessibility through user involvement, and Sauer et al. (2020) emphasizes the importance of user involvement for accessibility assessment.

Methods used to measure usability and user experience involve heuristic evaluations, cognitive walkthrough, interviews and focus groups, think aloud and questionnaires/scales (Keenan et al., 2022; Wronikowska et al., 2021). These methods yield different information and have both advantages and disadvantages. The direct observation methods such as cognitive walkthrough, think aloud and interviews provide in-depth information about the user perceptions and can identify severe and specific problems. However, these are time-consuming and ideally involve human factors professionals conducting the evaluation. On the other hand, questionnaires and scales are cost-effective, easy to use and easily accessible. Scales have also been developed to assess feasibility, appropriateness, and acceptability in implementation science (Weiner et al., 2017). The downside of using questionnaires is that they provide less specific information about the usability issues.

2.4.3 Usability and the Technology Acceptance Model

The Technology Acceptance Model (TAM), proposed by Davis in 1989 (Davis, 1989), has been used to predict and explain usage of technology-based systems/products by assessing perceived ease of use and perceived usefulness. Originally, TAM assessed perceived usefulness and perceived ease of use (PEU). Additional constructs were added in TAM 2 and TAM 3 to improve predictive ability and desire to explain adoption, respectively, including for example self-efficacy, intrinsic motivation, perceived enjoyment and computer anxiety (J. R. Lewis & Sauro, 2021). It has been proposed that the underlying construct of perceived usability is the same as one of the components in TAM, namely PEU, indicating that perceived usability may have the ability to predict future use of a product/system (J. R. Lewis, 2018b). Lewis & Sauro (2021) also reports that there is evidence showing that the System Usability Scale (SUS), measuring

perceived usability, and the PEU component of the TAM is in fact measuring the same thing.

Health Information Technology Acceptance Model (HITAM) is an extended version of TAM3 developed to describe health consumers' behavioral intentions (Kim et al., 2012). The HITAM was developed based on the Health Belief Model (Rosenstock, 1974), TAM3 and the Theory of Planned Behavior (TPB) (Ajzen, 1985). Items measuring perceived susceptibility, perceived seriousness, perceived threat, and behavioral intention were based on the Health Belief Model. Items measuring HIT reliability, perceived ease of use, HIT self-efficacy, perceived usefulness, attitude, and behavioral intention were derived from TAM3. Finally, items measuring health belief and concerns, subjective norm, attitude, and behavioral intention were derived from the TPB.

2.4.4 The Theoretical Framework of Acceptability (TFA)

The Theoretical Framework of Acceptability (TFA) was developed by Sekhon et al. (2017) and is concerned with the acceptance of an intervention. TFA reflects the extent to which people that deliver or receive an intervention consider it to be appropriate. It is a multi-faceted framework consisting of seven constructs: Affective Attitude, Burden, Ethicality, Intervention Coherence, Opportunity Costs, Perceived Effectiveness, and Self-Efficacy. Several of these constructs seem to be related to the constructs in usability testing, TAM and behavioral theories such as the Self-Determination Theory (SDT) and the Theory of Planned Behavior (TPB) which is presented below. For instance, perceived effectiveness and burden in TFA align with effectiveness and efficiency, respectively, in usability. Similarly, affective attitude and ethicality align with attitude and subjective norms in the TPB, whereas self-efficacy or confidence is included in both TFA, TAM, TPB and SDT. In contrast to usability, TAM and SDT, the TFA does not seem to emphasize satisfaction/enjoyment, which is important and related to intrinsic motivation.

2.4.5 Behavioral theories affecting use

There are two common theoretical frameworks within behavioral science that can affect people's intention and motivation to adopt and use a system/product. The first one is Self-Determination Theory (SDT), which proposes that behavior (e.g. implementing an exercise intervention) is affected and predicted by three basic needs; autonomy, competence and relatedness (Ryan et al., 2019). According to the SDT continuum, a person that is highly self-determined is considered to have high levels of intrinsic motivation (Ryan & Deci, 2000). Intrinsic motivation involves motivation that is regulated by and reflected in the person's excitement, enjoyment, and interest, whereas extrinsic motivation involves motivation regulated by pressure, avoidance of punishment, reward, and external demands. Intrinsic motivation is considered a positive predictor of (intention of) behavior change (Chan & Hagger, 2012). Previous literature has found that intrinsic motivation is positively correlated with perceived usability and the three basic needs from SDT (Brühlmann et al., 2018). Combining the three needs has been referred to as need satisfaction, and it seems likely that need satisfaction is related to the satisfaction component of usability, which is defined as "the extent to which the user's

physical, cognitive and emotional responses that result from the use of a system, product or service meet the user's needs and expectations" (International Organization of Standardization, 2018).

Enjoyment is considered the self-report measure of intrinsic motivation (Ryan & Deci, 2000). Enjoyment has been described as a critical factor for adherence to physical activity and exercise (Chen et al., 2020; Chen et al., 2017; Dishman et al., 2005; Jekauc et al., 2015), and a few scales has been developed to capture enjoyment in physical activity (e.g., the physical activity enjoyment scale and the enjoy scale) (C. Chen et al., 2021; Davidson et al., 2023; Kendzierski & DeCarlo, 1991).

The second theory is the Theory of Planned Behavior (TPB) which proposes that people's engagement in a behavior is influenced by their attitudes, subjective norms, and perceived behavioral control (Ajzen, 1985). In other words, what the person thinks about the behavior, whether the person finds the behavior socially appropriate and whether the person is confident in his/her ability to engage in the behavior are predicting his/her intention to perform the behavior in the future. Chan & Hagger (2012) found that intrinsic motivation from the SDT is positively associated with attitudes, subjective norms, and perceived behavioral control from the TPB, and that the three TPB variables positively predict intention to perform injury rehabilitation and prevention.

There have been scales and questionnaires developed based on SDT and TPB with promising psychometric evidence (e.g., The Behavioral Regulation in Sport Questionnaire, Intrinsic Motivation Inventory, User Motivation Inventory, Treatment Self-Regulation Questionnaire, Theory of Planned Behaviour Measures for Doing Adequate Physical Activity etc.) (Brühlmann et al., 2018; Chan et al., 2020; Chan & Hagger, 2012; Lonsdale et al., 2008; McAuley et al., 1989). There seems to be some overlap between items used in these questionnaires and the items seen in usability questionnaires. For example, enjoyment can be related to satisfaction, whereas attitudes and confidence can be related to perceived effectiveness.

2.4.6 Implementation outcome measures

As mentioned above, the goal of both usability and implementation science are to increase the likelihood of system use. Within implementation science, implementation outcome measures have been developed and psychometrically assessed for this purpose (Weiner et al., 2017). Acceptability, feasibility and appropriateness are proposed to be critical in the adoption phase, and associated with the adoption of an intervention (Proctor et al., 2011; Weiner et al., 2017). Proctor et al. (2011) has proposed the following definitions for acceptability, appropriateness, and feasibility. Acceptability is defined as the perception that a given treatment or practice is agreeable, palatable, or satisfactory. Appropriateness is the perceived fit, relevance, or compatibility of the intervention for a given practice setting and/or its perceived fit to address a particular issue or problem. Feasibility is defined as the extent to which the intervention can be successfully carried out within a given setting. The Acceptability of Intervention Measure (AIM), the

Intervention Appropriateness Measure (IAM), and the Feasibility of Intervention Measure (FIM) are three scales that aim to measure implementation outcomes. They have all been psychometrically tested and found to have sufficient validity evidence, in addition to being pragmatic (e.g., short, general and freely available) (Weiner et al., 2017). Factor analyses reveal that each item has a factor loading of above 0.74. The three-factor model revealed a good fit (CFI = 0.96 and RMSEA = 0.08 [CI, 0.06–0.09]). However, it is worth noting that the three factors are relatively highly correlated, especially appropriateness and feasibility and acceptability and appropriateness (0.66 and 0.77, respectively). Testing these scales with a bifactor model may help elucidate if these are in fact different constructs or mostly measuring a higher order construct (e.g. usability). It is also worth noting that the AIM measure includes a satisfaction item (e.g. "I like this intervention") even though satisfaction does not seem to be included in the TFA.

2.4.7 Existing measures from related concepts and theories

Table 2.4 displays scales and questionnaires developed based on concepts related to usability.

Table 2.4 Scales from theories and concepts related to usability.

Scale/Questionnaire	**Theory/Model**	**Constructs**	**Items**
Acceptability of Intervention Measure (AIM)	Implementation Science	Acceptability	This intervention meets my approval This intervention is appealing I like this intervention I welcome the use of this intervention
Intervention Appropriateness Measure (IAM)	Implementation Science	Appropriateness	This intervention seems fitting This intervention seems suitable This intervention seems applicable This intervention seems like a good match
Feasibility of Implementation Measure (FIM)	Implementation Science	Feasibility	This intervention seems implementable This intervention seems possible This intervention seems doable This intervention seems easy to use
Behavioral Regulation in Sports Questionnaire (BRSQ)	Self-Determination Theory	Amotivation External regulation	…but I question why I continue. …but I question why I am putting myself through this. …but the reasons why are not clear to me anymore. …but I wonder what's the point. …because people push me to play. …to satisfy people who want me to play. …because I feel pressure from other people to play.

		Introjected regulation Identified regulation Integrated regulation Intrinsic Motivation	…because if I don't, other people will not be pleased with me. …because I would feel guilty if I quit. …because I would feel ashamed if I quit. …because I feel obligated to continue. …because I would feel like a failure if I quit. …because the benefits of sport are important to me. …because I value the benefits of my sport. …because it teaches me self-discipline. …because it's a good way to learn things which could be useful to me in my life. …because it's an opportunity to just be who I am. …because it's a part of who I am. …because what I do in sport is an expression of who I am. …because it allows me to live in a way that is true to my values. …because I enjoy it. …because I like it. …because it's fun. …because I find it pleasurable.
Treatment Self-Regulation Questionnaire for Adequate Physical Activity	Self-Determination Theory	Autonomous motivation (Item 1-6) Controlled motivation (Item 7-12)	1.I want to do adequate physical activity because I feel that I want to take responsibility for my own health. 2.I want to do adequate physical activity because I personally believe it is the best thing for my health. 3.I want to do adequate physical activity because I have carefully thought about it and believe it is very important for many aspects of my life. 4.I want to do adequate physical activity because it is an important choice I really want to make. 5.I want to do adequate physical activity because it is consistent with my life goals. 6.I want to do adequate physical activity because it is very important for being as healthy as possible. 7.I want to do adequate physical activity because I would feel guilty or ashamed of myself if I did not.

		Amotivation (Item 13-15)	8.I want to do adequate physical activity because others would be upset with me if I did not. 9.I want to do adequate physical activity because I would feel bad about myself if I did not. 10.I want to do adequate physical activity because I feel pressure from others to do so. 11.I want to do adequate physical activity because I want others to approve of me. 12.I want to do adequate physical activity because I want others to see I can do it. 13.I really don't think about doing adequate physical activity. 14.I want to do adequate physical activity because it is easier to do what I am told than think about it. 15.I don't really know why I want to do adequate physical activity.
Theory of Planned Behaviour Measures for Doing Adequate Physical Activity	Theory of Planned Behavior	Intention (Item 1-3) Subjective norm (Item 4-6) Perceived behavioral control (Item 7-11) Attitude (Item 12)	1.I intend to do adequate physical activity in the forthcoming month. 2.I will try to put great effort into doing adequate physical activity in the forthcoming month. 3.I plan to do adequate physical activity in the forthcoming month. 4.Most people who are important to me think that I should do adequate physical activity in the forthcoming month. 5.It is expected of me that I do adequate physical activity in the forthcoming month. 6.The people in my life whose opinions I value would approve of me to do adequate physical activity in the forthcoming month. 7.It is possible for me to do adequate physical activity in the forthcoming month. 8.If I want to I could do adequate physical activity in the forthcoming month. 9.I have complete control over how to do adequate physical activity in the forthcoming month. 10.It is completely down to me to decide to do adequate physical activity in the forthcoming month.

			11.It is easy for me to do adequate physical activity in the forthcoming month. 12.For me doing adequate physical activity in the forthcoming month would be … • Worthless/Valuable • Harmful/Beneficial • Unpleasant/Pleasant • Unenjoyable/Enjoyable • Bad/Good
Activity Perception Questionnaire (part of the Intrinsic Motivation Inventory)	Self-Determination Theory	Interest/ enjoyment: 3, 5, 7, 11, 12(R), 15, 17, 23 Value/ usefulness: 1, 4, 6, 10, 13, 16, 19, 21, 25 Perceived choice: 2, 8(R), 9, 14(R), 18(R), 20(R), 22, 24(R)	1. I believe that doing this activity could be of some value for me. 2. I believe I had some choice about doing this activity. 3. While I was doing this activity, I was thinking about how much I enjoyed it. 4. I believe that doing this activity is useful for improved concentration. 5. This activity was fun to do. 6. I think this activity is important for my improvement. 7. I enjoyed doing this activity very much. 8. I really did not have a choice about doing this activity. 9. I did this activity because I wanted to. 10. I think this is an important activity. 11. I felt like I was enjoying the activity while I was doing it. 12. I thought this was a very boring activity. 13. It is possible that this activity could improve my studying habits. 14. I felt like I had no choice but to do this activity. 15. I thought this was a very interesting activity. 16. I am willing to do this activity again because I think it is somewhat useful. 17. I would describe this activity as very enjoyable. 18. I felt like I had to do this activity. 19. I believe doing this activity could be somewhat beneficial for me. 20. I did this activity because I had to. 21. I believe doing this activity could help me do better in school. 22. While doing this activity I felt like I had a choice.

			23. I would describe this activity as very fun. 24. I felt like it was not my own choice to do this activity. 25. I would be willing to do this activity again because it has some value for me.
Health Information Technology Acceptance Model (HITAM)	Technology Acceptance Model	Health belief and concerns Perceived susceptibility Perceived Seriousness Subjective norm HIT self-efficacy Perceived ease of use Perceived usefulness HIT reliability Attitude Behavioral intention	I am positive about using HIT to manage my health and to search for reliable health information I think it is beneficial to manage my health and search for reliable health information using HIT I am satisfied by and large with the use of HIT to manage my health and search for reliable health information using HIT. I will continue to use HIT to manage my health and to search for reliable health information I will regularly use HIT to manage my health and to search for reliable health information I will recommend use of HIT to other people to manage their health and to search for reliable health information. + 44 other items
The Adherence to Exercise for Musculoskeletal Pain Tool (ATEMPT) (Bailey et al., 2024)		Communication with expert Targets How exercise prescribed Patient knowledge and understanding Motivation and support Psychological approach and attitudes	1. I feel my individual needs were understood when the exercise therapist recommended my exercises 2. I am doing enough exercise to produce a positive change in my musculoskeletal pain 3. I am doing my exercises as instructed 4. I understand how my exercises will help with my musculoskeletal pain 5. I understand the consequences of not doing my exercises 6. I believe in the exercises that have been recommended to me
ENJOY	Self-Determination Theory	Pleasure Relatedness Competence	2. The activity was pleasurable to me 5. The activity made me feel happy 9. The activity was fun 17. I liked doing the activity 25. The activity made me feel good 4. I felt connected with others during the activity 8. I liked interacting with others during the activity

		Challenge/ Improvement Engagement	16. I cooperated with others during the activity 19. The activity was a shared effort with others 21. I felt close to others when I did the activity 6. I felt very capable during the activity 11. I am good at the activity 22. I felt like I did a good job the last time I did the activity 23. I was proficient in the activity 24. I felt competent at performing the activity 1. The activity allowed me to develop new skills 7. I felt challenged, but not over-challenged, during the activity 10. I improved my skills the last time I did the activity 15. During the activity I could get better at doing it 18. I felt challenged, but not under-challenged, during the activity 3. I lost track of what was going on outside of the activity 12. I forgot what was going on around me during the activity 13. I lost track of time during the activity 14. When I did the activity, I thought about nothing else 20. I lost track of what was going on around me during the activity
Physical Activity Enjoyment Scale		Enjoyment	I enjoy it; I hate it I feel bored; I feel interested I dislike it; I like it I find it pleasurable; I find it unpleasurable I am very absorbed in this activity; I am not at all absorbed in this activity It is not fun at all; it is a lot fun I find it energizing; I find it tiring It make me depressed; it makes me happy It is very pleasant; it is very unpleasant I feel good physically while doing it; I feel bad physically while doing it It is very invigorating; it is not at all invigorating I am very frustrated by it; I am not at all frustrated by it

			It is very gratifying; it is not at all gratifying It is very exhilarating; it is not at all exhilarating It is not at all stimulation; it is very stimulating It gives me a strong sense of accomplishment; it does not give me any sense of accomplishment It is very refreshing; it is not at all refreshing I felt as though I would rather be doing something else; I felt as though there was nothing else

2.5 Summary and implications

Based on the previous literature discussed, integrating usability and implementation sciences may be an appropriate way to assess and evaluate exercise program characteristics. However, many usability scales lack a clear description of what construct they are aiming at measuring and their evidence of validity and reliability. Most items in the identified scales also seem irrelevant to assess in an exercise intervention context. The System Usability Scale or the Intervention Usability Scale seem to be more generic scales with good psychometric properties and might be relevant to involve when assessing usability of exercise interventions. However, these scales do not seem to address satisfaction/enjoyability, which might be important for adherence to exercise programs. In addition, the implementation outcome scales/items have not been blended/tested with the typical usability items which aim to measure a similar construct of "use likelihood." This makes it challenging to navigate and select a comprehensive and appropriate scale.

The goal of usability testing is to assess whether a system is likely to be used and identify aspects that are not working sufficiently. Usability testing emphasizes user involvement and should be utilized throughout the design, development, and implementation process of an exercise intervention to increase the likelihood of sustained use. Developing a new comprehensive scale with sufficient psychometric properties for assessing the usability of exercise programs can be a cost-effective and efficient tool to help design, develop and evaluate program characteristics and implementation likelihood. The scale will potentially help bridge the gap between controlled research settings and implementation of exercise interventions in the wild with the goal to increase uptake and adherence.

3 Usability Considerations for Enhancing Exercise-Based Injury Prevention Interventions

3.1 Abstract

Our study centers around the importance of usability and usability testing in the context of exercise interventions. Despite the proven effectiveness of exercise programs in reducing injuries, adherence to these programs remains a challenge, and injury rates have not decreased significantly. This study explores the application of usability testing in the field of health and exercise medicine, specifically assessing the usability of an exercise-based injury prevention intervention. A team of female handball players was included in the study (n = 23). The study employed a modified usability questionnaire, including items from the System Usability Scale (SUS), to evaluate important program characteristics (e.g., learnability, enjoyability, perceived effectiveness, ease of use, and efficiency). The Likert scale-based questionnaire was distributed within two sessions into the intervention period and post intervention. Perceived effectiveness, enjoyability, and efficiency highly correlated with players' intention to use the program (r_s 0.50, p = 0.02, r_s 0.50, p= 0.02, r_s 0.65, p < 0.001, respectively), whereas learnability and ease of use did not. Hence, the SUS items were not correlated with intention to use the program. Program usability, especially enjoyability, significantly decreased over time (Cohen's d = 0.60, 95 % CI 0.13, 1.06, p = 0.01, and Cohen's d = 0.61, 95 % CI 0.18, 1.10, p = 0.01, respectively). Our findings underscore the potential significance of considering program characteristics and usability testing throughout the development and testing of exercise programs. Moreover, this study highlights the shortcomings of current scales such as SUS and advocates for the creation of a tailored usability assessment tool for exercise interventions.

Keywords. Usability; Usability testing; Injury prevention; System Usability Scale; Anterior cruciate ligament injury; User experience

Footnote: The material contained in this chapter is under review as a conference proceeding at *HFES International Symposium on Human Factors and Ergonomics in Health Care.*

3.2 Introduction

Beyond developing exercise interventions that are effective in improving health or reducing risk of injury and illnesses, it is imperative to develop something the target users would want to use. Usability testing and end-user engagement has become an important part of product development, particularly within technology and computer-based products, due to the ability to detect barriers and enhance user experience. Despite its increasing emphasis during product development, usability testing is not typically utilized in exercise intervention development. The International Organization of Standardization (ISO) defines usability as the 'extent to which a system, product, or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use' (International Organization of Standardization, 2018). Implicitly, end-user involvement is critical in usability assessments. Utilizing usability testing throughout the development and testing phase of an exercise intervention can potentially help mitigate the adherence issue seen in exercise medicine and injury prevention implementation research.

For instance, anterior cruciate ligament (ACL) injuries remain a large problem within the population of female athletes performing pivoting sports (e.g., soccer, handball, basketball, netball). While previous research has established the effectiveness of neuromuscular injury prevention programs (IPP) in reducing ACL injury risk (Foss et al., 2018; Halvorsen et al., 2023; Myklebust et al., 2003; E. J. Petushek et al., 2018; Waldén et al., 2012), adherence to these IPPs is very low (~4-20 %) (Joy et al., 2013; Norcross et al., 2016; Sugimoto et al., 2016) and we have yet to see this effect resulting in fewer ACL injuries among this population outside a research setting.

Several studies have been published on why evidence-based interventions are not being implemented as desired outside research settings, and how facilitators and barriers can be targeted to increase implementation rates of IPPs in sports (Ageberg et al., 2019, 2022; Donaldson et al., 2019; Finch, 2011; Hawkinson et al., 2022; McKay et al., 2016; Minnig et al., 2022; Moesch et al., 2022; O'Brien & Finch, 2016; Richmond et al., 2020). These studies have identified program design, exercises, and learnability as facilitators, whereas lack of knowledge about the program, lack of player enjoyability, and lack of resources available have been identified as barriers for program use. Several of these facilitators and barriers can be identified and addressed through usability assessments.

Based on recent findings related to ACL injury risk factors in female handball (Bill et al., 2022), we developed a novel ACL injury prevention intervention targeting reduction of knee abduction moment (KAM) during cutting tasks. In addition to exploring the programs' effectiveness in reducing KAM, a concurrent evaluation of usability characteristics such as the players' perceived effectiveness, efficiency and satisfaction with the program is important and efficient to conduct for exploring implementation likelihood.

The aim of this study was therefore to explore the relationship between perceived usability characteristics and players' intention to use the newly developed ACL IPP on a regular basis. Additionally, the study sought to elucidate whether perceived program usability characteristics underwent changes over time. This research will help inform IPP developers of exercise-based interventions about important program characteristics affecting sustained use and effectiveness, and how usability assessment of exercise-based interventions may be different from more traditional approaches (e.g. System Usability Scale).

3.3 Material and Methods

Data on usability-related factors affecting the target user's intention to use the program were collected from the intervention group participating in a cluster-randomized controlled trial (RCT) conducted by Müller-Kühnle et al. (NCT05643261, unpublished manuscript) during the Fall of 2022. The main purpose of the RCT was to investigate whether a novel ACL IPP could reduce KAM in female handball players' cutting performance. During and after the intervention period, data on usability of the new IPP and barriers and facilitators to further use were collected.

3.3.1 Sample

Twenty-three female handball players, aged 15 to 18, were included in the intervention group. All players went to the same elite sports high school which hosted the handball sessions where the injury prevention intervention was conducted. All players additionally played handball for different clubs competing in various divisions of the Norwegian handball leagues, including the elite, 1st, 2nd, and 3rd division. The ethical aspects of the study adhered to the principles outlined in the Declaration of Helsinki, and prior approval was obtained from the regional ethics committee at the Norwegian School of Sports Sciences. All participants provided informed consent.

3.3.2 Intervention

The intervention involved performing an exercise program including both a warm-up with cutting and landing exercises, and strength exercises for hip, core, and calves. The program incorporated progression for training volume, exercise load, complexity, and specificity. They were instructed to perform the IPP at least two times per week for 9 weeks. The IPP was supervised by the team's coaches and mainly performed during their handball session. The exercises were either performed in the gym or on the handball court, depending on what their training schedule looked like. The whole program, including warm-up with cutting and landing tasks and strength exercises, was designed to take about 20 minutes to complete. To increase coach and player comprehension and motivation, several measures were taken: 1) One of the main coaches was involved in designing the intervention, 2) A short and focused presentation about mechanisms for why each of the components were likely to reduce injury risk and improve performance were held for both players and coaches, and 3) During pretest, the players individually

received feedback on cutting technique. During this session, the rationale for why these technique modifications were beneficial was repeated.

3.3.3 Outcome measures

The primary outcome measure of the RCT was changes in KAM, whereas the main outcome measure in this study was the players' intention to use the program. Intention to use was measured using the Likert scale item "I think I will want to use the training program regularly", adapted from the System Usability Scale.

Surveys were distributed to the players in the intervention group after two sessions and after the intervention period was over. The PRE survey was distributed using pen and paper and the POST survey was distributed through Qualtrics (Qualtrics, Provo, UT, USA). The different survey distribution methods were due to the environment and timing in which the players received the survey. For the PRE survey they were at handball training with the research team present, whereas the POST survey was provided outside training time and after the intervention period was over, making it more efficient to collect this data through an online survey. The two usability surveys included 13 items that all were responded to through a five-point Likert scale. Six of the items were collected from the modified version of the System Usability Scale (Brooke, 1995) - the Intervention Usability Scale (IUS) (Lyon et al., 2021). Additionally, seven items that mapped more directly onto the player's enjoyment, perceived effectiveness and efficiency of the program were given, consistent with findings from previous literature on program adherence, and behavioral theories (Ajzen, 1985; Ryan & Deci, 2000). Hypothesized constructs of the usability scale included learnability, perceived overall effectiveness, ease of use, enjoyability and efficiency. See Figure 3.1 for the full list of usability scale items and their affiliation to the hypothesized constructs.

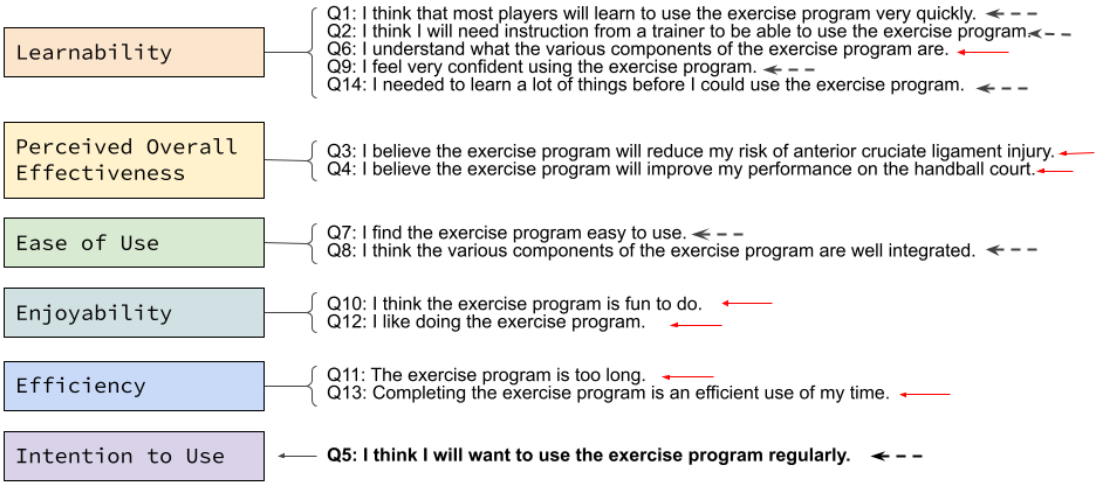


Figure 3.1. Usability scale items and affiliation to subcomponents. Note: Red arrows indicate items added by the research team. Black, dotted arrows indicate items from the System Usability Scale.

3.3.4 Statistical analysis

Power analysis and sample size justification was made for the main objective of the intervention study (e.g. changes in KAM), but not for the exploratory nature of the usability assessment. Paired sample's t-test were used to assess differences between pre and post responses on the usability scale items. Spearman's rho (r_s) was used to assess correlation between the hypothetical constructs and the outcome measure intention to use.

3.4 Results

Twenty-one players were included in the analyses. For demographic data of the players, see Table 3.1. Two players were excluded from survey data analyses, one due to early intervention exit due to injury and one due to aberrant responses.

Table 3.1 Demographics

Players	Age (years) mean (sd)	Body Mass (kg) mean (sd)	Height (m) mean (sd)	Level of Play	Number of Technique Sessions mean (sd)	Number of Strength Sessions mean (sd)
n=21	16.71 (1.10)	69.33 (8.20)	1.75 (0.06)	Level 1: 3 Level 2: 10 Junior: 2 Girls 2006: 6	17.25 (3.27)	18.25 (3.70)

3.4.1 Correlations between usability characteristics and intention to use

The total scale score was significantly correlated with intention to use the program (r_s 0.50, 95% CI 0.09, 0.77, p = 0.02). When assessing the correlation with the five hypothetical constructs, perceived overall effectiveness, enjoyability and efficiency were significantly correlated with intention to use the program (See Figure 2). The correlation between intention to use and their perceived effectiveness on performance was stronger than the correlation between intention to use and perceived effectiveness in reducing injury risk (r_s = 0.63, 95% CI 0.28, 0.84, p < 0.001 and r_s 0.53, 95% CI 0.12, 0.78, p = 0.01, respectively). There were no significant relationships between learnability or ease of use and intention to use the program.

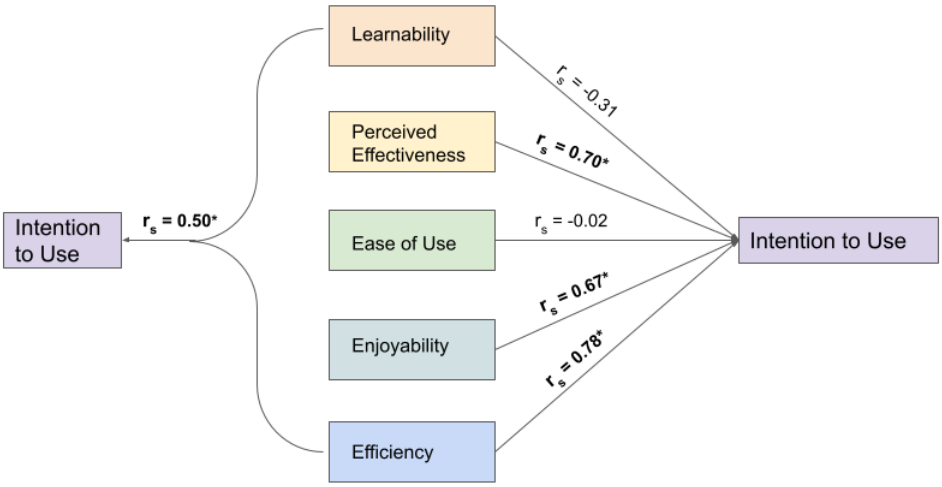


Figure 3.2. Correlation between usability characteristics and intention to use/implementation likelihood. The left side shows the correlations between the total scale score and intention to use. The right side shows the correlation between different usability characteristics and intention to use. Values in bold indicate significant correlations.

3.4.2 Changes in usability scores from pre to post intervention

The paired sample's t-test showed that the total usability scale score significantly decreased from pre and post intervention with enjoyability being the subcomponent that changed the most. See Table 3.2.

Table 3.2. Differences in pre and post intervention scale scores (n=21)

Constructs (items)	**Pre** (mean)	**Post** (mean)	**Mean Difference** [95% CI]	**Cohen's *d*** [95% CI]	**p-value**
Total Scale Score (13)	3.602	3.328	0.274 [0.07, 0.48]	**0.60** [0.13, 1.06]	**0.012**
Learnability (5)	3.924	3.733	0.19 [-0.05, 0.43]	0.36 [-0.09, 0.79]	0.118
Enjoyability (2)	3.0	2.405	0.595 [0.15, 1.03]	**0.61** [0.18, 1.10]	**0.011**
Efficiency (2)	3.071	2.619	0.452 [-0.10, 1.01]	0.37 [-0.08, 0.81]	0.106
Perceived Effectiveness on Injury Risk (1)	4.238	4.238	0 [-0.41, 0.41]	0 [-0.43, 0.43]	1

Perceived Effectiveness on Performance (1)	3.286	3.571	-0.286 [-0.79, 0.22]	-0.26 [-0.69, 0.18]	0.249
Ease of Use (2)	3.952	3.571	0.381 [0.03, 0.73]	**0.49** [0.03-0.94]	**0.035**

Players responded to the items using a Likert scale ranging from 1-5 (Strongly disagree-Strongly agree). Pre and post values refer to the mean value of the items belonging to the specific construct.

3.5 Discussion

The survey data revealed a significant correlation between the usability characteristics "perceived overall effectiveness", "efficiency", and "enjoyability", and the players' intention to use the program. This indicates that players' willingness to use the program regularly was affected by whether they believed the program would prevent injuries and/or increase their performance, whether they found it efficient and worth their time, and whether they enjoyed doing it.

Enjoyability and perceived overall effectiveness' positive impact on the players' intention to use the program is supported by previous literature on IPPs (Moesch et al., 2022; Møller et al., 2021) and behavioral theories such as the Self-Determination Theory and the Theory of Planned Behavior (Ajzen, 1985; Ryan & Deci, 2000). Enjoyment is considered the self-report measure of intrinsic motivation (Ryan & Deci, 2000), and intrinsic motivation has been suggested to predict athletes' intentions to perform injury prevention and rehabilitation (Chan & Hagger, 2012). This suggests that the development of IPPs should emphasize and assess the players' intrinsic motivation for performing the program. As shown in our study, enjoyability and perceived effectiveness levels can change over time and should therefore be monitored and addressed regularly.

Our results suggest that this intervention was easy to use and learn initially and was perceived effective in reducing ACL injury risk with scores above 3.9 (out of 5), whereas it was perceived less enjoyable, efficient, and effective in improving performance (3.0, 3.1, and 3.3, respectively). Interestingly, learnability and ease of use, which included items from the System Usability Scale (SUS)/Intervention Usability Scale (IUS), did not significantly correlate with the players' intention to use the program. The SUS/IUS is a highly acknowledged and widely used measure to assess usability of a broad range of systems/products (Brooke, 2013). These scales have been psychometrically tested for use in other healthcare interventions, such as internet-based interventions for depression (Mol et al., 2020) and Motivational Interviewing (Lyon et al., 2021). The fact that none of the items included from this scale correlated with the players' intention to use the program, suggests that the use of this generic scale is limited and does not seem appropriate for assessing usability of exercise interventions when the goal is to assess whether a program is likely to be used or not.

This study, with support from prior research and theoretical frameworks such as the Self-Determination Theory (SDT) and Theory of Planned Behavior (TPB), suggests that enjoyability, efficiency, and perceived effectiveness seem important to target when designing sustainable injury prevention programs. These usability components have not been particularly emphasized in other existing usability measures, where a product's ease of use and learnability have been of higher priority (e.g., IUS, Technology Acceptance Model, Usability Magnitude Estimation, Single Ease Question). Other frameworks and scales related to usability assessments of exercise interventions include the Theoretical Framework of Acceptability, implementation outcome measures (e.g., The Acceptability of Intervention Measure, the Intervention Appropriateness Measure, and the Feasibility of Intervention Measure), Behavioral Regulation in Sports Questionnaire, Activity Perception Questionnaire, Theory of Planned Behavior Measures for Doing Adequate Physical Activity, and Treatment Self-Regulation Questionnaire for Adequate Physical Activity (Chan et al., 2020; Lonsdale et al., 2008; McAuley et al., 1989; Sekhon et al., 2017; Weiner et al., 2017; Wilson et al., 2006). These scales all seem to measure a similar construct of "use likelihood" as the typical usability items and can provide valuable insights into constructs that are important to measure when assessing usability of an exercise program. However, none of the scales seem to assess the usability constructs effectiveness, efficiency, and satisfaction as defined by the ISO.

The domain of Human Factors has successfully utilized usability testing and human-centered design in product development. The use of usability testing and end-user involvement during program development and testing, seems to be valuable within the domain of exercise science as well, where the development of an exercise intervention usability scale with sufficient psychometric properties that to a greater extent captures the users' enjoyment, perceived efficiency and effectiveness of an exercise intervention might help bridge the gap between compliance in controlled research settings and long term adherence "in the wild".

3.5.1 Limitations

Our results are based on survey measures, which may affect the validity. For example, survey measures can be subject to careless responding or differences in interpretation between researchers and players or between players. Unfortunately, we did not collect timing data for the different survey completions, which is considered one of the most reliable methods to detect careless responders (Leiner, 2019; Wang et al., 2018). However, "longstring" methods was used to detect aberrant/careless responders. Secondly, a notable limitation was the relatively small sample size, which may have caused limited variability in responses and the precision of the correlation estimates. Finally, in this study we measured intention to future use. This is not equivalent to actual sustained use, which is the ideal outcome of implementing IPPs. However, intention to use has shown to be significantly associated with actual behavior (Chen et al., 2022; Fishman et al., 2020) and is considered the most proximal indicator of future behavior (Ajzen, 1985; Chen et al., 2022).

3.6 Conclusion

The willingness of players to engage in the IPP appears to depend on their enjoyment of it, its perceived efficiency, and their belief in its effectiveness for reducing ACL injury risk or enhancing performance. The study underscores the value of usability assessments – beyond typical constructs measured within the SUS - in evaluating the likelihood of program implementation and in identifying barriers to program use, ultimately contributing to the design of more effective and user-friendly exercise programs.

3.7 Acknowledgements

The study has received funding grants through a Health Research Institute Fellowship and a Finishing Fellowship from the Graduate School at Michigan Technological University.

4 Perspectives on Usability and Adoption of a New ACL Injury Prevention Program for Female Handball Players: A Mixed Methods Approach

4.1 Abstract

Effective injury prevention programs (IPPs) hold promise for mitigating anterior cruciate ligament (ACL) injuries in female handball players, yet adherence remains a challenge. This mixed methods study explores the relationship between program usability characteristics and potential effectiveness of a newly developed ACL IPP through survey and interview data from 23 female handball players aged 15-18 and their four coaches. Findings reveal significant correlations between reductions in knee abduction moments (KAM) and players' perceived effectiveness of the program in reducing ACL injury risk and their intention to use it (r_s -0.52, 95% CI -0.78, -0.1, p = 0.02, and r_s -0.46, 95% CI -0.75, 0.03, p = 0.04, respectively). Coaches and players identified efficiency and perceived effectiveness as key factors influencing program adoption. Concerns regarding program length and doubts about exercise efficacy and transferability emerged as barriers, while perceived effectiveness, efficiency, and enjoyability were facilitators for program use. Players' suggestions for program improvement included shortening duration and incorporating playful elements. Design thinking sessions yielded strategies to optimize time efficiency and integrate IPPs into existing training routines. The study underscores the importance of balancing program effectiveness with practicality and clear communication about its purpose and benefits to facilitate program adoption. Enhancing program usability and involving end-users in the design process are crucial steps toward promoting acceptance and adoption of IPPs among target populations.

Keywords: knee abduction moment, anterior cruciate ligament, injury prevention, handball, implementation outcomes

Footnote: The material contained in this chapter has been submitted to the *BMJ Open Sport & Exercise Medicine*.

What is already known on this topic

- Injury prevention programs (IPP) can reduce risk of sustaining an anterior cruciate ligament injury by more than 50 % in female athletes participating in pivoting sports. However, adherence to these programs is low, and the number of ACL injuries in female athletes is not decreasing. Several barriers and facilitators to implementation of IPPs have been suggested, including program design, player enjoyment, lack of knowledge and resources.

What this study adds

- Our results show that perceived effectiveness, efficiency, flexibility and enjoyability of using the program affects coaches and players' adoption and program acceptance.

- If players believe in the programs' effectiveness, it does not have to be perceived as enjoyable for it to reduce KAM.

- IPPs have a great potential to be made more enjoyable.

- Understanding of the rationale and potential value of IPPs should be heavily emphasized in communication between program developers and target users (e.g., between researcher and coaches/players and between coaches and players).

How this study might affect research, practice, or policy

- Our results indicate that collecting data on the coach and player perceptions on usability of exercise programs can add important insights during the development and testing phase of the program. Including end-users through program usability assessments should be emphasized early on in development of new injury prevention programs.

4.2 Introduction

Exercise interventions, widely acknowledged for their efficacy in preventing and treating various diseases and injuries, face a significant challenge in terms of adherence, thereby limiting their potential benefits. This issue is particularly evident in the context of anterior cruciate ligament (ACL) injuries among female athletes engaged in pivoting sports like soccer and handball (Gornitzky et al., 2016; Myklebust et al., 2003; Prodromos et al., 2007). Despite the proven effectiveness of injury prevention programs (IPP) in reducing ACL injury risk, adherence rates remain notably low (approximately 4-20%) (Joy et al., 2013; Norcross et al., 2016; Sugimoto et al., 2016).

The adherence issue is multifaceted, and a large number of different facilitators and barriers have been identified in previous research (Ageberg et al., 2022; Donaldson et al., 2019; Finch, 2011; Joy et al., 2013; McKay et al., 2014, 2016; Moesch et al., 2022; Møller et al., 2021; O'Brien & Finch, 2016). Knowledge, resources (e.g. time) and player enjoyability have been considered barriers to use, whereas program content/design, relevance of exercises and the ease of learning have been identified as facilitators. Additionally, behavioral theories such as the Self-Determination Theory and the Theory of Planned Behavior propose factors affecting individuals' engagement in a behavior (change) (e.g., intrinsic motivation, attitudes, subjective norms, and perceived behavioral control), and how these predict intention to perform injury prevention and rehabilitation (Ajzen, 1985; D. K. C. Chan & Hagger, 2012; Ryan & Deci, 2000).

Recently, Bill et al. (2022) found that female handball players with high knee abduction moments (KAM), a likely risk factor for ACL injuries (Boden et al., 2009; Koga et al., 2010; Krosshaug et al., 2007), show increased vertical center of mass excursions and knee valgus angles across cutting tasks. Based on these findings and other biomechanical analyses (Dempsey et al., 2007; Kristianslund et al., 2014), we developed and tested a new ACL IPP for female handball players specifically targeting cutting technique and hip abductors, hip external rotator, calf- and core muscles to reduce KAM. In addition to exploring the programs' effectiveness in reducing KAM, it is important and efficient to conduct a concurrent evaluation of usability characteristics early in the development and testing phase of new IPPs. This evaluation should focus on aspects such as the users' perceived effectiveness, efficiency, and satisfaction with the program, and aim to identify potential barriers to program use. This early evaluation is crucial because these programs will have very limited value if they are not being adopted outside research settings.

Usability testing and end-user engagement has become an important part of product development, particularly within technology and computer-based products, due to their ability to detect barriers and enhance user experience. Despite its increasing emphasis during product development, usability testing is not typically utilized in exercise intervention development. The International Organization of Standardization (ISO) defines usability as the 'extent to which a system, product, or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use' (ISO/TS 20282-2:2013, 2023). Other outcomes of use in the

context of usability includes user experience, accessibility, and avoidance of harm (International Organization of Standardization, 2018). Implicitly, end-user involvement is critical in usability assessments. Utilizing usability testing throughout the development and testing phase of an exercise intervention can potentially help mitigate the adherence issue seen in injury prevention implementation research.

The aim of this mixed methods study was to gain deeper insights into coaches and players' perceived barriers and facilitators for use of the newly developed ACL injury prevention program. Additionally, we explored the relationship between player's reduction in KAM and their perceptions of, and intention to use, the program. This research will help inform program developers about important program characteristics affecting use and the potential effectiveness of the program.

4.3 Methods and Materials

This mixed methods study included survey and interview data collected from the intervention group participating in a randomized controlled trial conducted by researchers at the Oslo Sports Trauma and Research Center during the Fall of 2022 (NCT05643261, Müller-Kühnle et al., Unpublished data) investigating whether a program particularly targeting cutting technique and muscle strength exercises for hip, calf and core could reduce KAM in female handball players. The ethical aspects of the study adhered to the principles outlined in the Declaration of Helsinki, and prior approval was obtained from the regional ethics committee of the Norwegian School of Sport Sciences (233–160622). All participants provided informed consent.

4.3.1 Sample

The intervention group consisted of 23 female handball players aged 15-18 years old. The players all went to the same high school where the intervention was executed. All players conducted the surveys. Six of the players were also included in a focus group/design thinking activity. The four coaches responsible for the execution of the IPP individually took part in semi-structured interviews with one female interviewer (AIM).

4.3.2 Intervention

The 9-week intervention involved performing a warm-up with cutting and landing exercises, and strength exercises for hip, core, and calves 2-3 times per week (NCT05643261, Müller-Kühnle et al., Unpublished data). The team's coaches supervised the program, and it was mainly performed during their handball session, in the gym or on the court. The whole program (e.g., warm-up and strength exercises) was designed to take about 20 minutes.

4.3.3 Outcome measures

The primary quantitative outcome measure in the original RCT was the players' reduction in KAM. KAM was evaluated using a 3D motion capture system (24 cameras, Qualisys, Gothenburg, Sweden, 200 Hz) and two floor-embedded force plates (AMTI,

Watertown, Massachusetts, USA,1,000 Hz, 1,200 × 600 mm) as players executed a sidestep cutting maneuver to bypass a stationary defender (NCT05643261, Müller-Kühnle et al., Unpublished data). See Figure 4.1 for a flow chart of assessing outcome measures.

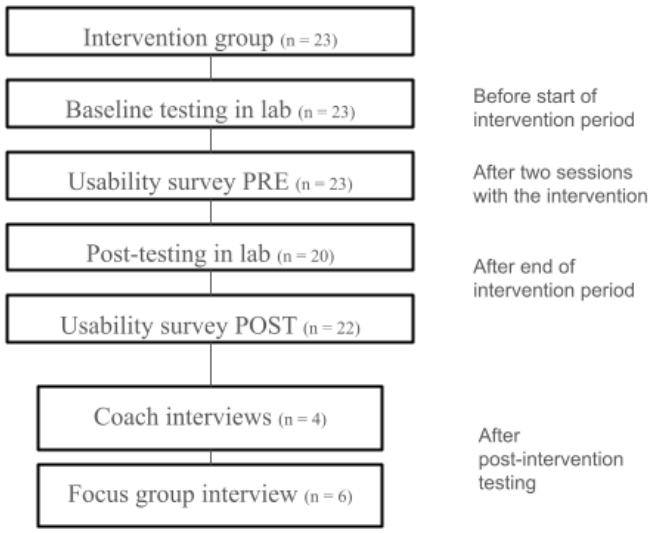


Figure 4.1. Flow chart of lab testing, survey distribution and interviews

Surveys were distributed to the players in the intervention group after two sessions and after the intervention period was over. The PRE survey was distributed using pen and paper and the POST survey was distributed through Qualtrics (Qualtrics, Provo, UT, USA). The two usability surveys included 13 items that all were responded to through a five-point Likert scale. Six of the items were collected from the modified version of the Intervention Usability Scale (IUS) (Lyon et al., 2021). Additionally, seven items that mapped more directly onto the player's enjoyment, perceived effectiveness and efficiency of the program were given, consistent with findings from previous literature on program adherence, and behavioral theories. Hypothesized constructs of the usability scale included learnability, perceived overall effectiveness, ease of use, enjoyability and efficiency. See Figure 4.2 for the full list of usability scale items and their affiliation to the hypothesized constructs. Surveys are included in the appendices.

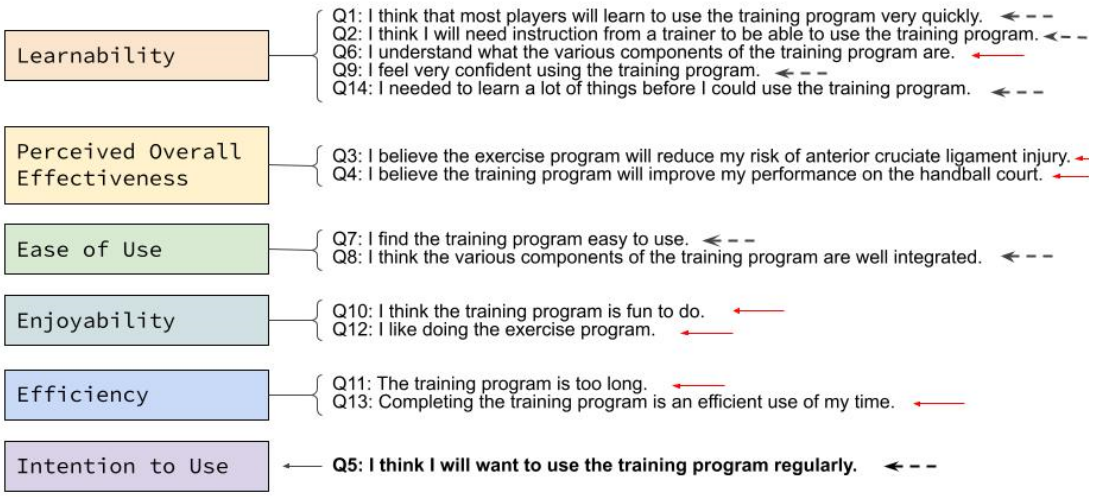


Figure 4.2. Usability scale items and affiliation to subcomponents. Note: Red arrows indicate items added by the research team. Black, dotted arrows indicate items from the System Usability Scale.

Semi-structured interviews were conducted individually with four of the coaches in the intervention group responsible for the execution of the program, each lasting between 25-30 minutes. They were invited to the interviews via emails, and all consented to participate. Additionally, six of the players participated in a 60 minutes, semi-structured focus group interview. They were selected and invited by the head coach based on various age and experience levels. All interviews were conducted by AIM (physiotherapist and PhD student), who had previous experience with conducting semi-structured interviews. The interviewees did not know of the interviewer in advance, but they received information about her role in the study and rationale for the interviews before interviews started. All interviews were conducted in-person at the Norwegian School of Sport Sciences with only the participants and the interviewer present. Both the coaches and the group of players were asked questions about what they liked/disliked about the program, what they would change about the program and general feedback about facilitators and barriers for implementation and use of injury prevention programs. The results from the lab testing and survey data were not presented to the coaches or players prior to the interviews. Sound recordings from interviews were transcribed and thematically synthesized subsequently. Interview guides (See appendix 8.3) were not pilot tested, and no repeat interviews were carried out. Data saturation was discussed between co-authors. Transcripts were not returned to interview subjects for further comments/corrections primarily due to logistical constraints and time limitations and given the straightforward nature of the interview process.

As part of the focus group interview, the six players were also involved in a design thinking activity (Plattner, 2016) used to pull out ideas on how to design an exercise-based injury prevention program players would like to utilize long term. The design

thinking method contains 5 different stages: 1) Empathize (what is the users' needs), 2) Define (state the users' needs and problems), 3) Ideate (challenge assumptions and create ideas), 4) Prototype (create your solution), and 5) Test (try out your solution) (Plattner, 2016). For this interview, the problem was already defined and stated to the players: 1) How can we, as researchers, coaches or others, help encourage players to perform IPT frequently, 2) How can this program become more enjoyable? and 3) Do you have any suggestions for how you can integrate IPT as part of your weekly training schedule?. They were given one problem at a time and asked to challenge assumptions and create ideas (stage 3) and create a solution (stage 4). All interviews were audio recorded.

4.3.4 Patient and public involvement

The players were not involved in the design, conduct, reporting, or dissemination plans of this research. However, the coaches from both the intervention and control group were involved in the planning and design of the intervention prior to baseline testing. Secondly, one of the coaches was invited to a meeting with TK to work out the intervention details. Prior to intervention start, the players took part in a lecture on study and exercise selection rationale.

4.3.5 Statistical analysis

Power analysis and sample size justification was made for the main objective of the randomized controlled trial (e.g. changes in KAM), but not for the exploratory and qualitative nature of the usability assessment. Spearman's rho (r_s) was used to assess correlations between the hypothetical constructs and changes in KAM. Qualitative Description methodology (Sandelowski, 2000) was used to thematically analyze the interview data and was conducted by AIM. Qualitative description methodology is a research approach used in qualitative research to explore and describe a specific phenomenon or topic in a detailed and comprehensive manner. For the coaches, themes emerged if two or more coaches reported on them. For the focus group interview with players, themes emerged based on individual contributions and discussion in the group. Interviewees did not provide feedback on findings.

4.4 Results

Twenty players were included in the analyses including changes in KAM. One player was excluded from the study early due to an injury, and two players were excluded from post lab testing analyses due to a broken wrist and post-Covid, respectively. Two players were excluded from survey data analyses, one due to early intervention exit due to injury and one due to aberrant responses. For demographic data of the players and coaches, see Table 4.1.

Table 4.1. Demographics

Players	Age (years) mean (sd)	Body Mass (kg)	Height (m) mean (sd)	Level of Play	Number of Technique Sessions	Number of Strength Sessions

		mean (sd)			mean (sd)	mean (sd)
n=21	16.71 (1.10)	69.33 (8.20)	1.75 (0.06)	Level 1: 3 Level 2: 10 Junior: 2 Girls 2006: 6	17.25 (3.27)	18.25 (3.70)
Coaches	Age (years) mean (range)	Sex		Years of Experience mean (range)		
n=4	40.5 (26-53)	Female: 2 Male: 2		12.5 (2-25)		

4.4.1 Correlations between usability characteristics and improvement in KAM

Reduction in KAM significantly correlated with perceived effectiveness on reducing ACL injury risk (r_s -0.52, 95% CI -0.78, -0.10, p = 0.02) and the players' intention to use the program (r_s -0.46, 95% CI -0.75, 0.03, p = 0.04), where a higher score on these items was correlated with a reduction in KAM. See Figure 3.3.

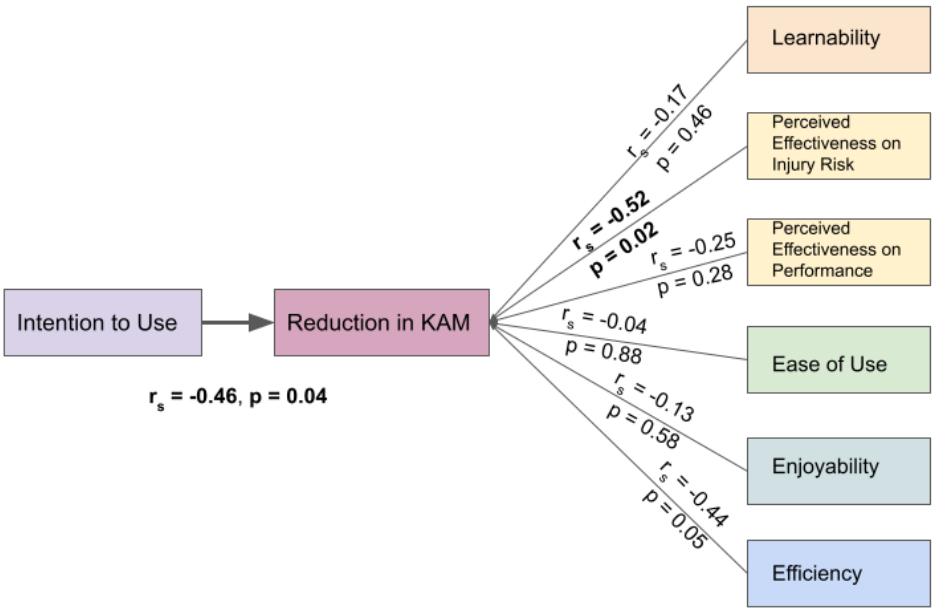


Figure 4.3. Correlation between usability characteristics and improvements in KAM.

4.4.2 Usability characteristics affecting use of program

Narrative findings and quotes from interviews with players and coaches are presented in Table 4.2. Players had mixed feelings about the ACL injury prevention intervention. They appreciated perceived improvements in some exercises, while expressing concerns about program length, and doubts regarding exercise effectiveness and transferability (e.g. cutting tasks). The perceived effectiveness and improvements in some exercises and tests motivated players, but they were generally not highly motivated to continue it without previously demonstrated proof of its effectiveness in reducing ACL injury risk. Coaches had similar concerns about program length and doubts about its effectiveness, with a reluctance to continue using it after the research period. They emphasized the importance of clear communication, information, and referencing research to motivate players and facilitate program adoption. Overall, both players and coaches highlighted the need to balance program effectiveness with practicality and clear communication about its purpose and benefits.

Table 4.2. Narrative findings and quotes from interviews with coaches (n = 4) and players (n = 6)

Questions	Coaches	Players	Quotes
What did you specifically like about the exercise program?	**Purpose and potential value:** Two of the four coaches said they liked the purpose of the program, namely preventing ACL injuries in handball. **Strength exercises:** Three coaches liked the strength exercises and particularly focusing on exercises for external rotation of the hip.	**Perceived effectiveness:** Players appreciated exercises that they felt were effective and those that showed noticeable improvement. They especially liked the clamshell exercise as they could feel that this exercise was very effective while they were doing it. **Integration into warm-up**: They appreciated that the program could be used as part of their warm-up, since they had to spend time on that anyway. Although, they did agree that playing soccer as a warm-up was more fun. **Program Content Variation:** The players liked that there were some exercise variations introduced during the intervention period. **Relatedness:** Most exercises were related to handball	Coach: *"I cannot say something that I really like about this program… I don't have any preconceptions about what I like when I am not sure if it will work or not."* Coach: *"I did like the strength exercises a lot. We saw that the boys fell far behind the girls [on those exercises] since they had not been a part of the 8 week intervention."* Player: *"The clam exercise was very good. …I felt that I worked hard with my knee in that exercise specifically, and felt I got stronger in that exercise specifically."* Player: *"Glad that it was a warm-up so we could easily slide into the rest of the handball session and be more efficient"*

		Exercise Purpose: The players found the purpose of the exercises easy to understand.	
What did you specifically dislike about the exercise program?	**Program length:** Three out of the four coaches said that the program was too long. They spent up to 40 minutes on the program in the beginning. This was time taken away from the 90 minutes the team had available to play handball or work on specific skills etc. **Doubts about cutting task effectiveness and transferability:** All four coaches questioned the potential value of the cutting tasks and their ability to change cutting technique in this age and with this method, in addition to the cutting tasks' transferability to actual gameplay. Three of the coaches said the cutting task seemed "artificial", not handball specific enough. Even though one might see an effect on the cutting task test in the lab, the coaches were skeptical one would see this technique change in a game-like environment where external factors are present. **Exercise feasibility:** The calf raise exercise was especially challenging to execute properly without proper equipment. The coaches responsible for the strength training reported that they did not believe they got the desired effect from this exercise.	**Program length:** They thought the program was too long and it substantially reduced their time to play handball. Players suggested that the program would be more feasible and motivating if it were shorter in duration or added in addition to the time originally designated to play handball. **Doubts about cutting task effectiveness and transferability:** The players had limited confidence that the cutting task would have any transfer to a game where the ultimate goal is to outmaneuver the opponent and score. The players discussed the challenge of focusing on proper technique during cutting tasks and how it might affect their game performance. They recognized the uniqueness of each player's technique and the difficulty in intentionally changing it. **Exercise feasibility:** The calf raise exercise was hard to get done properly and did not feel effective. There seemed to be an agreement that calf raises was easy and boring, the clam exercise was hard and great, and the side plank exercise was hard, but provoked pain for several. They needed to pair up to conduct some of the exercises, which made it harder to execute them twice weekly if one lost a session.	Coach: *"It took too much time. [To spend] 20 minutes is okay, but 40 minutes is too much."* Coach: *"The circumstances under which we, during a large part of this intervention period, exercised cutting technique seem to have a low transfer value to a game. It is not specific enough, it is too simple, technical features that I don't believe have any transfer value to when they are on the handball court….That was perhaps the most difficult, the technical aspects of the cut, how it should be executed, knowing what to look for and how effective that actually is with regard to the cut."* Coach: *"Calf raises were hard to execute properly. It was more tiring to be the person providing external load than to be the person performing the calf raises….It was hard to apply enough load. I mean, they did use their calf muscles, but not close to a maximum effort, if that was supposed to be the purpose."* Player: "*It ate up time to play handball, a large amount of the original handball time got lost, so perhaps this should have been added in addition to, as opposed to instead of.*"

		Perceived effectiveness: It was hard to notice any improvements in some exercises, which then felt meaningless and influenced their motivation for conducting them.	
What other factors affect your willingness to use this program? Barriers Facilitators	**Lack of proof:** Two coaches indicated that it was hard to say anything they actually liked about the program, that it was something they just had committed to do and that it all depended on the post test results. None of the coaches said that they would like to continue using the program after the research program was over. They were all interested in using some elements from it, but three of the coaches said these elements would need to be proven effective before they were to consider implementing them (e.g., strength exercises and/or cutting task). **Integration with existing training routines:** Two coaches mentioned that they would need to integrate parts of the program where they fit in and fall naturally based on the rest of the session and weekly schedule and focus on normalizing it. **Communicating purpose and potential value:** Coaches highlighted the importance of communicating the program's purpose and potential value in injury prevention. They believed that emphasizing the program's goal of preventing ACL injuries could motivate players	**Evidence and Testing:** Players expressed a desire for evidence of the program's effectiveness and frequent objective measures to enhance motivation. They emphasized the importance of testing during the intervention period to track their progress. While not highly motivated to continue with the same program, players expressed openness to it if the results of the intervention study showed its effectiveness. This indicated that perceived effectiveness played a significant role in their willingness to continue. **Attitude:** The players said they associate IP exercises with easier exercises that do not feel very effective immediately. The players indicated that IPT does not feel like "training". **Perceptions and communication of IPT:** The players said they know why they are doing IPT and emphasize the importance of getting in-depth information and understanding the purpose of the exercises. However, players acknowledged that they typically need to be told to do injury prevention training, which can have a negative association. They emphasized the need for a positive association with such training and stressed	Coach: *"I think it is really important that we look at the results. If we are doing something this specific for injury prevention it is important that we know exactly what works and not…If the results are really positive, that it is a gamechanger, then yes, we should use it, of course….then it is also a bit easier to sell it to the girls. Then they are not part of a project to see how it goes, then you can say that it is actually for them.*" Coach: *"It is not fun to do this program, that must be said…It is not fun exercises. But there is not much injury prevention training that is fun, you do it because it works and you want to be able to still play handball."* Player: "*We are doing it for our own sake, and it should be motivating enough in itself to not get injured. It just doesn't feel like we are 'training' [when doing IPT]*" Player: "*It would have been motivating if we knew that this program was actually really effective for preventing ACL injuries, that it was proven to help*"

	and serve as a facilitator to use. **Enjoyability:** Two coaches stated that injury prevention is never fun, and especially when the alternative is to spend time on playing handball. Two coaches also said that it does not have to be fun as long as the players know that it works and understand its importance, but at this point they did not yet know if it would be worth it. **Change of routines:** Three coaches emphasized that these players are very set in their routines, and that it is difficult for them to be told to do otherwise. They lost their weekly training structure, they had to add an additional strength session, and they had to do strength training before practice, making some players afraid of tiring their muscles before playing handball.	the importance of understanding the purpose.	Player: "*It's IPT, it has to be done regardless of how much time it takes*" Coach: "*It has to be implemented as a normality…I would avoid implementing it as one single block…I think there are rich opportunities to implement this, but it has to be implemented using parts of the program where they fall naturally in our training*." Coach: *"They have to know it is important. Many players don't like to go to the gym and do strength training, but they do it because they know it is important for them. They do it even though it is not very fun."*

4.4.3 Design Thinking

From the collaborative ideation and prototyping session, key themes in the generated ideas included optimizing time efficiency, breaking down the program into manageable parts, emphasizing the program's purpose and rationale, utilizing rewards and punishments, introducing exercise variations, incorporating IPT into warm-ups, addressing stakeholder attitudes, and adapting equipment. In total, 31 ideas were generated between two groups. The highest-rated ideas focused on reducing the time commitment, spreading IPT across multiple sessions, and introducing playful elements to the exercises. Prototyping suggestions included modifying sets and reps, integrating IPT into existing playful activities (e.g., cannonball, tag games), adding competitive elements within the exercises, and creating obstacle courses where IPT exercises are also included. See Table 4.3 for a summary of the procedure and results.

Table 4.3. Design Thinking Approach

CONTEXT
2 groups, 3-5 min per question, 1 idea per post-it, discuss and build off each other's ideas
Generate ideas for 1) *"How we, as researchers, coaches or others, can help encourage players to perform IPT frequently"* 2) *"How can this program become more enjoyable?"* 3) *"Do you have any suggestions for how you can include IPT as part of your weekly training schedule?"*
Prompts: *You have all the time, money, and resources in the world.* *Try to let go of the attitude "IPT is not fun, just something we have to do".* *Imagine it is possible to make IPT fun - how can we make this fun?*

IDEATE - RESULTS (CHALLENGE ASSUMPTIONS AND CREATE IDEAS)	
Group 1: 14 ideas generated, Group 2: 17 ideas generated	
THEMES EMERGED:	
TIME (n=7):	Shorter duration of IPT **and/or** add more time to each handball session.
SPLIT UP PROGRAM (n=6):	Implement a few exercises each session, not one long block.
PURPOSE/RATIONALE (n=6):	Focus on, and clearly explain, purpose, and importance of program and potential. consequences on injury risk and performance.
REWARD/PUNISHMENT (n=4):	Get a reward for completion, reward the one with greatest results, or use exercises as punishment in competitions.
VARIATION (n=3):	Vary and swap out exercises.
WARM-UP (n=3):	IPT as part of the warm-up
STAKEHOLDER ATTITUDE (n=3):	Communicate in a positive manner about IPT/program.
EQUIPMENT (n=2):	Exercises with elastic bands or weights instead of partners.
Assorted singles statements (n=8):	Add IPT to end of low-intensity sessions, plan training week individually and take responsibility to

get IPT done, use role models, inform the club, add playful activities, look at statistics of previous test results, fewer sets/reps, add strength exercises to current strength program.

RATE THE BEST AND SECOND BEST IDEA

"Our goal is to design an IP program that you find okay to use long term. We know that you didn't enjoy this program and was happy when the 8 weeks were over. So which idea can we build on to address this issue - making IPT feel quite alright to do and something that feels natural for you to implement in your weekly training schedule?"

BEST IDEA:

"Not spending so much time on it" - 4 best votes, 1 second best vote
"Do a little bit each session" - 1 best vote, 1 second best vote
"Add some playful components into the exercise" - 1 best vote
("focus on consequences" got two second best votes, and "work towards a reward" and "more variation" both got one second best vote)

PROTOTYPE - RESULTS

(FIND A SOLUTION)

"Design the program that you find reasonable with regard to time use"

1. Cut down on set and reps (especially the side push exercise, 6x6 reps)
2. Spread it out over three sessions instead of two
3. Remove the calf raise exercise

"If you think about the idea of spreading it out more, how would that look?"

1. Can do the same exercises, but split up and spread it out over all weekly sessions
2. 10 minute per session feels like a cut-off

Since we have to do a warm-up anyway, 10 minutes is fine

"If you think about the idea of making IPT part of a more playful activity, can you design a playful activity?"

1. The exercises can be added to most of the playful activities we are already doing (e.g., cannonball, tag games)
2. Include in competitive games, where you have to do IPT when you lose
3. Make the IPT exercise itself a competition (e.g., standing in the plank position the longest)
4. One group plays a game, the other do IPT, then we swap
5. If it does not have to be these exercises, we can learn from gymnastics where they jump around with weights and stuff
6. Make an obstacle course including the IP exercises, but also more fun stuff in between
7. Circuit training format

4.5 Discussion

A team of female handball players tested out a new injury prevention program for 9 weeks, and two factors stood out as important reasons for why this intervention will not be adopted by this team for long term use: efficiency and perceived effectiveness. Coaches and players expressed concerns about the program's length, which significantly encroached upon their practice time, and harbored doubts regarding the efficacy and transferability of certain components, such as the ability to alter players' cutting technique to prevent ACL injuries. Notably, none of the coaches favored continuing the program post-intervention period due to insufficient evidence of its effectiveness, the substantial time commitment, and its inflexible design.

Perceived effectiveness, efficiency and enjoyability were factors identified as barriers and facilitators for program use. This is in line with previous studies (Minnig et al., 2022; Moesch et al., 2022). In a literature review, Minnig et al. (2020) report that time was the biggest barrier for implementing evidence-based injury prevention programs noted by coaches. Lack of importance placed on the program and lack of player motivation were other barriers identified in their review, which can be considered part of intrinsic motivation and enjoyability. Similarly, the results of this study to a great extent mimic findings from Moesch et al. (2022). They reported that about 10-15 minutes of injury prevention training per session seem feasible, exercises that the players understood the purpose of, experienced improvements in, and considered handball-specific increased their motivation. Additionally, including competitive elements and variation in exercises was appreciated, whereas requiring specific resources was a perceived barrier.

The interviews identified program flexibility, or rather lack thereof, as an important factor for use among both coaches and players. The coaches indicated that if the program was already proven effective in reducing ACL injuries, they would be more inclined to adopt parts of it, but they would organize it differently and integrate specific elements of the injury prevention program where they fell naturally regarding the rest of their training session. Previous literature has also proposed that the intervention needs to be flexible

and delivered in a “propose”, rather than “impose” manner, allowing the coaches to adapt the intervention based on the needs and resources of the team (Bruder et al., 2024; Hawkinson et al., 2022; Minnig et al., 2022).

Some coaches and players appreciated the program's purpose of preventing ACL injuries and found value in certain exercises, particularly those related to strength and external hip rotation, while others had reservations. Overall, the coaches' feedback highlighted a need for balancing program effectiveness with practicality and ensuring clear communication about its purpose and benefits, stressing the importance of communication, information, and referencing research to motivate players and facilitate program adoption. While the players appreciated perceived improvements in certain exercises, they also desired more evidence of their effectiveness in reducing injuries and/or improving performance. Although lack of importance placed on the program was identified as a barrier in the review by Minnig et al. (2022), their recommendations for facilitating implementation of injury prevention programs does not explicitly include communication and information about rationale and evidence of effectiveness targeted to the players to increase their motivation. Nonetheless, the exceptionally high motivation to improve performance in this sample may have negatively influenced the ease with which the research team and coaches could convince them to incorporate such exercises before strong evidence for their effectiveness and value exists.

In general, injury prevention training was seen as a necessary but unenjoyable task among the players, highlighting the importance of communication about, and understanding of, its purpose. When asked for program design ideas, players leaned towards shortening the program, splitting it up into shorter blocks and making it more playful, indicating that time-efficiency and enjoyment affects adherence. Most of their ideas to improve enjoyability seem reasonable and possible to incorporate (e.g., obstacle courses, tag games, circuit training). Similar formats have been suggested in previous research where end-users were involved in program development (Ageberg et al., 2022). Additionally, incorporating some form of competition and feedback on improvements (e.g. force measures) seem to be motivating factors for these players. However, the players indicated that as long as they knew that it would be effective in reducing ACL injury risk, they would be inclined to do it even though it might not be fun.

On the same note, reduction in KAM was significantly correlated with players’ perceived effectiveness in reducing ACL injury risk. This may indicate that as long as stakeholders (e.g. coaches) are able to communicate to the players that the intervention will reduce their injury risk, the program does not have to be perceived as enjoyable to have an effect on injury risk. Coaches’ buy-in has in previous studies been reported as the crucial first step of injury prevention implementation and having the players do the program (Bizzini & Dvorak, 2015; Minnig et al., 2022; Padua et al., 2014; Root et al., 2019). For example, when coaches utilize the IPPs (e.g. compliance > 50%) in research settings, compliance rates among players are close to 90% (Sugimoto et al., 2012, 2017).

On the other hand, poor player buy-in has been reported as a barrier for implementation, indicating that coach buy-in alone is not enough for long-term implementation (Minnig et al., 2022; Owoeye et al., 2020). Behavioral theories such as the Self-Determination Theory and previous research on youth handball players suggest that IPPs need to be made more enjoyable to increase likelihood of use (Ageberg et al., 2022, 2024; Chan & Hagger, 2012; Moesch et al., 2022; Møller et al., 2021; Ryan & Deci, 2000). Although player enjoyment did not significantly correlate with injury risk reduction (e.g. reducing KAM), there is reason to believe that player enjoyability will affect long-term use of the program.

This study underscores the need for a holistic approach to exercise intervention development and design. This is in line with previous research (e.g., Prep-to-Play and I-PROTECT) where end-users at different organizational levels were extensively involved in the development of new injury prevention programs (Ageberg et al., 2024; Bruder et al., 2023). For instance, the Prep-to-Play PRO IPP yielded high implementation, likely affected by the program's flexibility and engagement of various stakeholders throughout the development and implementation phase (Bruder et al., 2024). However, such approaches are notably intricate, time-intensive, and resource-demanding. Alternative methods, like surveys or usability scales employed in this study, may offer a streamlined means of identifying barriers and gauging the likelihood of adoption. Nonetheless, the creation and validation of such tools must precede their deployment, ensuring their efficacy as efficient substitutes for more exhaustive methods such as workshops.

4.5.1 Limitations

A notable limitation was the relatively small sample size, only including one team and their coaches, which impacts the generalizability of the findings. The players participating in this study are ambitious and play at a relatively high level (Müller-Kuhne et al., Unpublished data), and the barriers and facilitators may be different for a different sample of players (e.g. younger, less ambitious

4.6 Conclusion

This study found that coaches and players did not want to continue using the newly developed ACL IPP outside a research setting, particularly due to their lack of belief in the program's effectiveness, enjoyability, flexibility and time. Furthermore, supporting previous literature, these factors seem to be important to target when designing sustainable injury prevention programs. Usability testing and end-users' involvement throughout the design and development process is essential to maximize sustained IPP use. Even in the presence of highly effective prevention programs, their value is diminished if we cannot facilitate acceptance and adoption by the target population.

4.7 Clinical implications

This study supports the need for a holistic approach to injury prevention program development and implementation, focusing not only on efficacy but also on usability, communication, and athlete engagement. The results of this study is in line with previous research on barriers and facilitators to IPP use, however, the importance of player enjoyability seemed less clear in this population. In this study, we discovered that employing the design thinking approach to co-create or modify injury prevention programs may be beneficial, suggesting its potential as an innovative and collaborative strategy for enhancing program effectiveness.

Funding

The study has received funding grants through a Health Research Institute Fellowship and a Finishing Fellowship from Michigan Technological University. Data used in this study was collected from a randomized controlled trial (NCT05643261) conducted with a research team at the Oslo Sports Trauma Research Center. The Oslo Sports Trauma Research Center has been established at the Norwegian School of Sport Sciences through generous grants from the Royal Norwegian Ministry of Culture, the South-Eastern Norway Regional Health Authority, the International Olympic Committee, the Norwegian Olympic Committee & Confederation of Sport, and Norsk Tipping AS.

Author contributions

TK was involved in designing the intervention study project and included participants. All authors designed the survey project. AIM collected and analyzed data and wrote the first draft of the paper. All authors contributed to the final manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to acknowledge the OSTRC research group involved in the RCT for providing valuable input survey item selection, and for distributing surveys and collecting responses from the players.

Conflicts of interest

The authors declare no conflict of interest.

5 EVALUATING THE PULSE OF EXERCISE PROGRAMS: DEVELOPING AND VALIDATING THE INTERVENTION USABILITY SCALE FOR EXERCISE (IUSE) TO ENHANCE IMPLEMENTATION AND ADHERENCE

5.1 Abstract

Objectives: This study aimed to develop and assess the psychometric properties of the Intervention Usability Scale for Exercise (IUSE).

Methods: Item generation and content validation involved cognitive interviews and feedback from eight exercise intervention stakeholders and ten target users from the general public. Subsequently, 526 target users from University, Qualtrics and Prolific participant panels assessed exercise programs through an online survey. Dimensionality was assessed using PCA, EFA/CFA, and bifactor models. Item reduction was conducted through Item Response Theory (IRT) graded response models, assessing discriminability, item information levels/patterns and differential item functioning. Test-retest reliability was assessed using Pearson's *r*. Subsequently, short vs. original scale versions were compared. Finally, evidence of criterion, convergent, and discriminant validity was assessed.

Results: Thirty-six items were initially included, with 16 removed due to low test-retest reliability and factor loadings ($r < 0.50$). A 3-factor structure emerged from EFA, PCA, and bifactor models: Value, Ease of Use/Learnability, and Social. Iterative IRT evaluation led to 12 item removals, resulting in 8 final items across three subscales. Validity analyses indicated good convergent (e.g., $r = 0.79$ with Intervention Appropriateness Measure, $r = 0.65$ with Intervention Usability Scale), criterion (e.g., $r = 0.71$ with Net Promoter Score and 0.75 with intention measures), and discriminant validity ($r = 0.42$ with External Motivation), along with satisfactory internal consistency (Cronbach's alpha: 0.79-0.84).

Conclusions: The IUSE scale demonstrated promising psychometric properties. Application of the scale with collection of actual uptake/adherence data is needed to assess predictive validity.

Keywords: implementation, exercise, uptake, psychometrics, physical activity

Footnote: The material contained in this chapter is in preparation for submission to the *British Journal of Sports Medicine*.

What is already known on this topic

- Usability testing is commonly used as part of product development and testing in domains such as technology, due to its ability to detect barriers to use and implementation likelihood.

- To our knowledge, no usability scales has been developed that sufficiently assesses the usability of exercise programs.

What this study adds

- This study introduces a new measure for assessing usability of exercise programs, namely the **I**ntervention **U**sability **S**cale for **E**xercise - IUSE scale.

How this study might affect research, practice or policy

- The IUSE can be an efficient and valuable tool for exercise program developers, both within research and clinical practice in that it can help identify potential barriers to program use. However, future studies are needed to add evidence to the predictive ability of the scale where actual adherence is assessed as opposed to intention to use the program.

5.2 Introduction

Extensive research has documented the positive effects of exercise interventions and physical activity on both preventing and treating a large number of diseases, conditions, and injuries (Pedersen & Saltin, 2015). Consequently, guidelines for physical activity have been provided for the world's population, recommending 150-300 minutes of moderate-vigorous aerobic exercise and two sessions of muscle-strengthening exercises weekly (U.S. Department of Health and Human Services, 2018; World Health Organization, 2022). Nonetheless, the World Health Organization and the Center for Disease and Control and Prevention states that about 1 in 4 people globally and in America meet the recommended guidelines for physical activity (< 20 % of women, < 30 % of men). Adherence to most types of prescribed exercise interventions remains an issue and substantially limits the value of evidence-based, highly effective exercise programs.

Targeting the development and design of exercise interventions by employing usability testing might be one way to increase adherence (Petushek & Donaldson, 2020). Usability is defined as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use." (International Organization of Standardization (ISO), 2018). For exercise intervention usability, the ISO definition applies by simply replacing "a product" with

“an exercise intervention”. Usability testing helps detect barriers to, and likelihood of, product use, and is commonly employed in product development and testing in other domains, such as technology (Hass, 2019; Keenan et al., 2022; Mator et al., 2021).

Several usability scales have been developed for different purposes. To our knowledge, usability testing within exercise medicine is a novel approach, and no existing scale seems directly applicable for assessing exercise interventions. Many existing usability scales do not incorporate behavioral/psychological theories such as the Self-Determination Theory and Theory of Planned Behavior (Ajzen, 1985; Ryan & Deci, 2000), which are proposed to predict intention to use rehabilitation and prevention programs (Chan & Hagger, 2012). Similar disciplines more familiar with uptake/adherence, such as Implementation Science, Theoretical Framework for Acceptability, and Health Information Technology Acceptance Model, are all concerned with factors affecting use/behavior. Hypothetical constructs and items seem to substantially overlap between usability scales and scales in behavioral disciplines, and each discipline may provide valuable insights to the usability assessment of exercise programs. Moreover, there is a need to integrate and adapt these various measures specifically for exercise interventions.

In an attempt to improve usability of prescribed exercise interventions, we introduce the development of the Intervention Usability Scale for Exercise (IUSE). The aims of this study were 1) to develop and assess construct coverage, content validity, and comprehension of IUSE, and 2) to further develop, refine and add validity evidence to the IUSE through assessment of dimensionality, reliability, item reduction and hypothesis testing (e.g., relations to other variables and intervention study).

5.3 Methods

5.3.1 Equity, diversity, and inclusion statement

Our study population included a diverse sample of the general US adult (>18 years) population (e.g., different ages, genders, ethnicities, and activity levels). The generalizability of this study does neither extend beyond English speakers nor adults.

5.3.2 Development of the first version of IUSE

The foundation of scale development is assessing and revealing its psychometric properties (e.g., reliability and validity evidence) (American Educational Research Association, 2018). Both the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) taxonomy of measurement properties (Mokkink, Terwee, Patrick, et al., 2010) and the Standards for Educational and Psychological Testing (2014 ed.) have developed guidelines for measuring the validity of interpretations of scale scores for the intended scale uses. Adding validity and reliability evidence to a scale requires multiple data collections and statistical analyses and it is typically divided into three phases: 1) familiarizing with, and identifying, the domain and subsequently generating and pretesting hypothetical constructs and items, 2) assessing the factor

structure or dimensionality of the scale, and reliability, and 3) adding validity evidence (e.g., convergent/discriminant and criterion validity) (Boateng et al., 2018).

Figure 5.1 briefly summarizes the methodological approach and data collections for this study. Data collection 1 and 2 were used in item generation and evidence of content validity, whereas data collection 3 was used for assessing dimensionality, further validity evidence (e.g., convergent, discriminant and criterion validity), internal consistency, test-retest reliability, and hypothesis testing.

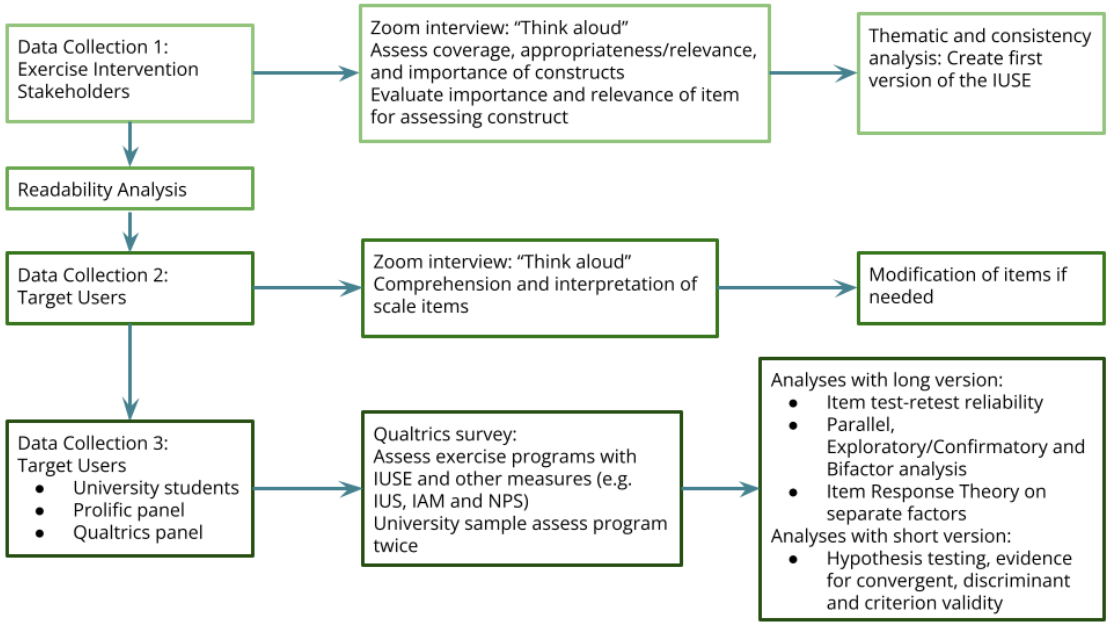


Figure 5.1. Flow chart of data collection and analyses.

For data collection 1, exercise intervention stakeholders were invited to a Zoom meeting with one researcher where they were 1) asked what characterizes a good exercise program, 2) given a pre-mortem case of an unsuccessful program adoption where they had to generate reasons it failed, 3) asked to evaluate the definition of exercise intervention usability, 4) evaluate the coverage of constructs (e.g., relevance, redundancy, missing) and items affiliated with them (e.g., clarity, relevance), and 5) rate the two most important constructs for assessing exercise intervention usability. The meeting was audio recorded and deleted once analysis was conducted. Thematic and consistency analysis were used to synthesize and incorporate stakeholder feedback to create a first version of the IUSE scale.

Subsequently, item readability analysis was conducted through Originality.ai. This artificial intelligence-generated, online tool automatically calculates a variety of readability statistics (e.g., Flesch-Kincaid Grade Level, Automated Readability Index, SMOG Index, Dale-Chall Readability Grade etc.) for texts. The readability was set at high-school level or below and items were modified if needed.

The items were further tested with a sample of target users through cognitive interviews conducted via Zoom to ensure response processes aligned with the item goal in addition to comprehensiveness. The participants were presented with an exercise program and asked to familiarize with and pretend it was prescribed to them. Subsequently, they used the first version of the IUSE scale to assess the usability of this exercise program. They were asked to "think aloud" during the whole interview to help ensure that they understood and interpreted the item as intended. Items were subsequently modified if needed.

5.3.3 Sample

For data collection 1, five stakeholders for developing and implementing/prescribing exercise interventions (e.g., coaches, physical therapists, athletic trainers and researchers) from a diverse population were recruited (Boateng et al., 2018; Macefield, 2009; Nielsen & Landauer, 1993), in addition to four stakeholders authoring this paper. For data collection 2, a sample from the target population (n = 10) was recruited for comprehension purposes, supporting the readability analysis. The target population refers to individuals that are prescribed an exercise intervention (e.g. Physical Activity Guidelines for Americans). The target users were purposely sampled, aiming at persons with a comprehension level of high schoolers or higher and with various experiences with exercise. Sample sizes were based on previous literature on scale development and usability testing (Boateng et al., 2018; Macefield, 2009; Nielsen & Landauer, 1993). Both stakeholders and target users were recruited through word of mouth and our personal network of colleagues within the healthcare, sports medicine and implementation science community, and everyone gave consent to participate prior to recording.

For data collection 3, participants were recruited from the student pool at a Midwestern Technological University (e.g. Introductory to Psychology) and from two online panels (e.g., Qualtrics and Prolific). With our three sampling sources, we aimed at recruiting 300 or more participants representative of the U.S. adult population (e.g., 50 % female, average age of ~38 etc.). The sample size is based on previous studies reporting sample sizes needed for conducting factor analyses and item response theory (IRT) (Boateng et al., 2018; W.-H. Chen et al., 2014; Comrey & Lee, 2013; Macefield, 2009; Nielsen & Landauer, 1993). Recommended sample sizes for scale development studies vary, with the lowest reported sufficient sample size detected in previous studies is n = 75-100 (W.-H. Chen et al., 2014; Houts et al., 2006). Comrey & Lee (2013) have suggested that sample sizes for scale development of 100 = poor, 200 = fair, 300 = good, 500 = very good and >1000 = excellent (Comrey & Lee, 2013). Other studies report that the rule of thumb in scale development is that the sample size should include a minimum of 10 respondents per item in the scale (Morgado et al., 2017; Nunnally & Bernstein, 1994). For IRT purposes, responses on each item should cover all response options (e.g. avoid "null" values) and a larger sample size is likely to cover all options.

5.3.4 Intervention Usability Scale for Exercise (IUSE)

Items for the IUSE scale were generated based on a literature review of existing scales and theories, known facilitators/barriers to exercise intervention uptake and adherence from previous studies, and stakeholder interviews. The response options were either presented in a semantic differential format with 5 points (e.g. Boring to Fun) or in a 5-point Likert scale format (Strongly Disagree-Strongly Agree). Scales of five or more points improve reliability, and when there are no strong floor or ceiling effects, they can be treated as continuous items in confirmatory factor analysis and structural equation modeling (Boateng et al., 2018).

5.3.5 Online survey

To collect responses on the IUSE scale, surveys were conducted through Qualtrics (Qualtrics, Provo, UT). All participants provided their consent to participate in the study before the survey started. Participants were randomized into two groups. Both groups were presented with an exercise program (see details about programs below) and asked to use the IUSE to evaluate the usability of it (See Figure 5.2 below). After the usability assessment, they were presented demographic questions (e.g., age, gender, exercise background), in addition to items from other scales measuring usability and implementation likelihood. The University sample was further presented with either the same program or the other program and asked to assess the usability of it using the IUSE scale before ending the survey. Figure 5.2 shows the flow of participation.

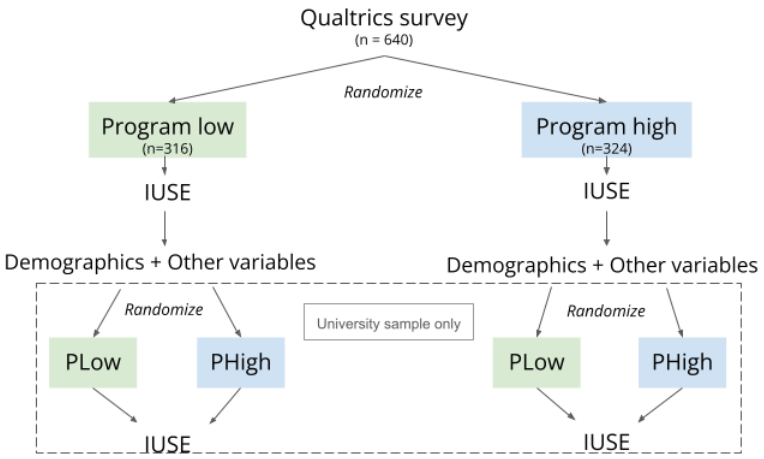


Figure 5.2. Flow chart of study participation.

5.3.6 Exercise program

One of two different exercise programs was presented to the participants in the survey, both of which were based on recommendations for adult Americans regarding weekly strength training (e.g. moderate/high intensity muscle-strengthening exercises involving all major muscle groups twice weekly). Strength training has positive effects on bone mass, lean mass and tensile strength, and are typically important in exercise interventions

prescribed for prevention, preoperative, postoperative, and treatment of musculoskeletal injuries (Shaw et al., 2016). One of the programs, Program Low (PLow), was designed to have low usability and the other program, Program High (PHigh), was designed to have high usability. The design was informed by known barriers and facilitators for adherence to exercise interventions (e.g., resources, time-commitment, flexibility, comprehension, feasibility). For instance, Plow was presented through a pamphlet with drawings and text descriptions. It was expected to take ~80 minutes to complete (8 exercises, 3 sets, 3 in rest between sets) and it required gym equipment. PHigh was presented through a ~6 min video with 3D animations and comprehensive explanations and rationale. It was expected to take ~20 minutes to complete (4 exercises, 2 sets, circuit format) and provided two options for each exercise based on difficulty level. The exercise programs were developed using ExorLive ® (an online tool for developing exercise programs) and animated materials from Muscle Animations AS. Both programs are attached in the appendices.

5.3.7 Other measures for assessing validity evidence

The Intervention Usability Scale (IUS) (Lyon et al., 2021), a modified version of the widely used System Usability Scale (Brooke, 1995), and the Intervention Appropriateness Measure (IAM) (Weiner et al., 2017) were used for assessing evidence of convergent validity. Extrinsic motivation items from the Situational Intrinsic and Extrinsic Motivation Scale (SIMS) were used for discriminant validity evidence. Criterion validity evidence was assessed through the Net Promoter Score (NPS) (Reichheld, 2003) and items assessing intention to use the program (e.g. “I intend to use this program frequently”).

5.3.8 Statistical analysis

Careless/aberrant responses were detected using timing data for each page in the survey. Participants with a speed factor above 3 were removed (e.g. (median(response time)/response time) > 3) (Leiner, 2019). Secondly, participants that did not watch the exercise video (e.g. spent less time than the duration of video/2) were excluded for analysis. A third measure was the inclusion of attention checks (e.g. “Please describe one thing you remember from the exercise program you were presented”) (Abbey & Meloy, 2017). For missing data rates of less than 5% for the scale items and other variables, the median item response was imputed (Cokluk & Kayri, 2011; Downey & King, 1998; Jakobsen et al., 2017).

5.3.8.1 Dimensionality

Dimensionality was assessed through parallel analysis followed by exploratory/confirmatory factor analysis (EFA/CFA) and bifactor analysis. Items revealing loadings >.5 were retained for inclusion in IRT analysis. Model fit criteria included root mean square error of approximation (RMSEA; ≤ 0.06) and comparative fit index (CFI; > 0.95) and Tucker-Lewis index (TLI; > 0.95) (Boateng et al., 2018).

5.3.8.2 Item response theory

The use of unidimensional IRT or multidimensional IRT was decided based on dimensionality analyses. Models were fitted using the *ltm* and *mirt* packages in R. Items that represented a wide range of usability levels, high level of information and high degree of discriminability were selected. Measurement invariance was assessed by evaluating Differential Item Functioning (DIF), exploring whether people from different groups (e.g., active vs non-active, male vs female) with the same level of functioning have different probabilities of giving a certain response to an item (Smit et al., 2020).

5.3.8.3 Reliability analysis

Internal consistency was assessed through Cronbach's alpha where an alpha of 0.7 or higher was considered acceptable. Test-retest reliability analysis was conducted by comparing the responses from the sample that were asked to assess the same program twice (e.g., PLow + PLow and PHigh + PHigh). Test-retest reliability was assessed through correlation coefficients (Pearson's *r*) and comparing mean difference and Cohen d_{av} (Lakens, 2017).

5.3.8.4 Relations to other variables

After refining the scale, the final scale version score was compared with the other variables for assessing evidence of criterion, convergent and discriminant validity using Pearson's *r*. We expected to see high correlations between IUSE and the IAM, IUS, NPS, and Intention to use-items, and lower correlations between IUSE and extrinsic motivation-items from the SIMS.

5.3.8.5 Hypothesis testing

For hypothesis testing, the exercise program designed to have high hypothetical levels of the defined constructs identified in the scale development process (PHigh) was compared to a second program that was designed to have low hypothetical levels (PLow). Our hypothesis was that the program designed to have high hypothetical levels will receive a higher score on usability compared to the other program. Mean difference and Cohen's d_{av} were assessed for dependent samples (Lakens, 2017).

5.4 Results

Figure 5.3 shows the flow chart of participants from the three sources. Additionally, the figure shows the amount of aberrant responses and duplicate responses that were removed before further analysis. Additionally, 221 missing item responses (<0.1% of total) were imputed with the item median before analyses. Five-hundred-twenty-six participants were included in the analysis. Table 5.1 shows the participant demographics.

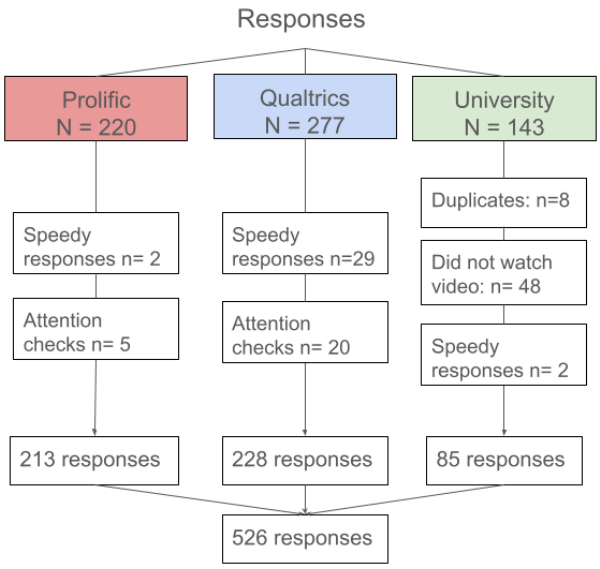


Figure 5.3. Flow chart of participant recruitment and data handling

Table 5.1. Participant demographics

	University	**Prolific**	**Qualtrics**	**Total**
N	85	213	228	526
Average Age (years)	19.4 (sd 1.2)	39.1 (sd 13.8)	53.9 (sd 16.4)	42.2 (sd 18.4)
Gender				
Man (n)	55	104	101	260 (49%)
Nonbinary (n)	5	4	0	9 (1.7%)
Transgender (n)	1	0	0	1 (0.19%)
Woman (n)	24	105	125	254 (48%)
Other	0	0	1	1 (0.19%)
Prefer not to respond	0	0	1	1 (0.19%)
Ethnicity				
American Indian/Native American/First Nation/Alaska Native (n)	1	3	2	5 (1%)

Black/African American (n)	4	22	24	50 (10%)
Asian/Asian American (n)	5	23	8	36 (7%)
Hispanic@/Latin@/Chican@ (n)	4	15	9	28 (5%)
Middle Eastern (n)	1	4	1	6 (1%)
South Asian (n)	1	0	0	1 (0.2%)
Native Hawaiian/Pacific Islander (n)	1	0	0	1 (0.2%)
White/European American (n)	75	158	191	424 (81%)
Other (n)	0	0	0	0
How many days do you perform vigorous-intensity activities?				
None (n)	14	52	70	136 (26%)
1-2 days of the week (n)	37	93	70	200 (38%)
3-4 days of the week (n)	23	44	50	117 (22%)
5-7 days of the week (n)	11	24	38	73 (14%)
How comfortable are you with performing muscle strengthening activities?				
From 1 to 10* (Mean)	7.65 (sd 1.89)	6.98 (sd 2.22)	7.29 (sd 2.47)	7.22 (sd 2.29)

*Ranging from Very uncomfortable (1) to Very comfortable (10).

Thirty-six items were included after the content validation phase. Table 5.2 shows the proportion of responses, the average score and item total correlation (r = 0.53-0.8) for each of the 36 items in the initial version. The last 10 items in the Table 5.2 revealed low test-retest reliability ($r < 0.5$) and were removed before further analysis.

Table 5.2 Item Responses

	Proportion endorsed by category						
Item	**1** Strongly Disagree	**2** Disagree	**3** Neither nor	**4** Agree	**5** Strongly Agree	**Mean (sd)**	**Item Total Correlation**
This program looks interesting	0.09	0.11	0.14	0.44	0.22	3.59 (1.2)	0.69

The program meets my needs	0.10	0.13	0.15	0.44	0.19	3.48 (1.22)	0.78
The program is valuable to me	0.10	0.11	0.20	0.41	0.18	3.47 (1.2)	0.75
This is an important program to me	0.13	0.17	0.26	0.30	0.14	3.16 (1.24)	0.75
I will enjoy doing this program	0.13	0.17	0.16	0.39	0.15	3.26 (1.27)	0.72
I like this program	0.07	0.12	0.14	0.44	0.23	3.63 (1.18)	0.68
I think performing this program is worth my time	0.08	0.12	0.15	0.42	0.24	3.62 (1.19)	0.8
This program required the use of equipment I like to use	0.13	0.13	0.16	0.41	0.17	3.36 (1.26)	0.5
The program is useful to me	0.07	0.10	0.12	0.48	0.24	3.72 (1.13)	0.77
I can do the program whenever I want to	0.04	0.08	0.07	0.46	0.35	3.99 (1.06)	0.51
This program required the use of equipment I have available	0.15	0.20	0.07	0.33	0.24	3.31 (1.42)	0.57
I can do the program wherever I want to	0.10	0.15	0.14	0.35	0.25	3.50 (1.29)	0.63
The program includes everything I expect it to have	0.10	0.13	0.18	0.39	0.20	3.45 (1.23)	0.73
The program seems exactly right for my goals	0.13	0.18	0.18	0.35	0.16	3.23 (1.27)	0.74
The difficulty of each exercise is exactly right for me	0.20	0.21	0.10	0.30	0.19	3.07 (1.44)	0.66
I have access to the equipment needed for this program	0.20	0.20	0.05	0.29	0.26	3.21 (1.51)	0.58
It is clear to me how this program will improve my strength	0.04	0.04	0.08	0.47	0.37	4.09 (0.97)	0.53
I can do this program with a friend	0.09	0.10	0.12	0.39	0.30	3.73 (1.23)	0.54

This program will allow me to exercise with friends	0.12	0.12	0.20	0.36	0.21	3.41 (1.27)	0.56
The exercise program seems Ineffective ------ Effective	0.02	0.02	0.14	0.44	0.38	4.15 (0.86)	0.56
The exercise program seems Boring ------ Fun	0.08	0.16	0.37	0.25	0.14	3.19 (1.12)	0.61
The exercise program seems Inefficient ------ Efficient	0.03	0.06	0.15	0.47	0.29	3.93 (0.97)	0.62
The exercise program seems Difficult to do ------ Easy to do	0.09	0.13	0.17	0.30	0.31	3.60 (1.29)	0.60
The exercise program seems Difficult to learn ------ Easy to learn	0.05	0.09	0.12	0.30	0.44	4.00 (1.16)	0.61
The exercise program seems Rigid ------ Flexible	0.04	0.16	0.18	0.33	0.29	3.66 (1.17)	0.62
The exercise program seems Complicated ------ Straightforward	0.04	0.07	0.14	0.32	0.43	4.03 (1.11)	0.64
The program is effective for improving strength	0.10	0.04	0.08	0.49	0.35	4.09 (0.96)	0.45
The program seems appropriate in length	0.04	0.082	0.09	0.46	0.34	3.98 (1.04)	0.63
It is clear how to use the program	0.05	0.08	0.08	0.38	0.41	4.02 (1.12)	0.66
The program is simple to understand	0.03	0.08	0.08	0.40	0.42	4.09 (1.03)	0.65
I will learn to use this program very quickly	0.06	0.14	0.12	0.39	0.28	3.68 (1.21)	0.67
I can adapt the program to my needs	0.09	0.10	0.10	0.41	0.29	3.72 (1.24)	0.63
The program provides flexible options	0.06	012	0.12	0.40	0.30	3.77 (1.16)	0.64

The exercise program seems Too long ------Too short*	0.12*		0.23*		0.65*	4.06 (1.40)	0.30
The exercise program seems Worthless ------ Valuable	0.02	0.03	0.20	0.44	0.31	3.99 (0.89)	0.65
The exercise program seems Harmful ------ Beneficial	0.03	0.04	0.09	0.43	0.41	4.13 (0.97)	0.58

*Item was transformed: 1 = Strongly agree + Strongly Disagree, 3 = Disagree + Agree, 5 = Neither nor

5.4.1 Dimensionality

Scree plot, eigenvalues and parallel analysis suggested a 3-factor structure for the 26-item scale ($IUSE_{Long}$). The three factors explained 31%, 16%, and 6% of the variance, respectively. An exploratory factor analysis (EFA) was conducted using the "oblimin" rotation and maximum likelihood with one and 3 factors. The factor correlations for the three factors (MR1, MR2, MR3) were examined. The correlation between MR1 and MR2 was 0.55, between MR1 and MR3 was 0.49, and between MR2 and MR3 was 0.47. Additionally, an exploratory bifactor model was fitted using the omega function from the *psych* package in *R*, indicating a multidimensional factor structure (Omega total = 0.96, Omega hierarchical = 0.71, explained common variance of $g = 0.58$). Table 5.3 shows the factor loadings for both the unidimensional and 3-factor structure, and the bifactor model. This table also shows that items load high on both the general and a specific factor. Three items were removed due to low loadings on the specific factor in the EFA and three items were removed due to low loading on the general factor ($r < 0.5$) in the bifactor model.

Table 5.3. Factor analyses

	Traditional factor model				**Bifactor model**			
	1-factor	**3-factor**			**3-factor**			
Item		**Value**	**Ease**	**Social**	**G**	**g1**	**g2**	**g3**
This program looks interesting	0.70	0.83	-0.11	0	0.55	0.54		
The program meets my needs	0.80	0.79	0.06	0	0.65	0.51		
The program is valuable to me	0.76	0.78	-0.05	0.10	0.62	0.50		
This is an important program to me	0.76	0.77	0	0.05	0.62	0.50		

I will enjoy doing this program	0.73	0.71	0.06	0.02	0.59	0.46	
I like this program	0.69	0.81	-0.09	-0.01	0.54	0.52	
I think performing this program is worth my time	0.81	0.71	0.20	-0.03	0.67	0.46	
This program required the use of equipment I like to use	0.50	0.52	-0.09	0.12	0.41	0.34	
The program is useful to me	0.78	0.80	0.02	0.01	0.63	0.52	
The program includes everything I expect it to have	0.73	0.57	0.26	-0.01	0.61	0.37	
The program seems exactly right for my goals	0.75	0.69	0.15	-0.04	0.61	0.44	
It is clear to me how this program will improve my strength	0.53	0.54	-0.02	0.06	0.44	0.35	
The exercise program seems Ineffective ------ Effective	0.65	0.63	-0.07	-0.02	0.45	0.41	
The exercise program seems Boring ------ Fun	0.61	0.67	-0.05	0.01	0.49	0.44	
The exercise program seems Inefficient ------ Efficient	0.61	0.60	0.07	-0.02	0.50	0.39	
I can do the program whenever I want to	0.50	0.16	0.31	0.17	0.46		0.22
This program required the use of equipment I have available	0.56	0.13	0.48	0.10	0.51		0.33
I can do the program wherever I want to	0.62	0.20	0.47	0.10	0.56		0.32
The difficulty of each exercise is exactly right for me	0.65	0.32	0.52	-0.09	0.56		0.36
I have access to the equipment needed for this program	0.56	0.09	0.61	0.01	0.52		0.42
The exercise program seems Difficult to do ------ Easy to do	0.58	-0.07	0.86	0	0.56		0.59
The exercise program seems	0.58	-0.05	0.73	0.12	0.57		0.50

Difficult to learn ------ Easy to learn							
The exercise program seems Rigid ------ Flexible	0.60	0.06	0.73	0	0.57	0.50	
The exercise program seems Complicated ------ Straightforward	0.62	0.03	0.73	0.06	0.59	0.50	
I can do this program with a friend	0.52	-0.05	0.11	0.8	0.56		0.62
This program will allow me to exercise with friends	0.54	0.09	-0.04	0.82	0.57		0.63

G = general factor, g = group factor, Value, Ease and Social = names of hypothetical constructs representing the three factors.

As one can observe in Table 5.3, the loadings on the general factor in the bifactor model tend to go down quite a bit relative to the traditional 1-factor model. The difference in loadings seen between the general factor in the bifactor model and the loadings in the 1-factor model supports the multidimensional factor structure.

5.4.2 Item Response Theory

Due to large sample size, (>300), number of items, and low number of missing data, the polytomous IRT model for each separate factor was conducted using the graded response model (Dai et al., 2021; Nguyen et al., 2014). Item information curves for each factor are presented in Figure 5.4. Twelve items were removed as a result of varying difficulty levels, low discriminative value, low information and/or identical information patterns. No additional items were flagged in DIF analyses. Following IRT analyses, the IUSE consisted of 8 items ($IUSE_{Overall}$) and 3 subscales ($IUSE_{Value}$, $IUSE_{Ease}$, and $IUSE_{Social}$). The correlation between the 26-item version ($IUSE_{Long}$) and $IUSE_{Overall}$ was $r = 0.94$. $IUSE_{Overall}$ is presented in Figure 5.5.

The scale score was based on a total score of the three factors as they each were unidimensional but had relatively strong factor intercorrelations (0.47-0.55). Factor analyses indicate that the three factors are unidimensional (low cross loadings; <0.12 in three factor EFA model) and each IRT model fit the data well. For interpretation and simplicity reasons, average scale scoring was considered more appropriate than IRT scoring until further analyses of the significant value of complex IRT scoring have been documented. Scores were instead scored based on the average responses and subsequently transformed to a 0-100 scale (e.g. (raw score -1) *25), as this range might be more intuitive than a 1-5 range and similar to legacy usability scales (e.g. System Usability Scale).

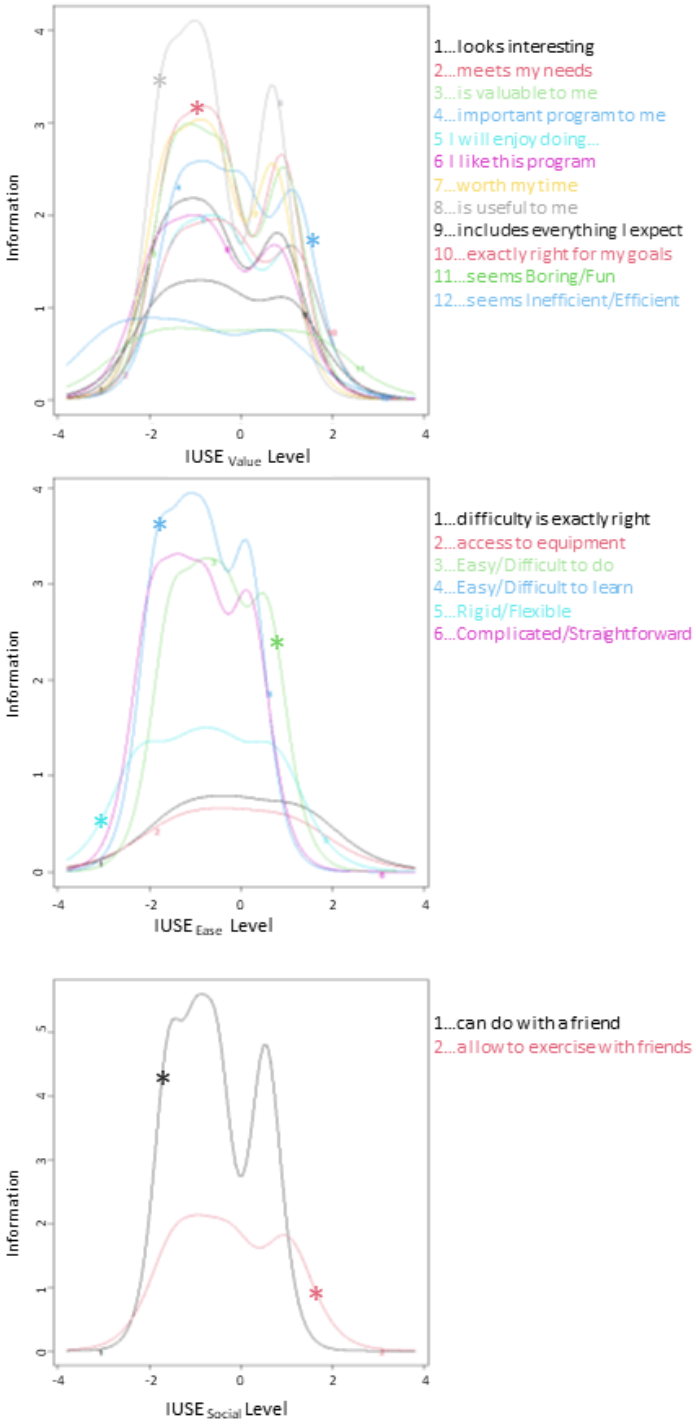


Figure 5.4. Item information curves for each of the subscales.
Footnote: *is denoting the items selected

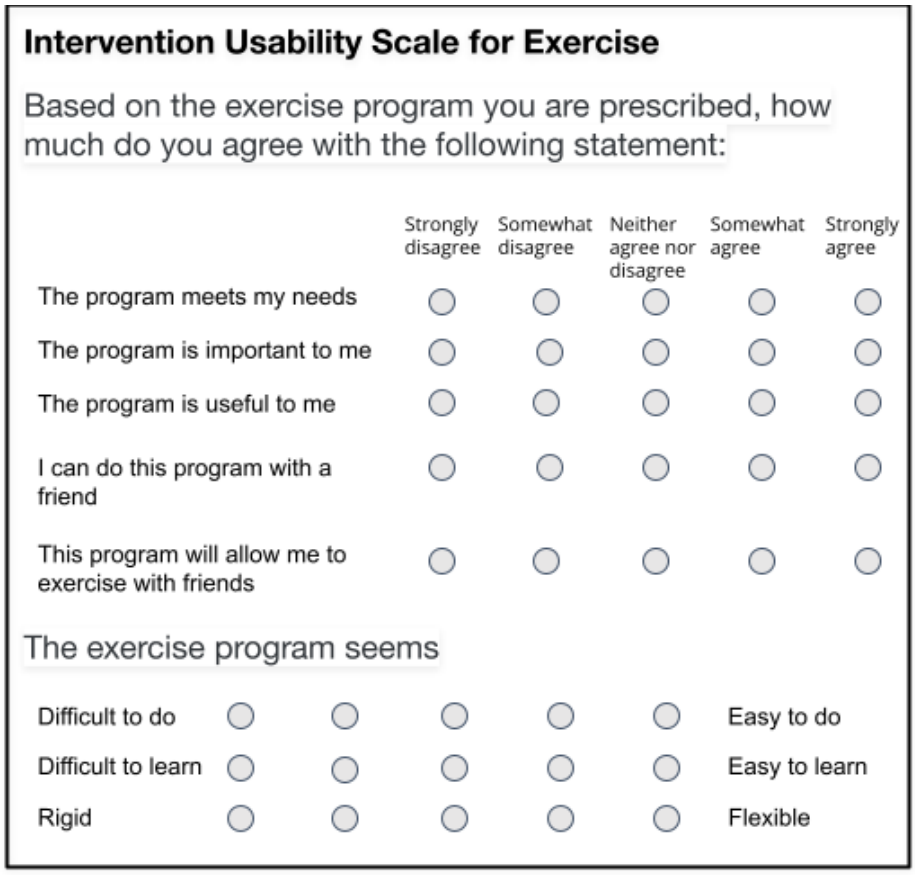


Figure 5.5. Mock-up of the final 8-item version ($IUSE_{Overall}$) of the Intervention Usability Scale for Exercise (IUSE)

5.4.3 Reliability analysis

Internal consistency for the $IUSE_{Overall}$ was Cronbach's α = 0.86, and α = 0.86, α = 0.84 and α = 0.83 for $IUSE_{Value}$, $IUSE_{Ease}$ and $IUSE_{Social}$, respectively. Test-retest descriptives and reliability measures for the $IUSE_{Overall}$ and the 3 subscales are presented in table 5.4.

Table 5.4. Test-retest analyses

	$IUSE_{Overall}$		**$IUSE_{Value}$**		**$IUSE_{Ease}$**		**$IUSE_{Social}$**	
	Test	**Retest**	**Test**	**Retest**	**Test**	**Retest**	**Test**	**Retest**
Descriptives								
Mean	67.2	69.1	55.9	58.7	72.6	74.8	76.0	76.2
SD	14.6	16.3	24.9	27.7	21.2	22.1	20.2	20.3
Range (min-max)	37.5-96.9	31.3-96.9	0-91.7	0-100	25-100	16.7-100	25-100	0-100

Reliability measures				
Test-retest correlation (*r*) [95% CI]	0.78 [0.65, 0.87]	0.79 [0.66, 0.87]	0.85 [0.75, 0.91]	0.83 [0.72, 0.90]
Typical Error	7.3	12.2	8.4	8.3
Mean Difference [95 % CI]	-1.9 [-4.7, 1.0]	-2.7 [-7.5, 2.1]	-2.1 [-5.4, 1.2]	-0.24 [-3.5, 3.0]
Cohen's d_{av}	-0.12	-0.10	0.04	-0.01

5.4.4 Relations to other variables

Criterion (intention to use items and NPS), convergent (IUS, IAM) and discriminant (external motivation items (SIMS)) validity evidence is presented through correlation coefficients in Table 5.5, in addition to correlation between the scale scores of the long and short versions and between subscales.

Table 5.5. Validity evidence

	Intention	**NPS**	**IUS**	**IAM**	**EM**	**IUSE Overall**	**IUSE Value**	**IUSE Ease**	**IUSE Social**
IUSE$_{Overall}$	0.75 (0.72, 0.79)	0.71 (0.66, 0.75)	0.71 (0.67, 0.75)	0.79 (0.76, 0.82)	0.42 (0.35, 0.49)	—	0.88	0.81	0.70
IUSE$_{Value}$	0.82 (0.79, 0.85)	0.72 (0.68, 0.76)	0.55 (0.48, 0.60)	0.80 (0.77, 0.83)	0.42 (0.35, 0.49)		—	0.51	0.48
IUSE$_{Ease}$	0.51 (0.45, 0.57)	0.51 (0.45, 0.57)	0.71 (0.66, 0.75)	0.58 (0.53, 0.64)	0.33 (0.26, 0.41)			—	0.44
IUSE$_{Social}$	0.42 (0.35, 0.49)	0.41 (0.34, 0.48)	0.41 (0.34, 0.48)	0.46 (0.40, 0.53)	0.23 (0.15, 0.31)				—
IUSE$_{Long}$						0.94	0.87	0.72	0.62

All correlation coefficients are Pearson's *r.* NPS = Net Promoter Score. IUS = Intervention Usability Scale. IAM = Intervention Appropriateness Measure. EM = Extrinsic motivation. IUSE$_{Long}$ = 26-item scale.

5.4.5 Hypothesis testing

Paired sample's t-test revealed significant differences in overall usability between PHigh and PLow (mean difference: 13.0, 95% CI 6.82, 19.15, p-value < 0.001, d_{av} = 1.05). Figure 5.6 shows the comparison of responses to PHigh and Plow for the independent sample. PHigh scored significantly higher than PLow on both total scale score and each of the subscales (p < 0.05). The largest difference was found for the IUSE_Ease subscale.

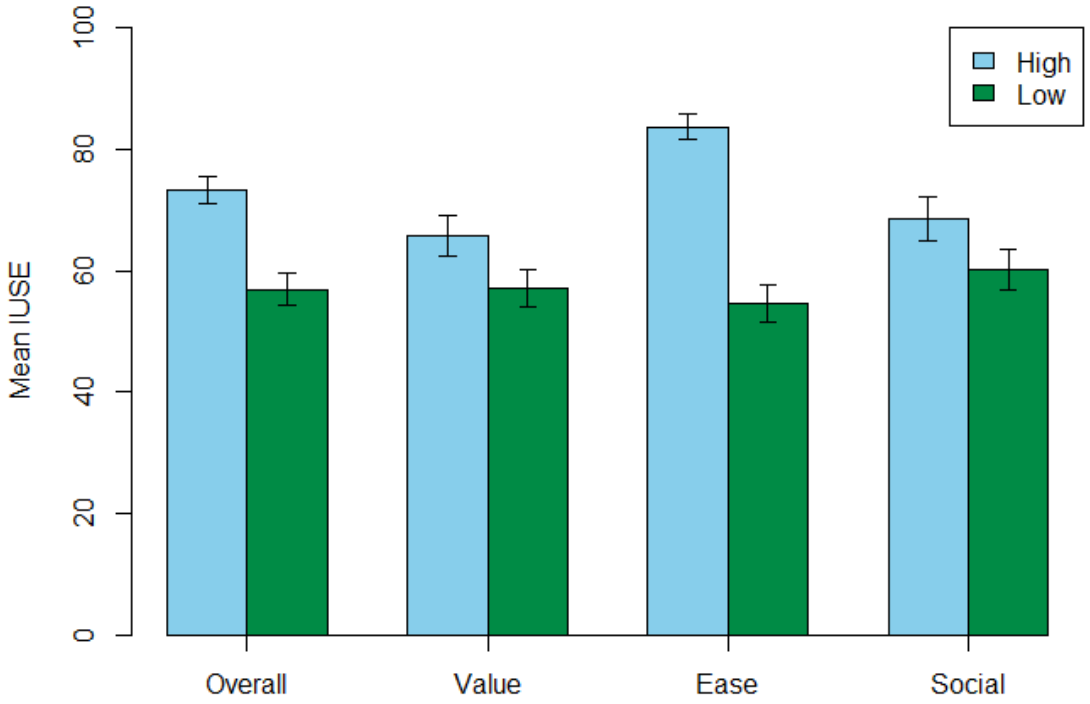


Figure 5.6. Average scores for PHigh and PLow (Error bars = 95 % CI).

5.5 Discussion

The purpose of this study was to develop and assess the psychometric properties of the Intervention Usability Scale for Exercise (IUSE), the first tool for assessing the usability of exercise programs. The 8-item IUSE scale revealed good psychometric properties. The scale has three factors/subscales with items related to either the programs' value/usefulness ($IUSE_{Value}$), ease of use/learnability ($IUSE_{Ease}$), and social aspects ($IUSE_{Social}$).

Items initially loading onto the $IUSE_{Value}$ subscale can be explained by behavioral theories like the Self-Determination Theory (Ryan & Deci, 2000), where the well-performing, remaining items to a large degree mimics some of the items capturing "Value/Usefulness" in the Activity Perception Questionnaire, which is part of the Intrinsic Motivation Inventory (McAuley et al., 1989). Intrinsic motivation serves as a favorable indicator for behavior change intention (Chan & Hagger, 2012). Prior research

by Brühlmann et al. (2018) found a positive association between intrinsic motivation, perceived usability, and the fundamental needs outlined in SDT (Brühlmann et al., 2018). This amalgamation of needs is often termed as need satisfaction, and it seems reasonable to compare this to the satisfaction facet of usability, defined by the ISO as the degree to which a system, product, or service fulfills the user's physical, cognitive, and emotional needs and expectations (International Organization of Standardization, 2018).

Items that make up the $IUSE_{Value}$ (e.g., “The program meets my needs”, “The program is important to me” and “The program is useful to me”) and $IUSE_{Ease}$ (e.g., “The program seems easy/difficult to use”, “The program seems easy/difficult to learn” and “The program seems rigid/flexible to use” are very similar to items included in the Feasibility Implementation Measure, Implementation Appropriateness Measure, and the Usefulness, Satisfaction, and Ease of Use Questionnaire (USE) (Gao et al., 2018; Lund, 2001; Weiner et al., 2017). The USE questionnaire has two subscales related to ease of use and ease of learning, including items similar to the ones included in the second subscale $IUSE_{Ease}$. However, in our scale, these items were delivered in a semantic differential format with two bipolar adjectives making up the far ends of the response option continuum (e.g. Difficult/Easy to use/learn), similar to the User Experience Questionnaire (Schankin et al., 2022). Previous research has suggested that the semantic differential format might perform better than the traditional Likert scale format (e.g. reduced acquiescence bias) (Friborg et al., 2006). Further analyses and research are needed to compare the use of the different modalities.

It is worth discussing the value of including the $IUSE_{Social}$ in assessing exercise program usability and/or for predicting program use. This construct seems less predictive of program use and has lower convergent validity evidence compared to the other subscales. Additionally, it explains about 6% of the variance which can be considered relatively low. However, the items load relatively high on the general factor (> 0.5), as well as being distinct (> 0.6). Additionally, previous research suggests that social support positively influences physical activity and exercise adherence (Anderson et al., 2016; Smith et al., 2023; Tian & Shi, 2022). We therefore believe the inclusion of items aiming to capture the influence of social support might have non-negligible value and should be included until further studies using the scale have been conducted on various groups. The inclusion of $IUSE_{Social}$ makes the IUSE unique relative to other related scales and questionnaires.

Some hypothetical constructs from the content validity phase were expected to perform better and be part of the final version of IUSE. For instance, as interest/enjoyment is considered the self-report measure of intrinsic motivation, which has been shown to affect intention of a behavior, it was expected that including items such as “I will enjoy doing this program”, “This program seems boring/fun” and “This program looks interesting” would have more value in the IUSE. However, there was a strong correlation between the three intrinsic motivation related items and $IUSE_{Value}$ ($r = 0.77$), and from a practical perspective these items can almost be used interchangeably.

Interestingly, when the exercise program stakeholders were asked to rate the importance of constructs to capture with the scale, “Self-efficacy/Confidence”, “Enjoyability/Intrinsic motivation” and “Perceived Effectiveness” were rated top 3 most frequently. Items like “It is clear to me how this program will improve my strength”, “The exercise program seems ineffective/effective” and “The program is effective for improving strength” (removed due to low test-retest reliability) was included to capture the participants’ perceived effectiveness, but none of these items had ideal psychometric properties.

In addition to satisfaction and perceived effectiveness, we aimed to capture perceived efficiency as these are the three components of “usability” as defined by the ISO. Efficiency is defined as “resources used in relation to the results achieved” and includes time, cost, materials, and human effort (International Organization of Standardization, 2018). Efficiency items derived from the content validity phase mostly loaded on the subfactor $IUSE_{Value}$, and not on $IUSE_{Ease}$ as the ISO definition might indicate. None of these items was retained after item reduction, despite previous research suggesting time and resources (e.g. equipment) are barriers for adherence to exercise programs. This might indicate that time and resources may not be as important in the concept of exercise usability.

5.5.1 Clinical and research implications

This study suggests that the IUSE seems to have adequate reliability and validity evidence for assessing usability of exercise programs. The IUSE can be a valuable tool in future exercise program development, both within research and clinical practice, by its ability to identify potential barriers to program implementation and long-term use. IUSE can also be used to test how different populations (e.g., sex, ethnicity, age) perceive different exercise programs. Future studies should utilize the IUSE to assess whether the scale can not only predict the intention/likelihood of using a program, but also actual adherence of the program. This will add evidence to the predictive ability of the scale. In addition, more validity evidence should be added by assessing different exercise interventions/modalities (e.g., prevention vs rehabilitation, cardio vs. strength etc.) and recruiting different populations (e.g., older vs. younger, athletes vs. non-athletes, injured/ill vs. healthy).

5.5.2 Strengths and limitations

This study has several strengths and limitations. To our knowledge, this is the first study to integrate the domain of Human Factors as a measure of addressing the uptake/adherence issue seen in exercise medicine through the development of an exercise intervention usability scale. The IUSE is a short questionnaire that does not demand much time to complete, as well as being easy for exercise intervention stakeholders to distribute to their target users. The IUSE also provides an item database in case some items are more or less relevant in a specific context. Another strength with this study is the large and diverse sample (age, gender, ethnicity) which can be considered

representative of the US population, in addition to the inclusion of individuals with varying levels of activity levels and comfortness with strength training.

The low number of scale items may be considered as both a strength and a limitation. Besides making the scale short, concise, and perhaps more applicable, a low number of items within each subscale can affect the stability and reliability of the factor/subscale. Future studies are needed to test stability and replicability across different samples. Another limitation is that this study did neither have participants actually performing the program nor collect data on actual adherence over time. Intention, although perhaps being the most proximal measure we have, is not the same as actual behavior (Bandura et al., 1999; Chen et al., 2022; Godin & Conner, 2008). Hence, to evaluate the scale's predictive value, future studies should apply the IUSE in prospective studies where data on the actual use of an exercise intervention is collected. Finally, future studies are needed to define cut-offs that indicate the level of program usability and implementation likelihood.

5.6 CONCLUSION

The 8-item IUSE revealed good psychometric properties, indicating it can serve as a reliable tool for assessing usability of exercise interventions. This study also indicated that the concept of exercise program usability is multidimensional and provides a good foundation for future measurement. The ultimate goal is that the IUSE can be a useful tool for exercise intervention developers/designers in enhancing program usability and increasing implementation likelihood. However, future studies are warranted to add validity evidence to the scale's predictive ability, in addition to assessing the psychometric properties of the scale across specific demographic groups and exercise modalities.

Funding: The study has received funding grants through the Health Research Institute at Michigan Technological University.

Acknowledgement: We would like to thank ExorLive AS and Muscle Animations AS for allowing us to use their material in the exercise programs used in this study, in addition to Åsta Lindstad Isum for help with developing the exercise video.

Competing interests: The authors declare no competing interests.

Patient and public involvement: Neither patients nor the public were not involved in the design, or conduct, or reporting or dissemination plans of this research.

6 Conclusion and Future Research

Adherence remains an issue in exercise medicine despite the wealth of research proving the effectiveness of physical activity and exercise interventions on a variety of health and injury outcomes. Adherence is influenced by variables such as exercise program characteristics, social support, knowledge, self-efficacy, and enjoyment (Collado-Mateo et al., 2021; Goddard et al., 2021; Moesch et al., 2022; Ng et al., 2021). Conducting research that is assessing the effectiveness of an intervention is crucial. Equally important is taking measures to ensure that the effective program will be used. Given the proposed 17-year lag in translation from research to practice (Arundale et al., 2022; Green et al., 2009; Morris et al., 2011; Rubin, 2023; Trochim, 2010), the design and assessment of exercise interventions need to adopt a more holistic approach. Through a literature review and three consecutive studies, this dissertation provided valuable insights and construct clarity towards the conceptualization of exercise program usability (e.g. what influences usability of an exercise program).

6.1 Theoretical contributions

Usability and usefulness may influence our acceptability of a new product, uncover errors or barriers, and predict use (Keenan et al., 2022). Nonetheless, there is very little consistency in the use of constructs within existing scales and questionnaires. Their psychometric properties are often insufficient or not reported, and terms like constructs, concepts, attributes, subconstructs, etc., are used interchangeably. Well-defined constructs and validity evidence are warranted to ensure more accurate interpretations of scale scores.

The term usability itself has been criticized for being an umbrella concept with ill-defined constructs (Tractinsky, 2018), whereas others emphasize evidence indicating a consistent underlying construct of usability (Lewis, 2018a). The definition of usability indicates three different constructs: effectiveness, efficiency, and satisfaction (International Organization of Standardization, 2018). In the context of exercise programs, one can think of the purpose of the program (preventing an injury, reducing depression, etc.) as being the “effectiveness” part, whereas the “efficiency” can be considered the time it takes to perform the program or reach the end-goal, or how much money it costs. Usability testing can be conducted through formative (e.g. cognitive task analysis) or summative methods (e.g. scales), both yielding advantages and disadvantages. With usability scales, we typically only measure subjective or perceived effectiveness and efficiency, which can be considered part of the “satisfaction” component. Satisfaction also involves factors like enjoyment, ease of use, and the extent to which these and all other physical, cognitive, and emotional responses that result from using the product meet the user’s needs and expectations (International Organization of Standardization, 2018). Whether there are sub-constructs in scales measuring perceived usability/satisfaction might depend on the context in which they are used.

The term usability and its affiliated scales coincide with other concepts, frameworks, and theories more commonly referred to in exercise medicine (e.g., Implementation Science, the Theoretical Framework of Acceptability, Self-Determination Theory). Existing scales developed based on these related concepts (e.g., Intervention Appropriateness Measure, ENJOY, Theory of Planned Behaviour Measures for Doing Adequate Physical Activity) (Chan et al., 2020; Davidson et al., 2023; Weiner et al., 2017) include items that completely match or substantially overlap with items identified in usability scales. Collectively, these scales provided a good starting point for the development of an exercise intervention usability scale. Several hypothetical constructs were included in the IUSE initially (e.g., usefulness, perceived effectiveness, enjoyability, ease of use, resources). These constructs might have some differences from a theoretical perspective (e.g., perceived effectiveness, value, usefulness). Nonetheless, based on our factor analyses and IRT, many of these constructs seem to substantially overlap and load onto one common latent trait and did not independently account for a considerable amount of the explained variance.

The IUSE was developed to synthesize other scales and items related to user perceptions and factors affecting use into one efficient scale assessing the usability of exercise programs. The items and constructs (subscales) included in the final version are items typically found in other scales, although not all are included in one specific scale like in IUSE. Nevertheless, these items and constructs are now psychometrically evaluated in the context of exercise programs, specifically strength training programs that meet the recommendations for physical activity among healthy adults. Particularly, the application of item response theory in item reduction is rarely reported in related scales (e.g., implementation outcome measures and USE; Gao et al., 2018; Weiner et al., 2017) and enhance the measurement rigor of the IUSE. We have yet to determine whether this scale can predict the actual use of an exercise program and appropriate cut-off scores and if modifications are needed in other subgroups or exercise program types (e.g., aerobic exercise, injury prevention, post-operative rehabilitation).

6.1.1 Practical contributions

The IUSE consists of three subscales that capture an exercise program's value to the user, ease of use, and opportunity for social support. The 8-item IUSE can be distributed to users of exercise programs. The overall usability score is calculated by averaging the response to the 8 items, subtract 1 and then multiply with 25 to get the final scale score on a 0-100 scale (e.g. average score -1) *25). Alternatively, subscales can be calculated independently to see if the score is mainly affected by either the value, ease, or social factor.

Criterion validity analyses indicate that an exercise program's perceived value or usefulness is more closely related to the users' intention to use it than whether it is easy to use/learn or can be conducted with others. Ease of use is still considered an important construct for assessing usability and might be particularly important for initial uptake of an exercise program. The social support construct did not correlate as strongly with

intention to use the program. However, previous research proposes that this is an important factor, and we recommend including it until further research has been conducted.

Surprisingly, items included to capture perceived effectiveness (e.g., “It is clear to me how this program will improve my strength,” “The exercise program seems ineffective/effective,” and “The program is effective for improving strength”) did not perform as well as initially expected. Previous literature, including studies on the ACL IPP, presented above, suggested that perceived effectiveness is an important factor for uptake and adherence. Exercise stakeholder evaluation also indicated that this construct would be critical to capture in the IUSE. Nonetheless, one can argue that perceived effectiveness would be highly correlated with value/usefulness, as it seems unlikely that someone would find a program particularly valuable/useful to them unless they believed it served its purpose.

Items aiming to capture intrinsic motivation/enjoyment (e.g., “I will enjoy doing this program,” “This program looks interesting,” “I like this program,” and “The program seems fun/boring”) did not make it to the final version of IUSE. Enjoyability is considered the self-report measure of intrinsic motivation, which further have been found to predict intention to do prevention and rehabilitation exercises. Previous research on exercise adherence has also emphasized enjoyment as a critical component affecting sustained use. The intrinsic motivation items loaded (highest) onto the same factor as the value/usefulness items in the bifactor model. Three of the items revealed adequate discriminability and information, but they were subsequently removed to shorten the scale as they did not cause substantial loss of information. These items can be considered part of an item bank that researchers/clinicians can choose from and add to the Value subscale if considered appropriate.

Lastly, resources such as time and use of equipment has been suggested as a barrier for use of exercise programs. None of the items assessing time (n = 3) and use of equipment (n = 3) in the initial 36-item scale made it to the final version. Like the intrinsic motivation items, the item “I think performing this program is worth my time” had acceptable discrimination and information but did not add substantial information to the scale. However, this item can also be considered part of an item bank.

Nonetheless, the results of the three studies presented in this dissertation are somewhat corresponding regarding conceptualization of exercise programs. In the first two studies, perceived effectiveness, efficiency, and enjoyability correlated highly with intention to use a novel injury prevention program. Items capturing these hypothetical constructs in IUSE mainly loaded onto the $IUSE_{Value}$, which also correlated highly with intention to use the strength training program. Similarly, items capturing ease of use and learnability had lower correlations with intention measures in both studies. Additionally, we learned that flexibility is an important factor for the implementation of an injury prevention program in the context of team sports. The IUSE item reduction process also indicated that the flexible/rigid item revealed valuable information. Although flexibility can be

considered both beneficial and harmful, it might be a valuable item to detect differences within groups/personas or for different contexts. Likewise, the unique contribution of the $IUSE_{Social}$ might be capturing an important factor for exercise adherence within certain groups.

6.1.2 Next steps

While the IUSE revealed good psychometric properties, we have yet to determine whether this scale can predict the actual use of an exercise program. Further research is needed to add predictive validity evidence to the scale. This can for instance be examined as part of an intervention study involving the use of an exercise program and that includes a follow-up assessment of adherence. Furthermore, an appropriate cut-off score for estimating usable programs or programs that are likely versus unlikely to be implemented would be valuable. Cut-off scores make it easier for program designers to determine whether modifications of the program are needed in order to enhance adherence.

Eventually, validation of IUSE in other languages and adding cross-cultural validity evidence is warranted to ensure that the scale is reliable across cultural contexts. This is an iterative process where the psychometric properties of the scale are assessed and compared across different demographic groups and languages. This process includes content validation (e.g. does one or more items need to be modified), factor analyses, IRT, reliability testing and criterion validity evidence. This process might be repeated until the results of comparisons are satisfactory. Similarly, comparing the psychometric properties across exercise modalities (e.g., aerobic exercise, injury prevention, post-operative rehabilitation) is needed to explore whether the same constructs (e.g., perceived value, ease of use and social support) are predicting adherence.

A few items from the initial 36-item IUSE that are not part of the final 8-item scale revealed good psychometric properties but were removed due to redundancy within each subscale. These items can be included in an IUSE item bank for alternative well-performing items. This allows clinicians or researchers to adapt the scale to their users or research questions. For instance, items can be added or replaced within each subscale, or one can choose to only use one subscale. The development of an item bank was not part of this project but is considered part of the next steps. The development of an adaptive scale version might also be worth exploring. This would imply that the IUSE dynamically adjusts its items or response options based on the users' characteristics or previous responses. For instance, if we find that the scale needs to be modified for a specific subgroup of users or specific exercise modalities, the adaptive scale will provide the targeted items based on premade decision rules.

In addition to program content, program delivery has been proposed as an important factor affecting uptake and adherence (McCall et al., 2014; O'Brien et al., 2017; O'Brien & Finch, 2016; Varnfield et al., 2014). For instance, in the scale development study, the exercise programs were delivered using an animated video or a text-based pamphlet,

whereas in study 1 and 2, the injury prevention program was delivered through in-person presentations and supervision by the research team. Other common delivery modalities include technology-based devices such as smartphone applications providing reminders, statistics, and outcome measures. These different modalities may affect individuals' motivation and behavior differently. Hence, assessing delivery aspects of the exercise program through a usability scale in addition to program content aspects may also provide valuable information on barriers and likelihood of use.

The development of the IUSE was an important first step in the integration of usability testing in exercise medicine. We believe that there is a great potential for the scale to help clinicians and researchers enhance program design and facilitate long term use. Further research is crucial to add predictive validity evidence and determine appropriate cut-off scores indicating if the exercise program is likely to be used or not.

7 Reference List

Abbey, J. D., & Meloy, M. G. (2017). Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management*, *53*, 63–70.

Ageberg, E., Brodin, E. M., Linnéll, J., Moesch, K., Donaldson, A., Adébo, E., Benjaminse, A., Ekengren, J., Granér, S., Johnson, U., & others. (2022). Cocreating injury prevention training for youth team handball: Bridging theory and practice. *BMJ Open Sport & Exercise Medicine*, *8*(2), e001263.

Ageberg, E., Bunke, S., Linnéll, J., & Moesch, K. (2024). Co-creating holistic injury prevention training for youth handball: Development of an intervention targeting end-users at the individual, team, and organizational levels. *BMC Sports Science, Medicine and Rehabilitation*, *16*(1), 10.

Ageberg, E., Bunke, S., Lucander, K., Nilsen, P., & Donaldson, A. (2019). Facilitators to support the implementation of injury prevention training in youth handball: A concept mapping approach. *Scandinavian Journal of Medicine & Science in Sports*, *29*(2), 275–285. https://doi.org/10.1111/sms.13323

Ajzen, I. (1985). *From intentions to actions: A theory of planned behavior*. Springer.

American Educational Research Association. (2018). *Standards for educational and psychological testing*. American Educational Research Association.

American Psychological Association. (2020). APA PsycTests methodology field values. *American Psychological Association*. https://www.apa.org/pubs/databases/training/psyctests-method-values

Anderson, P. J., Bovard, R. S., Wang, Z., Beebe, T. J., & Murad, M. H. (2016). A survey of social support for exercise and its relationship to health behaviours and health status among endurance Nordic skiers. *BMJ Open*, *6*(6), e010259.

Arnason, A., Andersen, T., Holme, I., Engebretsen, L., & Bahr, R. (2008). Prevention of hamstring strains in elite soccer: An intervention study. *Scandinavian Journal of Medicine & Science in Sports*, *18*(1), 40–48.

Arundale, A. J., Silvers-Granelli, H. J., & Myklebust, G. (2022). ACL injury prevention: Where have we come from and where are we going? *Journal of Orthopaedic Research®*, *40*(1), 43–54.

Assila, A., Ezzedine, H., & others. (2016). Standardized usability questionnaires: Features and quality focus. *Electronic Journal of Computer Science and Information Technology*, *6*(1).

Bailey, D. L., Bishop, A., McCray, G., Foster, N. E., & Holden, M. A. (2024). A new measure of exercise adherence: The ATEMPT (Adherence To Exercise for Musculoskeletal Pain Tool). *British Journal of Sports Medicine*, *58*(2), 73–80.

Bandura, A., Freeman, W. H., & Lightsey, R. (1999). *Self-efficacy: The exercise of control*. springer.

Batsis, J. A., Boateng, G. G., Seo, L. M., Petersen, C. L., Fortuna, K. L., Wechsler, E. V., Peterson, R. J., Cook, S. B., Pidgeon, D., Dokko, R. S., & others. (2019). Development and usability assessment of a connected resistance exercise band application for strength-monitoring. *World Academy of Science, Engineering and Technology*, *13*(5), 340.

Bauer, M. S., Damschroder, L., Hagedorn, H., Smith, J., & Kilbourne, A. M. (2015). An introduction to implementation science for the non-specialist. *BMC Psychology*, *3*(1), 1–12.

Bennie, J. A., Lee, D., Khan, A., Wiesner, G. H., Bauman, A. E., Stamatakis, E., & Biddle, S. J. (2018). Muscle-strengthening exercise among 397,423 US adults: Prevalence, correlates, and associations with health conditions. *American Journal of Preventive Medicine*, *55*(6), 864–874.

Bevan, N., Carter, J., & Harker, S. (2015). ISO 9241-11 revised: What have we learnt about usability since 1998? *Human-Computer Interaction: Design and Evaluation: 17th International Conference, HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015, Proceedings, Part I 17*, 143–151.

Bill, K., Mai, P., Willwacher, S., Krosshaug, T., & Kersting, U. G. (2022). Athletes with high knee abduction moments show increased vertical center of mass excursions and knee valgus angles across sport-specific fake-and-cut tasks of different complexities. *Frontiers in Sports and Active Living*, *4*, 983889.

Bizzini, M., & Dvorak, J. (2015). FIFA 11+: An effective programme to prevent football injuries in various player groups worldwide-a narrative review. *British Journal of Sports Medicine*, *49*(9), 577–579. https://doi.org/10.1136/bjsports-2015-094765

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Frontiers in Public Health*, *6*. https://doi.org/10.3389/fpubh.2018.00149

Boden, B. P., Torg, J. S., Knowles, S. B., & Hewett, T. E. (2009). Video analysis of anterior cruciate ligament injury: Abnormalities in hip and ankle kinematics. *The American Journal of Sports Medicine*, *37*(2), 252–259.

Brooke, J. (1995). SUS: A quick and dirty usability scale. *Usability Eval. Ind.*, *189*.

Brooke, J. (2013). SUS: a retrospective. *Journal of Usability Studies*, *8*(2), 29–40.

Bruder, A. M., Donaldson, A., Mosler, A. B., Patterson, B. E., Haberfield, M., Mentiplay, B. F., Clifton, P., OAM, N. D. L., & Crossley, K. M. (2023). Creating PreP to play pro for women playing elite Australian football: A how-to guide for developing injury-prevention programs. *Journal of Sport and Health Science*, *12*(1), 130–138.

Bruder, A. M., Patterson, B. E., Crossley, K. M., Mosler, A. B., Haberfield, M. J., Hägglund, M., Culvenor, A. G., Cowan, S. M., & Donaldson, A. (2024). If we build it together, will they use it? A mixed-methods study evaluating the implementation of Prep-to-Play PRO: an injury prevention programme for women's elite Australian Football. *British Journal of Sports Medicine*.

Brühlmann, F., Vollenwyder, B., Opwis, K., & Mekler, E. D. (2018). Measuring the "why" of interaction: Development and validation of the user motivation inventory (umi). *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–13.

Chan, D. K. C., & Hagger, M. S. (2012). Self-determined forms of motivation predict sport injury prevention and rehabilitation intentions. *Journal of Science and Medicine in Sport*, *15*(5), 398–406. https://doi.org/10.1016/j.jsams.2012.03.016

Chan, D. K., Stenling, A., Yusainy, C., Hikmiah, Z., Ivarsson, A., Hagger, M. S., Rhodes, R. E., & Beauchamp, M. R. (2020). Editor's Choice: Consistency tendency and the theory of planned behavior: A randomized controlled crossover trial in a physical activity context. *Psychology & Health*, *35*(6), 665–684.

Chen, C., Finne, E., Kopp, A., & Jekauc, D. (2020). Can positive affective variables mediate intervention effects on physical activity? A systematic review and meta-analysis. *Frontiers in Psychology*, *11*, 587757.

Chen, C., Weyland, S., Fritsch, J., Woll, A., Niessner, C., Burchartz, A., Schmidt, S. C., & Jekauc, D. (2021). A short version of the physical activity enjoyment scale: Development and psychometric properties. *International Journal of Environmental Research and Public Health*, *18*(21), 11035.

Chen, H., Sun, H., & Dai, J. (2017). Peer support and adolescents' physical activity: The mediating roles of self-efficacy and enjoyment. *Journal of Pediatric Psychology*, *42*(5), 569–577.

Chen, W.-H., Lenderking, W., Jin, Y., Wyrwich, K. W., Gelhorn, H., & Revicki, D. A. (2014). Is Rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An example using PROMIS pain behavior item bank data. *Quality of Life Research*, *23*, 485–493.

Chen, Y., Yao, S.-J., Ma, Q.-S., Shao, W., Liu, C., & Guo, K.-L. (2022). The relationship between exercise intention and exercise behavior of junior school students: An analysis of chain mediating effect. *Frontiers in Psychology*, *13*, 935264.

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment*, *31*(12), 1412.

Cokluk, O., & Kayri, M. (2011). The Effects of Methods of Imputation for Missing Values on the Validity and Reliability of Scales. *Educational Sciences: Theory and Practice*, *11*(1), 303–309.

Collado-Mateo, D., Lavín-Pérez, A. M., Peñacoba, C., Del Coso, J., Leyton-Román, M., Luque-Casado, A., Gasque, P., Fernández-del-Olmo, M. Á., & Amado-Alonso, D. (2021). Key factors associated with adherence to physical exercise in patients with chronic diseases and older adults: An umbrella review. *International Journal of Environmental Research and Public Health*, *18*(4), 2023.

Comrey, A. L., & Lee, H. B. (2013). *A first course in factor analysis*. Psychology press.

Dai, S., Vo, T. T., Kehinde, O. J., He, H., Xue, Y., Demir, C., & Wang, X. (2021). Performance of polytomous IRT models with rating scale data: An investigation over sample size, instrument length, and missing data. *Frontiers in Education*, *6*, 721963.

Davidson, S. S., Keebler, J. R., Zhang, T., Chaparro, B., Szalma, J., & Frederick, C. M. (2023). The development and validation of a universal enjoyment measure: The enjoy scale. *Current Psychology*, *42*(21), 17733–17745.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 319–340.

Dempsey, A. R., Lloyd, D. G., Elliott, B. C., Steele, J. R., Munro, B. J., & Russo, K. A. (2007). The effect of technique change on knee loads during sidestep cutting. *Medicine & Science in Sports & Exercise*, *39*(10), 1765–1773.

Dishman, R. K., Motl, R. W., Saunders, R., Felton, G., Ward, D. S., Dowda, M., & Pate, R. R. (2005). Enjoyment mediates effects of a school-based physical-activity intervention. *Medicine & Science in Sports & Exercise*, *37*(3), 478–487.

Donaldson, A., Callaghan, A., Bizzini, M., Jowett, A., Keyzer, P., & Nicholson, M. (2019). A concept mapping approach to identifying the barriers to implementing an evidence-based sports injury prevention programme. *Injury Prevention : Journal of the International Society for Child and Adolescent Injury Prevention*, *25*(4), 244–251. https://doi.org/10.1136/injuryprev-2017-042639

Downey, R. G., & King, C. V. (1998). Missing data in Likert ratings: A comparison of replacement methods. *The Journal of General Psychology*, *125*(2), 175–191.

Evans, H. E., Forbes, C. C., Galvão, D. A., Vandelanotte, C., Newton, R. U., Wittert, G., Chambers, S., Vincent, A. D., Kichenadasse, G., Girard, D., & others. (2021). Usability, acceptability, and safety analysis of a computer-tailored web-based exercise intervention (ExerciseGuide) for individuals with metastatic prostate cancer: Multi-methods laboratory-based study. *JMIR Cancer*, *7*(3), e28370.

Finch, C. F. (2011). No longer lost in translation: The art and science of sports injury prevention implementation research. *British Journal of Sports Medicine*, *45*(16), 1253. https://doi.org/10.1136/bjsports-2011-090230

Finstad, K. (2010). The Usability Metric for User Experience. *Interacting with Computers*, *22*, 323–327. https://doi.org/10.1016/j.intcom.2010.04.004

Fishman, J., Lushin, V., & Mandell, D. S. (2020). Predicting implementation: Comparing validated measures of intention and assessing the role of motivation when designing behavioral interventions. *Implementation Science Communications*, *1*(1), 1–10.

Foss, K. D. B., Thomas, S., Khoury, J. C., Myer, G. D., & Hewett, T. E. (2018). A school-based neuromuscular training program and sport-related injury incidence: A prospective randomized controlled clinical trial. *Journal of Athletic Training*, *53*(1), 20–28.

Friborg, O., Martinussen, M., & Rosenvinge, J. H. (2006). Likert-based vs. Semantic differential-based scorings of positive psychological constructs: A psychometric comparison of two versions of a scale measuring resilience. *Personality and Individual Differences*, *40*(5), 873–884.

Frøkjær, E., Hertzum, M., & Hornbæk, K. (2000). Measuring Usability: Are Effectiveness, Efficiency, and Satisfaction Really Correlated? *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 345–352. https://doi.org/10.1145/332040.332455

Gao, M., Kortum, P., & Oswald, F. (2018). Psychometric evaluation of the use (usefulness, satisfaction, and ease of use) questionnaire for reliability and validity. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 1414–1418.

Godin, G., & Conner, M. (2008). Intention-behavior relationship based on epidemiologic indices: An application to physical activity. *American Journal of Health Promotion*, *22*(3), 180–182.

Gornitzky, A. L., Lott, A., Yellin, J. L., Fabricant, P. D., Lawrence, J. T., & Ganley, T. J. (2016). Sport-Specific Yearly Risk and Incidence of Anterior Cruciate Ligament Tears in High School Athletes: A Systematic Review and Meta-analysis. *The American Journal of Sports Medicine*, *44*(10), 2716–2723. https://doi.org/10.1177/0363546515617742

Green, L. W., Ottoson, J. M., Garcia, C., & Hiatt, R. A. (2009). Diffusion theory and knowledge dissemination, utilization, and integration in public health. *Annual Review of Public Health*, *30*, 151–174.

Halvorsen, K. C., Marx, R. G., Wolfe, I., Taber, C., Jivanelli, B., Pearle, A. D., & Ling, D. I. (2023). Higher Adherence to Anterior Cruciate Ligament Injury Prevention Programs Is Associated With Lower Injury Rates: A Meta-Analysis and Meta-Regression. *HSS Journal®*, *19*(2), 154–162.

Hass, C. (2019). A practical guide to usability testing. *Consumer Informatics and Digital Health: Solutions for Health and Health Care*, 107–124.

Hassenzahl, M. (2007). The hedonic/pragmatic model of user experience. *Towards a UX Manifesto*, *10*, 2007.

Hawkinson, L. E., Yates, L., Minnig, M. C., Register-Mihalik, J. K., Golightly, Y. M., & Padua, D. A. (2022). Understanding youth sport coaches' perceptions of

evidence-based injury-prevention training programs: A systematic literature review. *Journal of Athletic Training*, *57*(9–10), 877–893.

Hawley-Hague, H., Tacconi, C., Mellone, S., Martinez, E., Ford, C., Chiari, L., Helbostad, J., Todd, C., & others. (2020). Smartphone apps to support falls rehabilitation exercise: App development and usability and acceptability study. *JMIR mHealth and uHealth*, *8*(9), e15460.

Heywood, R., McCarthy, A. L., & Skinner, T. L. (2017). Safety and feasibility of exercise interventions in patients with advanced cancer: A systematic review. *Supportive Care in Cancer*, *25*, 3031–3050.

Hodrien, A., Fernando, T., & others. (2021). A review of post-study and post-task subjective questionnaires to guide assessment of system usability. *Journal of Usability Studies*, *16*(3), 203–232.

Hornbæk, K., & Law, E. L.-C. (2007). Meta-Analysis of Correlations among Usability Measures. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 617–626. https://doi.org/10.1145/1240624.1240722

Houts, P. S., Doak, C. C., Doak, L. G., & Loscalzo, M. J. (2006). The role of pictures in improving health communication: A review of research on attention, comprehension, recall, and adherence. *Patient Education and Counseling*, *61*(2), 173–190. https://doi.org/10.1016/j.pec.2005.05.004

International Organization of Standardization. (2018). *Ergonomics of human-system interaction—Part 11: Usability: Definitions and concepts* (BS EN ISO 9241-11:2018). https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en

Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials–a practical guide with flowcharts. *BMC Medical Research Methodology*, *17*(1), 1–10.

Jekauc, D. & others. (2015). Enjoyment during exercise mediates the effects of an intervention on exercise adherence. *Psychology*, *6*(01), 48.

Joy, E. A., Taylor, J. R., Novak, M. A., Chen, M., Fink, B. P., & Porucznik, C. A. (2013). Factors influencing the implementation of anterior cruciate ligament injury prevention strategies by girls soccer coaches. *Journal of Strength and Conditioning Research*, *27*(8), 2263–2269. https://doi.org/10.1519/JSC.0b013e31827ef12e

Keenan, H. L., Duke, S. L., Wharrad, H. J., Doody, G. A., & Patel, R. S. (2022). Usability: An introduction to and literature review of usability testing for educational resources in radiation oncology. *Technical Innovations & Patient Support in Radiation Oncology*.

Kendzierski, D., & DeCarlo, K. J. (1991). Physical activity enjoyment scale: Two validation studies. *Journal of Sport & Exercise Psychology*, *13*(1).

Kim, J., Park, H.-A., & others. (2012). Development of a health information technology acceptance model using consumers' health behavior intention. *Journal of Medical Internet Research*, *14*(5), e2143.

Koga, H., Nakamae, A., Shima, Y., Iwasa, J., Myklebust, G., Engebretsen, L., Bahr, R., & Krosshaug, T. (2010). Mechanisms for Noncontact Anterior Cruciate Ligament

Injuries: Knee Joint Kinematics in 10 Injury Situations from Female Team Handball and Basketball. *The American Journal of Sports Medicine*, *38*(11), 2218–2225. https://doi.org/10.1177/0363546510373570

Kristianslund, E., Faul, O., Bahr, R., Myklebust, G., & Krosshaug, T. (2014). Sidestep cutting technique and knee abduction loading: Implications for ACL prevention exercises. *British Journal of Sports Medicine*, *48*(9), 779–783.

Krosshaug, T., Nakamae, A., Boden, B. P., Engebretsen, L., Smith, G., Slauterbeck, J. R., Hewett, T. E., & Bahr, R. (2007). Mechanisms of anterior cruciate ligament injury in basketball: Video analysis of 39 cases. *The American Journal of Sports Medicine*, *35*(3), 359–367.

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*(4), 355–362.

Leiner, D. J. (2019). Too fast, too straight, too weird: Non-reactive indicators for meaningless data in internet surveys. *Survey Research Methods*, *13*(3), 229–248.

Lewis, J. R. (2018a). Is the report of the death of the construct of usability an exaggeration? *Journal of Usability Studies*, *14*(1), 1–7.

Lewis, J. R. (2018b). The System Usability Scale: Past, Present, and Future. *International Journal of Human–Computer Interaction*, *34*(7), 577–590. https://doi.org/10.1080/10447318.2018.1455307

Lewis, J. R., & Sauro, J. (2021). Usability and user experience: Design and evaluation. *Handbook of Human Factors and Ergonomics*, 972–1015.

Lewis, J., & Sauro, J. (2017). Revisiting the Factor Structure of the System Usability Scale. *Journal of Usability Studies*, *12*, 183–192.

Lonsdale, C., Hodge, K., & Rose, E. A. (2008). The Behavioral Regulation in Sport Questionnaire (BRSQ): Instrument development and initial validity evidence. *Journal of Sport and Exercise Psychology*, *30*(3), 323–355.

Lund, A. M. (2001). Measuring usability with the use questionnaire12. *Usability Interface*, *8*(2), 3–6.

Lyon, A. R., & Koerner, K. (2016). User-centered design for psychosocial intervention development and implementation. *Clinical Psychology: Science and Practice*, *23*(2), 180–200.

Lyon, A. R., Pullmann, M. D., Jacobson, J., Osterhage, K., Al Achkar, M., Renn, B. N., Munson, S. A., & Areán, P. A. (2021). Assessing the usability of complex psychosocial interventions: The Intervention Usability Scale. *Implementation Research and Practice*, *2*, 2633489520987828.

Macefield, R. (2009). How to specify the participant group size for usability studies: A practitioner's guide. *Journal of Usability Studies*, *5*(1), 34–45.

Madan, A., & Kumar, S. (2012). Usability evaluation methods: A literature review. *International Journal of Engineering Science and Technology*, *4*.

Mator, J. D., Lehman, W. E., McManus, W., Powers, S., Tiller, L., Unverricht, J. R., & Still, J. D. (2021). Usability: Adoption, measurement, value. *Human Factors*, *63*(6), 956–973.

McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport*, *60*(1), 48–58.

McCall, A., Carling, C., Nedelec, M., Davison, M., Le Gall, F., Berthoin, S., & Dupont, G. (2014). Risk factors, testing and preventative strategies for non-contact injuries in professional football: Current perceptions and practices of 44 teams from various premier leagues. *British Journal of Sports Medicine*, *48*(18), 1352–1357.

McKay, C. D., Merrett, C. K., & Emery, C. A. (2016). Predictors of FIFA 11+ Implementation Intention in Female Adolescent Soccer: An Application of the Health Action Process Approach (HAPA) Model. *International Journal of Environmental Research and Public Health*, *13*(7). https://doi.org/10.3390/ijerph13070657

McKay, C. D., Steffen, K., Romiti, M., Finch, C. F., & Emery, C. A. (2014). The effect of coach and player injury knowledge, attitudes and beliefs on adherence to the FIFA 11+ programme in female youth soccer. *British Journal of Sports Medicine*, *48*(17), 1281–1286. https://doi.org/10.1136/bjsports-2014-093543

Mehra, S., Visser, B., Cila, N., van den Helder, J., Engelbert, R. H., Weijs, P. J., Kröse, B. J., & others. (2019). Supporting older adults in exercising with a tablet: A usability study. *JMIR Human Factors*, *6*(1), e11598.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741.

Minnig, M. C., Hawkinson, L., Root, H. J., Driban, J., DiStefano, L. J., Callahan, L., Ambrose, K. R., Spang, J. T., & Golightly, Y. M. (2022). Barriers and facilitators to the adoption and implementation of evidence-based injury prevention training programmes: A narrative review. *BMJ Open Sport & Exercise Medicine*, *8*(3), e001374.

Moesch, K., Bunke, S., Linnéll, J., Brodin, E. M., Donaldson, A., & Ageberg, E. (2022). "Yeah, i mean, you're going to handball, so you want to use balls as much as possible at training": End-users' perspectives of injury prevention training for youth handball players. *International Journal of Environmental Research and Public Health*, *19*(6), 3402.

Mokkink, L. B., Terwee, C. B., Knol, D. L., Stratford, P. W., Alonso, J., Patrick, D. L., Bouter, L. M., & De Vet, H. C. (2010). The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Medical Research Methodology*, *10*(1), 1–8.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, *63*(7), 737–745.

Mol, M., van Schaik, A., Dozeman, E., Ruwaard, J., Vis, C., Ebert, D. D., Etzelmueller, A., Mathiasen, K., Moles, B., Mora, T., & others. (2020). Dimensionality of the

system usability scale among professionals using internet-based interventions for depression: A confirmatory factor analysis. *BMC Psychiatry*, *20*(1), 1–10.

Møller, M., Zebis, M. K., Myklebust, G., Lind, M., Wedderkopp, N., & Bekker, S. (2021). "Is it fun and does it enhance my performance?" – Key implementation considerations for injury prevention programs in youth handball. *Journal of Science and Medicine in Sport*, *24*(11), 1136–1142. https://doi.org/10.1016/j.jsams.2021.04.017

Morgado, F. F., Meireles, J. F., Neves, C. M., Amaral, A., & Ferreira, M. E. (2017). Scale development: Ten main limitations and recommendations to improve future research practices. *Psicologia: Reflexão e Crítica*, *30*.

Morris, Z. S., Wooding, S., & Grant, J. (2011). The answer is 17 years, what is the question: Understanding time lags in translational research. *Journal of the Royal Society of Medicine*, *104*(12), 510–520.

Myklebust, G., Engebretsen, L., Brækken, I. H., Skjølberg, A., Olsen, O.-E., & Bahr, R. (2003). Prevention of anterior cruciate ligament injuries in female team handball players: A prospective intervention study over three seasons. *Clinical Journal of Sport Medicine*, *13*(2), 71–78.

Nguyen, T. H., Han, H.-R., Kim, M. T., & Chan, K. S. (2014). An introduction to item response theory for patient-reported outcome measurement. *The Patient-Patient-Centered Outcomes Research*, *7*, 23–35.

Nielsen, J., & Landauer, T. K. (1993). A mathematical model of the finding of usability problems. *Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems*, 206–213.

Norcross, M. F., Johnson, S. T., Bovbjerg, V. E., Koester, M. C., & Hoffman, M. A. (2016). Factors influencing high school coaches' adoption of injury prevention programs. *Journal of Science and Medicine in Sport*, *19*(4), 299–304. https://doi.org/10.1016/j.jsams.2015.03.009

Nunnally, J. C., & Bernstein, I. (1994). *Psychometric Theory. New York, NY: McGraw-Hall*. Inc.

O'Brien, J., & Finch, C. F. (2016). Injury prevention exercise programmes in professional youth soccer: Understanding the perceptions of programme deliverers. *BMJ Open Sport & Exercise Medicine*, *2*(1), e000075. https://doi.org/10.1136/bmjsem-2015-000075

O'Brien, J., Young, W., & Finch, C. F. (2017). The delivery of injury prevention exercise programmes in professional youth soccer: Comparison to the FIFA 11+. *Journal of Science and Medicine in Sport*, *20*(1), 26–31.

Padua, D. A., Frank, B., Donaldson, A., de la Motte, S., Cameron, K. L., Beutler, A. I., DiStefano, L. J., & Marshall, S. W. (2014). Seven steps for developing and implementing a preventive training program: Lessons learned from JUMP-ACL and beyond. *Clinics in Sports Medicine*, *33*(4), 615–632.

Pedersen, B. K., & Saltin, B. (2015). Exercise as medicine–evidence for prescribing exercise as therapy in 26 different chronic diseases. *Scandinavian Journal of Medicine & Science in Sports*, *25*, 1–72.

Petushek, E., & Donaldson, A. (2020). Injury Prevention in Sport–Design with Implementation in Mind. In *Human Factors and Ergonomics in Sport* (pp. 57–70). CRC Press.

Petushek, E. J., Sugimoto, D., Stoolmiller, M., Smith, G., & Myer, G. D. (2018). Evidence-Based Best-Practice Guidelines for Preventing Anterior Cruciate Ligament Injuries in Young Female Athletes: A Systematic Review and Meta-analysis. *The American Journal of Sports Medicine*, *47*(7), 1744–1753. https://doi.org/10.1177/0363546518782460

Plattner, H. (2016). *Institute of Design at Stanford. An introduction to DT PROCESS GUIDE.* https://web.stanford.edu/~mshanks/MichaelShanks/files/509554.pdf

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., & Hensley, M. (n.d.). Health, P., i, M., & Research, MHS (2011). Outcomes for implementation research: Conceptual distinctions. *Measurement Challenges, and Research Agenda*, *38*(2), 65–76.

Prodromos, C. C., Han, Y., Rogowski, J., Joyce, B., & Shi, K. (2007). A meta-analysis of the incidence of anterior cruciate ligament tears as a function of gender, sport, and a knee injury-reduction regimen. *Arthroscopy : The Journal of Arthroscopic & Related Surgery : Official Publication of the Arthroscopy Association of North*

America and the International Arthroscopy Association, *23*(12), 1320-1325.e6. https://doi.org/10.1016/j.arthro.2007.07.003

Reichheld, F. F. (2003). The one number you need to grow. *Harvard Business Review*, *81*(12), 46–55.

Richmond, S. A., Donaldson, A., Macpherson, A., Bridel, W., van den Berg, C., Finch, C. F., Hagel, B., & Emery, C. A. (2020). Facilitators and Barriers to the Implementation of iSPRINT: A Sport Injury Prevention Program in Junior High Schools. *Clinical Journal of Sport Medicine*, *30*(3). https://journals.lww.com/cjsportsmed/Fulltext/2020/05000/Facilitators_and_Barriers_to_the_Implementation_of.6.aspx

Root, H. J., Frank, B. S., Denegar, C. R., Casa, D. J., Gregorio, D. I., Mazerolle, S. M., & DiStefano, L. J. (2019). Application of a preventive training program implementation framework to youth soccer and basketball organizations. *Journal of Athletic Training*, *54*(2), 182–191.

Rosenstock, I. M. (1974). The health belief model and preventive health behavior. *Health Education Monographs*, *2*(4), 354–386.

Rubin, R. (2023). It takes an average of 17 years for evidence to change practice—The burgeoning field of implementation science seeks to speed things up. *Jama*, *329*(16), 1333–1336.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037//0003-066x.55.1.68

Ryan, R. M., Ryan, W. S., Di Domenico, S. I., & Deci, E. L. (2019). *The nature and the conditions of human autonomy and flourishing: Self-determination theory and basic psychological needs*.

Salmond, S. S. (2008). Evaluating the reliability and validity of measurement instruments. *Orthopaedic Nursing*, *27*(1), 28–30.

Sandelowski, M. (2000). Whatever happened to qualitative description? *Research in Nursing & Health*, *23*(4), 334–340.

Sauer, J., Sonderegger, A., & Schmutz, S. (2020). Usability, user experience and accessibility: Towards an integrative model. *Ergonomics*, *63*(10), 1207–1220.

Sauro, J., & Kindlund, E. (2005). A method to standardize usability metrics into a single score. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 401–409.

Sauro, J., & Lewis, J. R. (2009). Correlations among prototypical usability metrics: Evidence for the construct of usability. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1609–1618.

Schankin, A., Budde, M., Riedel, T., & Beigl, M. (2022). Psychometric properties of the user experience questionnaire (UEQ). *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–11.

Sekhon, M., Cartwright, M., & Francis, J. J. (2017). Acceptability of healthcare interventions: An overview of reviews and development of a theoretical framework. *BMC Health Services Research*, *17*(1), 1–13.

Shaw, I., Shaw, B., Brown, G., & Shariat, A. (2016). Review of the role of resistance training and musculoskeletal injury prevention and rehabilitation. *Int J Clin Exp Med*, *2016*, 1–5.

Smit, E., Bouwstra, H., Van Der Wouden, J., Hertogh, C., Wattel, E., Roorda, L., & Terwee, C. (2020). Development of a Patient-Reported Outcomes Measurement Information System (PROMIS®) short form for measuring physical function in geriatric rehabilitation patients. *Quality of Life Research*, *29*, 2563–2572.

Smith, G. S., Moyle, W., & Burton, N. W. (2023). The Relationship between Social Support for Physical Activity and Physical Activity across Nine Years in Adults Aged 60–65 Years at Baseline. *International Journal of Environmental Research and Public Health*, *20*(5), 4531.

Souza, A. C. de, Alexandre, N. M. C., & Guirardello, E. de B. (2017). Psychometric properties in instruments evaluation of reliability and validity. *Epidemiologia e Servicos de Saude*, *26*, 649–659.

Sugimoto, D., Mattacola, C. G., Bush, H. M., Thomas, S. M., Foss, K. D. B., Myer, G. D., & Hewett, T. E. (2017). Preventive Neuromuscular Training for Young Female Athletes: Comparison of Coach and Athlete Compliance Rates. *Journal of Athletic Training*, *52*(1), 58–64. https://doi.org/10.4085/1062-6050-51.12.20

Sugimoto, D., Myer, G. D., Barber Foss, K. D., Pepin, M. J., Micheli, L. J., & Hewett, T. E. (2016). Critical components of neuromuscular training to reduce ACL injury risk in female athletes: Meta-regression analysis. *British Journal of Sports Medicine*, *50*(20), 1259–1266. https://doi.org/10.1136/bjsports-2015-095596

Sugimoto, D., Myer, G. D., Bush, H. M., Klugman, M. F., Medina McKeon, J. M., & Hewett, T. E. (2012). Compliance with neuromuscular training and anterior cruciate ligament injury risk reduction in female athletes: A meta-analysis. *Journal of Athletic Training*, *47*(6), 714–723. https://doi.org/10.4085/1062-6050-47.6.10

Tanner, K. (2018). Chapter 6—Survey designs. In K. Williamson & G. Johanson (Eds.), *Research Methods (Second Edition)* (Second Edition, pp. 159–192). Chandos Publishing. https://doi.org/10.1016/B978-0-08-102220-7.00006-6

Tian, Y., & Shi, Z. (2022). The relationship between social support and exercise adherence among Chinese college students during the COVID-19 Pandemic: The mediating effects of subjective exercise experience and commitment. *International Journal of Environmental Research and Public Health*, *19*(18), 11827.

Tractinsky, N. (2018). The usability construct: A dead end? *Human–Computer Interaction*, *33*(2), 131–177.

Trochim, W. (2010). Translation won't happen without dissemination and implementation: Some measurement and evaluation issues. *3rd Annual Conference on the Science of Dissemination and Implementation*, 581–629.

U.S. Department of Health and Human Services. (2018). *Physical Activity Guidelines for Americans (2nd ed.)*. https://health.gov/our-work/nutrition-physical-activity/physical-activity-guidelines/current-guidelines

Varnfield, M., Karunanithi, M., Lee, C.-K., Honeyman, E., Arnold, D., Ding, H., Smith, C., & Walters, D. L. (2014). Smartphone-based home care model improved use of cardiac rehabilitation in postmyocardial infarction patients: Results from a randomised controlled trial. *Heart*, *100*(22), 1770–1779.

Waldén, M., Atroshi, I., Magnusson, H., Wagner, P., & Hägglund, M. (2012). Prevention of acute knee injuries in adolescent female football players: Cluster randomised controlled trial. *Bmj*, *344*.

Wang, C., Xu, G., Shang, Z., & Kuncel, N. (2018). Detecting aberrant behavior and item preknowledge: A comparison of mixture modeling method and residual method. *Journal of Educational and Behavioral Statistics*, *43*(4), 469–501.

Weichbroth, P. (2020). Usability of Mobile Applications: A Systematic Literature Study. *IEEE Access*, *8*, 55563–55577. https://doi.org/10.1109/ACCESS.2020.2981892

Weiner, B. J., Lewis, C. C., Stanick, C., Powell, B. J., Dorsey, C. N., Clary, A. S., Boynton, M. H., & Halko, H. (2017). Psychometric assessment of three newly developed implementation outcome measures. *Implementation Science : IS*, *12*(1), 108. https://doi.org/10.1186/s13012-017-0635-3

Wilson, P. M., Blanchard, C. M., Nehl, E., & Baker, F. (2006). Predicting physical activity and outcome expectations in cancer survivors: An application of Self-Determination Theory. *Psycho-Oncology: Journal of the Psychological, Social and Behavioral Dimensions of Cancer*, *15*(7), 567–578.

World Health Organization. (2022). *Physical activity*. https://www.who.int/news-room/fact-sheets/detail/physical-activity

Wronikowska, M. W., Malycha, J., Morgan, L. J., Westgate, V., Petrinic, T., Young, J. D., & Watkinson, P. J. (2021). Systematic review of applied usability metrics within usability evaluation methods for hospital electronic healthcare record systems: Metrics and Evaluation Methods for eHealth Systems. *Journal of Evaluation in Clinical Practice*, *27*(6), 1403–1416.

Ziegler, M. (2014). Stop and state your intentions! *European Journal of Psychological Assessment*.

8 Appendices

8.1 Strength training programs (Study 3)

8.1.1 Program low

80min Strength Training Program

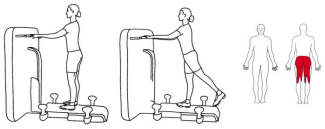
Based on recommendations for weekly physical activity.
Perform program 2 times per week.

Sit well back in the seat and support your upper body by grasping the handles. Push down the pad with your ankles, bending your knees fully. Slowly return to the start position.
Sets: 3 , Reps: 8-12 , Pause: 3 min 0 sec, Intensity: 85 % of max

Start in a handstand position with support from the appliance. Keep your body straight by tightening abdomen and back. Start the movement by lowering your head steadily down towards the ground. Press yourself steadily back up to starting position.
Sets: 3 , Reps: 8-12 , Pause: 3 min 0 sec, Intensity: 85 % of max

Raise the front plate to the required height. Stand on the front plate, on a single leg, with the heel of the foot off the step, grip the handles. Raise up on to the tip toe. Slowly lower the body until the calf muscle is on stretch.
Sets: 3 , Reps: 8-12 , Pause: 3 min 0 sec, Intensity: 85 % of max

Sit slightly bent forward with your inactive arm resting on your knee in support of your active arm. Turn shoulder of your active arm outward. Bend your elbow and raise the dumbbell to your chin. Slowly return to the start position and repeat.
Sets: 3 , Reps: 8-12 , Pause: 3 min 0 sec, Intensity: 85 % of max



Standing on the front plate, facing the machine, in an upright position. Grip the support handle, extend one leg backwards. Hold the upper body upright and head forward. Slowly return to start position.
Sets: 3 , Reps: 8-12 , Pause: 3 min 0 sec, Intensity: 85 % of max

Make sure that your lower back is well supported. Sit with your knees bent approximately 120 degrees. Then straighten your legs as much as possible. The top point may be accentuated. Slowly lower to the starting position.
Sets: 3 , Reps: 8-12 , Pause: 3 min 0 sec, Intensity: 85 % of max

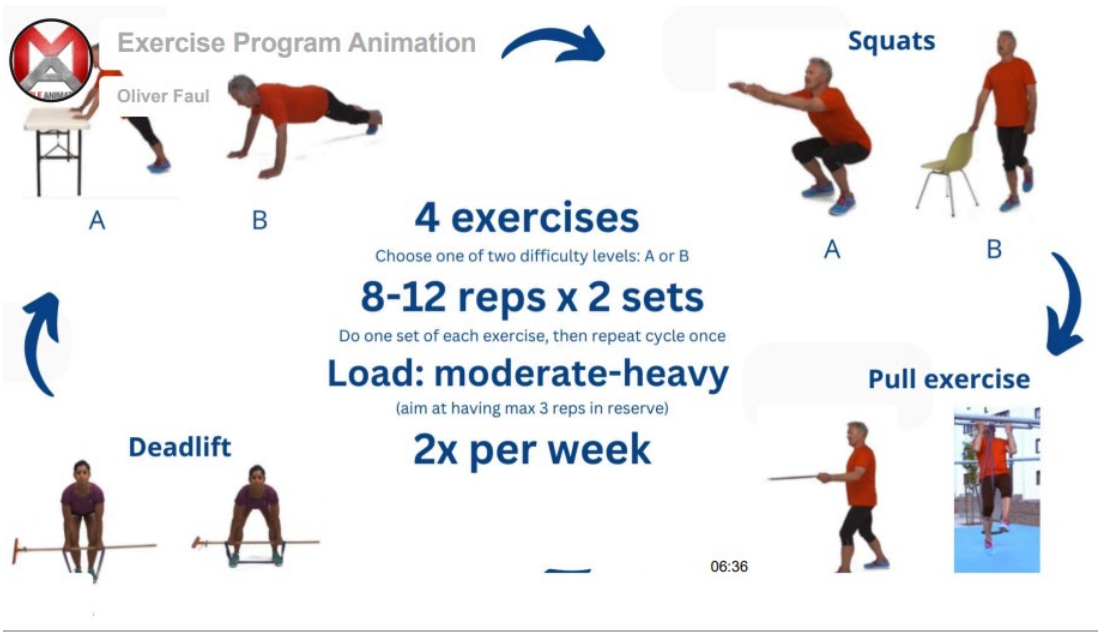
Place both feet in TRX facing away from the anchor point with the hands on the floor right under the shoulders. Lower the chest down to the floor. When you get back up, move both knees towards the elbows till the hip joint is in a 90 degrees angle. The exercise can also be done by moving the knees towards the elbows, while the upper body is pushed up.
Sets: 3 , Reps: 8-12 , Pause: 3 min 0 sec, Intensity: 85 % of max

Lie on your back with legs bent and hold a barbell over your hips. Flex your abs and raise your buttocks from the floor until there is a straight line from shoulders to knees. Hold a few seconds and slowly lower to starting position.
Sets: 3 , Reps: 8-12 , Pause: 3 min 0 sec, Intensity: 85 % of max

8.1.2 Program High

Video of the exercise program can be found here:
https://player.vimeo.com/video/884504025?badge=0&autopause=0&quality_selector=1&progress_bar=1&player_id=0&app_id=58479

The picture below summarizes the main components of the exercise program.



8.2 Qualtrics survey

Intro

Thank you for your interest in participating in our study about usability of exercise programs!

You are being asked to take part in a research study conducted by researchers from Michigan Technological University. The purpose of the study is to develop a tool for assessing the usability of exercise interventions.

If you choose to be in the study, you will complete an online survey. You will be presented with an exercise program and asked to provide responses of how much you agree to a series of statements. It will take you about **15 minutes to complete the study**.

The survey will be conducted via Qualtrics. The data will pertain to feedback on the usability of the exercise program, thus the nature of this data is not sensitive and wouldn't place you at risk if compromised. Although every reasonable effort has been taken, confidentiality during actual Internet communication procedures cannot be guaranteed.

The study is **completely voluntary** and you have the right to **withdraw from the study at any time without penalty**.

Compensation: You will receive $3 to compensate for your time for participating in this study.

If you have any questions or concerns about participating in this study, please contact the primary investigator, Anne Inger Mortvedt, at amrtvedt@mtu.edu.
If you want to participate in this study, please confirm below

- ○ Yes, I confirm that I want to participate in this study
- ○ No, I confirm that I do not want to participate in this study

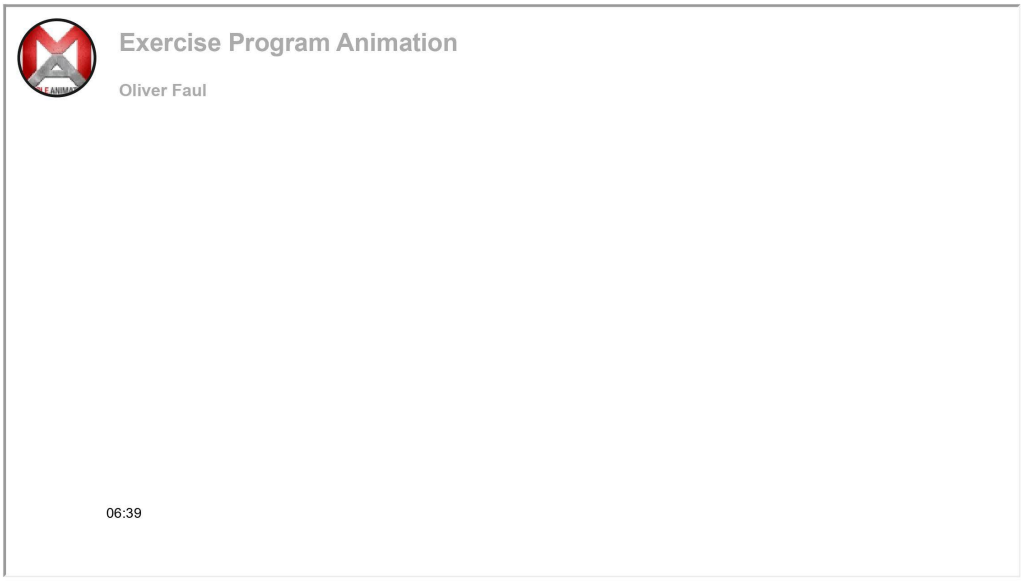
Prolific ID

What is your Prolific ID?
Please note that this response should auto-fill with the correct ID'

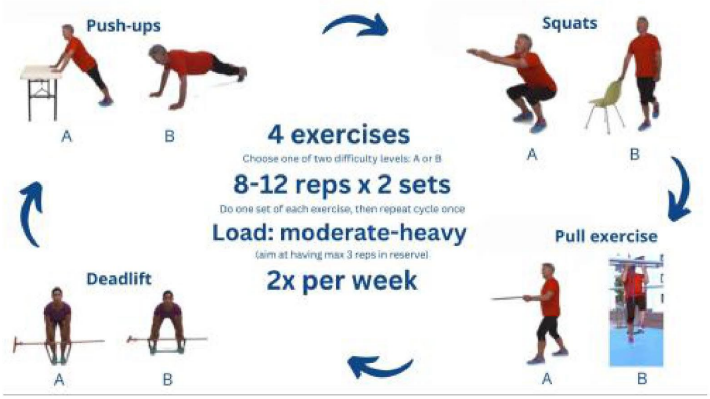
${e://Field/PROLIFIC_PID}

Exercise video

This video is presenting an exercise program



Click play button to start the video.



Pretend that you were prescribed this program.

How much do you agree with the following statement

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I will enjoy doing this program	○	○	○	○	○
I like this program	○	○	○	○	○
I will learn to use this program very quickly	○	○	○	○	○
The program meets my needs	○	○	○	○	○
This program required the use of equipment I like to use	○	○	○	○	○
The difficulty of each exercise is exactly right for me	○	○	○	○	○
The program is simple to understand	○	○	○	○	○
This is an important program to me	○	○	○	○	○
I have access to the equipment needed for this program	○	○	○	○	○
The program includes everything I would expect it to have	○	○	○	○	○
I can do this program with a friend	○	○	○	○	○
I think performing this program is worth my time	○	○	○	○	○
It is clear to me how this program will improve my strength	○	○	○	○	○
The program is useful to me	○	○	○	○	○
This program will allow me to exercise with friends	○	○	○	○	○
The program is valuable to me	○	○	○	○	○

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
The program provides flexible options	○	○	○	○	○
I can do the program wherever I want to	○	○	○	○	○
I can do the program whenever I want to	○	○	○	○	○
The program seems exactly right for my goals	○	○	○	○	○
This program looks interesting	○	○	○	○	○
It is clear how to use the program	○	○	○	○	○
This program required the use of equipment I have available	○	○	○	○	○
I can adapt the program to my needs	○	○	○	○	○
The program is effective for improving strength	○	○	○	○	○
The program seems appropriate in length	○	○	○	○	○

The exercise program seems

Ineffective	○	○	○	○	○	Effective
Boring	○	○	○	○	○	Fun
Too long	○	○	○	○	○	Too short
Inefficient	○	○	○	○	○	Efficient
Difficult to do	○	○	○	○	○	Easy to do
Difficult to learn	○	○	○	○	○	Easy to learn
Worthless	○	○	○	○	○	Valuable
Harmful	○	○	○	○	○	Beneficial
Rigid	○	○	○	○	○	Flexible
Complicated	○	○	○	○	○	Straightforward

Answer the following query in a specific manner and move on to the following question.

	1	2	3	4	5
Mark number 1 and 3	☐	☐	☐	☐	☐

Other measures

How much do you agree with the following statement

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I found this program very cumbersome to use	○	○	○	○	○
This program seems suitable	○	○	○	○	○
I believe the program will not hurt me	○	○	○	○	○
This program seems like a good match	○	○	○	○	○
I would like to try this program	○	○	○	○	○
I feel very confident doing the program	○	○	○	○	○
This program seems fitting	○	○	○	○	○
I would engage in this program because I don’t have any choice	○	○	○	○	○
I intend to do this program consistently	○	○	○	○	○
I would engage in this program because I feel that I have to do it	○	○	○	○	○
I am confident that I can do this program without the help of a therapist or trainer	○	○	○	○	○
I needed to learn a lot of things before I could get going with this program	○	○	○	○	○

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I would engage in this program because it is something that I have to do	○	○	○	○	○
I would like to use this program frequently	○	○	○	○	○
I will make time to do this program twice per week	○	○	○	○	○
I think that I would need the support of an expert consultant to be able to use this program	○	○	○	○	○
I would rather eat a piece of fruit than a piece of paper	○	○	○	○	○
I found the program unnecessarily complex	○	○	○	○	○
I thought there was too much inconsistency in this program	○	○	○	○	○
I found the various components of this program were well integrated	○	○	○	○	○
I feel confident I can do the program	○	○	○	○	○
This program seems applicable	○	○	○	○	○
I would engage in this program because I am supposed to do it	○	○	○	○	○

How likely are you to recommend this program to a friend?

Not at all likely 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10 ○ Extremely likely

Demographics

What is your age?

(Numeric entry only)

Gender

- ○ Man
- ○ Non-Binary
- ○ Transgender
- ○ Woman
- ○ [ ] A gender not listed here. Please specify:
- ○ Prefer not to answer

Education: What is the highest degree or level of school you have completed? If currently enrolled, highest degree received.

- ○ Some high school, no diploma
- ○ High school graduate, diploma or the equivalent (for example: GED)
- ○ Bachelor's degree
- ○ Master's degree
- ○ Professional degree
- ○ Doctorate degree
- ○ Other

Ethnic origin: Please specify your ethnicity

- ☐ American Indian/Native American/First Nation/Alaska Native
- ☐ Black/African American
- ☐ Asian/Asian American
- ☐ Hispanic/Latin@/Chican@
- ☐ Middle Eastern
- ☐ South Asian
- ☐ Native Hawaiian/Pacific Islander
- ☐ White/European American
- ☐ [ ] A racial/ethnic identity not listed here. Please specify:

How comfortable are you with performing muscle strengthening activities?

Very uncomfortable 0 1 2 3 4 5 6 7 8 9 10 Very comfortable

Vigorous activities refer to activities that cause large increases in breathing or heart rate like running, strength training or play sport for at least 10 minutes continuously. In a typical week, on **how many days do you do vigorous-intensity sports, fitness or recreational (leisure) activities**?

- None
- 1-2 days a week
- 3-4 days a week
- 5-7 days a week

Attention check3

Please describe one thing you remember from the exercise program you were presented:

End of survey message

Thank you for taking part in this study. Please click the button below to be redirected back to Prolific and register your submission.

Powered by Qualtrics

8.3 Study 2: Interview Guides

Post Intervention Interviews

Participants: Coaches (n = 4) and players (n = 6)
Where: In person at their school/training facility
Time period: December 2022
Method: Semi-structured interview

<u>Interviews with coaches</u>

- What is your overall impression of this injury prevention program?
    - What specifically did you like about this injury prevention program?
    - What specifically did you dislike or find challenging about this injury prevention program?

- Would you like to continue using the program outside of the research setting?
    - If yes: Why?
    - If no: Why?

- Based on the players' survey responses, they did not seem to enjoy this program. Why do you think the athletes did not enjoy doing the program?
    - What would you change to ensure that athletes are doing it but also are enjoying it?

- What were some reasons your team or individual players did not comply with the program?

- What specific ways did you try to increase the players motivation to perform the IPP?

- What are some specific strategies you would use to improve compliance with this injury prevention program?

- What does an ideal implementation process look like in your mind? (When, what, how - and why?)

- Who do you think has the main responsibility for implementing injury prevention programs?

Additional questions:

- As a former handball player, you might have some valuable insights and thoughts. How important is it that the players understand why they are investing in the IPT? Is it enough that the coach just tells them to do it?
- To increase likelihood of sustained use, how important do you think it is that they like to perform the program, or experience some kind of effect from performing it?

Focus group interviews with players (6)

- What did you think about the exercise program?
    - What did you like about it
    - What did you dislike about it
    - What things would improve the program?

- What would make you more interested or motivated to do this program?

Design Thinking Activity

- 3 people in each group (2 groups)
- Generate ideas for
    - "How we, as researchers, coaches or others, can help encourage players to perform IPT frequently"
    - "How can this program become more fun to do?"
    - "Do you have any suggestions for how you can include IPT as part of your weekly training schedule?"
- One idea on each post-it
- Vote on the best and second best idea
- “Prototype” the best ideas

- How do you feel about IPT being something you just "have to do", even though it might not be fun?
- Progression - see the relationship between the exercises and/or “distance” between levels of difficulty? (Progresjon - sammenheng mellom øvelsene, avstand mellom vanskelighetsgrad)
- Exercises - any exercises too easy or too hard (Øvelser - noen for lette og/eller kjedelige)

- How is it to focus on technique when performing a finte/cut of high quality including an opponent? Did you at all focus on technique in these exercises? Teknikkfokus i finteøvelse (Hvordan er det å fokusere på teknikk når man skal utføre en finte av god kvalitet mot en motspiller). Fokuserte dere på teknikk?
- Specificity - To what extent did you feel the exercises relate to handball? Spesifisitet opp mot håndball

8.4 IRB Documents

8.4.1 Study 1 and 2

Tron Krosshaug
Institutt for idrettsmedisinske fag

OSLO 21. juni 2022

Søknad 233 – 160622 - Redusere risikofaktorer for korsbåndskader hos unge kvinnelige håndballspillere

Vi viser til søknad, prosjektbeskrivelse, informasjonsskriv og innsendt melding til NSD.

I henhold til retningslinjer for behandling av søknad til etisk komite for idrettsvitenskapelig forskning på mennesker har komiteen i møte 16.juni 2022 konkludert med følgende:

Vurdering

I søknaden opplyses det om en viss risiko for skade under testene, herunder fremre korsbåndskade: Risikoen er imidlertid lav (1-2 ganger per 1000 spilletimer). I søknaden oppgis at skaderisikoen er vesentlig høyere i kamp enn på trening. Risikoen for skade som ledd i studien oppgis derfor som lavere i enn i reelle kampsituasjoner. Når det gjelder spillere med eksisterende korsbånd skade oppgis imidlertid risikoen for skade under planlagte studietester for å være tre ganger høyere, men at inklusjon av denne gruppen kan forsvares grunnet høy nytteverdi med tanke på ny kunnskap om sammenhengen mellom belastning og skaderisiko – også for deltakere med tidligere kjent kneskade som er klarert som kampklare. Komiteen støtter prosjektleders vurdering, men ber om at den særskilte risikoen gruppen med tidligere kjent kneskade utsettes for, er godt beskrevet i informasjonsskrivet.

Deltakere skal rekrutteres fra videregående skoler med toppidrett. Treningen vil gjennomføres på utøvernes respektive skoler. Dersom denne treningen inngår i ordinær undervisning må det påses at skolen har et alternativt opplegg for de om ikke ønsker å delta i prosjektet. Det bør opplyse om et ev alternativt opplegg i informasjonsskrivet. Komiteen ber videre om at det rettes opp i begrepene knyttet til anonyme og avidentifiserte data i informasjonsskrivet.

Vedtak
På bakgrunn av forelagte dokumentasjon finner komiteen at prosjektet er forsvarlig og at det kan gjennomføres innenfor rammene av anerkjente etiske forskningsetiske normer nedfelt i NIHs retningslinjer. Til vedtaket har komiteen lagt følgende forutsetning til grunn:

- *Vilkår fra NSD følges*
- *Informasjonsskrivet justeres i tråd med komiteens merknader*
- *Dersom testene/treningen inngår i ordinær undervisning må det påses at skolen har et alternativ opplegg for elever som ikke ønsker å delta*

NIH NORGES IDRETTSHØGSKOLE

Besøksadresse: Sognsveien 220, Oslo
Postadresse: Pb 4014 Ullevål Stadion, 0806 Oslo
Telefon: +47 23 26 20 00, postmottak@nih.no
www.nih.no

Komiteen gjør oppmerksom på at vedtaket er avgrenset i tråd med fremlagte dokumentasjon. Dersom det gjøres vesentlige endringer i prosjektet som kan ha betydning for deltakernes helse og sikkerhet, skal dette legges fram for komiteen før eventuelle endringer kan iverksettes.

Med vennlig hilsen

Professor Anne Marte Pensgaard
Leder, Etisk komite, Norges idrettshøgskole

NIH NORGES IDRETTSHØGSKOLE

Besøksadresse: Sognsveien 220, Oslo
Postadresse: Pb 4014 Ullevål Stadion, 0806 Oslo
Telefon: +47 23 26 20 00, postmottak@nih.no
www.nih.no

8.4.2 Study 3

Michigan Tech

Office of Compliance, Integrity, and Safety
Phone: 906-487-2902 E-mail: IRB@mtu.edu
1400 Townsend Drive
Lakeshore Center, 3rd Floor
Houghton, MI 49931

Reset Form

Exemption Request and/or Limited Review

Federal regulations (45 CFR 46) permit the exemption of some types of research from IRB review.
Exemption does not mean that you do not need to submit a study for review; our office requests information about your study and will determine the level of review required for approval.
If you have any questions, feel free to contact our office.

Eligible for Exemption: There are several classifications of research which may involve human subjects but their classification falls outside of the IRB's policies and jurisdiction.

Determination that research is exempt or requires limited Institutional Review Board (IRB) review is made through the Office of Compliance, Integrity, and Safety. Exemption from review is only available to certain categories of research as define by federal regulation. If you have questions about whether your project might qualify for exemption, please contact our office.

Project Title	Development of a Usability Scale for Exercise Prescription, Phase II		
Project Start Date	Nov 1, 2023	Project End Date	Jan 31, 2024
Principal Investigator	Erich J. Petushek	Department	Cognitive and Learning Sciences CLS
E-mail	ejpetush@mtu.edu	Phone	

I. Project Description

1. Purpose and goals of the research:
(text field will expand)

Exercise interventions have proven to be effective in treating and preventing numerous diseases and injuries. However, adherence to recommended exercise interventions is insufficient and their potential is far from being reached. One way to address the barriers and facilitators associated with adherence to exercise interventions is to consider the development process of the exercise interventions, where a human-centered approach and an assessment of the usability of the program to a much greater extent needs to be integrated. The domain of Human Factors Engineering has effectively succeeded in identifying and developing to meet the needs of the users through human-centered design and usability testing of the system. There are currently no existing comprehensive usability measures that efficiently evaluate exercise-based programs. There is also a need to understand what constructs and psychometric properties need to be included and whether they are valid and reliable measures.
The aim of this study is to further develop, refine and add validity evidence to the Usability Scale for Exercise Prescription (USERx) through assessment of dimensionality, reliability, item reduction and hypothesis testing (e.g. relations to other variables). The end goal of this study is the development of a reliable and valid usability scale for assessing usability of exercise programs. The USERx scale will ultimately function as a valuable tool in future

exercise program development, both within research and clinical practice, by its ability to identify potential barriers to program implementation and long term use. USERx can also be used to test how different populations (e.g. sex, ethnicity, age) perceive different exercise programs.

2. Methods and procedures: Describe in detail what subjects will be asked to do, what information will be collected about them, and when or how often research procedures will be conducted. You may also upload an attachment describing the methods including a graph, table, timeline of events.

To collect responses on the USERx scale, surveys will be conducted through Qualtrics. Participants will be randomized into two groups. Both groups will be presented with one out of two exercise programs and asked to use the USERx to evaluate the usability of it. One exercise program will be displayed through a 6 min animation (https://vimeo.com/882144001?share=copy) and the other exercise program will be displayed through a picture/ PDF. Both programs included for assessing usability with the USERx will be based on the physical activity guidelines for Americans regarding strength training recommendations. This means that the exercise program is highly relevant for the target population and participants in our sample, and should involve no greater risk of harm than is in what is recommended by the Department of Health and Human Services elsewhere. After the usability assessment, the participants will be presented some demographic questions (e.g. age, sex, exercise background) and items assessing their intention to use the program, the Intervention Usability Scale (IUS) and motivation items from the Situational Motivation Scale (SiMS). A subset of the respondents in each group (n=60) will be presented with either the same program or the other program and further asked to assess the usability of it using the USERx scale. Participation is estimated to take between 20-30 minutes to complete for this subset. The remaining respondents in each group will end their participation after responding to the demographic questions and the other variables (e.g. intention to use, IUS, SiMS). Participation is estimated to take between 15-20 minutes to complete.

3. Research Site:

Digital platforms (e.g. Qualtrics)

4A. Will you obtain identifiable private information about these individuals? ○ Yes ◉ No

Private Information includes information about behavior that occurs in a context in which an individual can reasonable expect that no observation or recording is taking place, or information provided for specific purposes which the individual can reasonably expect will not be made public (e.g. student record).

Identifiable means that the identity of the participant may be ascertained by the investigator or associated with the information (e.g. by name, code number, pattern of answers, etc.)

4B. Will data be collected and stored in a manner such that participants may be individually identified directly or indirectly? ○ Yes ◉ No

5. Does the study present more than minimal risk to the participants? ○ Yes ◉ No

Minimal risk means that the risks of harm or discomfort anticipated in the proposed research are not greater, considering probability and magnitude, than those ordinarily encountered in daily life or during performance of routine physical or psychological examinations or tests. Note that the concept of risk goes beyond physical risk and includes psychological, emotional, or behavioral risk as well as risks to employability, economic well being, social standing, and risks of civil and criminal liability.

If Yes, you can not use this form, please submit a Protocol Document

6. Is this a graduate level research project? ◉ Yes ○ No

II. Exemption Categories

Check the category or categories which apply and respond to the questions within that exemption section:

☐ **Category 1**: Research, conducted in established or commonly accepted educational settings that specificially involves normal educational practices that are not likely to adversely impact students' opportunity to learn required educational content or the assessment of educators who provide instruction. This includes most research on regular and special education instructional strategies, and research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods.

a) Describe the established or commonly accepted educational setting of the research:

b) Could the research adversely impact student achievement in anyway? ○ Yes ○ No

If Yes, the study does not qualify under this category

c) Could the research adversely impact the assessment of educators who provide instruction? ○ Yes ○ No

If Yes, the study does not qualify under this category

d) Does the research involve a comparison of a proven educational technique to a novel technique? ○ Yes ○ No

If Yes, the study does not qualify under this category

☒ **Category 2:** Research that only includes interactions involving education tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures, or observation of public behavior (including visual or auditory recording) if at least one of the following criteria are met:

☒ The information obtained is recorded by the investigator in such a manner that the identity of the human subjects cannot readily be ascertained either directly or through identifiers linked to the subjects (e.g., anonymous survey);

☐ Any disclosure of the human subjects' responses outside the research would not reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, education advancement, or reputation;

a) Does the research involve minor participants? ○ Yes ◉ No

b) If yes, does the research involve surveys? ○ Yes ○ No

If yes to b, exemption category 2 does not apply. Complete a Protocol Document and submit for expedited review.

☐ The information obtained is recorded by the investigator in such a manner that the identity of the human subjects can readily be ascertained either directly or through identifiers linked to the subjects, and disclosure has risks, **then an IRB limited review will be conducted to ensure privacy and confidentiality of subjects**. This category may NOT be applied to research with children.

a) Does the research involve an intervention? ○ Yes ○ No

Intervention is defined as, "manipulations of the subject or the subject's environment that are performed for research purposes."

If yes, exemption category 2 does not apply

☐ **Category 3:** Research involving benign behavioral *intervention** in conjunction with the collection of information from an adult subject through verbal or written responses (including data entry) or audiovisual recording if the subject prospectively agrees to the intervention and information collection and at least one of the following criteria is met:

A) The information obtained is recorded by the investigator in such a manner that the identity of the human subjects cannot readily be ascertained, directly or through identifiers linked to the subjects;

B) Any disclosure of the human subject's response outside the research would not reasonable place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, educational advancement, or reputation; or

C) The information obtained is recorded by the investigator in such a manner that the identity of the human subjects can readily be ascertained, directly or through identifiers linked to the subject, **and an IRB limited review will be conducted to ensure privacy and confidentially of subjects.**

***benign behavioral interventions** are brief in duration, harmless, painless, not physically invasive, not likely to have a significant adverse lasting impact on the subjects, and the investigator has no reason to think the subjects will find the intervention offensive or embarrassing. Provided all such criteria are met, examples of such benign behavioral interventions would include having the subject play an online game, having them solve puzzles under various noise condition, or having them decide how to allocate a nominal amount of cash received between themselves and someone else.

a) Describe the benign behavioral intervention:

If the research involves deceiving the subjects regarding the nature or purposes of the research, this exemption is not applicable unless the subject authorizes the deception through a prospective agreement to participate in research in circumstances in which the subject is informed that he or she will be unaware of or misled regarding the nature or purposes of the research.

b) Does the research involve deception? ○ Yes ○ No

c) If so, will subjects prospectively agree to be unaware of or misled regarding the nature of the research? ○ Yes ○ No

If Yes to B) but no to C), the research will not qualify under this category. You must complete and submit a Protocol Document, you cannot use this form.

Does the research involve minors? ○ Yes ○ No

If yes, the research does not qualify under this category. You must complete and submit a Protocol Document, you cannot use this form.

☐ **Category 4:** Secondary research for which consent is not required: Secondary research uses of identifiable private information or identifiable bio specimens. **Call our office for assistance.**

☐ **Category 5:** Research and demonstration projects that are conducted or supported by a Federal department or agency, or otherwise subject to the approval of department or agency heads (or the approval of the heads of bureaus or other subordinate agencies that have been delegated authority to conduct the research and demonstration projects), and that are designed to study, evaluate, improve, or otherwise examine public benefit or service programs, including procedures for obtaining benefits or services under those programs, possible changes in or alternatives to those programs or procedures, or possible changes in methods or levels or payment for benefits or services under those programs. Such projects include, but are not limited to, internal studies by federal employees, and studies under contracts or consulting arrangements, cooperative agrements, or grants. Exempt projects also include waivers of otherwise mandatory requirements using authorities such as section 1115 and 1115A of the Social Security Act, as amended.

NOTE: exemption under Category 5 is only permitted upon Federal Agency approval AND after being published on a federal website.

☐ **Category 6:** Taste & food quality evaluation and consumer acceptance studies: (a) if wholesome foods without additives are consumed; or (b) if a food is consumed that contains a food ingredient at or below the level and for a use found to be safe, or agricultural chemical or environmental contaminant at or below the level found to be safe by the Food and Drug Administration or approved by the Environmental Protection Agency or the Food Safety and Inspection Service of the U.S. Department of Agriculture.

Michigan Tech is not currently set up to use these two exemptions categories at this time. Call our office for assistance.

Category 7: Storage or maintenance for secondary research for which broad consent is required.
Category 8: Secondary research for which broad consent is required.

III. Participants, recruitment, and informed consent

1. Describe the proposed participants:

Participants will be recruited from the student pool at Michigan Technological University (e.g. Introductory to Psychology), through colleagues and word of mouth, and from online panels (e.g. Qualtrics, Prolific). The online platforms are reported to be cost-effective and provide high-quality respondents. We aim to recruit a representative sample of US adults (e.g. 50 % female, average age of ~38 etc.). We will aim to recruit 300 participants for the required analyses. Our sample size is based on previous studies reporting sample sizes needed for conducting factor analyses and item response theory (IRT).

2. Recruitment: Describe recruitment procedures. Include how participants will be initially identified, approached, or contacted regarding the research and in what setting. *Please provide a copy of any recruitment materials, advertisements, flyers, text of e-mails, etc. which will be used.*

The student sample will be recruited using the Psychology Subject Pool System. Participants will be able to review the available research projects that count toward course credit on the SONA system. Participants can choose to sign up for the study on the SONA system. Advertisement of the study will not occur. When participants sign up for the study they will be given, and they will read the informed consent form. At this time the participant may either choose to sign the informed consent and continue with the study or be given an alternate task, as stipulated by their course instructor, such as reading an article and writing a review.

Other participants will be identified and recruited through word of mouth/personal network, and by using online samples/market research panel (e.g. Qualtrics, Prolific). A market research panel is a group of people recruited to respond to a survey. They are typically chosen from a pre-arranged pool of respondents who've agreed to be contacted by a market research service in order to respond to surveys.

Email addresses from the general public will be obtained through our personal network and with help from colleagues. Participants from the general public will be sent an email with information about the study, compensation, what it requires and opportunities to participate. There will be a link to the survey in the email. See email recruitment letter attached.

Qualtrics: Panel members are sent an email invitation or prompted on the respective survey platform to proceed with a given survey. The typical survey invitation is generally very simple and generic. It provides a hyperlink which will take the respondent to the survey as well as mention the incentive offered. See attachment about how Qualtrics recruit and compensate their panel members.
Prolific: Participants create an account on Prolific and are then notified of future studies they are eligible for based on the demographic information they provide. When our study is posted, an email will be sent out to a random subset of all eligible participants. Participants can also see currently available studies that they are eligible for by going to their studies dashboard on Prolific. Prolific also send email invites to eligible participants for a study every 48 hours if it has not reached its maximum number of submissions within this time.

3. Describe procedures for informing participants about the research and how they will actively indicate their agreement to participate. *Please provide a copy of the oral script or information sheet which will be used.*

Participants will be provided with a link to the survey through email. In this email, they will also receive some information about the research, estimated length and what compensation they will receive. When they open the link to start the survey, the first page will repeat information about their rights and the need for their consent. See attachment. They will be provided a multiple choice question asking them whether they confirm that they have read the consent form and want to participate, or whether they would like to end the survey before any data is collected.

4. Compensation/incentives: Will participants or others be offered incentives for their participation (e.g., gifts, payment, reimbursement, services, extra or course credit, or other incentives? ○ Yes ○ No

If yes, please describe the amount, alternative ways to earn compensation (i.e., in cases of course/extra credit), and when compensation/incentives will be awarded. Please be sure to follow the guidance document, **Procedure for Compensation for Human Subject Participants (found on our website)**.

Students from the Introductory to Psychology class will be compensated through course credits (1 credit when less than 30 minute survey). The Introductory to Psych students will be given the survey that asks them to assess a second program as described in methods section of the Project Description, but the survey is still estimated to take less than 30 minutes for this population.

General population participants will be offered to sign up for a lottery where the prize will be $100. If they want to sign up for the lottery and opportunity to win $100, they will enter their name and email address in a separate page that is not linked to their survey responses. They will only be allowed to enter their name and email address once. To prevent duplicate responses, the Qualtrics survey is set to prevent multiple submissions bu flagging the responses. This means that they will be allowed to finish their response. However, their responses will be assigned a value under a field named Q_BallotBoxStuffing.

Participants recruited through online panels will be compensated based on the panel they belong to. Qualtrics states: Our panelists join from a variety of sources. They may be airline customers who chose to join in reward for SkyMiles, retail customers who opted into get points at their favorite retail outlet, or general consumers who participate for cash or gift cards, etc. When participants are invited to take a survey, they are informed what they will be compensated. See example in attachment about how Qualtrics compensate and inform their panel members.

Prolific only offers monetary compensation through PayPal. Participants recruited through Prolific will receive $12/h (e.g. $4 for 20 minutes).

5. Dual relationships: Does the investigator, co-investigators, or any member of the research team, or anyone assisting with the research have an authority relationship (e.g., instructor/student, employer or supervisor/employee, or other) with potential participants? ○ Yes ◉ No

If yes, describe the relationship, and indicate how the research will be conducted to avoid undue influence on participation

6. Will any aspect of the research be conducted in a classroom setting during class time? ○ Yes ◉ No

If yes, describe what those who choose not to participate will be doing, and provide justification for use of class time for research. *You may be asked to include the course syllabus.*

7. Will all participants, their parents/guardians and /or their legally authorized representative (as applicable) be fluent in English? ◉ Yes ○ No

If no, explain how informed consent will be obtained, and provide a copy of the translated documents(s) to be used.

8. If research will be conducted at an international site, indicate the investigator's familiarity with the culture and cultural norms, and how the research may affect an individual's standing in their community ☒ N/A

IV. Instruments

Be sure to upload the questionnaire(s), survey instrument(s), or list of interview or focus group questions to your irbnet.org submission package.

V. Privacy and Confidentiality

1. Privacy: Describe the conditions under which interaction with the subjects will occur (e.g., consent discussion occurs in a private room). Explain how these conditions adequately address the PRIVACY of subject:

2. Personally identifiable information: Will the researchers obtain any personally identifiable information (PII) from or about participants (e.g. names, address, telephone numbers, etc.)?

○ Yes ◉ No - (proceed to Question 3)

a) What direct identifiers will be obtained?

b) How long will the PII be maintained?

c) Why is it necessary to maintain direct identifiers?

d) Describe the coding system that will be used to protect against disclosure of these identifiers.

e) How long will the link between identifiers and code be maintained?

f) Explain how the research will mitigate a risk of participant responses that could place them at risks such as criminal or civil liability, or be damaging to their financial standing, employability, insurability, reputation, or be stigmatizing (e.g. limiting access to identifiers, obtaining a Certificate of Confidentiality from NIH, etc.). *If a Certificate of Confidentiality is obtained, provide a copy to the IRB once available.*

3. Will any demographic information be collected which could lead to a deductive disclosure of participant(s) indentifies? If so, how will participant privacy be addressed?

Internet protocol (IP) addresses will be collected. These will be deleted whenthe data is downloaded from Qualtrics.

4. In what format(s) will the data originate, be shared among team members/collaborators, and be maintained during the life of the study (e.g. paper, digital, electronic media, video, audio or photographic):

The data will be downloaded from Qualtrics in a excel file format. The file will be shared among the research team electronically (e.g. Google Drive).

5. Where will data be stored including security provisions that will be taken to protect the data (include both paper/hardcopy records and digital/electronic files).

Data will be stored on a local password protected computer and electronically and be saved indefinitely.

6. Are there potential ethical or legal circumstances when it would be necessary to break confidentiality (e.g., requirements for mandated reporting or other professional obligations to report)? If so, describe:

No

7. Final disposition: Please describe at what point in time PII and deductive identifiers will be removed from the dataset and/or the records retention plan for the research records:

IP addresses will be removed from the dataset immediately after the data is downloaded. No other PII or deductive identifiers will be collected/recorded.

Click here to read instructions on how to submit form

Michigan Technological University

Compliance, Integrity, and Safety

302 Lakeshore Center | 1400 Townsend Drive Houghton, MI 49931-1295

906-487-2902 | f. 906-487-2245

Michigan Tech is an EOE which includes protected veterans and individuals with disabilities.

DATE:	November 21, 2023
TO:	Erich J. Petushek, PhD
RE:	IRBNet #2118322-1
TITLE:	Development of a Usability Scale for Exercise Prescription, Phase II
SUBMISSION TYPE:	New Project
STATUS:	**APPROVED, Exempt Status**

Thank you for your submission of materials for this research study. Your application has been reviewed and determined to have Exempt status under Exempt Category #2(ii).

This approval is based on no greater than minimal risk to research participants. All research must be conducted in accordance with the approved submission. If any changes are made in the protocol or conditions set forth in the application, the principal investigator must obtain a separate approval before the changes are implemented, other than to avoid any immediate harm to the subject.

When a research project is determined by the Office of Compliance, Integrity, and Safety (CIS) to meet the criteria for an Exempt status, it means that it is exempt from annual continuing review by the Institutional Review Board (IRB). The investigator is required to obtain IRB approval for any changes, revisions, or modifications in the research activity including personnel changes. Any changes made may cause the research status to change from Exempt to Expedited or Full Review and must be determined and documented by the CIS Office.

All Unanticipated Problems/Serious Adverse Events to participants or other parties affected by the research must be reported to this office within two days of the event occurrence. All instances of noncompliance or complaints regarding this study must be reported to this office in a timely manner. Please use the INSTRUCTIONS and FORM: Unanticipated Problem/Serious Adverse Event Form found both on our web site and the IRBNet Library.

This approval applies only for this project and only under the conditions and procedures described in the application. If/when changes become necessary (and include but are not limited to changes in protocol, personnel, study location, participant recruitment, etc., as set forth in this approval), you must submit the FORM for Change Request during approval found in the IRBNet Library. You must receive notification of approval prior to implementing the change(s).

If you have any questions, please contact Mike Reay at mjreay@mtu.edu or Dorinda Williams at (401) 523-9022 or dorinda@peercg.com.