Measuring Quality of Occupational Performance Based on Self-report and Observation:

Development and Validation of Instruments to Evaluate ADL Task Performance

Eva Ejlersen Wæhrens

Department of Community Medicine and Rehabilitation, Occupational Therapy, Umeå University
901 87 Umeå, Sweden
2010

ISBN: 978-91-7264-993-4
ISSN: 0346-6612
Printed in Sweden by Arkitektkopia, Umeå, 2010

To my Family – the most important people in my life

Table of Contents

Abstract

Background

People with rheumatic or neurologic conditions are commonly referred for occupational therapy because of decreased ability to perform ADL tasks. Upon referral, occupational therapists use a client-centred, occupation-focused approach when evaluating a person's perceived and observed quality of ADL task performance to clarify the nature and extent of the person's problems, plan interventions and determine effectiveness of interventions. Evaluation of the extent of problems and change following intervention require the use of linear measures of ADL. The aim of this doctoral thesis was to develop and validate linear measures of perceived and observed quality of ADL task performance for occupational therapy clinical praxis and research.

Methods

To develop linear measures of observed quality of ADL task performance based on Rasch measurement methods, clients with ABI (n=70) were evaluated using a 3-category rating scale and the ADL Taxonomy. Similarly, to develop linear measures of perceived quality of ADL task performance women with rheumatic diseases (n=118) reported their ability on a questionnaire and in an interview based on the ADL Taxonomy and a 4-category rating scale. To further validate the ADL ability measures of the Assessment of Motor and Process Skills (AMPS) data were collected in two samples. In a sample of women with chronic widespread pain (CWP) or fibromyalgia (FM) (n=50) the ADL ability measures were validated with regard to discrimination, stability and sensitivity to changes. Discrimination was examined by comparing AMPS data from women with CWP/FM to those of healthy women. Stability and sensitivity to change were examined based on repeated AMPS observations pre and post intervention. In addition, sensitivity to change was examined in a sample of clients with ABI (n=36) using a pre- and post-test design. Finally, the validity of the AMPS ADL ability measures in relation to perceived ADL ability was examined in women with rheumatic diseases (n=118) and women with CWP/FM (n=50).

Results

Based on the results of the four studies in this thesis, we were able to conclude that it was possible to obtain linear measures of observed quality of ADL task performance based on a revised version of the ADL Taxonomy (ADL-O). It was also possible to obtain linear measures of perceived quality of ADL task performance based on revised versions of the ADL Taxonomy using the formats of questionnaires (ADL-Q) and interviews (ADL-I). Moreover, it was possible to establish further evidence of validity of the ADL ability measures of the AMPS. Thus, the AMPS ADL ability measures could discriminate between women with CWP/FM and healthy women. The ADL ability measures of remained stable when no intervention was provided and the ADL motor ability measures were sensitive to change among women with CWP/FM. Moreover, the ADL ability measures were sensitive to change among clients with low ADL ability following ABI. Finally, only low to moderate correlations were found between measures of perceived and observed ADL ability.

Conclusions

Rasch analyses of revised versions of the ADL Taxonomy suggest that the majority of the ADL tasks and rating scales focused on quality of ADL task performance can be used to obtain linear measures of quality of ADL task performance based on methods of questionnaire, interview or observation. Furthermore, the studies provided evidence of validity of the ADL ability measures of the AMPS in relation to discrimination, stability and sensitivity to change and relation to perceived ADL ability. The results may, therefore, enhance the use of ADL instruments that provide linear measures of perceived and observed quality of ADL task performance in occupational therapy clinical praxis and research.

Key words

Activities of daily living, Acquired brain injury, ADL evaluation, ADL task performance, Chronic disease, Chronic widespread pain, Interview, Observation, Occupational therapy, Observation-based evaluation, Questionnaire, Rasch measurement, Rheumatic diseases.

Abbreviations

ABI	Acquired brain injury
ADL	Activities of daily living
ADL-I	Revised ADL Taxonomy – Interview
ADL-O	Revised ADL Taxonomy – Observation
ADL-Q	Revised ADL Taxonomy - Questionnaire
AMPS	Assessment of Motor and Process Skills
CI	Confidence interval
CTT	Classical test theory
CWP	Chronic widespread pain
DIF	Differential item functioning
DTF	Differential test functioning
FIQ	Fibromyalgia Impact Questionnaire
FM	Fibromyalgia
GMF	General Motor Function Assessment Scale
IADL	Instrumental activities of daily living
M	Mean
MnSq	Mean square
OA	Osteoarthritis
OTIPM	Occupational Therapy Intervention Process Model
PADL	Personal activities of daily living
PCA	Principal component analysis
RA	Rheumatoid arthritis
SAH	Subarachnoid haemorrhage
SE	Standard error
TBI	Traumatic brain injury

Original Papers

This thesis is based on the following papers:

I Wæhrens, E.E., & Fisher, A.G. Developing linear ADL ability measures based on the ADL Taxonomy: a Rasch analysis. Scand J Occup Ther 2009;16:159-171.

II Wæhrens, E.E., Bliddal, H., Danneskiold-Samsøe, B., Lund, H., & Fisher, A.G. Differences between questionnaire- and interview-based measures of ADL ability and their association to observed ADL ability in women with rheumatoid arthritis, knee osteoarthritis and fibromyalgia. *Manuscript submitted for publication*.

III Wæhrens, E.E., Amris, K., & Fisher, A.G. Observation-based assessment of ADL ability among women with chronic widespread pain. *Manuscript submitted for publication*.

IV Wæhrens, E.E., & Fisher, A.G. Improving quality of ADL task performance after rehabilitation among people with acquired brain injury. Scand J Occup Ther 2007;14:250-257.

Preface

Formålet med ergoterapi er at forebygge og behandle ADL problemer. Undersøgelse og vurdering af ADL problemernes art, omfang, årsag og betydning for klienten går forud for iværksættelse af enhver form for ergoterapeutisk intervention og foretages i hele forløbet. . . . ADL vurdering bliver (dermed) fundamentet i det ergoterapeutiske arbejde.

[The aim of occupational therapy is to prevent and treat ADL problems. Examination and evaluation of the nature, extent, cause and meaning of the ADL problems to the client precedes implementation of any kind of occupational therapy intervention and is performed throughout the entire process. . . . ADL evaluation (therefore) becomes the foundation for occupational therapy praxis.]

Tjørnov, 1988, p. 73 & 75.

This quote is from the first occupational therapy theory text book published in Danish(1). It was published soon after I graduated from my occupational therapy education. As I was an inexperienced therapist, I studied the book and found limited overlap between occupational therapy as it was described in theory and my actual clinical praxis. Reading it made me question the types of evaluations that I performed. At that time, I worked in a neuro-geriatric rehabilitation setting, where I was expected to determine the clients' potential to benefit from rehabilitation and make recommendations for future living placement based on table-top tests of cognition and sensory motor assessments followed by nonstandardized ADL observations. The fact that there often was a lack of agreement between my body function test results and what I observed during my clients' performances of activities of daily living (ADL) bothered me. I trusted intuitively in my nonstandardized ADL observations, but I did not have available instruments to help me document what I saw.

Thus, it became a personal interest — an occupation — and later, a professional issue, to be involved in the search for, translation of, education in relation to, validation of and implementation of ADL instruments useful in clinical praxis. In the process, I learned about instrument development and measurement from the perspectives of classical test theory (CTT) and

Rasch measurement theory(2;3). I also gained experience with the implementation of diverse ADL instruments in clinical praxis. My introduction to the Occupational Therapy Intervention Process Model (OTIPM)(4;5) helped me to think about different types of ADL instruments for different purposes: evaluation of the client's perceived ADL ability, ADL ability as observed by professionals, evaluation of reasons for decreased ADL ability and, finally, evaluation of changes in perceived and observed ADL ability. After experience implementing the OTIPM and using two ADL instruments, the ADL Taxonomy(6;7) and the Assessment of Motor and Process Skills (AMPS)(8;9), in clinical practice at a neuro-rehabilitation centre, I had the opportunity to use the data for research purposes. My research questions related to the applicability of the AMPS as an outcome measure in persons with moderate to severe acquired brain injury and the possibility of developing linear measures of ADL ability based on the ADL Taxonomy. In both studies, data were based on professional observations of ADL task performances as these persons with acquired brain injury were not expected to be able to report their ADL ability in a reliable manner.

Subsequently, I became involved in research concerning people with rheumatic diseases at the Parker Institute and the Department of Rheumatology at Frederiksberg Hospital. In contrast to observation-based ADL evaluation within neuro-rehabilitation, evaluation of ADL ability within the field of rheumatology was predominantly based on self-report. Based on my clinical background, I expected that perceived and observed ADL ability represented different but complimentary aspects of ability. I decided, therefore, to build upon my previous experience with the AMPS and the ADL Taxonomy and investigate the utility of AMPS and the possibility of creating linear measures of perceived ADL ability based on the ADL Taxonomy in women with chronic rheumatic diseases.

The results of these studies are now united in this thesis. My hope is that it will strengthen understanding of ADL evaluation and documentation of outcomes as a foundation for occupational therapy praxis and research.

Eva Ejlersen Wæhrens

Rationale

The foundation for occupational therapy intervention is evaluation[1] of clients' abilities to perform everyday life tasks(1), including activities of daily living (ADL) tasks that the client wants to, needs to or is expected to perform(5;12). This can be a challenge when we work with client groups with fluctuating or very low ADL ability and/or limited potential to change due to long-term or chronic disabilities following rheumatic disease or acquired brain injury (ABI).

When occupational therapists evaluate their clients' ADL abilities, they evaluate not just whether or not their clients are independent or require assistance and if they require assistance, how much assistance they need. They also consider the safety risk, decreased time-space organization (i.e. decreased effiency) of the task performance and/or increased physical effort or clumsiness related to the ADL task performance. That is, occupational therapists evaluate ADL ability based on the quality of the ADL task performance. According to the Occupational Therapy Intervention Process Model (OTIPM)(5), the occupational therapist first evaluates the client's perceived ability to perform important ADL tasks, followed by professional observation of the client's ADL task performance. These evaluations of the client's perceived and observed ADL ability enable the occupational therapist to, in collaboration with the client, identify the nature and extent of the clients' ADL task performance problems, set goals and plan interventions focused on improving the quality of ADL task performance.

After intervention, the occupational therapist re-evaluates the client's quality of ADL task performance and documents change to establish effectiveness of the provided interventions. Whether re-evaluation is based on self-report or observation varies among client groups. While clients with ABI are not expected to report reliably and, therefore, observation-based instruments most often are used, clients with rheumatic diseases are generally evaluated based on self-report using questionnaires and/or

[1] In this thesis, the term *evaluation* is used synonomously with assessment(10) and refers to gathering and judging information about aspects related to a client's ADL ability. In the occupational therapy process, evaluations are performed to clarify the nature and extent of the person's performance problems, plan appropriate interventions, and determine the effectiveness of such interventions(11).

interviews. There is, however, limited evidence to support that evaluations based on self-report and observations can be used to generate similar information, suggesting that more than one method should be used when evaluating and documenting change(13-15) to "capture" a more complete picture of ADL ability. Still, the relationship between perceived and observed ADL ability among clients with rheumatic diseases needs to be further explored.

To determine the nature and extent of ADL task performance problems and document change, there is a need for instruments[2] that can be used to obtain linear measures[3] of ADL ability. To date, however, there is limited availability of such instruments. The Assessment of Motor and Process Skills (AMPS)(8;9) represents the only instrument that can be used to evaluate the observed quality of ADL task performance and obtain linear measures that can be used to clarify the nature and extent of the person's ADL task performance problems and document change. While the ADL ability measures of the AMPS have been found to be valid among a variety of client groups with long-term or chronic disability (18-24), they have not been validated among client groups with fluctuating or very low ADL ability and/or limited potential to change due to long-term or chronic disabilities following rheumatic disease or acquired brain injury (ABI). Moreover, no instruments have been developed to be used to obtain linear measures of perceived quality of ADL task performance. The ADL Taxonomy(6), however, has potential to become an instrument applicable for measuring perceived and observed quality of ADL task performance as (a) it is widely used, (b) evaluations can be done either through self-report or observation, and (c) it allows for application of a rating scale based on quality of performance.

Development and validation of instruments that can be used to obtain measures of perceived and observed quality of ADL task performance useful to clarify the extent of the person's ADL task performance problems will

[2] ADL evaluations can be performed in an informal, non-standardized manner or based on standardized administration criteria. In this thesis, an *ADL instrument* is defined as a tool used for the purpose of performing standardized ADL evaluations.

[3] Instruments can provide descriptive, categorical (ordinal) or quantitative (linear) information. While categorical data such as total scores based on summed ordinal data often are used to "measure", in order to have measurement, we need instruments that can be used to obtain quantitative information in the form of unidimensional linear measures based on additive numbers (2;16;17). To stress this, the term *linear measures* will be used in this thesis.

increase the opportunities to implement and grade interventions directed at such problems. Moreover, if such instruments also can be used to obtain measures useful to document the effectiveness of the occupational therapy interventions, such instruments can be used to increase knowledge within occupational therapy.

The purpose of this thesis, therefore, was to build upon the existing ADL Taxonomy and develop and validate revised versions that could be used to generate measures of perceived and observed quality of ADL task performance among client groups with fluctuating or very low ADL ability due to long-term or chronic disabilities following rheumatic disease or acquired brain injury (ABI). In addition, there was a need for further validation of the ADL ability measures of the AMPS with these groups. More specifically, there was a need to verify sound psychometric properties of the ADL ability measures of the AMPS related to discrimination, stability and sensitivity to change among clients with fluctuating or very low ADL ability. Furthermore, to determine if instruments based on questionnaires, interviews and observations can be used to obtain similar information about ADL ability among clients with rheumatic diseases, the relationship between data based on methods of self-report and observation needs to be further explored.

Introduction

The focus of this doctoral thesis is on the development and validation of instruments that can be used to obtain linear measures of clients' abilities to perform ADL tasks that can be used in occupational therapy clinical praxis and for research purposes. More specifically, my focus is on instruments that can be used to generate measures of perceived and observed quality of ADL task performance among clients with fluctuating or very low ADL ability and/or limited potential to change due to long-term or chronic disabilities following rheumatic disease or acquired brain injury (ABI). Therefore, in the following sections, I will clarify some concepts related to occupational therapy and ADL, and address issues related to ADL evaluation and instrument development.

ADL from an occupational therapy perspective

A fundamental premise underlying occupational therapy is that occupation is a basic human need and promotes health, well-being and quality of life(12;25-27). *Occupation* pertains to engagement in doing — the performance of tasks(4;5;28) — and occurs when a person perceives purpose and meaning related to the task performances(5). Human beings perform a variety of tasks in their everyday lives, in the home, at work or in school, during leisure time and as a citizen in a society(29). The types of everyday life tasks people perform are determined by the roles they have to fulfil(30;31) and are influenced by factors such as age, gender, habits and culture(5;28;32). Although most task performances serve certain purposes (e.g. ironing a shirt to look presentable at work), their importance varies among individuals. That is, whether performance of a task might be considered occupational performance (i.e. meaningful and purposeful doing) or "just" task performance (i.e. doing any kind of task), is not defined by the task itself, but is based on the purpose and meaning associated with the performance by the individual who performs the task(5). While people perform a variety of tasks, this thesis will focus on ADL tasks.

Definitions of ADL

Although the term *activities of daily living* is often used in relation to everyday life tasks, there is no general consensus as to how to define the term. In some definitions, the term *ADL* is used to refer to basic self-care tasks(33;34), whereas others use the term in a broader sense to refer to a variety of everyday life tasks that people can be engaged in including work and leisure(29;35). For example, in the Danish occupational therapy literature, the term *activities of daily living* is used synonymously with everyday life tasks as an umbrella term covering the areas of (a) self-care, (b) domestic, (c) work and (d) leisure(1;29). In contrast, in English-language occupational therapy literature, ADL most often is only one among several areas of everyday life tasks (e.g. work, school, leisure, play)(8;30;31;36). In addition, while ADL is often used to refer to only basic or personal ADL (PADL), it sometimes is used to also refer to domestic or instrumental ADL (IADL)(36-39).

More specifically, *PADL* concerns self-care tasks, which most people need to perform regardless of gender, culture, environmental conditions and interests. They include performance of tasks such as eating, grooming, bathing, dressing and toileting(31;40). *IADL* comprise more complex tasks which are essential for living an independent life in society. They include performance of tasks such as transportation, shopping, cooking, housework and yard work(31;40). In this thesis, the term *ADL* is used to refer to both PADL and IADL tasks which are considered a sub-category of everyday life tasks.

Decreased occupational performance among clients with long-term or chronic disability

When a person experience problems related to performance of ADL tasks of importance to his or her everyday life, referral for occupational therapy becomes appropriate(12). Among the more common groups to be referred for occupational therapy are clients with rheumatic disease, including rheumatoid arthritis (RA), osteoarthritis (OA), chronic widespread pain (CWP) or fibromyalgia (FM); or acquired brain injury (ABI) including stroke and traumatic brain injury (TBI)(27;41), as such conditions often result in long-term or chronic disability (i.e. problems performing everyday life tasks)(42-52). The types of problems include difficulties retaining a job(43;46;47), loss of ability to engage in leisure tasks(44;48;50;51) and, important to this thesis, decreased ability to perform ADL tasks(49;51;52), causing changes in the person's habits and roles(44).

When clients with long-term or chronic disability are referred for occupational therapy, the aim is to enable them to engage in performance of those everyday life tasks, including ADL tasks, that the clients perceive as important for fulfilling their current and future roles(5;12). When clients are referred to occupational therapy because of problems performing ADL tasks, occupational therapists need instruments for the purpose of evaluating clients' abilities to perform ADL tasks.

Purposes of evaluation in the occupational therapy process

Since occupational therapy services are implemented following a process of problem identification, intervention and re-evaluation, occupational therapists need ADL instruments for the purposes of (a) clarifying the nature and extent of the client's performance problems, (b) planning appropriate interventions, and (c) establishing effectiveness of the interventions provided.

According to the Occupational Therapy Intervention Process Model (OTIPM)(5), the first step in the occupational therapy process is to establish the client's performance context including gathering information about what ADL tasks are relevant to this client. That is, this step concerns a first insight into the nature of the client's ADL task performance problems. Such information can be gathered through varies sources including the client's self-report. In the next step, the client identifies which of the relevant ADL tasks are problematic to perform and which ADL tasks can be performed without any problems. In this process, the client also prioritizes tasks based on their importance to the client's everyday life. This step, therefore, gives further insight into the nature of the client's ADL task performance problems as well as the client's perceived ADL ability (i.e. the extent of perceived ADL task performance problems). Such information is gathered based on the client's self-report and is followed, in the next step, by professional observation of the client's performance of important and problematic ADL tasks to further evaluate the extent of the ADL task performance problems. Thus, the extent of the ADL task performance problems (i.e. ADL ability) is first evaluated from the insider's perspectice and then from the outsider's perspective. All the information gathered form the basis for intervention planning. To establish effectiveness of the intervention, the last step in the process is to re-evaluate the extent of perceived and observed ADL task performance problems (i.e. document change in perceived and observed ADL ability).

Characteristics needed for ADL instruments in occupational therapy

For instruments to be suitable for the purposes of occupational-therapy-specific evaluations, certain characteristics are needed. These will be described in more detail below.

Client-centred and occupation-focused evaluation

When occupational therapists use ADL instruments in clinical practice, it is important that they choose to use instruments which have been developed to take into consideration certain premises that underlie our profession(5;11;53;54). That is, according to contemporary occupational therapy theory(5;12;30;31), occupational therapy is characterized by being client-centred and occupation-focused, without consideration of specific diagnostic groups. Thus, the ADL instruments used by occupational therapists need to be client-centred, occupation-focused and generic with regard to diagnosis or other reason for diminished ADL task performance.

Performing ADL evaluations in a *client-centred* manner implies initial and ongoing collaboration with the client throughout the evaluation process to ensure that the client's perspectives and priorities are incorporated into the process(5;12;30). This can be addressed by implementing ADL evaluations based on self-report using methods of questionnaires and interviews to capture the client's perspective on the ADL task performance(30). When implementing observation-based ADL evaluations, another way to incorporate the client's perspective is to focus the observation on those ADL tasks that the client has identified as important. If the person is not able to self-report such information, questionnaires and interviews can be used for proxy-report, where relatives report on behalf of the person.

Following an *occupation-focused* approach (i.e. focusing on those task performances that the person perceives as purposeful and meaningful in his or her everyday life) implies that occupational therapists focus on occupation throughout the process of evaluation, intervention and documentation as well as on improved occupation as the outcome of the process. In contrast, an occupation-based approach to evaluation involves the use of occupation (i.e. the person's engagement in ADL task performances) as a therapeutic means for the purpose of evaluation(5;28;55). Thus, while self-reports can be client-centred and occupation-focused, only observation-based instruments can be occupation-based. In both cases, occupation-focused and occupation-based, an important premise is that the instrument is designed to allow for

evaluation only of those ADL task performances that the client perceives as relevant and meaningful within the context of daily life. Thus, occupation-focused and occupation-based approaches also incorporate the principle of being client-centred.

Evaluation based on self-report versus observation

Through the use of self-report methods like ADL questionnaires and/or ADL interviews, the occupational therapist can establish which ADL tasks are important in this client's everyday life, which of these ADL tasks the client experiences as having problems performing and the nature and extent of performance problems perceived by the client. By implementing observations of the client during performance of those ADL tasks reported as important but problematic to perform, the occupational therapist evaluates the client's quality of performance of important ADL tasks and the extent of performance problems observed.

Among other professionals, the preferred method for evaluation of ADL ability often varies for clients with long-term or chronic disability depending on their diagnosis. Self-report questionnaires have been the preferred method within rheumatology research(56-58), which can be explained by the fact that they generally are viewed as being time-efficient and inexpensive(59). In contrast, observation-based evaluations of ADL ability are commonly used within neuro-rehabilitation when evaluating clients with moderate to severe disability following ABI(60-62). One major reason for using observation rather than self-report might be that clients with ABI often have perceptual, cognitive or language disorders that result in problems communicating their perceived ADL abilities.

Still, the complementary use of both methods is supported by the fact that several researchers have concluded that data gathered through questionnaires, interviews and observations provide distinct but complementary information about ability (13;24;63-68). Yet, because ADL evaluations based on observation are rarely used in studies concerning clients with rheumatic diseases, very little is known about the extend to which perceived ADL ability reflects observed ADL ability in that population.

Generic versus disease-specific evaluation

Some ADL instruments are disease specific, focusing on particular ADL tasks thought to be affected by certain diagnoses. Often, the evaluation of ADL ability is included as a subscale within a more global, multidimensional,

disease-specific instrument like the physical function subscale of the Fibromyalgia Impact Questionnaire (FIQ)(69). Occupational therapists, however, work with clients with a variety of diagnoses and implement interventions directed towards enabling performance of ADL tasks across target groups. Thus, in order to be able to compare the nature and extent of ADL task performance problems and establish effectiveness of provided interventions across groups, occupational therapists need ADL instruments that are suitable for all target groups and clinical settings, rather than diagnosis-specific instruments.

Evaluating quality of ADL task performance

Traditionally, evaluations of clients' abilities to perform ADL tasks have been based on the need of assistance during the ADL task performances to determine the level of care needed, which is reflected in common ADL instruments developed for interdisciplinary use within rehabilitation(70;71). From that perspective, the level of ADL ability can be rated roughly as can or cannot perform the task or in more detail by grading the amount, frequency or type of assistance needed in order to perform the task (e.g. independent, verbal assistance, minimal physical assistance, maximal physical assistance). Assistance often is provided when the person is at risk during task performance, but other reasons for receiving assistance can be decreased quality of ADL task performance due to increased amount of time and/or increased physical effort while performing the task. Moreover, many clients are able to perform ADL tasks without assistance, but still demonstrate decreased quality of ADL task performance that impacts their overall ability to manage their everyday lives. ADL instruments that are scored based on need for assistance will, therefore, often display ceiling effects when used with such higher level functioning clients.

For example, clients with rheumatic diseases are often able to perform single ADL tasks independently, but report that they experience increase in effort and/or fatigue, decreased efficiency and/or safety risks(72). Such increased amount of time and/or increased clumsiness, effort or fatigue can be early signs of decreased ADL task performance and later need for assistance in clients with only mild disability(40). Also, rheumatic diseases often have an irregular course resulting in fluctuating ADL ability. This can mean that clients with rheumatic diseases one day or week might be independent and the following day or week need assistance. These clients, therefore, have to adapt by changing their habits and redefining their roles. For example the clients reduce the amount of hours they work, require a job

that is flexible or stop working to be able to independently perform all or most ADL tasks. Another possibility is to have family members take over parts of their IADL tasks (e.g. leaving all house cleaning to spouse) permanently or on "bad" days(72).

When employed with such clients, instruments used to evaluate ADL ability based solely on assistance are, therefore, inadequate both for clarifying the nature and extent of ADL task performance problems and for documenting outcomes related to change in the amount of time and effort they spend while performing ADL tasks. In addition, in order to document outcomes of interventions among a group with higher, but fluctuating, levels of ADL ability, the occupational therapist has to use ADL instruments sensitive to change in their quality of ADL task performance, but at the same time be able to determine how much of the "change" merely is a result of the irregular course of disease.

When occupational therapists work with clients with very low ADL abilities, such as clients with moderate to severe disabilities following ABI, evaluation and documentation of outcomes based on the need for assistance offers an additional challenge. That is, most of these clients need physical assistance to perform nearly all ADL tasks. The result is that such instruments will display floor effects as it is not possible to further grade their level of ability in order to discriminate between clients with higher and lower ADL ability(73). Yet another challenge is that independence in ADL can be hard to achieve for these clients. Yet, such clients, despite continued need for assistance, may now perform ADL tasks with decreased amount of time, decreased effort and/or increased safety. Thus, ADL outcomes based on instruments only evaluating need of assistance can leave the impression that no or non-needed improvements have occurred and that clients have not benefited from rehabilitation.

Considered together, traditional instruments used to evaluate ADL ability based on assistance can become inadequate for documenting outcomes as there is a risk for ceiling effects, floor effects and/or lack of sensitivity. These, in turn, create a risk for not being able to document improved quality of ADL task performance. Perhaps even more important from an occupational therapy perspective is that information only about the client's level of independence is gathered. Consequently, it is critical to use instruments covering a broad range of ADL tasks that enable the occupational therapist to evaluate the nature and extent of the client's ADL task performance

problems, discriminate between different levels of ADL ability and document change among clients with very high or low levels of ADL ability. Such instruments, therefore, must be designed to take into consideration other aspects besides assistance related to quality of ADL task performance (e.g. time, physical effort, safety).

Ordinal data versus interval measures

Most instruments developed for the purpose of evaluating ADL ability are based on ordinal rating scales which classify perceived and/or observed ADL task performance into categories along an ordinal continuum (e.g. 1 = unable, 2 = need physical assistance, 3 = need verbal assistance, 4 = able). While higher numbers may represent more ability, the numbers are not distributed at equal intervals (i.e. interval scale)(16). While we can assume that person A, who gets a "2", has less ability than person B, who gets a "4", we cannot conclude that person B has twice as much ability as person A. The symbols "1", "2" and so on are merely codes for the categories. As long as data analysis is based on the fact that ordinal data is qualitative data, it can be used to describe and classify changes in ADL ability (e.g. how many clients fall in the category of needing physical assistance when bathing).

Nevertheless, it has been the tradition to document ADL outcomes by treating ordinal data as if it were interval data, summing numbers from several categories across ADL tasks into a global total score. That is, although symbols like "1" and "2" are not linear data, they are often incorrectly treated as such when used in mathematical and statistical manipulations(3;16). Thus, they are commonly summed to create a total score and the total score is likewise assumed to be "the measure" of ADL ability. Conclusions about ADL ability based on ordinal data may, therefore, be misleading. For the purposes of evaluating the extent of clients' ADL task performance problems (i.e. measuring rather than describing and classifying the level of ADL ability), discriminating among clients with different levels of ADL ability and establishing effectiveness of provided interventions, we need instruments that can be used to obtain linear measures of ADL ability. The advantage of having linear measures is well documented in the literature(2;16;17;74;75).

In summary, for occupational therapists to evaluate their clients' abilities to perform ADL tasks, they need both self-report and observation-based ADL instruments that have been designed to:

1. Support a client-centred and occupation-focused approach that enables evaluation of those PADL and IADL tasks that are relevant to the client's everyday life
2. Promote a focus on quality of ADL task performance, not just need for assistance
3. Can be used across clients, target groups, clinical settings and with persons along the full range of ADL ability
4. Generate valid and reliable measures of quality of ADL task performance that are sensitive enough to be able to document change in the client's quality of ADL task performance.

Some ADL instruments have been developed within the field of occupational therapy in an attempt to address some of these needs.

Occupational-therapy-specific instruments used for ADL evaluation

Among the occupational-therapy-specific ADL instruments commonly used in the Nordic countries are the Canadian Occupational Performance Measure (COPM)(76;77), the ADL-focused Occupation-based Neurobehavioral Evaluation (A-ONE)(78;79), the ADL Taxonomy(6;7;80), and the Assessment of Motor and Process Skills (AMPS)(8;9). The COPM(76;77) is designed as an individualized, client-centred and occupation-focused semi-structured interview used to uncover the client's perspective of occupational performance problems and related goals. The everyday life tasks considered in the interview includes self-care tasks, productivity (including IADL) and leisure. After the client has identified, defined and prioritized tasks of importance to his or her everyday life, he or she rates how well they can perform up to five prioritized tasks and their satisfaction with the task performances using visual analouge scales from 1-10. Since the types of tasks included in the evaluation and the definition of the tasks varies among clients, information gathered among several clients can not be used to discriminate among clients with different levels of ADL ability. The A-ONE(78;79) is limited because it was not designed to evaluate both PADL and IADL. In contrast, a potential advantage of the A-ONE, at least when used to evaluate clients with neurological disorders (e.g. stroke, dementia), is that it is the only instrument developed to evaluate a client's PADL task performance and at the same time establish which underlying neurobehavioral dysfunctions cause diminished PADL task performance in the natural context of task performance (i.e. occupation-based).

The ADL Taxonomy(6;7;80) is a client-centred and occupation-focused instrument concerning performance of PADL and IADL tasks. In total, 47 ADL tasks have been defined and organized into hierarchies within 12 ADL domains. The ADL Taxonomy can be used to gather information based on either self-report or observation. In case of observation, the ADL Taxonomy is also an occupation-based instrument. When using the ADL Taxonomy, the occupational therapist is free to define and apply his or her own rating scale. While the ADL Taxonomy has been shown to be clinically useful to describe ADL ability, it cannot be used for the purpose of measuring ADL ability and documenting outcomes. The ADL Taxonomy has been applied among clients with a variety of diagnoses(66;81;82). The AMPS(8;9) is a client-centred, occupation-based instrument designed to evaluate quality of PADL and IADL task performance based on observation. Studies support the validity of the linear ADL ability measures of the AMPS across age groups(83), genders(84) and diagnostic groups(18-24;85-94) and is the only ADL instrument that fulfils all the requirements listed above. It would, however, further increase the utility of the ADL ability measures of the AMPS if they can be shown to be valid when used to evaluate change among clients with fluctuating or very low ADL ability and/or limited potential to change due to long-term or chronic disabilities following rheumatic disease or ABI. Therefore, the focus of this thesis will be on development of the ADL Taxonomy, and further validation of the ADL measures generated by the AMPS and revised versions of the ADL Taxonomy, when applied to clients with fluctuating or very low ADL ability following rheumatic disease or ABI.

Development and validation of ADL instruments

A guide for instrument development and validation

Almost three decades ago, Benson and Clark(95) described a detailed step-by-step process for planning, developing and validating new instruments. The process was divided into four phases concerning (a) planning, (b) construction, (c) quantitative evaluation and (d) further validation. While new methods since have been developed for the purpose of instrument development, Benson and Clark's four phases continue to provide a fundamental guide for the development process. The aim of the *planning phase* is to define the purpose and content of the instrument, and methods for data collection. In the *construction phase*, items reflecting the content are specified and evaluation criteria are developed. Then, after an

initial validation process related to the content, items and evaluation criteria are revised, if needed. Through the *quantitative evaluation phase*, one or more pilot studies are conducted for the purpose of item analysis and estimation of reliability. This phase is repeated until a final version of the instrument is developed. Finally, during the *further validation phase,* there is an ongoing process where different types of studies are implemented to further validate the measures obtained by using the instrument.

Validity and reliability

When developing an instrument, validity is the most fundamental issue to be considered. The concept of validity has evolved during the 20th century such that what was previously considered three different types of validity (i.e. content, criterion-related and construct validity) now is unified into one concept of construct validity. Based on this current view, validity is defined as the degree to which evidence and theory support the interpretation of the evaluation results (i.e. measures)(96). Thus, it is not the instrument in itself that is validated, but the interpretation of the results (i.e. the measure obtained by using the instrument). Evidence of validity is based on five types of evidence: (a) test content, (b) response processes, (c) internal structure, (d) relation to other variables and (e) consequences of testing(96).

Another concern is reliability of measures. Reliability refers to the consistency of measures when an evaluation is repeated on individuals or groups(96). Some degree of variation between repeated measures is always expected due to random, unpredictable errors of measurement. Reliability studies are undertaken to determine (a) the major sources of error, (b) the size of such errors and (c) the degree of generalizability (i.e. reproducability) of measures across formats, raters and over time. The difference between the client's observed measure and the client's "true" measure is the measurment error. Thus, the smaller the measurment error, the higher the reliability of the obtained measures. Sources of measurement error can be categorized as those related to the client (e.g. fluctuating level of ability) and those external to the client (e.g. physical environment, rater subjectivity) (96). Reliability is traditionally reported in terms of standard deviations of measurement errors (i.e. standard error [*SE*]) and in terms of reliability coefficients (96).

The development of the ADL Taxonomy and the AMPS, to various extents, followed the phases described by Benson and Clark(95). In the following sections, I will describe these two instruments in more detail. I will also briefly summarize the development of each instrument and the existing

evidence supporting the validity and reliability of the ordinal ADL Taxonomy scores and linear ADL ability measures of the AMPS.

ADL Taxonomy

The ADL Taxonomy is an instrument developed to be used by occupational therapists to describe ADL task performance across diagnostic groups and clinical settings(6;7;80;97). The ADL Taxonomy covers 12 ADL domains[4] within PADL and IADL. Each domain comprises a group of tasks intended to be hierarchically organized from the easiest to the most demanding task(80). The number of ADL tasks within each domain varies between two to six, with a total of 47 ADL tasks. When occupational therapists use the ADL Taxonomy, they consider the client's needs, habits and roles, and commonly evaluate only those ADL domains and ADL tasks that are important to the client. Evaluations can be done either through self-report or based on observations. The evaluation results can be recorded on either an ADL Taxonomy Circle or an assessment form. When using the ADL Taxonomy Circle, the occupational therapist decides what kind of information to record and uses any combination of symbols, lines or colours on the Circle to visualize the information recorded (e.g. use a bold line around the tasks the person performs independently and "X" to mark all tasks that interventions are directed towards). In the assessment form, the occupational therapist records what the person "can do", "does do" and/or "wants to be able to do" using a dichotomous scale (yes/no)(6).

Two research studies have been published describing the development of the ADL Taxonomy(80;81). Initially, the ADL tasks and domains of the ADL Taxonomy were identified through a literature review and verified through expert opinions to confirm evidence of validity based on content and internal structure(80). This resulted in the ADL Taxonomy Circle and a manual with definitions and suggestions for how to describe information in the Circle. Sonn, Törnquist and Svensson(81) further confirmed content validity of the ADL Taxonomy in relation to the suggested hierarchies of tasks within each domain. Based on a dichotomous scale of ability to perform the ADL tasks (actually do/do not), they established an ordered categorical structure within the ADL domains. Based on the findings of that study(81), some revisions were made and the latest version of the ADL Taxonomy was published(6;81).

[4] To use a more uniform terminology across descriptions of ADL instrument, I have made some systematic changes to the original ADL Taxonomy terminology, developed by Törnquist and Sonn(80). *"Activity domain"* has been changed to "*ADL domain"* and "*action"* to *"task"*.

This version also included the assessment form. Subsequently, ordinal data based on the ADL Taxonomy have been used in a number of research studies.

The ADL Taxonomy has been used to evaluate ADL based on self-report for the purposes of describing and classifying ADL ability related to the nature of ADL task performance problems(98), gender differences(82) and differences between clients and healthy controls(99). Furthermore, the ADL taxonomy has been used to examine differences between modes of self-report(66), change (82;100-103) and early signs of disability(99) among geriatric clients(98), and clients with Parkinson's disease(82;99-101;104), hip fractures(102) and neurologic diseases(66;103).

Scoring has been based on a variety of rating scales including dichotomized scales of independent (can)/dependent (can not)(66;102;105), do/do not do(66;98;102) and want to/do not want to(98;102) and scales grading the level of assistance and/or effort in ADL task performance into three to five rating scale categories(99-101;104). In some of these studies, the results have been reported based on summed ordinal scores within(82;99;100) or across ADL domains(103;105) or illustrated using profiles of mean (*M)* values and standard errors (*SE*) for each ADL task(99-101;104). In other studies, the ADL Taxonomy results have been reported based on the number of people in certain categories (i.e. "can and wants to", "can but does not want to", "cannot but wants to" and "cannot and does not want to"(98;102)) on each ADL task(98;102). In addition, data based on ADL Taxonomy interviews has been used to provide validity evidence based on relations to other variables of the General Motor Function Assessment Scale (GMF), an instrument designed to evaluate observed dependence, and self-reported pain and insecurity related to motor function (i.e. physical movement and arm/hand function); the results revealed moderate to high correlations between four out of 12 ADL Taxonomy domains (mobility, going to the toilet, dressing and personal hygiene) and the areas of dependence and insecurity in the GMF(105).

To summarize, the occupational therapist can use the ADL Taxonomy for the purposes of categorical, descriptive evaluation of perceived and observed ADL task performance among different diagnostic groups, focusing on those ADL tasks of relevance to the client's everyday life. When using the ADL Taxonomy Circle, occupational therapists are free to define their own rating scales, but so far, no studies have reported ADL ability data based on

evaluation of the quality of ADL task performance related to amount of time, physical effort, safety risk and/or need for assistance. Also, no studies have reported linear measures of ADL ability based on the ADL Taxonomy. Based on Benson and Clark's guide for instrument development, initial evidence of validity based on content and internal structure of the ADL Taxonomy has been confirmed in previous studies(80;81). Therefore, in the studies that comprise this thesis, we will continue the development process, with a focus on quantitative evaluation and further validation. Since Rasch measurement methods are used specifically for the development of instruments that can provide linear measures of constructs related to human behaviour (e.g. ADL ability)(3), Rasch measurement methods will be employed to develop revised versions of the ADL Taxonomy that can be used to generate linear measures of perceived and observed quality of ADL task performance (Rasch measurement methods are described in more detail below).

Assessment of Motor and Process Skills (AMPS)

The AMPS(8;9) is a standardized evaluation of occupational performance used to measure the quality of a person's observed performance of PADL and IADL tasks. A unique feature of the AMPS in relation to the focus of this thesis is that the AMPS was developed using Rasch measurement methods. Fisher(5) has defined decreased quality of performance as an observable increase in clumsiness, effort and/or fatigue; decrease in efficiency (timely and well-organized behaviour); decrease in safety and/or need for verbal or physical assistance.

When the AMPS is administered, the person chooses and performs two culturally-relevant and familiar ADL tasks of appropriate challenge. During each ADL task performance, an occupational therapist, who has calibrated as a valid and reliable rater, observes 36 ADL skills and rates the quality of the person's performance of each skill. A 4-point ordinal scale is used to rate each ADL skill based on the ease, efficiency, safety and independence observed (1 = markedly deficient, 2 = ineffective, 3 = questionable, 4 = competent). After scoring, the AMPS raters use their personal copies of a many-faceted Rasch-based AMPS computer-scoring software(106) to convert the person's ordinal ADL scores into two overall linear ADL ability measures, one for ADL motor ability and one for ADL process ability. These ability measures are expressed in logistically transformed probability units (logits)(8;107). The many-faceted Rasch model for the AMPS used to generate the linear ADL motor and ADL process ability measures adjusts

each measure to account for rater severity, skill item difficulty and ADL task challenge(8).

The overall ADL motor ability measure is an indication of how much clumsiness, effort or fatigue the person was observed to demonstrate when moving him- or herself and task objects. The overall ADL process ability measure is an indication of how timely and well-organized (efficient) the person was observed to be when choosing and using task objects and organizing the spatial-temporal actions of the task performance. Both the ADL motor and ADL process scales also reflect safety and independence in ADL task performance. ADL ability measures above the 2.0 logit cutoff on the ADL motor scale and above the 1.0 logit cutoff on the ADL process scale indicate effortless, efficient, safe and independent ADL task performance in everyday life. ADL ability measures below the 2.0 logits cutoff on the ADL motor scale and below the 1.0 logit cutoff on the ADL process scale indicate ADL skill deficits that reflect diminished quality of ADL task performance, whereas ADL motor and ADL process ability measures below the 1.5 ADL motor cutoff and/or below the 1.0 ADL process cutoff indicate a need for assistance for community living(108). Finally, changes in the ADL motor or ADL process ability measures of at least 0.30 logits have been suggested to be clinically meaningful(8).

The AMPS has been standardized on more than 100000 persons between 3 to 103 years of age internationally and cross-culturally(8;109). Studies support the validity of the AMPS ADL ability measures across age groups(83), genders(84) and diagnostic groups within neurology(18;19;85-89), geriatrics(20;21;90-92), psychiatry(22;23;93), rheumatology(24) and musculoskeletal diseases(94). Studies of test-retest(8;110) and alternate forms reliability(8;111) support stability of the ADL measures of the AMPS over time and between different pairs of ADL tasks among well adult persons as well as persons with neurologic, psychiatric, musculoskeletal and medical diagnoses. The ADL measures of the AMPS have also been shown to be sensitive when used to evaluate outcomes of intervention studies concerning interdisciplinary neuro-rehabilitation(86;87), occupational therapy (19;21;88;89;91) and pharmacological treatment(92;93).

To summarize, the occupational therapist can use the AMPS for the purposes of measuring the quality of ADL task performance based on observation focusing on the ADL tasks of relevance in the client's everyday life. While numerous studies support the use of the AMPS to evaluate ADL

ability and measure outcomes across diagnostic groups and clinical settings, it would add further evidence to support the validity of the ADL ability measures of the AMPS if they can be shown to have good psychometric properties among clients with long-term or chronic disability following ABI or rheumatic disease. Such studies can be related to the phase of *further validation* in Benson and Clark's guide for instrument development.

Development of ADL instruments based on Rasch measurement models

The family of Rasch measurement models offers an approach to instrument development and data analysis that results in linear measures(3). The models are based on the idea that in order to create instruments for the purposes of measurement, we need to evaluate only one construct at a time (i.e. the construct evaluated by the instrument must be unidimensional). The construct (e.g. ADL ability) should be represented by a list of items forming a hierarchically ordered continuum, a line, from more to less difficult items. In the case of an instrument designed to measure ADL ability, this line will also represent a continuum in the amount of ADL ability the persons being measured demonstrate, from more to less overall ADL ability(2;3;107).

When the constructed line is based on the Rasch rating scale model, two facets are considered (items and persons): (a) the more able a person, the more likely that person is to receive higher ratings on harder items than is a less able person; and (b) the easier an item, the more likely all persons are to receive higher ratings than they are on harder items(3). When the data meet these expectations, the items and the persons fit the measurement model, supporting validity in relation to internal structure (i.e. internal scale) and response processes (i.e. person response), respectively. When measurement involves additional facets beyond item difficulty and person ability, a many-faceted Rasch model applies(3). The many-faceted Rasch model of the AMPS is based on assertions not only about person ability and item difficulty (as noted above), but also ADL task difficulty and rater severity(8).

Rasch analysis computer programs are used to analyse whether data meet these assertions mathematically by converting raw ordinal scores into linear measures of person ability and item difficulty. These measures are based on the logarithm of the odds of passing an item when a person with a given ability is scored on a given item(3). In this process, the items are arranged on the line based on the number of persons passing the item, and persons are placed on the same line based on the number of items passed(3). The

measures of item difficulty and person ability are expressed in equal-interval units (log-odds probability units – logits) of either difficulty or ability, respectively.

Whether the generated measures are valid depends on how well the data fit the model assertions. Rasch analysis computer programs generate several statistics that can be used to evaluate the validity and reliability of the generated measures. When performing a Rasch analysis, prior to examining other forms of reliability or validity, it is recommended that the performance of the rating scale of the instrument be evaluated to ensure that it demonstrates sound psychometric properties(112). The analysis of the rating scale properties is followed by analyses to confirm unidimensionality. This process can be based on several types of analyses related to internal scale validity including (a) principal component analysis (PCA) of the standardized residuals, (b) examination of goodness-of-fit statistics for items, and (c) evaluation for differential item functioning (DIF)(113;114). The PCA of the standardized residuals (i.e. the difference between what the Rasch model predicts and what was observed) is performed to identify possible secondary dimensions within the data. Likewise, the purpose of examination of goodness of fit is to detect item responses that do not agree with the model expectations(3). Examination of fit is based on two types of statistics, mean squares (*MnSq*) statistics and the standardized (*z*) statistics. While the *MnSq* is the mean of the squared difference between what was expected and observed (i.e. the size of the misfit), the standardized *z* value shows the probability of the *MnSq* value (i.e. the significance of the misfit)(115). *MnSq* and standardized *z* values are provided for both outfit and infit misfits. While outfit statistics are more sensitive to unexpected responses on items that are very easy or very hard, infit statistics are more sensitive to unexpected patterns of responses on items that are roughly targeted to the people(116). Finally, DIF occurs when measures of item difficulty vary between groups. Items that display DIF, therefore, can represent risks to the requirement of unidimensionality(113).

More specifically, while unidimensionality theoretically is considered a prerequisite for measurement, in reality it is considered impossible to obtain as other factors will influence a person's performance beyond the construct intended to be measured(2;114). Rather, as suggested by Smith(114), unidimensionality should be viewed on a continuum and the question then becomes, "at what point on the continuum does multidimensionality threaten the interpretation of the item and person measures?". Therefore, if

items considered clinically useful have been identified to display misfit and/or DIF, the impact on the measurement system of retaining them can be further investigated by evaluating for differential test functioning (DTF)(113). DTF analyses can be performed by constructing 95% control lines on a plot of person ability measures based on separate analyses of two different versions of the instrument (e.g. one version with all items and one version only containing items that fit the measurement model). This is equivalent to conducting a series of independent *t* tests between the person ability measures based on the two versions to determine if the measures varied between versions, indicating a threat to the measurement system. More recent versions of Rasch analysis computer programs such as Winsteps®(117) also provide DTF results. Ultimately, the decision to retain or omit an item displaying misfit or DIF should be based on consideration of all available evidence including the theoretical and/or practical impact of item removal. Moreover, it is generally preferred to attempt to revise items rather than to omit them(118).

In the Rasch model, reliability is estimated both for items and for persons. The precision and reproducibility of the item difficulty and person ability measures can be determined based on overall separation reliability estimates and separation indices(3;116;119). The *separation reliability estimate* is equivalent to the traditional reliability coefficient based on the ratio of "true" measure variance to observed measure variance and reflects the part of the observed measure variance that is reproducible(116;119). The item separation reliability estimate is an indication on how reproducible is the order of item difficulty measures for this set of items for this sample of persons. Similarly, the person separation reliability estimate provides information about how reproducible is the order of person ability measures of this sample of persons for this set of items. High person reliability require person ability measures well targeted to the items and a large spread of ability across the sample, whereas high item reliability depends on a large spread of item difficulty in the instrument and a large sample of persons(3;116). The *separation index* is based on the ratio of the "true" measure variance to the average measurement error (i.e. the part of the total variance that is not accounted for by the Rasch model). The person separation index indicates how well the items separate the sample into significantly different levels of ability (119). Similarly, the item separation index specifies how well the sample separates the items into significantly different levels of difficulty.

Further validation

When a final version of an instrument has been developed, the last and ongoing step is *further validation*(95). This phase might include validation of the linear measures related to (a) discrimination, (b) stability and sensitivity to change, and (c) relation to other measures obtained by similar or different methods of evaluation. For example, for the purpose of evaluating the extent of clients' ADL task performance problems (i.e. the clients' levels of ADL ability), it is important that the measures can be used to discriminate between healthy persons and clients with decreased ADL ability. There are several inferential statistics available for comparing measures among groups including analyses of variance (ANOVAs), *t* tests and *z* tests.

For the purpose of establishing effectiveness of provided interventions, it is important that measures can be demonstrated to remain stable when no intervention is provided and at the same time be able to identify clinically important changes even when they are small(120). Stability (i.e. consistency) is a form of reliability related to whether measures remain stable over time, when no change in ability has occurred. Evidence of reliability of measures related to stability are required to be able to detect real change. On the other hand, sensitivity to change refers to the efficiency of the measures when used to detect actual changes that have occured. To develop measures that are both stable and yet, at the same time, sensitive to change can be difficult. That is, too much stability can result in not being able to detect actual changes that have occurred, but too much sensitivity can result in capturing changes of no real importance to the client or the intervention.

There are several approaches to evaluation of stability and sensitivity to change. Two ways of verifying stability are to determine if two sets of measures are highly correlated (Pearson product moment correlation coefficient) and to confirm that the mean ability measures do not differ significantly over time (*t* tests or ANOVA). Similarly, to evaluate for sensitivity to change, ANOVA and *t* tests can be used to determine if mean ability measure actually do differ significantly, when interventions have been provided. Another way of examining the sensitivity to change of ADL ability measures is to calculate effect sizes. Effect sizes can be applied to evaluate interval data and the larger the effect size the more sensitive are the measures(121).

Finally, *further validation* might include studies to examine if ADL instruments based on questionnaires, interviews and observations can be used to obtain related information about ADL ability and thereby establish evidence of validity based on relation to other variables. More specifically, convergent and divergent evidence of construct validity can be determined based on the relationship between measures obtained through different methods of evaluation. Convergent evidence is indicated by high agreement between measures obtained by different instruments which are assumed to measure the same construct. The degree of agreement can either be estimated based on correlation coefficients for paired measures for each person, or examined based on analysis of differences and similarities among mean ability measures obtained by similar methods of evaluation across and within groups by means of ANOVA or *t* test statistics.

Summary

There is a need for client-centred and occupation-focused ADL instruments, based on self-report and observation, suitable for the purpose of evaluating the nature and extent of ADL task performance problems (i.e. the quality of ADL task performance) and the effectiveness of provided interventions across client groups in occupational therapy clinical praxis and research. Evaluation of the extent of a client's problems, and evaluating change following intervention, requires the use of linear measures of ADL ability. As the ADL Taxonomy already has shown clinical utility as a client-centred, occupation-focused ADL instrument, it is relevant to further explore the possibility of obtaining linear measures of perceived and observed quality of ADL task performance based on the ADL Taxonomy. In addition, there is a need for further validation of the ADL ability measures of the AMPS. More specifically, there is a need to verify sound psychometric properties of the ADL ability measures of the AMPS related to discrimination, stability and sensitivity to change among clients with fluctuating or very low ADL ability. Furthermore, to determine if evaluations based on questionnaires, interviews and observations can be used to obtain similar information about ADL ability among clients with rheumatic diseases, the relationship between data based on methods of self-report and observation needs to be further explored.

Aims of this Thesis

The overall aim of this doctoral thesis was to develop new ADL instruments based on the ADL Taxonomy that can be used to generate linear measures of quality of ADL task performance based on self-report and observation, and validate the ADL ability measures of both the newly developed versions of the ADL Taxonomy and the AMPS for use in occupational therapy clinical praxis and research, with a special focus on clients with fluctuating or very low ADL ability and/or limited potential to change due to long-term or chronic disabilities following rheumatic disease or ABI. The specific research aims were:

Specific aims

1. To develop a revised version of the ADL Taxonomy that can be used to obtain linear measures of quality of ADL task performance based on observations of clients with chronic or long-term disability following ABI and validate the ADL ability measures (Study I).

2. To develop revised versions of the ADL Taxonomy that can be used to obtain linear measures of quality of ADL task performance based on self-report using questionnaire and interview formats with clients with chronic or long-term disability following rheumatic and validate the ADL ability measures (Study II).

3. To further validate the ADL ability measures of the AMPS with regard to
   - Discrimination between healthy persons and clients with fluctuating ADL ability (Study III)
   - Stability among clients with fluctuating ADL ability when no intervention is provided (Study III)
   - Sensitivity to change among clients with fluctuating ADL ability (Study III) and clients with very low levels of ADL ability and limited potential for change (Study IV)
   - Relationship to measures of ADL ability based on self-report among clients with rheumatic diseases (Study II-III).

Methods

Participants

The participants in this thesis were adults with chronic or long-term disabilities following rheumatic disease or ABI. An overview of the participant demographics in Studies I–IV is presented in Table 1. Additional information about the participants is provided in more detail below.

***Table 1*. Overview of Participant Demographics in Studies I–IV**

Study	I	II	III	IV
Number of participants	70	118	50	36
Percentage of women	36	100	100	33.3
Mean age in years (range)	44.0 (17-69)	50.9 (21-79)	43.4 (21-64)	48.1 (17-69)
Diagnoses	ABI including stroke, SAH, TBI	RA, knee OA, FM	CWP/FM	ABI including stroke, SAH, TBI

ABI = acquired brain injury, CWP = chronic widespread pain, FM = fibromyalgia, OA = osteoarthritis, RA = rheumatoid arthritis, SAH = subarachnoid haemorrhage, TBI = traumatic brain injury.

Study I

For the purpose of the first study, retrospective observation-based ADL Taxonomy data were obtained on adults with moderate to severe ABI admitted at an inpatient neuro-rehabilitation clinic in Denmark between January 1999 and December 2003. In total, 70 participants had been evaluated between one and four times during the rehabilitation (Table 1). Forty-four participants were evaluated one time, 21 participants two times, three participants three times, and two participants four times, resulting in a total of 103 evaluations (gender distribution: female 38 [37%], male 65 [63%]).

Study II

For the second study, consecutive female clients ≥18 years of age diagnosed with rheumatoid arthritis (RA) or fibromyalgia (FM) were recruited when they were referred to occupational therapy from a rheumatologic outpatient clinic at a Danish hospital between March 2007 and October 2009. During the study period, clinicians were asked to also recruit clients with knee osteoarthritis (OA) from the same clinic. In total, 54 women with FM, 40 with RA and 24 with knee OA were included in the study (Table 1). The women with FM were significantly younger (*M*=45.2, *SD*=11.3) than the women with RA (*M*=54.5, *SD*=12.0)(Difference = 9.3, 95% CI = 4.8–13.8) and OA (*M*=57.9, *SD*=7.4)(Difference = 12.7, 95% CI = 7.4–18.0). In contrast, there were no significant differences in disease duration among the women with FM (*M*=7.3, *SD*=7.9), RA (*M*=8.7, *SD*=8.9) and OA (*M*=9.8, *SD*=9.4)(F[2,118] = 0.72, p = 0.4882). The women with FM had significantly lower ADL motor ability measures of the AMPS compared to women with RA and knee OA (Table 7). Furthermore, the women with knee OA obtained significantly lower ADL process ability measure than did the women with RA. Age had no effects on differences in ADL motor (F[2,118] = 0.38, p = 0.6863) or ADL process (F[2,118] = 0.43, p = 0.6491) ability measures among diagnostic groups.

Study III

For the third study, participants were women, diagnosed with chronic widespread pain (CWP) or FM, who were referred to a two-week, outpatient, interdisciplinary rehabilitation program in a clinical setting at a Danish hospital between November 2007 and August 2008. In total, 50 women were referred to the rehabilitation program (Table 1). Forty-one (82%) participants completed the rehabilitation program and the repeated AMPS observations. Forty-three (86%) of the total sample and eight (89%) of those who dropped out of the study were diagnosed with FM. There were no significant differences in demographic data and baseline clinical measures between participants who completed the program and those who dropped out.

Study IV

The fourth study was based upon AMPS observations performed between October 2000 and December 2003 as part of clinical practice at the same neuro-rehabilitation centre as in Study I. Adults with moderate to severe disability following ABI were included retrospectively if they had been evaluated with the AMPS at least twice during their rehabilitation, with the

last AMPS observation occurring within three weeks before discharge. Thirty-six participants fulfilled the inclusion criteria (Table 1). Injury had occurred between three to 24 months (*M*=6.9, *SD*=3.75) prior to admission to the rehabilitation clinic and their length of stay at the clinic varied between three to 16.5 months (*M*=7.8, *SD*=3.2).

Instrumentation and procedures

The evaluations of the participants' quality of ADL task performance in this thesis were based on self-report (i.e. questionnaire and interview) and observation, using the ADL Taxonomy and/or the AMPS. In Study III, the Fibromyalgia Impact Questionnaire (FIQ)(69;122) also was administered. An overview of the methods for data collection in Studies I–IV is presented in Table 2.

***Table 2*. Overview of Methods for Data Collection in Studies I–IV**

Study	I	II	III	IV
ADL instruments	ADL-Taxonomy (ADL-O)	ADL-Taxonomy (ADL-Q, ADL-I) AMPS	FIQ AMPS	AMPS
Data collection methods	Observation by inter-disciplinary team	Self-report questionnaire (ADL-Q) Interview by occupational therapists (ADL-I) Observation by occupational therapists (AMPS)	Self-report questionnaire (FIQ) Observation by occupational therapists (AMPS)	Observation by occupational therapists
Design	Retrospective register research	Prospective, consecutive, cross-sectional data collection	Prospective pre-, pre-, post- and follow-up test	Retrospective pre- and post-test

ADL-I = ADL Taxonomy-Interview, ADL-O = ADL Taxonomy-Observation, ADL-Q = ADL Taxonomy-Questionnaire, AMPS = Assessment of Motor and Process Skills, FIQ = Fibromyalgia Impact Questionnaire.

Additional information about the ADL instruments and procedures is provided in more detail below. Furthermore, an overview of evaluations employed to determine evidence of validity based on Rasch measurement methods and classical test theory is displayed in Table 3. The classification of Rasch measurement methods in relation to different types of validity evidence is based on the classification by Árnadottír(123).

***Table 3*. Overview of Evaluations Employed to Determine Evidence of Validity**

Types of validity evidence	ADL-O	ADL-Q/ADL-I	AMPS
Evidence based on internal structure	Evaluation of the psychometric properties of the rating scale Evaluation of unidimensionality using PCA Evaluation of item goodness of fit		
	Evaluation for DIF		
Evidence based on consequences of testing	Evaluation for test bias (i.e. DTF)		
Evidence based on test content	Evaluation of spread and range of items (targeting)		
	Evaluation of hierarchical ordering of items		
Evidence based on relations to other variables	Evaluation of sensitivity to change	Evaluation of agreement between measures assumed to measure the same construct (convergent evidence)	Evaluation of discriminative properties Evaluation of stability and sensitivity to change Evaluation of agreement between measures assumed to measure the different constructs (divergent evidence)

ADL-I = ADL Taxonomy-Interview, ADL-O = ADL Taxonomy-Observation, ADL-Q = ADL Taxonomy-Questionnaire, AMPS = Assessment of Motor and Process Skills, DIF = Differential Item Functioning, DTF = Differential Test Functioning, PCA = Principal Component Analysis.

ADL Taxonomy linear measures of quality of ADL task performance

The first and second specific aims in this thesis concerned the development and validation of new revised versions of the ADL Taxonomy such that it can provide linear measures of quality of ADL task performance based on observation and self-report using questionnaires and interviews among persons with chronic or long-term disability. Data to address these aims were collected in Studies I and II. In both studies, the person's needs, habits and roles were considered. Therefore, only ADL tasks which were important to the clients were evaluated. Hence, all evaluations were client-centred and occupation-focused.

ADL Taxonomy linear measures of observed quality of ADL task performance (ADL-O)

ADL Taxonomy data were extracted from the participants' medical records in Study I. Data had been recorded on ADL Taxonomy Circles by occupational therapists using a colour-coded (green, yellow and red) three-category ordinal rating scale (Table 4). Ratings were based on the occupational therapists' own observations and proxy-reports of interdisciplinary team-members direct observations of each participant's ADL task performances. Rating scale criteria were inspired by the ratings of quality of ADL task performance based on the degree of observed clumsiness, effort, efficiency, safety and need of assistance in the AMPS(8;9). As the participants' ADL abilities sometimes varied over time, ratings were assigned based on their worst performances within a week. Tasks not relevant to the person were left blank. All data were re-coded from colours into numbers to prepare for Rasch analysis (Table 4). In this way, the three-category rating scale was retained by using the scores of zero, 1 and 2 instead of colours.

Rasch measurement methods were employed to develop the ADL-O and to evaluate aspects of validity and reliability related to the linear measures of observed quality of ADL task performance. Rasch measurement statistics are described in more detail below. Moreover, the linear measures of quality of ADL task performance based on the ADL-O were further validated in relation to their sensitivity to change by comparing the initial and second evaluation of the 26 participants evaluated more than one time (Table 3).

Table 4. **Recoding the ADL-O Data from Colours into Numbers in Study I**

Colour code	ADL-O Rating Criteria	Scale
Green	Independent, safe and efficient in time and demonstrates no increase in effort (competent)	2
Yellow	Able to participate actively and in a positive way, but needs assistance (physical or verbal), uses extra time or effort or is at risk during the performance of the action or activity (inefficient)	1
Red	Not able to participate actively, or participates in a negative way that aggravates the performance of the action (deficient)	0
Blank	Not relevant	Blank

ADL-O = ADL Taxonomy-Observation

ADL Taxonomy linear measures of perceived quality of ADL task performance (ADL-Q & ADL-I)

In Study II, the 47 ADL tasks of the ADL Taxonomy formed the basis for an evaluation form designed to be used both as an ADL questionnaire (ADL-Q) and for the purpose of an ADL interview (ADL-I) (Appendix A). A new rating scale structure was developed based upon the rating scale used in Study I (Tables 4 and 5). The participants were instructed to rate the PADL tasks based on their perceived quality of ADL task performance within the last 24 hours, and the IADL tasks based on perceived quality of ADL task performance within the last week. The participants were asked to mark ("X") all applicable answers on the rating scale for each of the ADL tasks of importance to their everyday lives.

Instructions were given in writing for the ADL-Q and again verbally before the ADL-I interview. The ADL-Q was mailed out to be rated by the participants within one week prior to an occupational therapy interview based on the ADL-I. The interviews were performed by occupational therapists, who had received written instructions as well as a one-hour introduction to the ADL-I. All data were re-coded from X marks into numbers to prepare for Rasch analysis (Table 5). Numbers were based on the lowest category marked on the scale for each of the ADL tasks (e.g. if the participant had marked both "spends extra time" = 4 and "need verbal or

physical assistance" = 1 for the ADL task "prepare a hot meal", the mark was recoded into the score of "1" for that task).

***Table 5*. Rating Scales Used in ADL-O, ADL-Q and ADL-I**

	ADL-O	Scale	ADL-Q & ADL-I	Scale
Competent	Independent, safe and efficient in time, and demonstrates no increase in effort.	2	Independent, safe and efficient in time, and demonstrates no increase in effort.	6
Inefficient	Able to participate actively and in a positive way, but needs assistance (physical or verbal), uses extra time or effort or is at risk during the performance of the action or activity.	1	Able to participate actively and in a positive way, but -uses helping aids	5
			-spends extra time	4
			-experiences effort or fatigue	3
			-is at risk	2
			-needs verbal/physical assistance	1
Deficient	Not able to participate actively, or participates in a negative way that aggravates the performance of the action.	0	Does not/cannot participate actively	0
Blank	Not relevant	Blank	Not relevant	Blank

ADL-I = ADL Taxonomy-Interview, ADL-O = ADL Taxonomy-Observation, ADL-Q = ADL Taxonomy-Questionnaire

To be able to determine it the the two formats for self-report, ADL-Q and ADL-I, could be used to obtain similar information about perceived ADL ability, we needed to generate comparable linearized ADL-Q and ADL-I ability measures. In the Rasch analyses we, therefore, chose an approach where the two version were analysed in parallel. Thus, based on initial analyses we first removed items that displayed misfit in both instruments. In the following analyses, items displaying misfit in either one or both instruments were removed and the process continued until no further misfit could be detected. To investigate the relationship between ADL-Q and ADL-I ability measures assumed to measure the same construct (i.e. perceived ADL

ability) across diagnostic groups (Table 3), ANOVA was implemented to compare mean ADL-Q and ADL-I ability measures within and across groups. Furthermore correlations were used to determine the relationship between paired ADL-Q and ADL-I ability measures across and within diagnostic groups.

Further validation of the ADL ability measures of the AMPS

The final aim in this thesis was to further validate the ADL ability measures of the AMPS with regard to discrimination, stability, sensitivity to change and relationship with measures of ADL ability based on self-report among clients with rheumatic diseases (Table 3). For that purpose, AMPS data were collected in Studies II–IV (Table 2). In all three studies, participants were evaluated with the AMPS by trained and calibrated occupational therapists according to the standardized procedures outlined in the AMPS manual(8;9). In Studies II and III, participants were offered the choice of being evaluated at home or in the clinic. Prior to the AMPS observations (in Study III only initial AMPS observations) the participants had filled in an ADL-Q questionnaire and gone through an ADL-I interview.

The information gathered through the ADL-I interview about tasks perceived as important to the participant formed the basis for the occupational therapist's decision about which AMPS tasks to offer to the participant for the purpose of the AMPS observation. In Study IV, all AMPS observations were performed at the rehabilitation centre. The majority of these participants had problems engaging in a formal interview due to perceptual, cognitive and/or communication disorders following the ABI. Often, the information from the interview was supplemented with information based on informal observation of the person before reaching a final decision about what tasks to offer. When setting up the environment and making the final contract, the occupational therapist used any type of communication needed, including pointing, mimicking and showing/performing parts of the task, to make sure that the person understood what was expected of him or her. In some instances, assistance for parts of the task performance was agreed upon before hand and enacted when the person requested help.

Discrimination

Evidence related to discriminative properties of the ADL ability measures of the AMPS was examined in Study III. We compared mean ADL motor and ADL process ability measures of women with CWP/FM based on initial

AMPS observations to the respective mean ADL ability measures of a Nordic sample of healthy women (n = 393) between 21–64 years of age extracted from the AMPS International database.

Stability

Stability of the ADL ability measures of the AMPS was examined in Study III in a sample of women with fluctuating ADL ability following CWP/FM. By means of two AMPS observations scheduled at least two weeks apart, we examined if the ADL ability measures of the AMPS remained stable when no intervention was provided.

Sensitivity to change

The psychometric properties of the ADL ability measures of the AMPS related to sensitivity to change were examined among women with CWP/FM (Study III) and among clients with very low ADL ability and limited potential for change (Study IV). In Study III, we used a design based on repeated AMPS observations (pre-, pre-, post- and follow-up test). The initial and second AMPS observations were scheduled at least two weeks apart, with the second observation scheduled within two weeks prior to a rehabilitation program. The third AMPS observation was scheduled within the first week after intervention. The fourth and final AMPS observation was scheduled as follow-up four weeks after the intervention. Data from all AMPS observations were used to analyse whether the ADL ability measures of the AMPS were sensitive to change.

In Study IV, a pre- and post-test design was implemented. The initial AMPS observation was performed as soon after admission as possible based on when the client was able to actively engage in performance of ADL tasks. The second AMPS observation was performed within the last three weeks before discharge.

Relationship to measures of ADL ability based on self-report

Evaluations of agreement between measures of ADL ability based on self-reported and observation were performed in Study II and III to establish convergent evidence of validity. In Study II, we examined the overall relationship as well as the relationships within each of the three groups with rheumatic diseases between (a) ADL-Q and AMPS ADL motor and AMPS ADL process ability measures and (b) ADL-I and AMPS ADL motor and AMPS ADL process ability measures.

In addition, in Study III we examined the relationship between self-reported ADL ability based on the physical function subscale of the FIQ and observed ADL ability based on the initial AMPS observations. The FIQ(57;69) is a self-administered, disease specific instrument designed to evaluate components of health status in persons with FM. The FIQ is considered the main instrument for self-reported functional ability and health status in CWP/FIQ studies(57;124) and is composed of 10 items. More specifically, the first item on the FIQ (the FIQ physical function subscale) contains 10 sub-questions related to ADL ability — each question is rated on a four-category ordinal scale. Only ADL items relevant to the person are scored. The scores are summed and a mean is calculated based on the number of tasks relevant to the person(122). The mean is multiplied by 3.3 to obtain a possible maximum score of 10. The FIQ was mailed out to participants one week prior to the AMPS observation. Further procedures related to data collections in Studies II and III have been reported in previous sections.

Statistical analyses

In this thesis, several approaches for statistical analyses were employed including descriptive statistics, Rasch measurement statistics and inferential statistics.

Descriptive statistics

Descriptive analyses of demographics and clinical measures were based on mean, *SD,* range and percentages. We employed the SAS® statistical software version 9.1 in Studies I–III, whereas the Statistical Package for the Social Sciences (SPSS) version 9.0 was applied in Study IV. In addition, in Study III, we calculated the percentage of participants who had ADL ability measures that were more than 2.0 *SD* below the mean ADL motor and ADL process ability measures of a healthy Nordic sample.

Rasch measurement statistics

Development of three versions of the revised ADL Taxonomy that could be used to provide linear measures of quality of ADL task performance was based on the Rasch rating scale model(125). The Rasch computer program Winsteps® version 3.63.0(126) and version 3.68.2(117) were used in Studies I and II, respectively, to implement logarithmic conversions of the ordinal

scores into measures of the person's overall perceived or observed ADL ability. Winsteps® was also used to generate statistics used to evaluate aspects of validity and reliability, including fit of the data to the Rasch model assertions(3;125).

The psychometric properties of the rating scales used in ADL-O, ADL-I and ADL-Q were evaluated based on Linacre's guidelines(116;127-129). According to Linacre's guidelines(127-129), five properties need to be evaluated. First, frequency distributions across categories should be either uniform, or peak in central or extreme categories, to signal optimal category use. Second, to avoid imprecise measures of rating scale threshold calibrations, at least 10 observations of each category are required. Third, average category measures should advance monotonically up the rating scale, indicating that persons with higher ability have higher item ratings(129). Fourth, scale category outfit mean square (*MnSq*) values should be ≤ 2.0. Finally, threshold calibrations should also advance monotonically, with no threshold disordering, and thresholds should increase by at least 1.4 logits to show distinction between categories, but by no more than 5 logits to avoid large gaps in the variable (128;129).

Evidence of validity

To determine if the 47 ADL tasks defined a single unidimensional construct (ADL ability) we performed principal components analyses (PCA) of the standardized residuals and examined detailed item goodness-of-fit statistics for each of the three versions of the revised ADL Taxonomy. According to Linacre's guidelines(116) for interpretation of PCA results (a) a proportion of variance explained by the measures > 50%, (b) unexplained variance explained by first contrast < 5% and (c) variance explained by items > four times the size of the first contrast all support unidimensionality.

Assuming the PCA of the standardized residuals does not support the presence of a secondary dimension in the data, a measure is generally considered to be unidimensional when no more than 5% of the items fail to fit the Rasch model ($p \leq .05$)(114). When we analyzed goodness of fit, items with infit or outfit *MnSq* values >1.4 (130), combined with *z* values ≥ 2.0 (131) were considered to misfit, and were removed one at the time, in the order of highest *MnSq* values. While we evaluated both infit and outfit statistics, we focused first on high infit *MnSq* values, as infit misfit is a greater threat to measurement(116). In the analysis of the ADL-O data in Study I, removal of ADL items was stopped when all items met the criteria

for acceptable goodness of fit. In Study II, we wanted to make sure that the final versions of ADL-Q and ADL-I would have identical items, so that measures based on ADL-Q and ADL-I could be compared directly. Therefore, we initially analysed goodness of fit related to the ADL-I and removed misfitting items. When all items on the ADL-I met the criteria for acceptable goodness of fit, we initiated the analysis of the ADL-Q by first removing all items that had misfit on the ADL-I. We then analysed goodness of fit of the items remaining in ADL-Q and removed any additional misfitting items. Finally, we returned to the ADL-I and removed those items found to misfit on the ADL-Q. This process was repeated until all items in both formats met the criteria for acceptable goodness of fit.

As the ADL Taxonomy includes IADL actions that have been shown to vary in difficulty among men and women(132) (e.g. preparing meals and doing laundry(133)) we chose to evaluate for DIF based on gender in the analysis of ADL-O. An item was considered to display DIF when the difference in item difficulties between groups was ≥ 0.50(113) and statistically significant ($p \leq 0.05$)(134).

As the clinical relevance of the 47 ADL tasks of the ADL Taxonomy already had been verified in previous studies (80;81), in Study I we also investigated whether retaining misfitting ADL tasks and/or tasks displaying DIF would disrupt the measurement system by evaluating for differential test functioning (DTF)(113). The evaluation of DTF related to inclusion or omission of misfitting items was performed by constructing 95% control lines in a plot of (a) person ability measures based on a version only containing items with acceptable goodness of fit to the Rasch rating scale model and (b) person ability measures based on a version containing all items. Similarly, we also investigated whether retaining the items displaying DIF for gender would disrupt the measurement system. By plotting (a) person ability measures based on gender-specific item calibrations with (b) person ability measures based on common item calibrations, the variance of ability measures across a gender-specific version and a common version could be explored, for males and females, respectively.

Additionally, we verified the logical ordering of items along the scale by comparing our item difficulty hierarchy to the hierarchy originally developed by Sonn et al (81). Targeting of the items to the abilities of the participants was explored by comparing the mean difficulty measures for the ADL items

to the mean ability measures of the participants. The mean person ability measure would be close to zero for a well-targeted test(3).

Evidence of reliability

To evaluate precision and reproducibility of the item and person measures, we examined the overall separation and reliability indices. The separation indices should be at least 2.0 to obtain desired separation reliability coefficients of 0.80 for replicability of person and item ordering(3), and the closer the reliability index is to 1.0 (range 0.0 – 1.0) the better (119).

Rasch analysis of AMPS data

For the purpose of Rasch analysis of the ADL motor and ADL process item raw scores in Studies II–IV, a many-faceted Rasch model incorporated into the AMPS computer-scoring software(106) was used to convert the raw ordinal ADL skill item scores into linear ADL ability measures, adjusting for task challenge, ADL skill item difficulty and the severity of the rater.

Inferential statistical analyses

Inferential statistical analyses were employed across all studies. In all cases the level of significance was set at $p \leq .05$. The analyses performed are summarized below.

Comparisons among groups

In Study I, to determine if inclusion of data from participants evaluated more than once would result in local dependence, we used *z* tests to analyse for differences between the two sets of item difficulty measures generated from analyses of the 70 initial evaluations and all 103 evaluations. Furthermore, a paired sample *t* test was performed to evaluate for significant differences between measures based on the initial and second ADL-O evaluation among the 26 participants evaluated more than one time.

In Study II, to examine if measures of perceived ADL ability based on ADL-Q and ADL-I yielded similar results, we used mixed models analysis of variance (ANOVA) and analyzed for a main format effect. To determine if the magnitude of the difference between the two formats varied among the diagnostic groups, we examined for a format by diagnosis interaction effect. Furthermore, mixed models ANOVAs were used to evaluate for differences in perceived (ADL-Q and ADL-I) and observed ADL ability (AMPS) among the women with FM, RA and OA. As the women with FM were significantly

younger than the women with RA and OA, we included examinations for age-related interaction effects in the analyses.

In Study III, we used *z* tests to compare our sample's mean ADL motor and ADL process ability measures (pre-test1) to the respective mean ADL ability measures of a Nordic sample of healthy women extracted from the AMPS International database. Stability and change in ADL motor and ADL process ability measures were analyzed using mixed model ANOVAs with time as a repeated measure. As some participants were evaluated in their home and others in the clinic, we also included examinations for time by setting (clinic/home) interaction effects. Finally, in Study IV, paired samples *t* tests were performed to evaluate if the participants demonstrated significant differences in ADL ability between their pre- and post-test AMPS observations.

Relations between variables

In Study I, we investigated the risk for local dependence by calculating Pearson product moment correlations (*r)* between the two sets of item calibration measures generated from analyses of the 70 initial evaluations and all 103 evaluations. Furthermore, Pearson product-moment correlations were computed to examine relationships across as well as within diagnostic-groups between (a) ADL-Q and ADL-I ability measures, (b) ADL-Q and ADL motor and ADL process ability measures and (c) ADL-I and ADL motor and AMPS ADL process ability measures in Study II. In Study III, Pearson product moment correlation analysis were performed to examine the relationship between pre-test1 and pre-test2 for ADL motor and ADL process ability measures of the AMPS, respectively and Spearman rho correlations (r_s) were computed to examine the relationship between ordinal scores based on the physical function subscale of the FIQ and the ADL motor and ADL process ability measures of the AMPS at pre-test1. A high correlation between measures of the same construct indicates high convergent evidence. Correlation coefficients ≥ 0.70 were considered acceptable as convergent evidence of validity, while coefficients ≤ 0.30 were considered low convergent evidence of validity(135).

Sample size and effect size

Sample size calculation for Study III was based on a clinically meaningful difference of 0.30 logit(8) in ADL motor ability with a power of 0.80 using ADL motor ability measures of the AMPS for 19 women with CWP/FM collected pre- and post-rehabilitation. In Study III–IV, effect sizes were

evaluated by calculating standardised mean differences (*SMD*) (i.e. Cohens *d)* based on mean differences divided by the pooled standard deviations(136). Results were interpreted based on the guidelines of Cohen(137). In Study III, effect sizes were evaluated based on the ADL ability measures for pre-test1 and post-test1, as well as pre-test1 and follow-up. In Study IV, effect sizes were evaluated based on the mean ADL ability measures for pre- and post-tests.

Ethical Considerations

Approval was obtained from the Danish Data Protection Agency and the Danish National Board of Health before data for Studies I and IV were extracted from the participants' medical records. Protocols for prospective data collections, including written information for participants, were approved by the local Ethical Committee (H-B-2007-084) before initiation of Studies II–III. Participants in Studies II–III were carefully informed about their rights before they agreed to participate in the study. More specifically, they were informed both verbally and in writing about the purpose of the study, the type of evaluations they would go through, expected use of time and their rights related to withdrawal from participation. All participants in Studies II–III received AMPS Narrative and Graphic Reports based on their AMPS observation within one week after the evaluation. Participants in Study III received AMPS Narrative and Graphic Reports for each time they had been evaluated. Participants with RA in Study II, who were evaluated prior to individual occupational therapy interventions, were offered the possibility of sharing the results from the ADL-Q, the ADL-I and the AMPS observation with their occupational therapist in the clinic for the purpose of intervention planning.

Results

Linear measures of observed quality of ADL task performance based on the ADL-O

Evidence of validity based on internal structure

Psychometric properties of the rating scale

The evaluation of the psychometric properties of the rating scale revealed no category or threshold disordering in the rating scale and no rating scale misfit, indicating sound psychometric properties of the 3-category rating scale structure.

Internal scale validity

In the initial analysis of all 47 ADL items of ADL-O (Table 6), the PCA of the standardized residuals revealed that 61% of the total variance[5] was explained by the Rasch dimension, and 4.1% of the unexplained variance was defined by the first contrast supporting unidimensionality. Furthermore, the variance explained by items (16.8%) was more than four times higher than the variance defined by the first contrast. We found that 10 items (21.3%) did not display acceptable goodness of fit (high infit and outfit *MnSq* combined with high *z* values), suggesting a potential risk to unidimensionality. Following removal of these 10 misfitting items, the results of a second PCA of the standardized residuals revealed that the Rasch dimension in the 37 item version now explained 65.5% of the total variance and 3.6% of the unexplained variance was defined by the first contrast, but the variance explained by items (11.2%) were now less than four times the variance defined by the first contrast, suggesting decreased unidimensionality. Initial analysis of DIF based on gender revealed that one of the transfer items was easier for males, whereas three communication items and a cooking item were easier for females (Table 6).

5 The PCA results are updated from those reported in Study I based on re-analysis of data using Winsteps® version 3.68.2(117)

***Table 6*. Results from the Rasch Analyses of the ADL-O, ADL-Q and ADL-I**

		ADL-O 47 items	**ADL-O 37 items**	**ADL-Q 40 items**	**ADL-I 40 items**
Evidence of internal scale validity	PCA				
	Variance explained by measures	61%	65.5%	52.7%	51%
	1st contrast	4.1%	3.6%	4.7%	4%
	Variance explained by items	16.8%	11.2%	27.7%	33.6%
	Items misfitting	Manicuring Pedicuring Washing by hand Calling for attention Using the phone Reading Taking part in conversation Writing by hand Bowel & urine elimination Driving car		Pedicuring* Driving car* Riding bicycle/moped$^{\Delta}$ Shaving/ make-up¤ Going by train$^{\Delta}$ Daily shopping¤ Getting food/ cutting up/ preparing food$^{\Delta}$	
	Items displaying DIF	Walking or moving in/out of the house Using the phone Reading Writing by hand Preparing a cold meal		-	
Evidence of reliability	Mean item difficulty (SD)	zero (2.20)	zero (2.38)	zero (1.44)	zero (1.46)
	Mean person ability (SD)	0.23 (2.58)	0.55 (3.35)	3.28 (1.75)	2.50 (1.23)
	Person separation index (reliability)	5.94 (0.97)	6.01 (0.97)	3.40 (0.92)	3.19 (0.91)
	Item separation index (reliability)	6.35 (0.98)	6.30 (0.98)	6.05 (0.97)	6.91 (0.98)

ADL-I = ADL Taxonomy-Interview, ADL-O = ADL Taxonomy-Observation, ADL-Q = ADL Taxonomy-Questionnaire, DIF = Differential Item Functioning, PCA = Principal Component Analysis

*Misfit on both versions, $^{\Delta}$Misfit on ADL-I, ¤ Misfit on ADL-Q,

Evidence of validity based on consequences of testing

The DTF analysis of the variance in ability measures between the 47 item and the 37 item versions of ADL-O revealed no significant difference in person ADL ability measures between the two versions, indicating that the 10 misfitting items were not a threat to the measurement system. Similarly, DTF analyses of the variance in ability measures between the common and gender-specific versions of the ADL-O for males and females, respectively, indicated no significant differences in person ADL ability measures between the common version and the gender-specific versions. Thus, the five items displaying DIF were also not a threat to the measurement system.

Evidence of validity based on test content

To further validate the Rasch generated hierarchy of the ADL-O item difficulty measures, it was compared to the original ADL Taxonomy hierarchy reported by Sonn et al.(81). We found that only five items changed location in the item difficulty hierarchy compared to what Sonn et al.(81) reported. The differences between the original hierarchy and our hierarchy remained stable after removal of the 10 misfitting items. The targeting of the 47 item version of the ADL-O to the participants' ADL ability indicated that the participants had a slightly higher mean level of ADL ability than the mean item difficulty measure (Table 6). No participants had maximum scores, and only one person displayed minimum scores, indicating no ceiling effect and essentially no floor effect. The items and participants were well distributed along the scale with item and person measures well targeted to each other. After removal of misfitting items, the mean *SD* of the item difficulty measures increased and the mean person ability measure increased, indicating slightly diminished targeting of the 37 item version of the ADL-O to the participants (Table 6). Furthermore, the number of persons with minimum scores increased to three, signalling a potential floor effect.

Evidence of reliability

The initial person separation index of 5.94 indicated that the items separated the persons into seven or eight different ability levels(119). Moreover, there was essentially no change in the person separation index or the item separation index after removal of misfitting items (Table 6).

In summary, since our DTF analyses revealed that the retention of items displaying misfit and/or DIF did not disrupt the measurement system, and as the PCA results and reliability measures did not improve by removal of

misfitting items, we tentatively recommended retaining all items until future Rasch analyses based on larger samples can verify that they indeed do represent a different construct and/or disrupt measurement.

Evidence of validity based on relations to other variables

Sensitivity to change

A paired sample *t* test of ADL-O quality of ADL task performance measures based on the initial and second evaluation of the 26 participants evaluated more than one time suggested that the participants measures had changed significantly between evaluations ($t[25] = 3.49$, $p = .002$). Thus, we concluded that the observed quality of ADL task performance measures based on the ADL-O were sensitive to change in a sample of clients with very low levels of ADL ability following ABI.

Linear measures of perceived quality of ADL task performance based on ADL-Q and ADL-I

Evidence of validity based on internal structure

Psychometric properties of the rating scale

After collapsing the seven categories of the common ADL-Q and ADL-I rating scale into a 4-category rating scale structure, the evaluation of the psychometric properties of the rating scale revealed no category or threshold disordering in the rating scale and no rating scale misfit, indicating sound psychometric properties in both instrument formats. The categories of the final rating scale were 3 = competent (original categories 6 and 5), 2 = minimally inefficient (original categories 4 and 3), 1 = moderately inefficient (original categories 2 and 1) and 0 = deficient (original category 0) (see Table 5).

Internal scale validity

In the initial analysis of all 47 ADL items of ADL-Q and ADL-I, the PCA of the standardized residuals revealed that 49.7% and 47.4% of the total variance was explained by the Rasch dimension, respectively. Furthermore, 24.4% and 30.3% of the variance was explained by the items and 4.6% and 4.1% of the unexplained variance was defined by the first contrast. That is, in

both the ADL-Q and the ADL-I, the variance explained by items was more than four times higher than the variance defined by the first contrast.

As the intention was to generate comparable linearized ADL-Q and ADL-I ability measures, we first removed items that displayed misfit in both instruments (i.e. pedicuring, driving a car). In the following analyses, five more items displayed misfit in either the ADL-Q or the ADL-I. In total, we found that seven items (14.9%) did not display acceptable goodness of fit, suggesting a risk to unidimensionality (Table 6). Following removal of these seven misfitting items, the results of repeated PCAs of the ADL-Q and ADL-I data revealed that 52.7% and 51% of the variance was explained by the Rasch dimension, 27.7% and 33.6% of the variance was explained by items and 4.7% and 4.0% by the first contrast, indicating improved unidimensionality.

Evidence of validity based on test content

The targeting of the 40 item versions of the ADL-Q and ADL-I to the participants' ADL ability indicated that the participants' mean level of ADL ability in each version was much higher than was the mean of the item difficulty calibrations (Table 6). Thus, while the items and participants were well distributed along the scales, the item and person measures were not well targeted to each other (Figures 1 and 2). ADL-I, however, had better targeting of items to people with higher ADL ability compared to the ADL-Q. As a consequence, 13 women had maximum scores on ADL-Q (2 with FM, 4 with OA and 7 with RA) compared to two women with RA receiving maximum scores on the ADL-I. One of the women with RA received maximum scores on both ADL-Q and ADL-I. This indicates a ceiling effect for the ADL-Q and a potential ceiling effect for ADL-I.

Evidence of reliability

The person separation indicies of 3.40 and 3.09 for ADL-Q and ADL-I respectively indicated that the items separated the persons into at least four different ability levels (Table 6) (119).

Evidence of validity based on relations to other variables

Differences in perceived ADL ability among groups

Initial mixed models ANOVA revealed no diagnosis by age interaction effects for ADL-Q ($F[2,118] = 0.11$, $p = 0.8919$) or ADL-I ($F[2,118] = 1.28$, $p = 0.2806$) ability measures. In the subsequent analyses, therefore, we only analyzed for effects related to diagnosis.

Figure 1. **ADL-Q Map of the Location of Persons and Items**

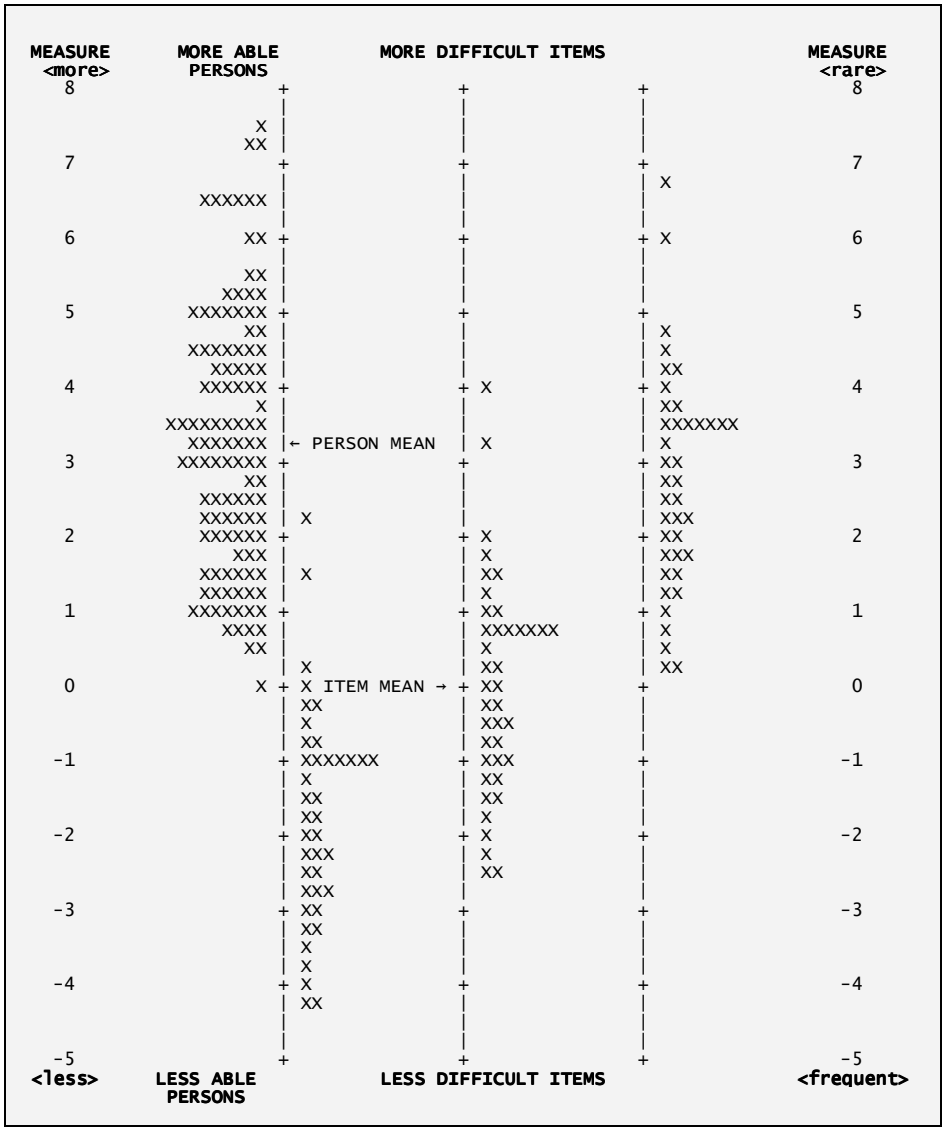
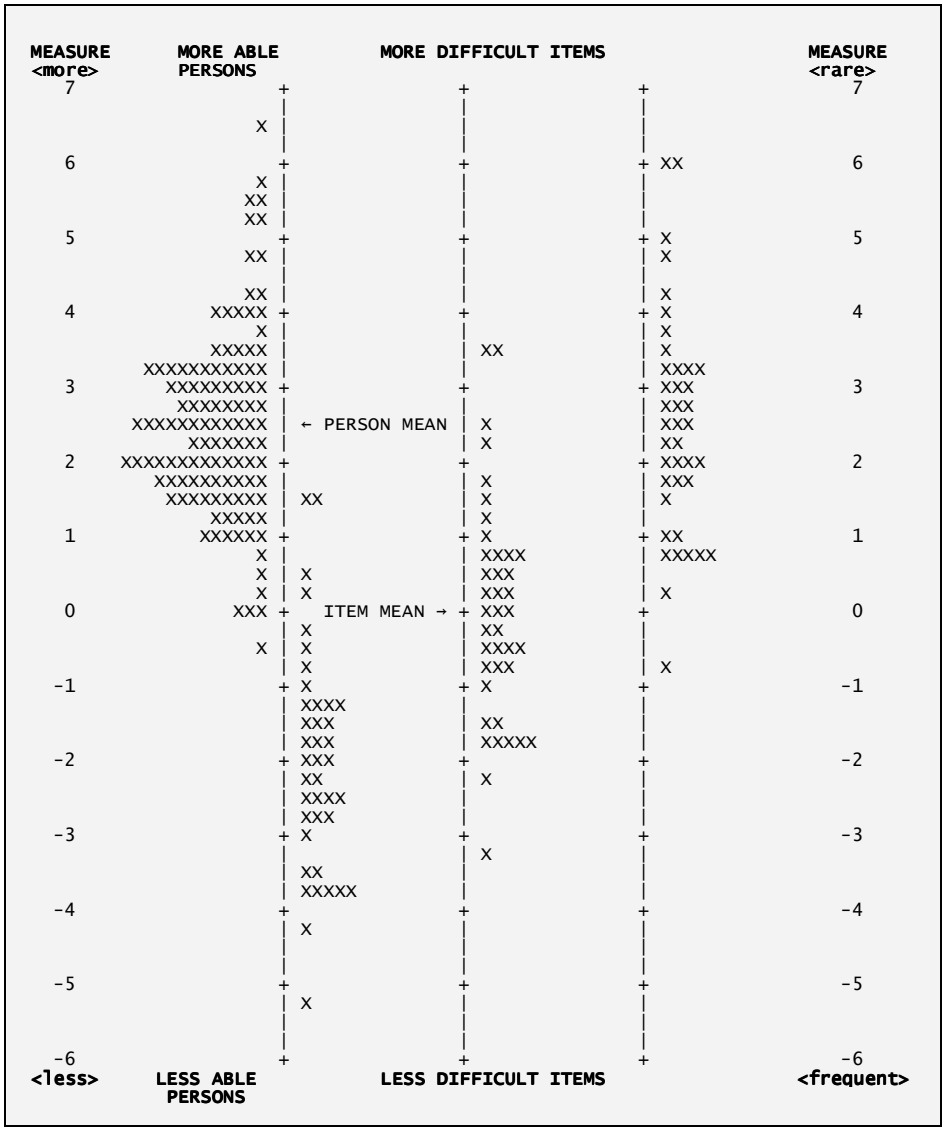


***Figure 2*. ADL-I Map of the Location of Persons and Items**



The one-way ANOVA and post hoc *t* tests revealed significant differences between perceived quality of ADL task performance measures based on the ADL-Q among the women with knee OA, RA and FM (Table 7). Similarly, the perceived quality of ADL task performance measures based on ADL-I were significantly lower in the women with FM compared to women with RA and knee OA.

***Table 7*. Mean Measures Based on the ADL-Q, the ADL-I and the AMPS Across Groups with Rheumatic Diseases**

	Fibro-myalgia (*n* = 54)	**Rheumatoid arthritis (*n* = 40)**	**Knee osteo-arthritis (*n* = 24)**	**Significant differences between groups**		
	M (SD)	*M (SD)*	*M (SD)*	*Groups*	*Dif*	*95% CI*
ADL-Q	2.67 (1.71)	3.46 (1.74)	4.33 (1.33)	FM/RA*	0.79	0.11–1.47
				FM/OA*	1.66	0.86–2.46
				RA/OA*	0.87	0.02–1.71
ADL-I	1.95 (1.00)	2.84 (1.30)	3.17 (1.09)	FM/RA*	0.88	0.41–1.35
				FM/OA*	1.21	0.66–1.76
				RA/OA	0.33	-0.24–0.91
AMPS ADL motor	1.03 (0.44)	1.40 (0.36)	1.38 (0.49)	FM/RA*	0.37	0.19–0.55
				FM/OA*	0.35	0.14–0.56
				RA/OA	0.02	-0.23–0.19
AMPS ADL process	1.09 (0.23)	1.20 (0.27)	1.00 (0.22)	FM/RA	0.11	0.00–0.20
				FM/OA	0.09	-0.20–0.03
				RA/OA*	0.20	0.07–0.31

ADL-I = ADL Taxonomy-Interview, ADL-O = ADL Taxonomy-Observation, ADL-Q = ADL Taxonomy-Questionnaire, AMPS = Assessment of Motor and Process Skills, Dif = difference, FM = Fibromyalgia, OA = Osteoarthritis, RA = Rheumatoid Arthritis

* Significant differences between groups

Relationships between ADL-Q and ADL-I ability measures

Initial ANOVAs revealed no significant interaction effects for format by diagnosis by age (F[2,224] = 0.15, p = 0.8641), format by diagnosis (F[2,226] = 0.46, p = 0.6319) or format by age (F[2,230] = 0.62, p =0.4329) to explain the difference between mean ADL-Q and ADL-I ability measures (Table 7).

Likewise, the results indicated that age did not have a significant impact on perceived ADL ability (F[1,231] = 0.03, p = 0.8670). In the subsequent analyses, therefore, we only analysed for a main format effect. We found a significant difference in perceived ADL ability between the ADL-Q and ADL-I formats (F[1,232] = 17.69, p < 0.0001) adjusted for significant differences in perceived ADL ability among diagnostic groups (F[2,232] = 19.24, p < 0.0001). More specifically, measures of perceived ADL ability based on ADL-Q were significantly higher than measures based on ADL-I across all diagnostic groups.

The Pearson product moment correlation analyses revealed significant correlation coefficients (p < 0.001) between ADL-Q and ADL-I across (r = 0.70) as well as within diagnostic groups (FM: r = 0.67, RA: r = 0.65, OA: r = 0.66). Only the overall correlation coefficient between ADL-Q and ADL-I reached the minimum of ≥ 0.7 to suggest evidence of convergent validity (136).

Evidence of validity of the ADL ability measures of the AMPS

Evidence of validity based on relations to other variables

Discrimination between healthy women and women with CWP/FM

Analyses based on z tests revealed that our sample of women with CWP/FM in Study III had a significantly lower mean ADL motor ability measure at pre-test1, compared to the mean ADL motor ability measure of the Nordic sample of healthy women (z = 3.01, p = 0.0027) (Table 8). In addition, our sample still had a significantly lower mean ADL motor ability at follow-up (z = 2.48, p = 0.0131), when compared to the normal sample, despite of a significant increase in ADL motor ability post intervention.

In contrast, their mean ADL process ability measure at pre-test1 was within normal limits compared to the mean ADL process ability measure of the healthy women (z = 1.85, p = 0.0643). Forty-six (86%) of the women with CWP/FM had ADL motor ability measures and 19 (38%) had ADL process ability measures more than 2.0 *SD* below the mean ADL motor and mean ADL process ability measures of the healthy Nordic sample.

***Table 8*. ADL Ability Measures in a Nordic Sample of Healthy Women and in a Sample of Women with CWP/FM**

	Nordic sample of healthy women	**Women with CWP/FM** *M* (*SD*) *			
	M (*SD*) (*n* =393)	Pre-test1 (*n* =50)	Pre-test2 (*n* =34)	Post-test (*n* =41)	Follow-up (*n* =41)
AMPS ADL motor	2.69 (0.56)	0.99 (0.59)	0.96 (0.47)	1.26 (0.43)	1.29 (0.43)
AMPS ADL process	1.92 (0.49)	1.02 (0.29)	1.14 (0.26)	1.14 (0.31)	1.16 (0.34)

AMPS = Assessment of Motor and Process Skills, CWP = Chronic Widespread Pain, FM = Fibromyalgia

* Unadjusted means and standard deviations

Stability in a sample with a fluctuating level of ability

Analyses based on *t* tests revealed no significant differences between pre-test1 and pre-test2 of the ADL motor (t[81.2] = 0.89, p = 0.3748) and ADL process ability measures (t[91.2] = -1.75, p = 0.0836) (Table 8). This indicates that the mean ADL ability measures of the AMPS remained stable over time when no intervention had been provided. Furthermore, Pearson product moment correlation analysis revealed a moderate correlation between pre-test1 and pre-test2 for ADL motor ability measures (r = 0.63, p < 0.0001) but a low correlation for ADL process ability measures (r = 0.17, p = 0.3387). Further inspection of stability of the individual pre-test1 and pre-test2 ADL motor ability measures of the AMPS (Figure 3) revealed that despite overall stability of measures at group level, individual measures can vary over time.

Absolute differences in ADL motor ability measures of the AMPS varied from 0.03 to 1.13 logits between pre-test1 and pre-test2. However, for 17 (50%) of the participants, the difference between measures was less than the 0.30 logit considered to reflect a clinically meaningful change(8). Inspection of the ADL process ability measures of the AMPS (Figure 4) revealed some, but less extreme variation between measures. Absolute differences in ADL process ability measures varied from 0.00 to 0.65 logits between pre-test1 and pre-test2 and, for 19 (56%) of the participants, the difference between measures was less than the 0.30 logits.

***Figure 3*. Individual ADL Motor Ability Measures at Pre-test1 and Pre-test2 Among Women with CWP/FM**

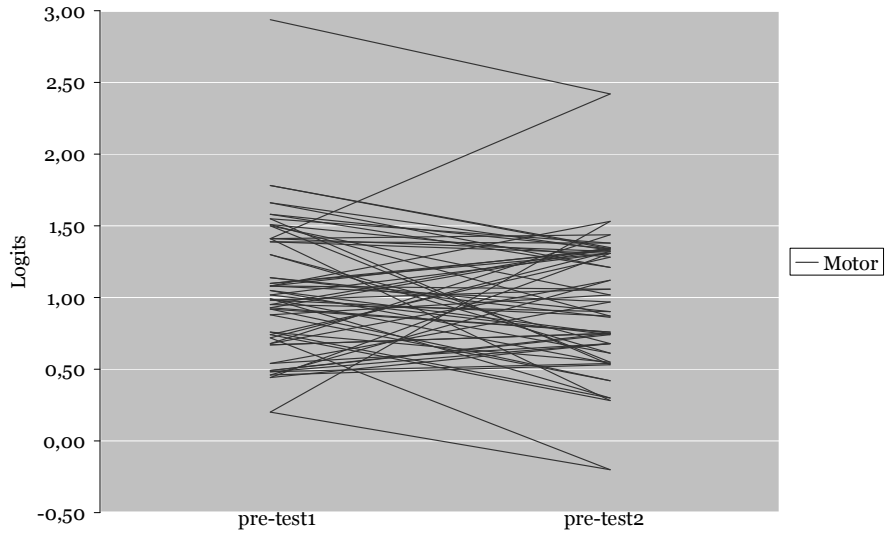
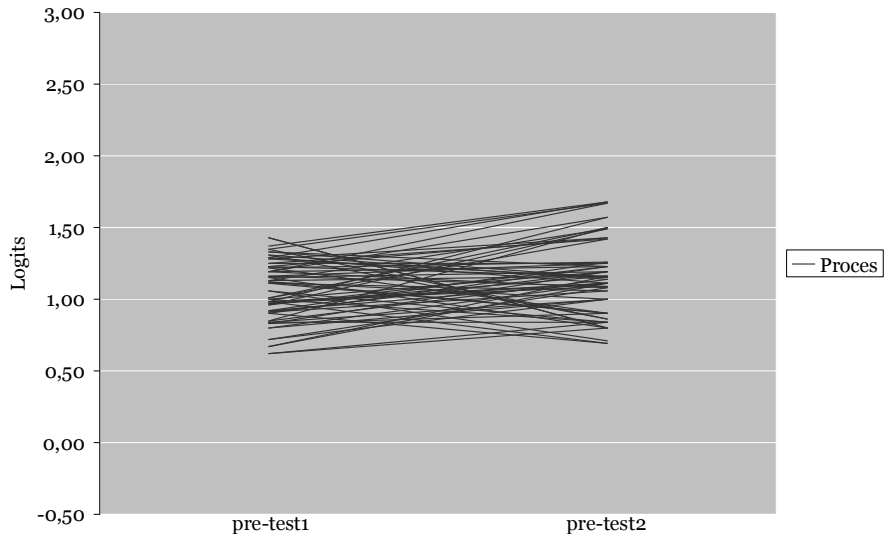


***Figure 4*. Individual ADL Process Ability Measures at Pre-test1 and Pre-test2 Among Women with CWP/FM**



Sensitivity to change in a sample with fluctuating ADL ability

While thirty-two (64%) of the participants were evaluated in their own homes, the initial ANOVAs revealed no significant time by setting interaction effect (F[3, 76.19] = 0.28, p = 0.8419) for the ADL ability measures of the AMPS. In the subsequent analyses, therefore, we only analysed for time effects. For the ADL motor ability measures of the AMPS, we found a significant change over time (F[3,75.2] = 9.59, p < 0.0001). The post hoc t tests procedure revealed significant differences between pre-test1 and the post-test (t[87.3] = -3.30, p = 0.0014) and between pre-test1 and follow-up (t[87.8] = -3.74, p = 0.0003); no difference was found between post-test1 and follow-up (t[72.7] =-0.42, p = 0.6782).

This suggests that the ADL motor ability measures of the AMPS increased over time when intervention was provided and then remained stable until at least four weeks after the end of intervention. Effect sizes for ADL motor ability between pre-test1 and post-test as well as pre-test1 and follow-up were medium. We found no significant change over time (F[3,85] = 2.53, p = 0.0627) in the ADL process ability measures of the AMPS.

Sensitivity to change in sample expected to have limited potential for change

In our sample of clients with ABI, the mean ADL motor ability measure increased from 0.68 to 1.20 logits (Table 9), indicating a decrease in the amount of clumsiness and effort, and possibly increased safety and decreased dependence demonstrated when performing ADL tasks.

***Table 9*. Pre- and Post-test ADL Ability Measures in a Sample with ABI**

		M	*SD*	Min	Max
AMPS ADL motor	Pre-test	0.68	1.03	-2.30	2.56
	Post-test	1.20	1.09	-0.76	3.78
AMPS ADL process	Pre-test	0.43	0.66	-1.37	1.67
	Post-test	0.86	0.64	-0.30	2.41

AMPS = Assessment of Motor and Process Skills

The mean ADL process ability measure increased from 0.43 to 0.86 indicating an increase in efficiency and likely increased safety and decreased dependence when performing ADLs. The results of the paired *t* tests revealed significant increases in both the ADL motor ($t[35] = 4.06$, $p < 0.001$) and ADL process ($t[35] = 5.45$, $p < 0.001$) ability measures of the AMPS. Effect sizes (*d*) were 0.5 for ADL motor ability and 0.6 for ADL process ability. We concluded, therefore, that the observed changes in ADL ability reflected at least medium effects(137). Based on a clinically meaningful change of at least .30 logit, we found that 29 participants (80.5%) improved in ADL motor and/or ADL process ability. Twenty-two (61.1%) participants improved in ADL motor ability and 20 (55.5%) in ADL process ability.

Relationship to measures of ADL ability based on self-report

Pearson product moment correlation analysis revealed low to moderate correlations overall and within diagnostic groups between measures of perceived ADL ability based on ADL-Q and ADL-I and measures of observed ADL motor ability based on the AMPS in Study II (Table 10). Only measures based on ADL-Q and ADL motor ability for women with knee OA reached the minimum of ≥ 0.70 determined to support evidence of convergent validity (i.e. support acceptable aggrement between measures of ADL ability based on self-report and observation). The fact that only one out of eight correlations between ADL-Q/ADL-I and ADL motor ability measures of the AMPS reached the predetermined level of ≥ 0.70 suggest that the measures of ADL based on ADL-Q/ADL-I and the ADL motor ability measures of the AMPS measure different but related aspects of ADL ability.

In contrast, the correlation coefficients between measures of ADL ability based on ADL-Q/ADLI and the ADL process ability measures of the AMPS were very low. Again, measures based on women with knee OA obtained the highest correlation between self-report (ADL-Q) and observation. Since seven out of eight correlations were ≤ 0.30, the results suggest that the measures of ADL-Q/ADL-I and the ADL process ability measures of the AMPS measure different aspects of ADL ability.

***Table 10*. Correlations between ADL-Q, ADL-I, AMPS ADL motor and AMPS ADL process ability measures**

		AMPS ADL motor	AMPS ADL process
		r	*r*
ADL-Q	All (*n* = 118)	.48***	.16
	FM (*n* = 54)	.36**	.09
	RA (*n* = 40)	.34*	.22
	OA (*n* = 24)	.72***	.52**
ADL-I	All (*n* = 118)	.52***	.23*
	FM (*n* = 54)	.49***	.27
	RA (*n* = 40)	.30	.27
	OA (*n* = 24)	.58**	.27
FIQ	CWP/FM (*n* = 50)	-.35*	-.02

ADL-I = ADL Taxonomy-Interview, ADL-O = ADL Taxonomy-Observation, ADL-Q = ADL Taxonomy-Questionnaire, AMPS = Assessment of Motor and Process Skills, FM = Fibromyalgia, OA = Osteoarthritis, RA = Rheumatoid Arthritis.

*$p < .05$. **$p < .01$. ***$p < .001$

In Study III, Spearman rho correlation analysis revealed a weak correlation between the scores of the self-reported physical function subscale of the FIQ and the observation-based ADL motor ability measures of the AMPS (Table 10). We found no correlation between the physical function subscale of the FIQ and the ADL process ability measures. The results suggest that perceived ADL ability as evaluated based on the physical function subscale of FIQ and observed ADL ability evaluated based on the ADL ability measures of the AMPS overall represent different aspects of ADL ability.

Summary

Based on the results of the four studies in this thesis, we were able to conclude that it was possible to obtain linear measures of observed quality of ADL task performance based on the revised version of the ADL Taxonomy (ADL-O). It was also possible to obtain linear measures of perceived quality of ADL task performance based on the revised versions of the ADL Taxonomy using the formats of questionnaires (ADL-Q) and interviews (ADL-I). Furthermore, we were able to establish preliminary evidence of validity of the ADL-O quality of performance measures in relation to sensitivity to change in clients with ABI. Similarly, preliminary evidence of validity of the ADL-Q and ADL-I quality of performance measures was established in relation to differences in mean ADL ability measures between ADL-Q and ADL-I and to convergent evidence of the relationship between ADL-Q and ADL-I. Moreover, it was possible to establish further evidence of validity of the ADL ability measures of the AMPS in relation to discrimination, stability and sensitivity to change among clients with fluctuating or very low ADL ability due to long-term or chronic disabilities following rheumatic disease or ABI and relationship to measures of ADL ability based on self-report.

Discussion

In the introduction, I proposed that occupational therapists practising in accordance with contemporary occupational therapy theory need ADL instruments that support a client-centred and occupation-focused approach when evaluating the nature and extent of the client's performance problems and establishing effectiveness of the interventions provided. Furthermore, I argued that for ADL instruments to be applicable for such purposes, they should be useful across target groups for evaluating perceived and/or observed ADL ability based on the quality of ADL task performance, and provide linear measures of ADL ability. Thus, this doctoral thesis was written with the overall purposes of developing new ADL instruments based on the ADL Taxonomy that could be used to generate linear measures of perceived and observed quality of ADL task performance, and validating the ADL measures of both the newly developed versions of the ADL Taxonomy and the AMPS for such occupational therapy clinical praxis, focusing on clients with long-term or chronic disability following ABI and rheumatic diseases.

Development of revised versions of the ADL Taxonomy

Two studies related to the third phase of instrument development (i.e. quantitative evaluation phase) of the original ADL Taxonomy had previously provided some evidence of validity related to content based on expert judgement(80) and evidence based on internal structure related to the hierarchies among ADL tasks within the 12 ADL domains(81). We initiated our contribution to the instrument development process by taking one step backwards to the *construction phase* and developed rating scales based on quality of ADL task performance. We then reentered the *quantitative evaluation phase* by addressing item analysis and reliability from a Rasch measurement perspective. Through this process, we were able to develop three revised versions of the ADL Taxonomy that could be used to obtain linear measures of quality of ADL task performance, the 47 item ADL-O based on observation, the 40 item ADL-Q based on self-report using a questionnaire format and the 40 item ADL-I based on an occupational therapy interview.

Evidence of validity of the quality of ADL task performance measures

In general, our results supported the previous validity evidence based on test content and internal structure of the ADL Taxonomy. However, as part of the Rasch analyses, we identified some ADL Taxonomy tasks that displayed misfit in the ADL-O and/or in the combined analysis of ADL-Q and ADL-I. Only two of these tasks were found to misfit in all three versions of the ADL Taxonomy – *Pedicuring* and *Driving car*. In the case of the ADL-O, *Driving a car* was too hard a task for most of the clients with ABI and rarely rated. A few clients in the sample, however, obtained a maximum score on driving despite the fact that they all were diagnosed with moderate to severe ABI rarely associated with ability to drive. Therefore, high scores on driving might result from administration or scoring error causing item misfit due to unpredictably high scores on a difficult item. In contrast, some of the more able women with rheumatic diseases reported need for assistance for driving on the ADL-Q and ADL-I, resulting in unexpectedly low scores causing the item of driving to misfit. One possible reason could be that the women, when indicating need for assistance, instead of reporting their ability to drive a car, reported that somebody else typically drives, when they are going somewhere. That is, they reported *what they do* or do not do rather than their *ability to do*. The fact that someone else drives can be based on routines and roles in the family (e.g. that the spouse typically drives) rather than the woman's need for assistance to drive. Similarly, low ratings for more able clients on the rather easy grooming item *Pedicuring* could cause the item to misfit. Again, a low rating might be given because the person does get professional help with pedicure even though he or she is able to perform the task independently.

Five out of 10 misfitting items in the ADL-O were related to the communication domain. While items concerning communication are frequently used in ADL instruments, they have often been found to misfit or to form a secondary dimension when such scales are subjected to Rasch analyses (138;139). However, since we found no misfit related to the communication items on the ADL-Q/ADL-I, it seems that the misfit on the ADL-O also could be explained by the fact that some, but not all, of the clients evaluated on the ADL-O had ABI associated with language dysfunctions (e.g. aphasia). Thus, it is likely that the communciation items that were found to be relatively easy for persons with rheumatic diseases and those with ABI but no language dysfunctions were unexpectedly hard for those clients with ABI associated with language disorders. That is, the misfit

could be due to diagnostic DIF which would have to be further explored in future studies. Yet another reason for the communcation items to display misfit on the ADL-O could be that three of these items displayed DIF based on gender (i.e. they were easier for women than for men). Since the ADL-Q and ADL-I were only used with women, DIF related to gender needs to be further investigated by also including ADL-Q and ADL-I data for male clients with rheumatic diseases in future studies.

In contrast, tasks requirering fine motor movement and within hand manipulation such as *Shaving/make-up* (ADL-Q) and *Getting food and liquid and cutting up/preparing food* (ADL-I) displayed misfit in the sample of women with rheumatic diseases. One example that many of the women reported during the ADL-I interviews was that they often received assistance to cut up meat on their plate to avoid pain, even though they otherwise were independent performing PADL tasks. Such a pattern will result in misfit as the women are performing unexpectedly low on a fairly easy ADL task, while they are independent in more difficult tasks such as dressing. Again, the reason for this misfit can be administration and scoring error if the women reported what they typically do (i.e. do get assistance) rather than their ability to do (i.e. need assistance). Still, these results are based on a combined analysis of the ADL-Q and ADL-I data, which caused removal of items from one version because they displayed misfit on the other version. The removal of "good" items in turn resulted in further items to misfit. We, therefore, do not know if these items would display misfit in separate analyses of the ADL-Q and ADL-I.

When an item displays misfit, it suggests that this item does not belong to the same construct as the other items(3). It was, therefore, a surprise to find that the items displaying misfit varied across the three versions of the revised ADL Taxonomy. Still, as discussed above, there can be a number of reasons for items displaying misfit besides lack of unidimensionality, including sample characteristics resulting in DIF and administration and/or rater scoring error. The former suggests the need for further research related to DIF across age, gender and diagnostic groups. The latter suggests that instructions for raters might have been unclear. For example, it may be needed to improve the verbal and written instructions for the ADL-Q and ADL-I to stress that the clients should report their ability to do (can do), rather than what they actually do and do not do as part of daily rutines. As the 47 tasks of the ADL Taxonomy are regarded to be of clinical relevance(80), such strategies should be implemented to try to resolve

problems of misfit before making decisions about final removal of tasks from the ADL-O, ADL-Q and ADL-I.

Another approach is to investigate if multidimensionality (i.e. misfitting items and items displaying DIF) threaten the measurement system by implementing DTF as suggested by Smith(114). We performed DTF analyses of the ADL-O which revealed that the retention of misfitting items and items displaying gender-related DIF did not disrupt the measurement system. Similarly, DTF analyses should be performed if future separate Rasch analyses of the ADL-Q and ADL-I data reveal misfitting items and/or DIF. For now, we tentatively recommend retaining all 47 items in the three revised versions of the ADL Taxonomy until future Rasch analyses based on larger samples verifies that they indeed do represent a different construct and/or disrupt measurement.

In all three versions we found sound psychometric properties of the rating scale structure reflecting quality of ADL task performance based on effort, efficiency, safety and independence. Traditional ADL instruments such as the Barthel Index(70) scored based on independence have been found to be insensitive to small changes and to display severe floor effects in people with severe stroke(140). One idea behind incorporating other aspects besides independence in evaluation of ADL task performance was, therefore, to create an instrument that was sensitive enough to differentiate between clients with very low levels of ADL abilities needing assistance in most everyday life tasks. Based on the 47 item ADL-O with its 3-category rating scale, we developed an instrument that could separate the persons with ABI into at least seven different ability levels with no ceiling effect and essentially no floor effect. In addition, we were able to document changes over time in a small sub-sample. These results are indications that the measures of observed quality of ADL task performance based on the ADL-O are sensitive measures of ADL ability among clients with long-term or chronic disabilities following ABI.

Another idea behind rating ADL task performance based on quality of performance was to create an instrument that could identify diminished ADL ability among clients with higher levels of ADL ability who seldom need assistance in everyday life tasks. Based on the 40 item ADL-Q and/or ADL-I with 4-category rating scales, we were able to document significant differences in mean perceived quality of ADL task performance measures among women diagnosed with FM, RA and knee OA and separate the sample

of women with rheumatic diseases into at least four different ability levels with no floor effect, but a ceiling effect for the ADL-Q and a potential ceiling effect for ADL-I. Again, these are indications that the measures of perceived quality of ADL task performance based ADL-Q and the ADL-I are sensitive measures among women with long-term or chronic disabilities following rheumatic diseases.

When we examined the relationship between measures of perceived quality of ADL task performance based on the ADL-Q and ADL-I among women with rheumatic diseases, we found that the two modes of self-report resulted in different but related information about perceived ADL ability. But contrary to previous studies of people with RA(67) and OA(68), we found a pattern of significantly higher perceived ADL ability based on ADL-Q compared to ADL-I across diagnostic groups. One reason for this difference could be that neither of the previous studies used the same instrument when collecting questionnaire and interview data. Instead, to be able to compare information gathered through questionnaires and interviews, information was extracted retrospectively from records based on semi-structured interviews, to give answers to the same questions as answered by the client in the questionnaire. This method increased the risk that the questions asked during questionnaires and interviews were not identical. Like us, Daving et al (66) used the items of the ADL Taxonomy with a dichotomized scoring based on dependent/independent performance to compare questionnaire and telephone interview data for each of the 12 ADL domains in a sample of persons with stroke. They found, similar to our results, moderate to good agreement between methods (i.e. kappa > 0.50 for 10 of the 12 ADL domains) and that the participants overall reported more dependence in the interviews. In our study, the pattern of higher perceived ADL ability reported in the ADL-Q was constant across all three diagnostic groups suggesting that the clients responded in the same way to the two modes of self-report.

Another reason for the difference between measures of perceived ADL ability could be related to the context in which the questionnaire and interview data were collected. In general, the two modes of self-report vary both in terms of the anonymity afforded the client and the opportunity to make a connection with the interviewer, both of which might affect clients' willingness to disclose information. Furthermore, the clients only had written instructions to follow when filling in the ADL-Q, whereas they were verbally instructed before the ADL interview and had the opportunity to ask

the occupational therapist for further clarification if needed prior to or during the interview. Getting additional instruction and potential correction might have impacted on how the clients understood the questions and thereby influenced their answers, resulting in reporting less quality of ADL task performance on the ADL-I. While we found no real difference in how the two modes of self-report overall related to actual ability to perform ADL tasks, it is reasonable to think that the ADL-I can be used to obtain more information about the types of problems perceived by the client related to quality of task performance than the ADL-O. Therefore, occupational therapist would probably find the ADL-I more suitable when obtaining information for the purpose of intervention planning.

While we have implemented the ADL-O with a sample of clients with ABI receiving inpatient interdisciplinary rehabilitation and the ADL-Q and ADL-I with women with rheumatic diseases receiving outpatient occupational therapy and interdisciplinary rehabilitation, the intention is to develop the instruments into generic ADL instruments useful across age, gender, diagnostic groups and clinical settings. Therefore, although these first studies provide evidence to support that linear quality of ADL task performance measures can be obtained for clients with ABI on ADL-O and for women with rheumatic diseases on the ADL-Q and ADL-I, more heterogenious data on all three versions will need to be collected to confirm our results before generic instruments can be validated for clinical purposes.

For now, it requires the use of Rasch measurement methods to convert the ordinal scores of the ADL-O, ADL-Q and ADL-I into linear measures. Thus, by replicating our methods, it is possible to obtain linear measures of perceived and observed quality of ADL task performance based on the three new versions of the ADL Taxonomy. However, to make future generic versions useful in clinical practice, there is a need to develop either conversion tables or computer software that can be used to obtain linear measures based on the ordinal scores of the ADL-O, ADL-Q and ADL-I. Conversion tables can be generated by Winsteps®(117), where ordinal total scores are converted to linear measures. The advantages of conversion tables are that they can easily and at low costs become available to clinicians, but the drawback is that conversion tables only can be used for ordinal total scores on a complete test. Thus, the client needs to be rated on all items to obtain a linear measure. This is in conflict with a client-centred and occupation-focused approach to evaluation, where occupational therapists only focus on ADL tasks of importance for their client. Therefore, computer

software that allows for missing data for some of the ADL tasks might be needed in the future.

Further validation of the ADL ability measures of the AMPS

Three studies in this thesis are concerned with *further validation* of the ADL motor and ADL process ability measures of the AMPS. They were enacted to provide further evidence of validity of the ADL ability measures of the AMPS among clients with very low and/or fluctuating ADL ability following ABI or rheumatic diseases.

Evidence of validity related to discrimination

Rheumatic diseases are known to have a fluctuating course with day-to–day or week-to-week variations in symptoms such as pain(141). These variations, in turn, result in fluctuations in clients' abilities to perform everyday life tasks(142). Therefore, when evaluating the extent of ADL task performance problems among such clients, we need instruments that can provide measures that discriminate between healthy people and clients with rheumatic diseases. Although numorous studies have supported the validity of the ADL ability measures of the AMPS across diagnostic groups (18-23;85-94), only one previous study has been concerned with clients diagnosed with a rheumatic disease (i.e**.** systemic lupus erythematosus, SLE)(24). Like us, Poole at al.(24) compared the ADL motor and ADL process ability of 15 women with SLE (ADL motor: M = 2.37, ADL process: M = 1.47) to the ability of 15 healthy women (ADL motor: M = 3.74, ADL process: M = 2.64). They found significant differences between groups on both the ADL motor and ADL process ability measures.

While our results provided evidence that the ADL motor ability measures of the AMPS could be used to discriminate between healthy women and women with CWP/FM, similar evidence could not be obtained for the ADL process ability measures. This was unexpected, as the mean ADL process ability measure of our sample of women with CWP/FM was 0.45 logits lower than the mean ADL process ability of the women with SLE. The most obvious explanation for the difference between Poole et al.'s and our results is related to how mean ADL ability measures for healthy people were obtained. Unlike Poole et al.(24), we obtained the mean ADL ability

measures of a Nordic sample of healthy women from the AMPS International database. These mean ADL ability measures were somewhat lower than those reported by Poole et al. but similar to the mean ADL ability measures for healthy adults (n = 2063) reported in the AMPS manual (ADL motor: M = 2.6 – 3.2, ADL process: M = 1.8 – 2.2)(8). The mean ADL ability measures from the AMPS International database represent a larger sample size, which ensures a more heterogenious sample of healthy women. Furthermore, the data is based on several raters limiting the risk of systematic error due to rater error (e.g. rater leniency). This might explain why we got results different from that of Poole et al(24).

Although our sample did not have significantly lower ADL process ability compared to healthy women, their mean ADL process ability measure was very close to the ADL process scale cutoff (1.0 logits), indicating problems performing familiar and life relevant ADL tasks in a timely and well-organized (efficient), safe and independent manner. Furthermore, more than one third of the women with CWP/FM had ADL process ability measures lower than those of healthy women of same age, suggesting that ADL process ability measures can be used to identify those women with CWP/FM with lower levels of ADL ability.

Evidence of validity related to stability

To establish effectiveness of interventions, we need measures that are stable and yet, at the same time, sensitive to change. Again, this is a challenge with clients with fluctuations in their abilities to perform everyday life tasks. Previous studies of test-retest(8;110) and alternate forms reliability(8;111) have reported high to very high correlation coefficients supporting the stability of the ADL measures of the AMPS over time and between different pairs of ADL tasks among well adult persons as well as persons with various diagnoses. Still, Kirkley and Fisher(111) identified that actual fluctuations in ADL ability associated with certain diagnoses, including arthritis, can result in day to day variations in ADL ability measures (i.e. differences > 0.50 logits). Thus, they conclude that the ADL ability measures of the AMPS are sensitive to actual fluctuations in ADL ability.

Our results support their conclusions. Specifically, we also concluded that the AMPS can be used to provide stable mean ADL motor and ADL process ability measures at the group level for clients with fluctuating abilities following CWP/FM, despite fluctuations at the individual level. Although

correlations were low to moderate, we found that 50% of the paired ADL motor and 56% of the paired process ability measures remained stable within ± 0.30 logits. In comparison, when applying the criteria of ± 0.50 logits used by Kirkley and Fisher, we found that 76% of the paired ADL motor and 91% of the paired ADL process ability measures remained stable across AMPS observations. These results are similar to those reported by Kirkley and Fisher (80% and 81%, respectively). Still, we had a few client with very large differences. However, the fact that those few clients with the largest differences were those with the most extreme scores at time of pre-test1 or pre-test2 may indicate that the evaluations of these participants, especially those with high ADL motor ability measures, were associated with AMPS administration or scoring error.

Since the ADL ability measures of the AMPS are sensitive to single client's fluctuating ADL ability, they may not be suitable for the purpose of establishing evidence of effectiveness of intervention provided to single clients with very fluctuating levels of ability. That is, in everyday practice, the occupational therapist commonly evaluates the person before and after the intervention. Yet, persons with daily fluctuations can show significant differences in ADL ability from day to day. Thus, if the occupational therapist finds that the person's ADL ability measures increase or decrease upon retest, the occupational therapist has no way of knowing if it was the intervention or the fluctuation that caused the changes. Yet, in both cases, the differences are "real" (i.e. actual differences in ability).

Evidence of validity related to sensitivity to change

The ADL ability measures of the AMPS have already been shown to be sensitive when used to evaluate outcomes in intervention studies(19;21;86-89;91-93). Two studies in this thesis further increased the amount of evidence that support that the ADL ability measures of the AMPS can be used to document effectiveness of interventions. These studies addressed issues related to measuring outcomes among clients with fluctuating ADL ability and clients with very low ADL ability and limited potential for change.

Sensitivity to changes among clients with fluctuating ADL ability

When we examined if the ADL ability measures of the AMPS were sensitive to change among women with CWP/FM, we were able to document a significant improvement in mean ADL motor ability. A similar change, however, was not seen in mean ADL process ability. There might be several reasons for these results. First of all, the types of intervention could be

argued to mainly affect ADL motor ability. Teaching the participants to use energy saving techniques and helping aids was targeted at reducing the amount of effort and fatigue when performing daily life tasks. Secondly, some energy saving techniques, like pausing to take a break, results in lower scores in some ADL process skills due to "less efficient" use of time. In other words, adaptations to reduce effort or fatigue might be at the cost of reduced efficiency which is reflected in the ADL process ability measures of the AMPS. Similar results were reported by Fisher et al.(21) who found improved ADL motor ability while ADL process ability did not change in a sample of frail older adults primarily receiving interventions that were adaptive and compensatory in nature. In both studies, the intervention periods were short, perhaps too short to affect ADL process ability. That is, it can be argued that methods designed to compensate for inefficient use of time and space while performing ADL tasks take more than four weeks post intervention to implement in an automatic manner in everyday life routines. Such a possibility is supported by the fact that among the studies that did document improved ADL process ability, outcomes were assessed between eight weeks and 12 months after the intervention(19;87;89).

The change in mean ADL motor ability was found to be significant and effect size was medium; the mean change was 0.27 logits between pre-test1 and post-test and 0.30 logits between pre-test1 and follow-up. Even small changes in mean ADL motor and ADL process ability have been reported to be significant in studies concerning clients with ABI(86;88), suggesting that small but significant changes can be seen across diagnostic groups. We, therefore, conclude that the ADL motor ability measures of the AMPS were sensitive to small changes in ADL motor ability after short term interventions focused on adaptations in everyday life among clients with fluctuating ADL ability.

Sensitivity to changes among clients with very low ADL ability

When we examined if the ADL ability measures of the AMPS were sensitive to change among clients with very low ADL ability following moderate to severe ABI, we were again able to document significant improvements in mean ADL motor and ADL process ability. This is not the first study to provide evidence for the sensitivity of the ADL ability measures of the AMPS among clients with ABI. In a study of younger stroke clients, Björkdahl et al.(86) found significant changes in ADL motor and ADL process ability measures after three weeks of occupational therapy and physical therapy home rehabilitation. Furthermore, Lindén et al.(87), in a

study of younger clients with traumatic brain injuries, reported significant improvements in ADL motor and ADL process ability after three months of interdisciplinary rehabilitation. In both studies, however, initial ADL motor ability was considerably higher (*M* = ~1.45 logits) than the initial ADL motor ability of our clients (*M* = 0.68 logits). Actually, as an indication of the very low level of ability in our sample, their overall ADL motor ability never increased to the level of ADL motor ability of the samples included in the other studies. Similarly, the initial ADL process ability of the younger stroke clients (*M* = 1.00 logits) was higher than the mean ADL process ability of our sample at any point in time. In contrast, the sample with traumatic brain injuries had initial ADL process ability measures similar to our sample (*M* = 0.26 logits), but increased within three months to a much higher level (M = 1.02 logits). The fast increase in the mean ADL process ability can be explained by the fact that the clients with traumatic brain injuries were evaluated during the early phases of rehabilitation, a short time after their brain injury. In contrast, our sample had been injuried on average 7.5 months prior to their first AMPS observation and, therefore, were in a phase where less spontanious recovery occurs. Based on these results we, therefore, concluded that the ADL motor and ADL process ability measures of the AMPS are sensitive to change among clients with very low ADL ability and/or limited potential for change following ABI.

Evidence of validity based on relation to measures of perceived ADL ability

Previous studies have shown that data based on self-report and observation provide different information about ADL ability, suggesting that perceived and observed ability are two different but related constructs (14;24). The results in this thesis added to that evidence by documenting that the observation-based measures of the AMPS were only low to moderately related to measures of perceived ADL ability based on the ADL-Q and ADL-I and to scores based on the physical function subscale of the FIQ.

Possible reasons why instruments can be found to provide different information can be related to the content and internal structure of the instruments(143;144). One issue related to internal structure is differences in scoring criteria among the ADL instruments. While the physical function subscale of the FIQ is rated based on *how often* they were able to perform the task, the AMPS and the ADL-Q and ADL-I were rated based on *quality* of ADL task performance. Still, this did not influence our results. That is, we saw similar low correlations to observed ADL ability across measures of

perceived ADL ability. Another issue is related to the items included in the instrument. While AMPS, ADL-Q, ADL-I and the physical function subscale of the FIQ all are concerned with PADL and/or IADL, there is one major difference between the AMPS and the other ADL instrument. During an AMPS observation, the client only performs two ADL tasks of relevance and appropriate challenge to the client and the client's ability to perform each of the ADL task is rated based on 36 specific ADL motor and ADL process skill items. In contrast, when the clients rate their perceived ADL ability on either the ADL-Q, ADL-I and the FIQ, the clients only assign one score for each task. Instead they rate their ability on several ADL tasks. While these differences to some extent can explain why the measures of observed ADL ability were only low to moderately correlated to measures of perceived ADL ability, it does not explain why the ADL process ability measures were less correlated to the measures of perceived ADL ability than the ADL motor ability measures of the AMPS.

The correlations between measures of observed ADL process ability and perceived ADL ability indicated divergent evidence of validity, suggesting that the ADL process ability measures do measure something else. The ADL process ability measure is an indication of how timely and well-organized the person was observed to be during the task performance. It was our intention to also incorporate this aspect in the ADL-Q and ADL-I measures of perceived ADL ability by including the category of "I perform the task independently but it takes me *extra time*" on the rating scale. Unfortunately, this category was, in the Rasch analysis, collapsed with the category, "I perform the task independently but I use *extra effort/get tired* faster", intended to reflect parts of the ADL motor ability measures of the AMPS. Future studies will be needed to determine if correlations between the ADL process ability measures and the measures of the ADL-Q and ADL-I can be improved by keeping the original categories of the rating scale separate.

Finally, the most straight foreward explanation to the modest relationship between measures of perceived and observed ADL ability is that they represent two distinct but complementary constructs. Thus, perceived ADL ability – the person's experience of task performance – is influenced by other factors besides the actual observable ability to perform. Hewlett(145) discussed this in a review concerned with convergence and divergens among outcomes measures based on self-report and professional judgement in rheumatology. She suggests that clients might focus on other aspects than professionals related to needs, expectations, experiences and priorities when

reporting their ability. This is in agreement with situations during our data collection, where the women reported differently during the ADL-I interview compared to what was later observed during the AMPS observation. For example, one woman said during the interview that she liked to iron and that she considered it an easy task. During the following observation, however, she displayed signs of decreased quality of task performance such as taking pauses and sitting to rest. We, therefore, concluded that evaluations of ADL ability based on methods of self-report do not provide the same information as evaluations of ADL ability based on actual observable ADL task performance among women with rheumatic diseases.

Implications for clinical practice

Since the research questions I sought to answers were founded in occupational therapy clinical praxis, I will now discuss the clinical implications of the results of this thesis focusing on five main issues: (a) measuring perceived or observed quality of ADL task performance, (b) measuring the nature and extent of perceived quality of ADL task performance, (c) measuring the nature and extent of observed quality of ADL task performance, (d) establishing effectiveness of the provided interventions and (e) using ADL evaluations in an interdisciplinary rehabilitation context.

Measuring perceived or observed quality of ADL task performance?

In accordance with previous findings, the examination of relationships between measures ADL ability based on self-report and observation did provide evidence to support that perceived and observed ability to perform ADL tasks are distinct but complimentary constructs(13;24;63-65). That is, different types of information about ADL ability will be obtained depending on the evaluation method used. Our findings, therefore, support implementation of both self-reported and observation-based evaluations as part of the occupational therapy process. This is in agreement with the guidelines in the OTIPM concerning the types of occupational therapy evaluations needed before planning and implementing interventions(5). In clinical praxis, however, occupational therapists often choose only one of these evaluation methods depending on the client groups with which they work. That is, there is a discrepancy between occupational therapy

intervention process guidelines for evaluation based on occupational therapy theory and existing evidence on one side and clinical praxis on the other side. One reason for this can be that occupational therapists are more influenced by diagnosis-specific approaches to evaluation of ADL task performance. For example, observation-based evaluations of ADL task performance are commonly used when evaluating clients with moderate to severe disability following ABI in interdisciplinary rehabilitation settings(60-62), as it is often argued that these clients are not expected to report reliably due to perceptual or cognitive deficits and/or language disorders. Working in a context based on this view might cause the occupational therapist not to use interviews or questionnaires with his or her clients. And vice versa, if working in a rheumatology unit, where self-report questionnaires or interviews have been the preferred method to obtain information about ADL ability, the occupational therapist might not perform observation-based evaluations.

Approaches based on the client's diagnosis, however, can be critizised, as we cannot expect to obtain similar information about ADL task performance based on methods of self-report and observation. The reason for evaluating the client's ADL ability using self-report is not to obtain information similar to what can be observed. Instead, the purpose of using methods of self-report is to obtain information about how the client perceives his or her ADL performance problems. More specifically, in order to implement client-centred occupational therapy, we need to know how the client perceives the nature and extent of his or her ADL task performance problems even if it is very different from what can be observed. Similarly, the reason for evaluating clients using observation is to obtain information about actual ADL task performance, which we now know is different from perceived ADL task performance even in clients with no cognitive deficits and/or language disorders. Therefore, as the ADL evaluations based on methods of self-report and observation serve different and complementary purposes, both methods of evaluation should be implemented to fully understand the nature and extent of the client's perceived and observed ADL task performance problems.

Measuring the nature and extent of perceived quality of ADL task performance problems

As stated previously, ADL instruments based on methods of self-report exist and can be used to identify strengths and weaknesses in occupational performance by providing information about the nature of ADL task performance problems. However, implementation of the ADL-Q and ADL-I

in clinical practice will make it possible to obtain more detailed information about the perceived quality of ADL task performance. Such information can guide the occupational therapist in planning interventions targeted at the nature of the performance problems. The ADL-Q and ADL-I can be used in combination or separatly. For example, clients may be asked to think about their ADL ability and rate it on the ADL-Q prior to meeting the occupational therapist. This can prepare the clients for the types of problems they can expect the occupational therapist to address. Furthermore, the information from the ADL-Q might help the occupational therapist to decide if she needs to also go through a thorough interview, the ADL-I, or just rely on the ADL domains indicated by the client as problematic but important on the ADL-Q. Hopefully, measures of perceived quality of ADL task performance based on the ADL-Q and/or ADL-I will be available in the future, which will make it possible for the occupational therapists to also measure the extent of perceived ADL ability. Availability of such measures will make it possible to compare perceived ability among clients and over time and thereby further increase the utility of the instruments in clinical practice.

Measuring the extent of observed ADL task performance problems

Similar to the ADL-Q and ADL-I, the ADL-O cannot yet be used to obtain measures of observed ADL ability, but will need to be further developed and validated before it can be used to measure the extent of ADL task performance problems and document change. Still, it can be used to rate problems related to quality of ADL task performance based on observation and thereby guide intervention planning. As it is time-consuming to observe all the ADL tasks of relevance to the client, the ADL-O is probably better used in the context of inpatient interdisciplinary rehabilitation, where the interdisciplinary team can observe the client over time. That is, the ADL-O can be used in a way similar to how the FIM™ and the Barthel Index are used in rehabilitation settings. But in contrast to these instruments, which provide information about level of assistance, the ADL-O can be used to obtain information about quality of ADL task performance. Still, evaluations based on observation by interdisciplinary team members requires supervision by occupational therapists skilled in observation of quality of ADL task performance, to ensure uniformity of collected data.

Still, a more time-efficient way to obtain information about the quality of ADL task performance based on observation is to implement the AMPS, as it only requires observation of two tasks and only involves the occupational

therapist in the evaluation process. While observation-based ADL task performance evaluations have been rarely used among clients with rheumatic diseases and clients with very low ADL ability following ABI, the AMPS has been shown to be a useful instrument to measure the extent of observed ADL task performance problems and establish effectiveness of interventions among such client groups. We do, therefore, recommend that the AMPS be implemented, when occupational therapy is provided to such clients in clinical practice.

Establish the effectiveness of provided interventions

Measures of ADL ability or quality of ADL task performance can be used to demonstrate whether or not our interventions are effective. When ADL instruments are implemented so that data are collected routinely, it will allow us to form a clearer idea over time about what aspects of our practice are effective and what aspects need to be changed so that we can base future interventions with similar clients on the results of our findings. Evidence of the effectiveness of our interventions can be established if we implement evaluations systematically on at least two occasions, pre- and post-interventions. For now, we can recommend the use of the AMPS for such purposes. If measures of observed quality of ADL task performance based on the ADL-O can be obtained in the future, this might be an alternative way to measure observed change among clients with chronic or longterm disability in rehabilitation settings.

If measures of perceived quality of ADL task performance based on the ADL-Q and ADL-I will be available in the future, they could supplement the AMPS, by measuring the client's perceived change in ADL task performance. The possibility of establishing effectiveness of interventions based on both perceived and observed ADL ability will provide further insight into the effective mechanisms of our interventions.

ADL evaluations in an interdisciplinary rehabilitation context

Occupational therapy for clients with long-term disability is rarely provided in isolation. Rather, occupational therapy services are offered in rehabilitation settings, where occupational therapists serve as one of several disciplines working in interdisciplinary teams. Rehabilitation for clients with long-term or chronic disability is directed towards improving the client's possibilities for participation in everyday life tasks of importance to the client(62). Each professional contributes to obtain this outcome based on his or her professional background. Occupational therapists can play a unique

role in the team related to evaluation of the extent of ADL task performance problems for the purpose of goal-setting and intervention planning and re-evaluation to document change and thereby establish effectiveness of the interdisciplinary interventions provided.

Methodological considerations

One general methodological limitation related to the samples included in the present studies includes the fact that all clients were living in Denmark. Therefore, generalization of the results to other cultures, especially cultures known to perform some of the ADL tasks differently, should be done cautiously until future studies have examined the potential for cross-cultural applications further.

Issues related to the development and validation of revised versions of the ADL Taxonomy

In general, the idea was to develop three new version of the ADL Taxonomy to be useful among clients with chronic or long-term disability across diagnostic groups. Therefore, the results based on Studies I–II can only be considered preliminary.

Considerations related to ADL-O

First of all, the development and validation of the ADL-O was solely based on clients with moderate to severe ABI. Furthermore, the ADL-O data were collected retrospectively based on medical records. While retrospective data have limitations when it comes to ensuring quality of data (including ensurance that all relevant ADL tasks are rated), they probably do resemble better the quality of future data gathered in clinical practice. Due to the severity of the clients' brain injuries, and the fact that the clients only were rated on items of relevance for their present everyday lives, we had a limited number of observations on harder items. These issues suggest the need to replicate our findings on a larger and more diverse sample representing other diagnoses and more variation in ability to further develop the ADL-O.

While the use of repeated ADL-O evaluations on a small number of clients with ABI might have affected the results, we felt that the risks were minimal, especially given the explorative nature of this pilot study. That is, in accordance with similar studies (13), the use of repeated evaluations of the

clients was effective to increase the sample size and ensure more stable measures of item difficulty (146;147). Moreover, several factors led to the conclusion that inclusion of these data was a minimal threat to the internal validity of this study. First of all, the item calibration measures remained stable between those based on all 103 evaluations and the 70 initial evaluations for all participants. Secondly, the 26 participants, who were repeatedly evaluated, had significant changes in the quality of their ADL ability over time. Furthermore, the number of clients tested more than twice was very small, and the analysis revealed that those clients were well distributed along the logit distribution and into several levels of ADL ability.

In addition, during the time since the Rasch analysis of the ADL-O was performed, updates in Rasch computer software and changes in criteria for interpretation of statistical results has resulted in the need to reevaluate previous results. This was done in part, when presenting the ADL-O results in the thesis, but when comparing results from future studies to the results first reported in Paper IV, there is a need to take these changes into consideration.

Considerations related to ADL-Q and ADL-I

The data for the ADL-Q and ADL-I were gathered by occupational therapists in clinic practice to imitate how data could be collected in the future. Still, the data were limited to women with rheumatic diseases receiving services in an outpatient rheumatology department. The fact that we were able to obtain more complete data (i.e. ratings on most of the 47 items for the majority of the participants) could result from the prospective design instructing the clients to rate all tasks of importance in their everyday lives. Or it could be due to the fact that the sample presented with higher ADL ability than the sample with ABI. In future studies, data will need to be collected among both genders, across diagnostic groups and also among clients with lower levels of ADL ability to further develop the ADL-Q and ADL-I.

Issues related to further validation of the ADL ability measures of the AMPS

Using the AMPS with clients with fluctuating ADL ability

The sample of clients with fluctuating ADL ability included in Study III were limited by the fact that it only included women independent in PADL. While women dependent of assistance in PADL were referred for

occupational therapy, they were not offered the possibility to participate in the interdisciplinary rehabilitation program in the clinic. Furthermore, due to the low number of men referred every year, only one course a year was conducted specifically for the male clients. As CWP and FM affect women more frequently than men(148), we do however consider the findings relevant for the CWP/FM population. We used a repeated measures study design to validate the measures of ADL ability of the AMPS in relation to stability and sensitivity to change. Such a design is limited when trying to establish if the changes in ADL ability was due to the intervention as it lacks of control for potential intervening factors. This suggest the need to replicate our findings during a randomized controlled trial.

Using the AMPS with clients with very low ADL ability

While the AMPS observations of the clients with ABI had been performed solely for clinical purposes, there was no systematic pattern related to the time of evaluation and the types of clients evaluated in the available retrospective data. We, therefore, had to determine criteria for inclusing of clients based on the need of data from at least two AMPS observations, one being close to admission and the other close to discharge, which influenced the sample size and the representativness of the sample. That is, we do not know if our sample represented clients with higher or lower ADL ability admitted to the rehabilitation center. Furthermore, the use of a pre- and post-test study design limited the possibility of controlling for potential intervening factors. Based on this study, we can not conclude conclusively if the observed change resulted from the interdisciplinary rehabilitation only.

Recommendations for future research

As already addressed in parts in the discussion, the studies in this thesis have raised many questions that need to be answered in future studies. To continue the development of the new revised versions of the ADL Taxonomy, the following issues should be further explored:

- First of all, separate Rasch analyses of the ADL-Q and ADL-I based on the existing data should be performed, including analyses for diagnostic DIF. Furthermore, DTF analyses should be performed to explore if misfitting items and items potentially displaying DIF significantly impact on the measures obtained.

- Before more data be collected on the ADL-Q and ADL-I, the written instructions need to be examined and possibly improved to stress the focus of the evaluation related to ability to do (can do) and avoid ratings based on what the person does or does not do. During analysis of data collected based on the revised instructions, evaluation should be made to determine if such revised instructions have reduced misfit related to items such as *Pedicuring* and *Driving a car*.

- The fact that the items displaying misfit vary across versions needs to be further investigated to determine the source or sources of misfit. Specifically, more data on more heterogenious samples need to be collected on all versions to further examine if misfit to some extent can be explained by gender- or diagnosis-related DIF.

- Another aspect of perceived quality of ADL task performance that has not been address in the present versions of the ADL-Q and ADL-I is the client's satisfaction with their perceived ADL ability. The development of such a scale will have to be explored in future studies to determine if it is possible to add this aspect into the ADL-Q and ADL-I.

- We found a very low relationship between measures of perceived ADL ability based on the ADL-Q and the ADL-I and the ADL process ability measures of the AMPS despite our attempts to include aspects of efficiency in the rating scale. There is a risk that the lack of correlation was due to the fact that we collapsed categories on the ADL-Q and ADL-I rating scale. Future studies will, therefore, be needed to determine if correlations between the ADL process ability measures and the measures of the ADL-Q and ADL-I can be improved by keeping the original categories of the rating scale separate.

Furthermore, the following issues should be explored in future studies:

- Further research is needed to determine if the ADL process ability measures of the AMPS are sensitive to change after longer term interventions among person with CWP/FM. As it can be reasoned that development of routine use of compensatory strategies to increase efficient ADL task performance takes more than four weeks post intervention, future studies, designed as randomised controlled trials, should evaluate long term changes in quality of ADL task performance among clients with fluctuating ADL ability.

- Finally, we need more research to establish effectiveness of occupational therapy interventions related to improvements in perceived and observed quality of ADL task performance among clients with chronic or longterm disability.

Conclusions

The following conclusions can be drawn from the results of the studies included in this thesis:

- By means of Rasch measurement methods, ordinal data based on the ADL Taxonomy using a rating scale focused on professional observation of quality of ADL task performance (ADL-O) can be converted into linear measures of observed quality of ADL task performance among clients with long-term or chronic disability following ABI.

- Furthermore, by implementing Rasch measurement methods, ordinal data based on the ADL Taxonomy using a rating scale focused on perceived quality of ADL task performance based on a questionnaire (ADL-Q) and an interview (ADL-I) can be converted into linear measures of perceived quality of ADL task performance among clients with long-term or chronic disability following rheumatic disease. The ADL-Q provides significantly higher measures of perceived quality of ADL task performance than ADL-I across diagnostic groups among women with RA, OA and FM.

- The AMPS, as a measure of observed quality of ADL task performance, has sound psychometric properties in relation to discrimination, stability and sensitivity to change in a sample of clients with long-term or chronic disabilities following ABI or rheumatic disease. The AMPS ADL ability measures can be used to discriminate between healthy women and women with CWP/FM. In addition, the AMPS can be used to provide ADL ability measures that are sensitive to change in quality of ADL task performance among clients with very low levels of ADL ability and limited potential for change as well as clients with a fluctuating level of ADL ability. While the AMPS ADL ability measures, when used with single clients, are sensitive to change caused by a fluctuating level of ADL ability, the instrument still can be used to provide stable ADL ability measures at group level among clients with a fluctuating level of ADL ability.

- ADL evaluations based on self-report and observation provide distinct but complementary information about a client's ADL ability. Therefore, the occupational therapist will need to evaluate perceived as well as observed quality of ADL task performance, before intervention planning and later for the purpose of documenting outcomes of intervention, in order to fully measure the client's quality of ADL task performance.

Acknowledgements

My motivation for entering into the doctoral studies was the possibility of being engaged in one of my favorite occupations — improving the basis for occupational therapy evaluations of our clients' ADL abilities — while developing new skills and improving some of my old skills. The journey towards the PhD degree has been full of challenges, learning experiences and moments of extreme frustration, enthusiasm and pure joy.

The creation of this thesis has only been possible through the generous contributions of many people. I wish to express my gratitude to everyone who in various ways has helped me with the studies and those who have been my support in everyday life.

My warmest gratitude to

The participants in the studies, the clients with acquired brain injuries and the women with rheumatic diseases. Meeting you all and for a short while sharing your experiences gave me an insight into how important it becomes to be engaged in all the common everyday life tasks, when they no longer can be taken for granted.

Anne G. Fisher, my main supervisor, for your sincere and firm guidance through the methodological fields of research and for sharing your incredible wit. When we met back in 1996, it was not on my agenda to pursue a PhD degree. Our mutual interest in an occupation-focused approach to occupational therapy evaluation and intervention provided the basis for a professional relationship that slowly evolved into a friendship. When I finally decided to move on, you were my first choice as a mentor on this journey and I couldn't have picked a more dedicated person, extremely ambitious on my behalf.

Hans Lund, my co-supervisor, for your generosity and warm support and unfailing belief in my abilities. You took me under your wings at the Parker Institute, supported my ideas and made sure that the road was cleared. In the most frustrating and discouraging periods of my work, I could always call you and you would listen, provide solutions and renew hope.

Bente Danneskiold-Samsøe, my co-supervisor and Director at The Parker Institute, for inviting me, as the first occupational therapist, to work at the Institute and providing me optimal conditions and facilities to conclude my thesis. Your encouragement and your commitment to research has rendered you my greatest appreciation.

Birgitta Bernspång, professor and Head of the Department of Occupational Therapy at Umeå University, for making me feel welcome from the very first day. Your way of raising critical questions gently and with a touch of humour is an example for me to follow.

Stine Amris, my colleague and co-author, for sharing your enthusiasm and knowledge about chronic pain and for offering professional collaboration and friendship. I deeply appreciate your constructive support and the trust you have shown me.

Mike Linacre, a wonderful educator and pedagogue, who helped me to understand some of the mysteries of Rasch measurement methods.

Henning Bliddal, my co-author and professor at the Parker Institute, for sharing ideas and knowledge within the field of rheumatology research.

Britta Löfgren, for bringing thoughtful input to my work and for accepting to physically nail this thesis on my behalf.

Elisabeth Bandak, my colleague, for sharing your positive attitude to life, for your warm and valuable support and interest throughout my studies and for being a friend to rely on.

Fellow PhD students, teachers and researchers at the Department of Occupational Therapy in Umeå, for sharing professional and personal issues in a warm and friendly atmosphere during seminars and "fika".

Colleagues at the Parker Institute, for showing interest in my work and for taking the time to discuss professional and personal issues. Especially, I want to direct my gratitude to *Robin Christensen, Sara Mortensen, Claus Bomhoff, Mette Gad* and *Else Marie Bartels* for each providing me support within your areas of expertise.

Occupational therapy colleagues at Frederiksberg Hospital, – especially *Lone Hoffman, Mette Hedeboe, Cecilie von Bülow* and *Christoffer Møller Jensen* – for your contribution to the collection of data among clients with rheumatic diseases, for bringing up issues of importance to the interpretation of data and for fun and laughter along the way.

Colleagues at Kurhus, who inspired me to implement process models and systematic evaluations in clinical praxis, who followed my guidelines and dared to disagree with me. Through our discussions I have learned so much.

Ingeborg Nilsson, Michaela Munkholm and Maria Lindström, who generously invited me into your family lives. Aside from the professional discusssions I will remember you for hours spent on serious shopping, happy laughter and Swedish christmas traditions.

Susi Kirkeskov, for guiding me through a process of personal growth during the times of my studies. By providing throughtfull input you helped me to gain further insight in to the complex mechanisms of social interactions among human beings.

Gudrun Árnadottír, my friend and colleague and companion on this journey. The days and nights we have spend together sharing laughter, knowledge and frustrations have meant the world to me and I treasure our friendship.

The Mimoses – Mette, Ulla, Jette, Britta and *Lia* – dear friends, who also happen to be occupational therapists. I feel very fortunate to have such caring friends. Your interest, support and encouragement throughout the process has been invaluable.

My precious family – my husband *Steen,* my children *Kristian and Nina,* my granddaughter *Julia,* my mother, father and brother – *Karen, Leif* and *Ib* – who always have been there for me, providing love and encouragement and reminded me of the importance of family life.

Mocca, Sussi and Tanja, for insisting on my attention and forcing me to join them on daily walks in the park, on the beach or in the woods.

Finally, the studies presented in this thesis would have been impossible to perform without the generous financial support of *The Oak Foundation*,

Frederik & Emma Kraghs Mindefond, Direktør E. Danielsen & Hustrus Fond, The Health Insurance Foundation, The Danish Rheumatism Association and *The Research Foundation of the Occupational Therapy Association*.

References

(1) Tjørnov J. Ergoterapi. baggrund og udvikling. [Occupational therapy. Background and development]. Copenhagen: FADL; 1987.

(2) Wright BD, Linacre JM. Observations are always ordinal; measurements, however, must be interval. Arch Phys Med Rehabil 1989;70:857-60.

(3) Bond TG, Fox CM. Applying the Rasch model: fundamental measurement in the human sciences. 2nd ed. Mahwah, NJ: Lawrence Erlbaum; 2007.

(4) Fisher AG. Uniting practice and theory in an occupational framework. 1998 Eleanor Clarke Slagle lecture. Am J Occup Ther 1998;52:509-21.

(5) Fisher AG. Occupational Therapy Intervention Process Model: a model for planning and implementing top-down, client-centred, and occupation-based interventions. Fort Collins, CO: Three Star Press, Inc.; 2009.

(6) Törnquist K, Sonn U. ADL Taxonomien - bedömning av aktivitetsförmåga. [The ADL Taxonomy - evaluation of ADL ability]. Nacka: Förbundet Sveriges Arbetsterapeuter; 2001.

(7) Wæhrens EE. ADL Taxonomien [The ADL Taxonomy]. Copenhagen: Ergoterapeutforeningen; 1998.

(8) Fisher AG. Assessment of Motor and Process Skills: development, standardization, and administration manual. 6th ed. Fort Collins, CO: Three Star Press; 2006.

(9) Fisher AG. Assessment of Motor and Process Skills: user manual. 6th ed. Fort Collins, CO: Three Star Press; 2006.

(10) Short-DeGraff M, Fisher AG. A proposal for diverse research methods and a common research language. Am J Occup Ther 1993;47:295-7.

(11) Fisher AG, Short-DeGraff M. Improving functional assessment in occupational therapy: recommendations and philosophy for change. Am J Occup Ther 1993;47:199-201.

(12) Townsend E, Polatajko H. Enabling occupation II: advancing an occupational therapy vision for health, well-being & justice through occupation. 1st ed. Ottawa: CAOT Publications ACE; 2007.

(13) Hidding A. Comparison between self-report measures and clinical observations of functional disability in ankylosing spondylitis, rheumatoid arthritis and fibromyalgia. J Rheumatol 1994;21:818-23.

(14) van den Ende CHM, Hazes JMW, Cessie SL, Breedveld FC, Dijkmans BAC. Discordance between objective and subjective assessment of functional ability of patients with rheumatoid arthritis. Br J Rheumatol 1995;34:951-5.

(15) Walsh EG, Khatutsky G. Mode of administration effects on disability measures in a sample of frail beneficiaries. Gerontologist 2007 Dec 1;47:838-44.

(16) Merbitz C, Morris J, Grip JC. Ordinal scales and foundations of misinference. Arch Phys Med Rehabil 1989;70:308-12.

(17) Wright B., Linacre J.M. Differences between scores and measures. Rasch Measurement Transactions 1989;3:63.

(18) Månsson E, Lexell J. Performance of activities of daily living in multiple sclerosis. Disabil Rehabil 2004;26:576-85.

(19) Kottorp A, Hällgren M, Bernspång B, Fisher AG. Client-centered occupational therapy for persons with mental retardation: implementation of an intervention program in activities of daily living tasks. Scand J Occup Ther 2003;10:51-60.

(20) Liu KPY, Chan CCH, Chu MML, Ng TYL, Chu LW, Hui SL, et al. Activities of daily living performance in dementia. Acta Neurol Scand 2007;116:91-5.

(21) Fisher AG, Atler K, Potts A. Effectiveness of occupational therapy with frail community living older adults. Scand J Occup Ther 2007;14:240-9.

(22) Girard C, Fisher AG, Short MA, Duran L. Occupational performance differences between psychiatric groups. Scand J Occup Ther 1999;6:119-26.

(23) McNulty MC, Fisher AG. Validity of using the Assessment of Motor and Process Skills to estimate overall home safety in persons with psychiatric conditions. Am J Occup Ther 2001;55:649-55.

(24) Poole JL, Atanasoff G, Pelsor JC, Sibbitt WL. Comparison of a self-report and performance-based test of disability in people with systemic lupus erythematosus. Disabil Rehabil 2006;28:653-8.

(25) Meyer A. The philosophy of occupational therapy. Arch Occup Ther 1922;1:1-10.

(26) Reilly M. Occupational therapy can be one of the great ideas of 20th century medicine. Am J Occup Ther 1962;16:1-9.

(27) Yerxa E. Health and the human spirit for occupation. Am J Occup Ther 1998;52:412-8.

(28) Trombly C. Occupation: purposefulness and meaningfulness as therapeutic mechanisms. Am J Occup Ther 1995;49:960-72.

(29) Borg T, Brandt Å, Madsen A, Runge UTJ. Introduktion til ergoterapi [Introduction to occupational therapy]. In: Borg T, Runge U, Tjørnov J, Brandt Å, Madsen A, editors. Basisbog i ergoterapi. aktivitet og deltagelse i hverdagslivet.[Basic book in occupational therapy. activity and participation in everyday life]. 2 ed. Copenhagen: Munksgaard Danmark; 2007. p. 18-25.

(30) Kielhofner G. Model of Human Occupation.Theory and application. 3rd ed. Baltimore: Lippincott Williams & Wilkins; 2002.

(31) Trombly Latham CA. Conceptual foundations for practice. In: Radomski MV, Trombly Latham CA, editors. Occupational therapy for physical dysfunction. 6th ed. Philadelphia: Lippincott Williams & Wilkins; 2008. p. 1-20.

(32) Kielhofner G. Motives, patterns and performance of occupation: basic concepts. In: Kielhofner G, editor. Model of Human Occupation. Theory and application. 3rd ed. Baltimore: Lippincott Williamsn & Wilkins.; 2002. p. 13-28.

(33) Wade DT. Measurement in neurological rehabilitation. 7th ed. New York: Oxford University Press Inc.; 1992.

(34) Agency for health-care research and quality health resources and services administration. Access to quality health services: healthy people 2010 - objectives for improving reproductive health. Department of Health and Human Services; 2010.

(35) MedicineNet.com. Webster's new world™ medical dictionary. 3rd ed. Hoboken, New Jersey: Wiley Publishing; 2008.

(36) Hagedorn R. Occupational therapy: perspectives and processes. 1st ed. Edinburgh: Churchill Livingstone; 1995.

(37) Fasoli SE. Assessing roles and competence. In: Radomski MV, Trombly C, editors. Occupational therapy for physical dysfunction. 6 ed. Lippincott Williams & Wilkins; 2007.

(38) Gitlin LN. Measuring performance in instrumental activities of daily living. In: Law M, Baum CM, Dunn W, editors. Measuring occupational performance: supporting best practice in occupational therapy. 2nd ed. SLACK Incorporated; 2005. p. 227-48.

(39) Letts L, Bosch J. Measuring occupational performance in basic activities of daily living. In: Law M, Baum CM, Dunn W, editors. Measuring occupational performance: supporting best practice in occupational therapy. 2nd ed. SLACK; 2005. p. 179-226.

(40) Avlund K. Disability in old age. University of Copenhagen; 2004.

(41) Aas RW, Grotle M. Clients using community occupational therapy services: sociodemographic factors and the occurence of diseases and disabilities. Scand J Occup Ther 2007;14:150-9.

(42) Burckhardt CS, Liedberg GM, Henriksson CM, Kendall S. The impact of fibromyalgia on employment status of newly-diagnosed young women: a pilot study. J Musculoskel Pain 2005;13:31-41.

(43) Henriksson C.M, Liedberg G.M., Gerdle B. Women with fibromyalgia: work and rehabilitation. Disabil Rehabil 2005;27:685-95.

(44) Henriksson CM. Living with continuous muscular pain--patient perspectives. Part II: strategies for daily life. Scand J Caring Sci 1995;9:77-86.

(45) Liedberg G.M., Henriksson C.M. Factors of importance for work disability in women with fibromyalgia: an interview study. Arthritis Rheum 2002;47:266-74.

(46) de Croon EM, Sluiter JK, Nijssen TF, Dijkmans BAC, Lankhorst GJ, Frings-Dresen MHW. Predictive factors of work disability in rheumatoid arthritis: a systematic literature review. Ann Rheum Dis 2004 Nov 1;63:1362-7.

(47) Johansson U, Högberg H, Bernspang B. Participation in everyday occupations in a late phase of recovery after brain injury. Scand J Occup Ther 2007;14:116-25.

(48) Parker C, Gladman JR, Drummond A. The role of leisure in stroke rehabilitation. Disabil Rehabil 1997;19:1-5.

(49) Dawson DR, Chipman M. The disablement experienced by traumatically brain-injured adults living in the community. Brain Inj 1995;9:339-53.

(50) Wikstrom I, Book C, Jacobsson LTH. Difficulties in performing leisure activities among persons with newly diagnosed rheumatoid arthritis: a prospective, controlled study. Rheumatology 2006 Sep 1;45:1162-6.

(51) Thomas SG, Pagura SM, Kennedy D. Physical activity and its relationship to physical performance in patients with end stage knee osteoarthritis. J Orthop Sports Phys Ther 2003;33:745-54.

(52) Thyberg I., Hass UAM, Nordenskiold U, Gerdle B., Skogh T. Activity limitation in rheumatoid arthritis correlates with reduced grip force regardless of sex: the Swedish TIRA project. Arthritis Rheum 2005;53:886-96.

(53) Trombly C. Anticipating the future: assessment of occupational function. Am J Occup Ther 1993;47:253-7.

(54) Gillette N. Research directions for occupational therapy. Am J Occup Ther 1991;45:563-5.

(55) Trombly Latham CA. Occupation as therapy: selection, gradation, analysis, and adaptation. In: Radomski MV, Trombly Latham CA, editors. Occupational therapy for physical dysfunction. 6th ed. Baltimore: Lippincott Williams & Wilkins; 2008. p. 358-81.

(56) Bellamy N. Osteoarthritis clinical trials: candidate variables and clinimetric properties. J Rheumatol 1997;24:768-78.

(57) Carville SF, Choy EH. Systematic review of discriminating power of the outcome measures used in clinical trials of Fibromyalgia. J Rheumatol 2008;35:2094-105.

(58) Buchbinder R, Bombardier C, Yeung M, Tugwell P. Which outcome measures should be used in rheumatoid arthritis clinical trials? Clinical and quality-of-life measures' responsiveness to treatment in a randomized controlled trial. Arthritis Rheum 1995;38:1568-80.

(59) McHorney CA, Kosinski M, Ware JE. Comparisons of the costs and quality of norms for the SF-36 Health Survey collected by mail versus telephone interview: results from a national survey. Med Care 1994;32:551-67.

(60) Haigh R. The use of outcome measures in physical medicine and rehabilitation within Europe. J Rehabil Med 2001;33:273-8.

(61) Roberts L, Counsell C. Assessment of clinical outcomes in acute stroke trials. Stroke 1998;29:986-91.

(62) Duncan PW, Jorgensen HS, Wade DT. Outcome measures in acute stroke trials : a systematic review and some recommendations to improve practice. Stroke 2000;31:1429-38.

(63) Wittink H, Rogers W, Sukiennik A, Carr DB. Physical functioning: self-report and performance measures are related but distinct. Spine 2003;8:2407-13.

(64) Latham NK, Mehta V, Nguyen AM, Jette A, Olarsch S, Papanicolaou D, et al. Performance-based or self-report measures of physical function: which should be used in clinical trials of hip fracture patients. Arch Phys Med Rehabil 2008;89:2146-55.

(65) Doble SE, Fisk JD, Fisher AG, Ritvo PG, Murray TJ. Functional competence of community-dwelling persons with multiple sclerosis using the Assessment of Motor and Process Skills. Arch Phys Med Rehabil 1994;75:843-51.

(66) Daving Y, Claesson L, Sunnerhagen KS. Agreement in activities of daily living performance after stroke in a postal questionnaire and interview of community-living persons. Acta Neurol Scand 2009;119:390-6.

(67) Spiegel JS, Hirshfield M, Spiegel TM. Evaluating self-care activities: comparison of a self-reported questionnaire with an occupational therapist interview. Rheumatology 1985;24:357-61.

(68) Campbell R, Quilty B, Dieppe P. Discrepancies between patients' assessments of outcome: qualitative study nested within a randomised controlled trial. BMJ 2003;326:252-3.

(69) Burckhardt CS, Clark SR, Bennett RM. The Fibromyalgia Impact Questionnaire: development and validation. J Rheumatol 1991;18:728-33.

(70) Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. Md State Med J 1965;14:61-5.

(71) Guide for the Uniform Data Set for Medical Rehabilitation (including the FIM™ instrument). Version 5.1. Buffalo, NY 14214-3007: State University of New York at Buffalo; 1997.

(72) Henriksson C, Gundmark I, Bengtsson A, Ek A-C. Living with fibromyalgia: consequences for everyday life. Clin J Pain 1992;8:138-44.

(73) Lue YJ, Lin RF, Chen SS, Lu YM. Measurement of the functional status of patients with different types of muscular dystrophy. Kaohsiung J Med Sci 2009;25:325-33.

(74) Tesio L. Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. J Rehabil Med 2003;35:105-15.

(75) Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Rheum 2007;57:1358-62.

(76) Law M, Baptiste S, McColl M, Opzoomer A, Polatajko H, Pollock N. The Canadian Occupational Performance Measure: manual. 4th ed. Ottawa: CAOT Publications ACE; 2006.

(77) Law M, Baptiste S, Carswell A, McColl M, Polatajko H, Pollock N. The Canadian Occupational Performance Measure. [Danish version: Canadian Occupational Performance Measure]. 4th ed. Copenhagen: Ergoterapeutforeningen; 2007.

(78) Árnadottir G. The brain and behavior: assessing cortical dysfunction through activities of daily living. 1st ed. St. Louis, MI: C.V. Mosby; 1990.

(79) Árnadottir G, Fisher AG, Lofgren B. Dimensionality of nonmotor neurobehavioral impairments when observed in the natural contexts of ADL task performance. Neurorehabil Neural Repair 2009;23:579-86.

(80) Törnquist K, Sonn U. Towards an ADL Taxonomy for occupational therapists. Scand J Occup Ther 1994;1:69-76.

(81) Sonn U, Tornquist K, Svensson E. The ADL Taxonomy - from individual categorical data to ordinal categorical data. Scand J Occup Ther 1999;6:11-20.

(82) Hariz GM, Lindberg M, Hariz MI, Bergenheim AT. Gender differences in disability and health-related quality of life in patients with Parkinson's disease treated with stereotactic surgery. Acta Neurol Scand 2003;108:28-37.

(83) Hayase D, Mosenteen D, Thimmaiah D, Zemke S, Atler K, Fisher AG. Age-related changes in activities of daily living ability. Aust Occup Ther J 2004;51:192-8.

(84) Merritt BK, Fisher AG. Gender differences in the performance of activities of daily living. Arch Phys Med Rehabil 2003;84:1872-7.

(85) Bernspång B, Fisher AG. Differences between persons with right or left cerebral vascular accident on the Assessment of Motor and Process Skills. Arch Phys Med Rehabil 1995 Dec;76:1144-51.

(86) Björkdahl A, Lundgren-Nilsson Å, Grimby G., Sunnerhagen KS. Does a short period of rehabilitation in the home setting facilitate functioning after stroke? A randomized controlled trial. Clin Rehabil 2006;20:1038-49.

(87) Linden A, Boschian K, Eker C, Schalen W, Nordstrom CH. Assessment of Motor and Process Skills reflects brain-injured patients' ability to resume independent living better than neuropsychological tests. Acta Neurol Scand 2005;111:48-53.

(88) Goverover Y, Johnston MV, Toglia J, DeLuca J. Treatment to improve self-awareness in persons with acquired brain injury. Brain Inj 2007;21:913-23.

(89) Tham K, Ginsburg E, Fisher AG, Tegnér R. Training to improve awareness of disabilities in clients with unilateral neglect. Am J Occup Ther 2001;55:46-54.

(90) Doble SE, Fisk JD, MacPherson KM, Fisher AG, Rockwood K. Measuring functional competence in older persons with Alzheimer's disease. Int Psychogeriatr 1997;9:25-38.

(91) Graff MJL, Vernooij-Dassen MJM, Thijssen M, Dekker J, Hoefnagels WHL, Rikkert MGMO. Community based occupational therapy for patients with dementia and their care givers: randomised controlled trial. BMJ 2006 Dec 9;333:1196.

(92) Oakley F. Assessment of Motor and Process Skills as a measure of IADL functioning in pharmacologic studies of people with Alzheimer's disease: a pilot study. Int Psychogeriatr 1997;9:197-206.

(93) Oakley F, Khin NA, Parks R, Bauer L, Sunderland T. Improvement in activities of daily living in elderly following treatment for post-bereavement depression. Acta Psychiatr Scand 2002;105:231-4.

(94) Åberg M, Ljungberg C, Edin E, Jenmalm P, Millqvisth NE, Wiberg M. Considerations in evaluating new treatment alternatives following peripheral nerve injuries: a prospective clinical study of methods used to investigate sensory, motor and functional recovery. J Plast Reconstr Aesth Surg 2007;60:103-13.

(95) Benson J, Clark F. A guide for instrument development and validation. Am J Occup Ther 1982;36:789-800.

(96) American Educational Research Association, American Psychological Association, National Council on Measurement in Education. Standards for education and psychological testing. Washington DC: American Educational Research Association; 1999.

(97) Törnquist K. Att fastställa og mäta förmåga til dagliga livets aktiviteter (ADL): en kritisk granskning av ADL-instrumenten och arbetsterapipraksis. [Verifying and measuring the ability to perform activities of daily living]. [dissertation] Göteborg University; 1995.

(98) Borell L, Lilja M, Carlsson-Alm S, Törnquist K, Ståhl E. Community-based occupational therapy. Scand J Occup Ther 1995;2:138-44.

(99) Hariz GM, Forsgren L. Activities of daily living and quality of life in persons with newly diagnosed Parkinson's disease according to subtype of disease, and in comparison to healthy controls. Acta Neurol Scand 2010;DOI: 10.1111/j.1600-0404.2009.01320.x.

(100) Hariz GM, Lindberg M, Bergenheim AT. Impact of thalamic deep brain stimulation on disability and health-related quality of life in patients with essential tremor. J Neurol Neurosurg Psychiatry 2002;72:47-52.

(101) Hariz GM, Blomstedt P, Koskinen LO. Long-term effect of deep brain stimulation for essential tremor on activities of daily living and health-related quality of life. Acta Neurol Scand 2008;118:387-94.

(102) Sirkka M, Bränholm IB. Consequences of a hip fracture in activity performance and life satisfaction in an elderly Swedish clientele. Scand J Occup Ther 2003;10:34-9.

(103) Sandin-Aldehag A, Jonsson H. Evaluation of a hand-training programme for patients with Welander Distal Myopathy. Scand J Occup Ther 2003;10:188-92.

(104) Hariz GM, Lindberg GM. Does the ADL part of the unified Parkinson's Disease Rating Scale measure ADL?: an evaluation in patients after pallidotomy and thalamic deep brain stimulation. Mov Disorder 2003;18:373-81.

(105) Gustafsson U, Grahn B. Validation of the General Motor Function Assessment Scale: an instrument for the elderly. Disabil Rehabil 2007;29:1-8.

(106) AMPS2005 computer-scoring software. Fort Collins, CO: Three Star Press; 2005.

(107) Fisher AG. The asssessment of IADL motor skills: an application of many-faceted Rasch analysis. Am J Occup Ther 1993;47:319-29.

(108) Merritt BK. Utilizing AMPS ability measures to predict level of community dependence. Scand J Occup Ther 2009;17:70-6.

(109) Goldman SL, Fisher AG. Cross-cultural validation of the Assessment of Motor and Process Skills (AMPS). Br J Occup Ther 1997;60:77-85.

(110) Doble SE, Fisk JD, Lewis N, Rockwood K. Test-retest reliability of the Assessment of Motor and Process Skills (AMPS) in elderly adults. Occup Ther J Res 1999;19:203-15.

(111) Kirkley KN, Fisher AG. Alternate forms reliability of the Assessment of Motor and Process Skills. J Outcome Meas 1999;3:53-70.

(112) Lopez W. Communication validity and rating scales. Rasch Measurement Transactions 1996;10:482-3.

(113) Linacre J.M. Differential item and test functioning (DIF & DTF). Rasch Measurement Transactions 2002;16:889.

(114) Smith EV. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Applied Measurement 2002;3:205-31.

(115) Schulz M. Standardization of the mean squares. Rasch Measurement Transactions 2002;16:879.

(116) Linacre J.M. A users guide to Winsteps(R) Ministep Rasch-model computer programs. http://www winsteps com/winman 2010 [cited 2010 Mar 31];

(117) Linacre J.M . WINSTEPS® Rasch measurement computer program [computer program]. Version 3.68.2. Chicago, IL: 2009.

(118) Bohlig M, Fisher WPJr, Masters GN, Bond TG. Content validity, construct validity and misfitting items. Rasch Measurement Transactions 1998;12:607.

(119) Fisher WP. Reliability statistics. Rasch Measurement Transactions 1992;6:238.

(120) Fawcett AL. Principles of assessment and outcome measurement for occupational therapists and physiotherapists. theory, skills and application. 1st ed. West Sussex, England: Wiley & Sons Ltd; 2007.

(121) Cohen J. A power primer. Psych Bulletin 1992;112:155-9.

(122) Bennett R. The Fibromyalgia Impact Questionnaire (FIQ): a review of its development, current version, operating characteristics and uses. Clin Exp Rheumatol 2005;23:S154-S162.

(123) Árnadottir G. Measuring the impact of body functions on occupational performance: validation of the ADL-focused Occupation-based Neurobehavioral Evaluation (A-ONE) [dissertation] Department of Community Medicine and Rehabilitation, Occupational Therapy, Umeå University, Sweden; 2010.

(124) Carville SF, Arendt-Nielsen S, Bliddal H, Blotman F, Branco JC, Buskilla D, et al. EULAR evidence based recommendations for the management of fibromyalgia syndrome. Ann Rheum Dis 2007 Jul 20;67:536-41.

(125) Wright B., Master GN. Rating scale analysis. Chicago, IL: MESA Press; 1982.

(126) Linacre J.M. WINSTEPS® Rasch measurement computer software [computer program]. Version 3.63.0. Chicago, IL: 2006.

(127) Linacre J.M. Optimizing rating scale category effectiveness. J Applied Measurement 2002;3:85-106.

(128) Linacre J.M. Category disordering vs. step (threshold) disordering. Rasch Measurement Transactions 1999;13:675.

(129) Linacre J.M. Investigating rating scale category utility. J Outcome Measurement 1999;3:103-22.

(130) Wright B., Linacre J.M. Reasonable mean-square fit values. Rasch Measurement Transactions 1994;8:370.

(131) Wilson M. Constructing measures: an item response modeling approach. Mahwah, NJ: Lawrence Erlbaum; 2005.

(132) Fleishman JA, Spector WD, Altman BM. Impact of differential item functioning on age and gender differences in functional disability. J Gerontol B Psychol Sci Soc Sci 2002;57:S275-S284.

(133) Niti M, Ng T-P, Chiam P-C, Kua E-H. Item response bias was present in instrumental activity of daily living scale in Asian older adults. J Clin Epidemiol 2007;60:366-74.

(134) Draba R. The identification and interpretation of item bias.Research Memorandum No 25. Available from http://www rasch org/memo25 htm 1977 [cited 2008 Sep 8];

(135) Hinkle DE, Wiersma W, Jurs SG. Applied statistics for the behavioral sciences. 5th ed. Boston, Mass: Houghton Mifflin; 2003.

(136) Higgins JPT, Deeks JJ, Altman DG. Special topics in statistics: standardized mean difference. In: Higgens JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions. Version 5.0.1. The Cochrane Collaboration. Available at: http://www.cochrane-handbook.org/; 2008.

(137) Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hilldale, NJ: Lawrence Erlbaum; 1988.

(138) Árnadottir G, Fisher AG. Rasch analysis of the ADL Scale of the A-ONE. Am J Occup Ther 2007;62:51-60.

(139) Tsuji T, Sonoda S, Domen K, Saitoh E, Liu M, Chino N. ADL structure for stroke patients in Japan based on the Functional Independence Measure. Am J Phys Med Rehabil 1995;74:432-8.

(140) Martinsson L, Eksborg S. Activity index - a complementary ADL scale for the Barthel Index in the acute stage in patients with severe stroke. Cerebrovasc Dis 2006;22:231-9.

(141) Harris RE, Williams DA, McLean SA, Sen A, Hufford M, Gendreau RM, et al. Characterization and consequences of pain variability in individuals with fibromyalgia. Arthritis Rheum 2005;52:3670-4.

(142) Henriksson C, Burckhardt CS. Impact of fibromyalgia on everyday life: a study of women in the USA and Sweden. Disabil Rehabil 1996;18:241-8.

(143) Coman L, Richardson J. Relationship between self-report and performance measures of function: a systematic review. Can J Aging 2006 Sep 1;25:253-70.

(144) Pepin V, Alexander JL, Phillips WT. Physical function assessment in cardiac rehabilitation: self-report, proxy-report and performance-based measures. J Cardiopulm Rehabil 2004;24:287-95.

(145) Hewlett S. Patients and clinicians have different perspective on outcomes in arthritis. J Rheumatol 2003;30:877-9.

(146) Duncan PW, Bode RK, Min Lai S, Perera S. Rasch analysis of a new stroke-specific outcome scale: the Stroke Impact Scale. Arch Phys Med Rehabil 2003;84:950-63.

(147) Hermansson L, Fisher A, Bernspång B, Eliasson AC. Assessment of capacity for myoelectric control: a new Rasch-built measure of prosthetic hand control. J Rehabil Med 2005;37:166-71.

(148) Neumann L, Buskila D. Epidemiology of fibromyalgia. Curr Pain Headache Rep 2003;7:362-8.