

From DEPARTMENT OF NEUROBIOLOGY, CARE
SCIENCES AND SOCIETY
Karolinska Institutet, Stockholm, Sweden

**MEASURING ACTIVITIES OF DAILY LIVING IN
CHILDREN:
VALIDATION OF THE GERMAN VERSION OF THE
PEDIATRIC EVALUATION OF DISABILITY
INVENTORY (PEDI-G)**

Christina Schulze



**Karolinska
Institutet**

Stockholm 2015

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet.

Printed by Eprint AB 2015

© Christina Schulze, 2015

ISBN 978-91-7549-897-3

Measuring Activities of Daily Living in Children: Validation of the German Version of the Pediatric Evaluation of Disability Inventory (PEDI-G)

THESIS FOR DOCTORAL DEGREE (Ph.D.)

by

Christina Schulze

Principal Supervisor:

Associate Professor Anders Kottorp
Karolinska Institutet
Department of Neurobiology,
Care Sciences and Society
Division of Occupational Therapy

Co-supervisor(s):

Professor Julie Page
Zurich University of Applied Sciences
School of Health Professions
Institute of Occupational Therapy

Professor Margareta Lilja
Luleå University of Technology
Department of Health Sciences
Division of Health and Rehabilitation

Opponent:

Professor Anita Bundy
Sydney University
Department of Health Sciences
Division of Occupational Therapy

Examination Board:

Associate Professor Erika Franzen
Karolinska Institutet
Department of Neurobiology,
Care Sciences and Society
Division of Physiotherapy

Associate Professor Lars E. Eriksson
Karolinska Institutet
Department of Learning, Informatics,
Management and Ethics (LIME)
Medical Management Centrum
(MMC)/Innovative Care

Associate Professor Liselotte
Norling Hermansson
Örebro County Council
Department of Prosthetics and Orthotics
Centre for Rehabilitation Research

ABSTRACT

Activities of daily living (ADL) are crucial for children with and without activity limitations because they enable them to participate in everyday life. For the evaluation of children's ADL performance, health professionals such as occupational therapists use standardized ADL assessments, e.g. the Pediatric Evaluation of Disability Inventory (PEDI). The general aim of this thesis was to evaluate evidence for the validity and the reliability of the German version of the PEDI (PEDI-G) for use in children with and without activity limitations in Austria, Germany, and Switzerland.

In Study I, a qualitative approach was chosen to gain knowledge about the process of translating and adapting an assessment in a review committee. Members from all three countries were recruited to represent a range of competencies and knowledge. The findings describe the adaptation of the PEDI for use in a new context as a question of striking a triple balance between (i) staying close to the original assessment while adapting the translation to the new context, (ii) making decisions and reaching consensus, (iii) catering for group dynamics while sticking to the planned schedule of the review committee.

The precision (inter-rater and test-retest reliability) of the PEDI-G was examined in Study II which included children with ($n=64$) and without activity limitations ($n=53$) from Austria, Germany, and Switzerland. The results of Study II demonstrate that the PEDI-G produces reliable results across raters and time points. As the test-retest reliability of the Modification Scale is overall insufficient, this scale may not be useful for measuring change.

The evidence of validity of the internal structure of the PEDI-G was evaluated in children with ($n=198$) and without ($n=64$) known activity limitations in Study III. Exploratory factor analysis was used on the subscales of the Functional Skills Scale and the items of the Caregiver Assistance Scale of the six PEDI-G domains. The results of the analysis suggest that a one-factor model should be used for all six PEDI-G domains.

The evidence relating to the validity of the PEDI-G was further examined with a Rasch model evaluating goodness-of fit, differential item functioning by country, and differential test functioning at item level of the Functional Skills Scale and the Caregiver Assistance Scale. Although a larger proportion of differences than

expected exists at item level, this does not impact the generated individual PEDI-G measure of the PEDI-G domains.

In conclusion, the results of these studies contribute to the evidence for the reliability and the validity of the PEDI-G domains in evaluating ADL of children with and without activity limitations living in Austria, Germany, and Switzerland. Still, it remains necessary to further examine the psychometric properties of the PEDI-G within larger samples and including more details about the parents, caregivers and contexts.

Key words: Activity limitations, interview, rehabilitation, occupational therapy, evaluation, psychometric properties, validity and reliability, Rasch model application

LIST OF SCIENTIFIC PAPERS

- I. Schulze, C., Page, J., Kottorp, A., Lilja, M. (2013). Adapting functional assessments for use in a new context: A balancing act. *Scandinavian Journal of Occupational Therapy*, 20(5), 336-342. doi: 10.3109/11038128.2013.802010.
- II. Schulze, C., Kottorp, A., Meichtry, A., Lilja, M., Page, J. (2014). Inter-rater and Test-retest Reliability of the German Pediatric Evaluation of Disability Inventory (PEDI-G). *Physical and Occupational Therapy in Pediatrics*, Nov 6. Epub ahead of print doi:10.3109/01942638.2014.975311
- III. Schulze, C. Kottorp, A., Meichtry, A., Page, J. Unidimensionality of the German Pediatric Evaluation of Disability Inventory (PEDI-G). *Submitted*.
- IV. Schulze, C., Page, J., Lilja, M., Kottorp, A. Cross-cultural validity of the German version of the Pediatric Evaluation of Disability Inventory (PEDI-G) - a Rasch model application. *Submitted*.

Published papers were reproduced with kind permission from the publisher.

CONTENTS

PREFACE.....	1
INTRODUCTION.....	5
Children’s activities of daily living	5
The role of context	6
Occupational therapy and activities of daily living.....	8
Evaluation of activities of daily living	9
Translation and adaptation of assessments.....	10
Pediatric evaluation of disability inventory (PEDI).....	11
Psychometric properties of the PEDI	17
Evaluation of ADL in German-speaking countries.....	21
Summing up	22
Aims	23
METHODS.....	24
Overview of the translation procedure of the PEDI-G	26
Investigation of test content (Study I)	28
Data collection procedures.....	29
Data analysis.....	29
Investigation of precision of measurement (Study II)	30
Participants	30
Data collection procedures.....	31
Data analysis.....	32
Investigation of internal structure (Study III & IV)	34
Participants	34
Data collection procedures.....	35
Data analysis.....	36
Ethical considerations	38
FINDINGS.....	40
Evidence based on test content (Study I)	40
Evidence based on precision of PEDI-G (Study II).....	42
Evidence based on test-retest reliability	43
Evidence based on inter-rater reliability.....	44

Evidence based on test-retest and inter-rater reliability of the Modification Scale	45
Evidence based on internal structure (Study III & IV)	45
Evidence based on exploratory factor analysis (EFA)	47
Evidence based on Rasch analysis	48
Evidence based on validity of the added German items	50
GENERAL DISCUSSION	51
Test content (Study I)	51
Precision (Study II)	53
Test-retest reliability of the Functional Skills Scale and Caregiver Assistance Scale	54
Inter-rater reliability of the Functional Skills Scale and Caregiver Assistance Scale	54
Test-retest and inter-rater reliability of the Modification Scale	55
Internal structure (Study III & IV)	55
Internal structure using EFA	56
Internal structure using Rasch analysis	57
Validity of added German items	60
Methodological considerations	61
Investigation of reliability and validity on item, subscale, and domain level with different approaches	62
Sample size and demographics in Study I-IV	63
Language issues	65
Clinical implications	66
CONCLUSIONS	68
FUTURE RESEARCH	69
ACKNOWLEDGEMENTS	71
REFERENCES	74
APPENDIX	89

LIST OF ABBREVIATIONS

A	Austria
ABI	Acquired brain injury
AERA	American Educational Research Association
ADL	Activities of daily living
Af	Acceleration factor
AMPS	Assessment of Motor and Process Skills
ANOVA	Analysis of variance
AOTA	American Occupational Therapy Association
APA	American Psychological Association
BWS	Between within subject mean square
CAT	Computer adaptive testing
CFA	Confirmatory factor analysis
CH	Switzerland
CI	Confidence interval
CP	Cerebral palsy
CTT	Classical test theory
DCD	Developmental Coordination Disorder
Df	Degree of freedom
DIF	Differential item functioning
DTF	Differential test functioning
<i>F</i>	Variance of the group means
EFA	Exploratory factor analysis
FIM	Functional Independent Measure
G	Germany
GMFCS	Gross Motor Function Classification System
GMFM	Gross-Motor Function Measure
I-ADL	Instrumental activities of daily living
ICC	Intraclass correlation coefficients
ICF	International Classification of Functioning, Disability and Health
IRT	Item response theory
KMO	Kaiser Meyer Olkin measure

ML	Maximum Likelihood
MnSq	Mean square
MTT	Modern test theory
N	Number of cases
NCME	National Council on Measurement in Education
oc	Optimal coordinates
OTPF	Occupational Therapy Practice Framework
pa	Parallel analysis
P-ADL	Personal activities of daily living
PEDI	Pediatric Evaluation of Disability Inventory
PEDI-CAT	Pediatric Evaluation of Disability Inventory- computer adaptive testing
PEDI-C	Pediatric Evaluation of Disability Inventory-Chinese version
PEDI-G	Pediatric Evaluation of Disability Inventory-German version
SEM	Standard error of measurement
SE	Standard error
SD	Standard deviation
SDD	Smallest detectable difference
Wee-FIM	Functional Independent Measure for Children
WMS	Within subject mean square
WFOT	World Federation of Occupational Therapy
WHO	World Health Organization
ZHAW	Zurich University of Applied Sciences

PREFACE

This thesis emanates from one of the very first research projects at the Institute of Occupational Therapy, the Zurich University of Applied Sciences (ZHAW). This is also one of the first research project performed in collaboration with occupational and physical therapists in Austria, Germany and Switzerland, others following. To others, this may hardly amount to anything special, but coming from the clinical side of occupational therapy and continuing working as an occupational therapist before, during, and probably also after my PhD graduation, this is, indeed, for me something very special.

Growing recognition of the need to use standardized, valid, and reliable assessments¹ in health care in general has sparked a rise in the use of various assessments in child rehabilitation. My interest in this field dawned early and inspired me to work with the “School Function Assessment” already during my Master thesis. At the heart of my current interest in child rehabilitation lies how occupational therapy intervention may improve activity and participation specifically in the context of activities of daily living (ADL); and my attention was therefore directed towards assessments suitable for this purpose when I became employed as a research associate at the ZHAW. Although developed and published already in 1992 as a functional assessment designed for use in the United States to evaluate ADL in children with a disability from 6 months to 7.5 years of age, the Pediatric Evaluation of Disability Inventory (PEDI) continues to be a preferred clinical and research assessment used worldwide. The choice of this assessment was therefore a natural one when I embarked on the present project, the unique characteristics of which are that:

- The research project was suggested by a peadiatrician and its results should be transferable to clinical practice.
- The research project represents an extension of my Master thesis which involved a pilot translation and validation of an assessment

¹ In this thesis, I use the term assessment to specify a broad range of methods in the literature referred to as tests, evaluation tools, scales, outcome measures, assessment tools or, simply, measures

tool; albeit, the present project of course has a much wider and much more scholarly focus.

- The project allowed me to apply the PEDI directly in practice while working as an occupational therapist.

Although I was and am still exploring which assessments serve my purpose and the child and family best, my belief in the use of assessments has changed during the course of the present study. In the beginning, I thought it is only about finding and using the one, right assessment and finding this one would solve everything (client's evaluation, intervention planning, and demonstrate the effectiveness of intervention). In fact and in all fairness, I must admit that the use of assessments like the PEDI is not a magic bullet resolving all challenges related to measuring ADL in children. While reliable and valid assessments are important for clinical practice they cannot replace the clinical experience of a therapist working with the children. Still, the thesis hopefully serves to advance this field of research and to provide health professionals with a valid and reliable assessment for measuring ADL.

During the work on my thesis, I struggled much with a plethora of terms, frameworks, and theoretical perspectives. I found it challenging to use terms consistently throughout the whole work and to justify unequivocally the choice of one particular framework or model over another. In this preface, I therefore wish to share some of my reflections on the use of the term “activity limitations”, the combination of the International Classification of Functioning Disability and Health (ICF) offered by the World Health Organization (WHO, 2001) and the Occupational Therapy Praxis Framework (OTPF) of the American Occupational Therapy Association (AOTA, 2014) as a framework, my reflections concerning reliability and validity and, lastly, my thoughts concerning the use of “German speaking countries” versus a “new context”.

The past decades in general and the period during which the ICF was developed and introduced in health care in particular have seen a change in our understanding of the concept of disability. The concept of disability has changed from being defined solely in medical terms to being acknowledged as having a strong social dimension, too. Still, the majority of health professionals might continue to see disability as being oriented towards clinical diagnosis and being

attributable to individual, functional impairments. However, every day people with disabilities encounter barriers to their daily activities that are attributable not to their personal impairments, but to the failure of their environments to duly cater for their impairment. Therefore, in the ICF, the term '*disabilities*' has been replaced by '*activity limitations*' which are defined as difficulties an individual may have in executing activities (WHO, 2001). I have accordingly decided to use the term '*children with and without activity limitations*', adopting the definition of the ICF, and I acknowledge that the inability to perform an activity can be rooted in various sources. This implies that the PEDI is here used for evaluating activity limitations irrespective of whether they are due to the child's functional impairments or ascribable to the child's context.

The ICF was developed by the WHO to provide a comprehensive framework of definitions and structures for rehabilitation. As the ICF has become an important and well-known classification system in rehabilitation, I decided to use it as one framework in the present thesis. However, after almost 20 years of experiences in occupational therapy involving children, it was my ambition also to place this thesis within an occupational therapy framework. I therefore welcomed my main supervisor's suggestion to use the OTPF (AOTA, 2014) for this thesis. I henceforth refer to both frameworks to underline the importance of ADL in children's health care. On the one hand, the ICF serves as a framework to present a more general perspective on ADL; on the other hand, the OTPF illustrates ADL from a more focused occupational therapy perspective.

A further aspect highlighted in the present thesis is the need to use valid and reliable assessment in health care to meet the need for evidence-based practice in child rehabilitation. There is a profusion of aspects of definitions concerning validity and reliability in the literature. Rather than discuss these definitions individually, I will use the commonly agreed definitions offered by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (AERA, APA, & NCME, 2014) so as to avoid any confusion about the meaning of these core concepts.

Lastly, although this thesis focuses on the validation of the PEDI in German-speaking countries (PEDI-G), the results of the four studies collectively

constituting the present thesis also invite recommendations and assumptions that may be applicable to assessments used in a new context in general; and they are by no means limited exclusively to the PEDI-G or to ‘German-speaking countries’. Therefore, the terms ‘new context’ and ‘German-speaking countries’ are used interchangeably. When the results apply exclusively to the German version or to German-speaking countries, the term ‘German-speaking countries’ is used; otherwise, the term ‘new context’ is used.

INTRODUCTION

CHILDREN'S ACTIVITIES OF DAILY LIVING

The term 'activities of daily living', also referred to as ADL, describes things we normally do during daily living and encompasses any such activity. According to the ICF, the term ADL is used in rehabilitation as an umbrella term relating to self-care, comprising those activities or tasks that people undertake routinely in their everyday life (WHO, 2001). ADL are tasks that are fundamental to supporting participation across school, home, and community environments (James, Ziviani, & Boyd, 2014; Simeonsson et al., 2003).

ADL are commonly classified as either personal ADL (P-ADL), which are orientated towards self-care (e.g. grooming or bathing), or instrumental ADL (I-ADL) (e.g. preparing meals, taking care of pets), which are orientated towards sustaining independence. I-ADL are usually learnt during the teenage years. Most I-ADL tasks require a higher level of physical and cognitive competency than P-ADL tasks, which are usually less challenging and learnt earlier (Spector, Katz, Murphy, & Fulton, 1987). The present thesis focuses on children who have not yet acquired the skills necessary to perform more complex I-ADL tasks. Therefore, when I use the term ADL in the following, I refer to P-ADL.

For most children, these ADL are learnt during their daily life and become a part of their daily routine (Dunn, Magalhaes, & Mancini, 2014; Lollar & Simeonsson, 2005; Mancini et al., 2002; Shepard, Procter, & Coley, 1996). Many children learn these skills by imitating adults, and children are sometimes motivated to acquire these skills through positive appeals like being a "big kid". In the western societies, children are expected to develop independence in their ADL; and sometimes young children even demand such independence, saying, "I can do it myself". Being independent in terms of ADL is one of the main expectations children with and without activity limitations need to fulfill because independence in this respect influences whether or not they will attend regular kindergarten or school (Berg, 2008; Brown et al., 1991; Haley, Coster, & Binda-Sundberg, 1994; Haley, Coster, Ludlow, Haltiwanger, & Andrellos, 1992). The majority of children develop the skills necessary to perform ADL, e.g. washing their face and body, without much parental help or assistance. Children with

activity limitations are more likely to experience difficulties and therefore need more parental or caregiver assistance, which may be a challenge for the parents (Cairney et al., 2005; Jasmin et al. 2009; Linderman & Stewart; 1999; Smyth & Anderson, 2000). If children with activity limitations face difficulties in their ADL performance, this may have serious consequences as illustrated in the following example.

At a specific age, children are expected to perform ADL, e.g., to dress themselves independently for playing outside, and most children without activity limitations meet these expectations with little effort and in time. In contrast, children with activity limitations often need parental help and/or more time to perform such activities. The activity “getting dressed” often includes several physical and cognitive tasks like choosing seasonally appropriate clothing, and for outside activities in winter time, these activities will include ADL like putting a zipper together, tying shoelaces and much more. Yet, children with activity limitations may find it difficult to perform these tasks and their participation in outdoor activities may therefore be restricted (Bundy, 1995; Cordier, Bundy, Hocking, & Einfeld, 2009; Rodger & Ziviani, 2006).

In school and kindergarten or in playgrounds, children with activity limitations often miss the breaks or the playtime because it takes them too long to get ready. Delay and inability to perform ADL can therefore be barriers for participation in family life, kindergarten or school, and in society in general (Cairney et al., 2005; Mandich, Polatajko, & Rodger, 2003; Ostensjø, 2003; Ponte-Allan & Giles, 1999). As illustrated above, the context obviously plays an important role for the child’s ability to perform ADL.

THE ROLE OF CONTEXT

Extant literature argues that ADL cannot be isolated from the environmental contexts in which they occur (Bonder, Martin, & Miracle, 2004; Bronfenbrenner, 1979; Dunn, 1993; Dunn, Brown, & McGuigan, 1994; Summers, Larkin, & Dewey, 2008). Indeed, both the physical environment as well as social, cultural and temporal factors significantly impact children’s ability to perform ADL. ADL performance is described as being contingent upon multiple relations such

as the unique attributes of the individual, the context of the physical and socio-cultural environment, and the specifics of the activity itself (Gannotti & Handwerker, 2002; Holsbeeke, Ketelaar, Schoemaker, & Gorter, 2009; Law et al., 1996). Although the extent to which the contextual factors affect ADL performance varies among children and families, they should always be considered when appraising children's ability to perform a specific ADL (Coster, 1998; Coster, 2006). Weisner (1984) defines a number of such contextual settings (e.g. home or school) and illustrates how context conditions performance. For example, children will perform ADL like eating by themselves differently and more or less successfully depending on their familiarity with the context in which the ADL is performed. Likewise, younger children usually rely on parental support, and they may feel more comfortable learning new tasks like using a spoon during lunch time and performing unknown activities like exploring different rooms if their parents are there to provide support. Therefore Weisner (1984) suggests that each ADL performance may be studied separately and may be compared within and across different settings and contexts (Weisner, 1984).

The following example may illustrate the influence of context on children's ability to perform ADL. In a comparison of Norwegian and children from the US, the most significant difference between the Norwegian and the American sample was the use of diapers. The Norwegian children were found to be continent 12-18 months later than the US normative group. Several explanations for this difference were offered; the main one was that the Scandinavian cultural values related to toilet training emphasize not stressing the child and parents, and the Scandinavian parents reported that they did not commence toilet training until they perceived that the child was "ready" for this (Berg, Aamodt, Stanghelle, Krumlinde-Sundholm, & Hussain, 2008; Schum et al., 2002). As a child's performance of ADL is conditioned by the context in which the ADL are performed, any appraisal of activity limitations must consider both perspectives, i.e. the child's functional limitations and the contextual restrictions bearing on the activity (Haley et al., 1994; Smits et al., 2010; Young, Williams, Yoshida, Bombardier, & Wright, 1996).

From this follows that the interpretation of the outcome of an ADL assessment shapes both the conclusion and any subsequent intervention. In order to assess

and measure aspects of ADL, the measurement constructs need to be defined carefully. Unfortunately, there is little tradition for assessing children's ability to perform ADL with the use of valid and reliable ADL assessments in German-speaking countries and almost no standardized assessments are available for this purpose. Therefore, the present thesis hopes to fill this gap in knowledge as well as clinical assessments.

OCCUPATIONAL THERAPY AND ACTIVITIES OF DAILY LIVING

In occupational therapy, the development and enhancement of children's ADL performance is usually undertaken in the natural context in which the child lives. Occupational therapy also often involves adaptation of equipment and seating, e.g., to allow better upper extremity use and to promote functional independence (Steultjens et al., 2004). Of much importance is also parental counselling to enhance parents' ability to support their child's learning to perform ADL.

In the OTPF (AOTA, 2014), ADL are defined as occupations in the sense of activities in which a person is engaged that are central to a client's identity and sense of competence and have particular meaning and value to that client. Among other definitions in the OTPF, the following three definitions of occupations referring to ADL have been singled out because they each illustrates at least one important aspect of the relation between occupation and children's ADL.

- “In occupational therapy occupations refer to the everyday activities that people do as families in families and with communities to occupy time and bring meaning and purpose to life. Occupations include things people need to, want to and are expected to do” (WFOT, 2012)
- “Activities ... of everyday life, named, organized, and given value and meaning by individuals and a culture. Occupation is everything people do to occupy themselves, including looking after themselves ... enjoying life ... and contributing to the social and economic fabric of their communities”(Law, Polataiko, Baptiste, & Townsend, 1997, p. 32)
- “Occupation is used to mean all the things people want, need or have to do, whether of physical, mental, social, sexual, political or spiritual

nature and is inclusive of sleep and rest. It refers to all aspects of actual human doing, being, becoming and belonging” (Wilcock & Townsend, 2014, p. 542).

The first definition refers to children’s contextually embedded occupations, e.g., their families which play an important role in the development of their ADL. The second definition takes into consideration that the activities are culturally and contextually dependent and are important requirements for participation in social contexts. The third definition highlights the needs and expectations related to ADL performance, and it captures all aspects of the actual doing, being, becoming, and belonging. These three definitions consider ADL as occupation, and they illustrate the lenses we as occupational therapists use to ascribe meaning to occupations like ADL (Hocking, 2001). These definitions of occupation also highlight the importance of considering ADL while appraising the client’s strength and weaknesses and they go to the core of my claim that children’s ADL performance should be considered by occupational therapists during their evaluation and intervention procedures. During these procedures, occupational therapist like health professionals in general should use valid and reliable ADL assessments.

EVALUATION OF ACTIVITIES OF DAILY LIVING

In rehabilitation, the child’s actual ability to perform ADL is often used as a practical measure to evaluate and predict the level of activity limitations (Jette & Haley, 2005; Lollar & Simeonsson, 2005; Ottenbacher et al., 1999). To evaluate children’s ADL performance such as getting dressed - to stay with the example – health professionals, including occupational therapists, use standardized assessments like the Gross-Motor Function Measure (GMFM), Wee-FIM (Functional Independent Measure for children), Assessment of Motor and Process Skills (AMPS), Pediatric Evaluation of Disability Inventory (PEDI) as well as non-standardized assessments such as self-developed questionnaires or checklists (Case-Smith & O’Brien, 2014; Majnemer, 2010). Yet, standardized assessments should be prioritized over non-standardized to ensure high-quality health care and to assess the effectiveness of interventions (Debusse & Brace, 2011; Law, 2003). Moreover, the use of outcome measures that have been

demonstrated to be valid and reliable for the purpose and population in question is highly recommended, especially in light of the scarcity of health care resources and the growing demand faced by health care professionals to document outcomes as a means of demonstrating that the therapy provided is effective (AERA et al., 2014; Fawcett, 2007). Concern over the psychometric properties is not just a prerogative of the researcher, but is as important to clinicians who employ outcome measures to obtain baseline information, to assess progress, and to inform treatment planning. Clinicians and clients want to be confident that the improvement in functional performance detected by an outcome measure does, indeed, represent a true change and is not just the result of random error (Jerosch-Herold, 2005).

TRANSLATION AND ADAPTATION OF ASSESSMENTS

If assessments are not available in the target language and context where they are intended to be used, the researcher or clinician may opt for one of three possibilities. First, an existing assessment may be translated and adapted for use in a new context. Second, a new assessment may be developed; or, third, the use of an assessment may be abandoned altogether. Obviously, the choice of one of these options over the other invites certain problems; but, generally, in order to obtain applicable assessments in a new context, a process of translation, adaptation, and validation of an existing assessment is recommended (Hambleton, 2001). Furthermore, an advantage of translating and adapting an existing validated assessment is that this enhances international comparability of clinical and research results and may be less costly than developing a new assessment (Guillemin, Bombardier, & Beaton, 1993; Herdman, Fox-Rushby, & Badia, 1997; Streiner & Norman, 2008). Processes and principles pertaining to cross-cultural adaptation and translation of assessments have been deliberated, among others, by AERA et al. (2014). Among issues raised in these organizations are aspects of fairness of testing and responsiveness to individual characteristics and testing contexts to ensure that test scores yield valid interpretations for the particular uses in question. “Fairness” in the sense used here implies that assessments do not place those assessed at an advantage or disadvantage. For this reason, cross-cultural translation and adaptation

encompass an extensive process that focuses on both language (translation) and cultural adaptation issues. In most cases, the translation and adaptation of assessments and the distribution of these assessments to participants in other cultures follow a procedure of translation(s), back-translation(s), review committee and pre-testing. Extensive guidelines and recommendations for translating and adapting an assessment exist (Beaton, Bombardier, Guillemin, & Ferraz, 2002; Brislin, 1970; Flaherty et al., 1988; Lee, Li, Arai, & Puntillo, 2009; WHO, 2012). The goal of translation is to achieve equivalence between the original and the translated version of the scale. Guillemin et al. (1993) recommend at least two independent translations and state that it is even better if each translation is done by a team rather than by individual translators.

PEDIATRIC EVALUATION OF DISABILITY INVENTORY (PEDI)

A standardized assessment often used by health professionals like occupational therapists is the Pediatric Evaluation of Disability Inventory (PEDI). It is an interdisciplinary assessment deployed to measure ADL in children with and without activity limitations. The PEDI is performed as a standardized interview based on a questionnaire with a parent or caregiver. It was developed as a functional assessment and an evaluative tool for children with a disability from 6 months to 7.5 years of age and was originally designed for use in the United States (Haley et al., 1992).

The construct measured with the PEDI is defined as capability² and performance in the English manual. The PEDI consists of three main scales: the Functional Skills Scale (Table 1 and 2); the Caregiver Assistance Scale (Table 3); and the Modification Scale (Table 3) in three domains, self-care, mobility and social function. These scales each measure different constructs.

According to the PEDI manual, the Functional Skills Scale measures capability. As the PEDI was developed before the ICF, and as the term ‘capability’ was

² In 1992 when the PEDI manual was published, ‘capability’ rather than ‘capacity’ was the preferred term. Only in this section, where I am referring to the PEDI manual, I will use the word ‘capability’ instead of ‘capacity’.

replaced by the word ‘capacity’ in the ICF, some clarification is needed here concerning these constructs. According to the ICF, capacity reflects what a child can do in a standardised environment or situation, whereas performance reflects what a child does in everyday settings (Hollenweger & Kraus, 2011). The PEDI manual states that “capability refers to the performance of tasks in either a standardized or an ideal situation and provides knowledge of the child’s best performance”(Haley et al., 1992, p.7). The criteria for capability in the PEDI are scored: 0 = unable or limited in capability to perform activity in most situations; 1 = capable of performing activity in most situations, or activity has been previously mastered and functional skills have progressed beyond this level (Haley et.al., 1992). In total, the Functional Skills Scale of the PEDI consists of 197 items, which are each given a score of 1 (capable) if the activity is mastered, or a score of 0 (unable) if not. According to these score definitions, ‘capability’ in the PEDI carries the same meaning as ‘capacity’ in the ICF. It was therefore here decided to use the term ‘capacity’ (instead of ‘capability’) in regard to the outcome of the Functional Skills Scale of the PEDI to stay closer to the terminology of the ICF, which serves as one of the main frameworks in this thesis. However, the wording “in most situations” in the scoring of the PEDI interferes somewhat with how the term ‘capacity’ is defined in the ICF, and it is hence a broader concept which relates more to the ICF’s definition of ‘performance’ (Sigrid Ostensjo, Bjorbaekmo, Carlsberg, & Vollestad, 2006).

The Caregiver Assistance Scale measures the degree or the amount of assistance required from the caregiver in order for the child to accomplish self-care, mobility, and social function activities. The Caregiver Assistance Scale consists of 20 items. The items of the Caregiver Assistance Scale are scored on a six-point ordinal scale ranging from “total assistance required” (score 0) to “independent” (score 5) and serves as indicator for a child’s need for assistance in ADL performance.

The Modification Scale evaluates the assistance the child needs to perform the ADL, and assistance is assessed with 20 items identical to the ones of the Caregiver Assistance Scale on a four-point ordinal scale ranging from “none” (score 0) to “extensive modifications” (score 3) needed. The Modification Scale is a nominal scale reporting the type and extent of environmental modifications needed for the child to be able to perform the activity.

In summary, the PEDI measures capacity and performance for selected activities within the domains of self-care, mobility, and social function on three scales: Functional Skills Scale, Caregiver Assistance Scale (the extent of help the caregivers provide) and Modification (i.e. environmental or technical modifications) needed to enhance the child's function (Haley et al., 1992). The following tables give an overview over the PEDI and provide information on the contents of the measurement scales and item examples of all three PEDI scales (see Tables 1-3).

Table 1: Overview of the content of the PEDI domains of the Functional Skills Scale

	<i>Self-care</i>	<i>Mobility</i>	<i>Social function</i>
A	Types of food texture (4)	Toilet Transfer (5)	Comprehension of word meanings (5)
B	Use of utensils (5)	Chair/Wheelchair transfers (5)	Comprehension of sentence complexity (5)
C	Use of drinking containers (5)	Car transfers (5)	Functional use of communication (5)
D	Tooth brushing (5)	Bed mobility/transfers (4)	Complexity of expressive communication (5)
E	Hair brushing (4)	Tub transfers (5)	Problem resolution (5)
F	Nose care (5)	Indoor locomotion methods (3)	Social interactive play (adults) (5)
G	Hand washing (5)	Indoor locomotion-distance/speed (5)	Peer interaction (child of similar age) (5)
H	Washing body and face (5)	Indoor locomotion- pulls/carries objects(5)	Play with objects (5)
I	Pullover/front opening garments (5)	Outdoor locomotion methods(2)	Self-information (5)
J	Fasteners (5)	Outdoor locomotion- distance/speed (5)	Time orientation (5)
K	Pants (5)	Outdoor surfaces (5)	Play with objects (5)
L	Shoes /socks (5)	Upstairs (5)	Self-protection (5)
M	Toileting task (5)	Downstairs (5)	Community function (5)
N	Management of bladder (5)	-	-
O	Management of bowel (5)	-	-
	15 subscales in self-care domain with 73 items	13 subscales in mobility domain with 59 items	13 subscales in social function with 65 items

Note: The numbers in brackets are the numbers of the items in each subscale

Table 2: Example of a PEDI item from the Functional Skills Scale

PART I : Functional Skills Scale
SELF-CARE DOMAIN

Place a check corresponding to each item:
 Items scores: 0 = unable; 1 = capable

A Food textures	Unable 0	Capable 1
1. Eats pureed/blended/strained foods		
2. Eats ground/lumpy foods		
3. Eats cut up/chunky/diced foods		
4. Eats all textures of table food		

Table 3: Example of PEDI items of the Caregiver Assistance Scale and the Modification Scale

Part II and III		Caregiver Assistance						Modification				Comment
Domain		5	4	3	2	1	0	N	C	R	E	
<i>Self-care 8 items</i>												
A	Eating											
B	Grooming											
C	Bathing											
D	Dressing upper body											
E	Dressing lower body											
F	Toileting											
G	Bladder management											
H	Bowel management											
<i>Mobility 7 items</i>												
A	Chair/toilet transfers											
B	Car transfers											
C	Bed mobility/transfers											
D	Tub transfers											
E	Indoor transfers											
F	Outdoor transfers											
G	Stairs											
<i>Social function 5 items</i>												
A	Functional comprehension											
B	Functional expression											
C	Peer play											
D	Joint problem-solving											
E	Safety											

5= Independent; 4= Supervision; 3= Minimal; 2= Moderate; 1= Maximal; 0= Total; N= None; C= Child, R= Rehab; E= Extensive

In recent papers, the development of a computer adaptive testing (CAT) version of the PEDI, the PEDI-CAT, has been described (Dumas et al., 2012; Haley et al., 2011; Kramer, Coster, Kao, Snow, & Orsmond, 2012). The authors report that in spite of the addition of some new items which extend the functional assessment and make the new tool appropriate for adolescents up to age 21, the replacement of the dichotomous scale by a 4-point scale to reflect difficulty may be highly welcomed. The PEDI-CAT includes illustration of all mobility and self-care items and is less time consuming to administer than the PEDI paper and pencil version. A responsibility domain replaces the Caregiver Assistance Scale to reflect the bigger age range to which the tool will be applied. The use of the PEDI-CAT in research has just recently been acknowledged. One reason why the PEDI is recommended is that it captures an individual's contextually embedded ADL performance, i.e. the PEDI assesses the child's environment task interaction, which is seen as a special strength of this assessment. It is also favoured over other assessments owing to its sound psychometric properties (Ketelaar & Wassenberg-Severijnen, 2010). This research will be discussed in the following section.

Psychometric properties of the PEDI

The following section describes the literature on evidence of the reliability and the validity of the PEDI and its translated versions. Reliability and validity cannot be seen as entirely different aspects of the psychometric properties of an assessment (AERA et al., 2014; Árnadóttir, 2010). Hence, the level of reliability/precisions of test scores has implications for validity. Whether classic test theory (CTT) or modern test theory (MTT) is used, evaluation of an assessment must include examination of aspects of reliability and validity (DeVellis, 2006). CTT and MTT offer different perspectives on how to use ordinal data statistically. In CTT, for example, an important statistical assumption is that the data used for analysis are interval data, not ordinal data, as specific mathematical manipulations would otherwise not be possible. However, the assessments used within rehabilitation are mostly based on ordinal scales (Árnadóttir, 2010). Because of this fundamental difference between CTT and MTT, many instruments used in health care developed using CTT have been criticized because they are in discord with the underlying statistical assumptions

and fail to prove them (Grimby, Ekholm, Fisher, & Sunnerhagen, 2005). However, even if a specific approach may be criticized for its limitations in respect of comparing results internationally and of relating findings to existing evidence, it may, nevertheless, be both feasible and applicable for the research question. It is therefore important to appreciate the strengths and weaknesses of either methodological choice (CTT and MTT) in relation to the findings, to discuss how they complement or contradict one another, and to use them synergistically.

Table 13 in the Appendix provides a general overview of existing evidence of the reliability and the validity of the PEDI. Studies including PEDI as an outcome assessment to evaluate the effectiveness of interventions have also been published (e.g., Duncan et al. 2012; Law et al., 2011; Rennie, Attfield, Morton, Polak, & Nicholson, 2000; Wiley, Meinen-Derr, Grether, Choo, & Hughes, 2012) but they are not summarized in Table 13, albeit their use in intervention studies may, indeed, also contribute with evidence of validity and/or reliability. The following sections offer a summary only of evidence relating specifically to the aspects explored in the present thesis.

Evidence of reliability of the PEDI

Reliability or precision is defined as the degree to which test scores for a group of test takers are consistent over repeated applications of a measurement procedure and hence are inferred to be dependable and consistent for an individual test taker (AERA et al., 2014). Specifically, reliability refers to the property of consistency in a measure.

The PEDI consists of altogether three scales (Functional Skills Scale, Caregiver Assistance Scale and Modification Scale); and within these scales, three domains are embedded (self-care, mobility and social function). As the Modification Scale is not a measurement Scale (it evaluates type and extent of environmental modifications the child depends on to support performance), after accomplishing the PEDI interview, the interviewer will get six different scores; three for each of the two scales (Functional Skills Scale and the Caregiver Assistance Scale).

During the interview and especially during the scoring, it is important to have some indication of the assessment's reliability (AERA et al., 2014). The evaluation of an assessment's reliability investigates whether it generates stable data; for instance, in the case of the PEDI, whether it produces stable data across two different time points (test-retest reliability) or across one or more therapists (inter-rater reliability). More specifically, the coefficients obtained by administration of the same form on separate occasions and the coefficients derived from the administration of alternate forms need to be evaluated to ensure that the data gathered and the conclusions drawn from the PEDI do not depend on the interviewer or the situation when the PEDI was performed.

Evidence of the reliability of alternate forms of the PEDI, its use on separate occasions and among different persons has been obtained for the PEDI (Feldman, Haley, & Coryell, 1990; James et al., 2014; Knox & Usen, 2000; McCarthy et al., 2002; Nichols & Case-Smith, 1996; Ziviani et al., 2001; Worth, Darrah, Magill-Evans, Wiart, & Law, 2014). Specifically, the reliability of translated versions of the PEDI has been investigated for the use in many countries (Berg, Jahnsen, Froslic, & Hussain, 2004; Chen, Hsieh, Sheu, Hu, & Tseng, 2009; Erkin, Elhan, Aybay, Sirzai, & Ozel, 2007; Elad et al., 2012; Gannotti, 1998; Stahlhut, Christensen, & Aadahl, 2010; Vos-Vromans, Ketelaar, & Gorter, 2005; Wassenberg-Severijnen, Custers, Hox, Vermeer, & Helders, 2003). These studies conclude that the PEDI is a reliable assessment for the tested contexts and populations in which it was used. For more detailed description of the evidence of PEDI, please consult Table 13.

Whether a scale which has been shown to be reliable with one group of individuals in a particular context is also reliable with other people and in different situations is an empirical issue that must be evaluated (Beaton et al., 2002; Streiner, 1993; Streiner & Norman, 2008; Vacha-Haase, 1998). No studies have so far examined the reliability of the PEDI translated into German or the reliability for use among participants from the German-speaking part of Europe.

Evidence of validity for the PEDI

Validity is the degree to which accumulated evidence and theory support a specific interpretation of test scores for a given use of an assessment. If multiple interpretations of a test score for different uses are intended, evidence of validity for each interpretation is needed (AERA et al., 2014). Rather than distinguishing between forms of validity (e.g. construct, criterion, and content validity) validity is described as a unity of concepts where different sources of evidence contribute to the intended interpretation of test scores. According to AERA et al. (2014) validity evidence is based on a) test content, b) response processes, c) internal structure, d) relations to other variables, and e) consequences of testing. From Table 13 (see in the Appendix) it may be seen that many studies evaluating different aspects of validity already exist, and one may question what new or additional validity studies with the PEDI or translated versions might contribute to the existing body of evidence.

In conclusion, the studies have shown satisfactory levels of validity of the PEDI (Ketelaar, Vermeer, & Helders, 1998; Reid, Boschen, & Wright, 1994; Vos-Vromans et al., 2005). However, the internal structure of the PEDI has only been sparsely evaluated, e.g. in relation to unidimensionality. The analysis of the internal structure of a test can indicate the degree to which the relationships among test items and test components conform to the construct on which the proposed test score interpretations are based (AERA et al., 2014). Internal structure has only been investigated for a Chinese PEDI version (PEDI-C) (Chen, Tseng, Hu, & Koh, 2010), a Spanish translation (Gannotti & Cruz, 2001) and a Turkish (Erkin et al., 2007) PEDI translation. For example, Gannotti et al., (2001) used the Spanish PEDI in the context of Puerto Rican families and performed confirmatory factor analysis (CFA) to investigate the internal structure of the PEDI (Gannotti & Cruz, 2001). Additionally, Chen et al. (2009) investigated the unidimensionality of PEDI-C using Mokken analysis. Another study using modern test theory performed by Erkin et al. (2007) investigated the internal structure of the PEDI in relation to goodness of fit at item level.

Based on AERA et al. (2014), studies of the internal structure of tests are designed to show whether particular items may function differently for identifiable subgroups of test takers (differential item functioning (DIF)). DIF

occurs when different groups of test takers with similar overall ability or similar status on an appropriate criterion on average have systematically different responses to a particular item. To our knowledge, no studies have yet been performed assessing the DIF of the PEDI as an aspect of its internal structure.

Evidence of cross-cultural validity of the PEDI

The previous section presented evidence of the reliability and the validity of the PEDI and its translated and adapted versions. The following section highlights the PEDI studies which evaluated the cross-cultural validity of the PEDI.

Over the past 20 years, the PEDI has been widely used in both clinical practice and research (Berg et al., 2008; Berg, Fräslie, & Hussain, 2003; Chen et al., 2009; Chen et al., 2010; Custers, Hoijtink, van der Net, & Helders, 2000; Custers, Wassenberg-Severijnen, Hoijtink, Vermeer, & Helders, 2003; Erkin et al., 2007; Gannotti & Cruz, 2001; Groleger, Vidmar, & Zupan, 2005; Mancini, 2005; Nordmark, Orban, Hägglund, & Jarnlo, 1999; Stahlhut et al., 2010; Stahlhut, Gard, Aadahl, & Christensen, 2011).

Not surprisingly, the evaluation of the applicability of the PEDI in various age groups in different countries did not provide consistent knowledge whether the normative age values from the US-PEDI are also applicable in the new target context. The studies from Norway, Sweden, the Netherlands, and Slovenia provide inconsistent evidence and point to possible cultural differences (Berg et al., 2008; Berg et al., 2003; Groleger et al., 2005; Nordmark et al., 1999; Wassenberg-Severijnen, 2005).

EVALUATION OF ADL IN GERMAN-SPEAKING COUNTRIES

Occupational therapists in German-speaking countries are often traditionally concerned with the restoration of body functions and body structures, and they tend to formulate therapeutic goals at the impairment level rather than at the activity and participation levels (Costa & Schönthaler, 2011; Gantschnig, 2014; Page et al., 2015; Romain, 2004).

While the use of standardized ADL assessment in pediatric health care has increased internationally (Dekker, Dallmeijer, & Lankhorst, 2005; Jerosch-Herold, 2005; Majnemer, 2010; Schädler et al., 2005) not many health professionals in the German-speaking part of Europe use standardized ADL assessments because only few such assessments are available. German speaking countries hence facing a lack of pediatric ADL assessments developed and/ or validated for use in German-speaking countries (Projektgruppe ergotherapeutische Befundinstrumente in der Pädiatrie, 2004).

SUMMING UP

Rehabilitation teams in which many health professionals are involved usually aim to achieve maximal activity and participation in everyday life for their clients such as children with activity limitations. Assessments that measure, e.g., ADL are therefore needed to document rehabilitation outcomes.

While there is evidence that supports the valid and reliable use of the PEDI in some contexts, no studies have explored the validity and reliability of the PEDI-G in German-speaking countries. Some of the applicability studies of the PEDI evaluated the use of different language versions of the PEDI, but none of the studies monitored differential item functioning to evaluate whether or not the different language versions of the PEDI do work in the same way. The living conditions in the German-speaking countries may not differ much from those of the US western life style; yet, the development of children's skills in everyday life is strongly influenced by differences in personal, social, cultural, and environmental factors. The aim of the present thesis is therefore to evaluate aspects of validity and reliability of the PEDI-G for use in children from Austria, Germany, and Switzerland

The knowledge provided with this thesis is important for different groups such as children who have difficulties performing ADL, health professionals working with those children, and their families. This thesis contributes with knowledge and methodological reflections on central aspects of validation of assessments in general and of pediatric assessments in particular, and it is therefore also relevant for researchers who are planning to translate and validate assessments for cross-cultural use.

AIMS

The general aim of the thesis was to evaluate aspects of validity and reliability of the PEDI-G for use in children with and without activity limitations in Austria, Germany, and Switzerland. This thesis also provides critical knowledge about the translation and adaptation process of assessments for use in new contexts. The overall aims of the specific studies included in this thesis were:

- To describe issues that influence the process of translating and adapting an assessment for use in a new context.
- To evaluate whether the PEDI-G used in an Austrian, a German, and a Swiss sample provide stable results when the assessment is used on different occasions or by two different interviewers (test-retest and inter-rater reliability).
- To evaluate whether the six domain scores of the PEDI-G can be used as unidimensional scores in research and practice.
- To evaluate the cross-regional validity of the PEDI-G in Austria, Germany, and Switzerland.

METHODS

The following sections describe the methods of the four studies in relation to participants, data collection procedures, data collection assessments, and data analysis. Study I (investigation of test content) is a qualitative study, and its structure varies slightly from the structures of the other three studies. Study II is summarized under the heading investigation of precision of measurement. Studies III and IV both evaluate the internal structure and use the same data for analysis; therefore, they are summed together under the heading investigation of internal structure. Table 4 presents the four studies contained in this thesis at a glance.

Table 4: Overview of methods and aims of the four studies comprised in the thesis

Study	Type of evidence ³	Method / Design	Specific aims of the studies
I	Evidence based on test content	<i>Qualitative approach</i> <ul style="list-style-type: none"> • Content analysis 	Identify issues that influence the process of translating and adapting an assessment for use in a new context when performing a review committee.
II	Evidence based on precision of measurement (test-retest and inter-rater reliability)	<i>Quantitative approach</i> <ul style="list-style-type: none"> • Intraclass correlation coefficients (ICC) • Kappa 	Examine test-retest and inter-rater reliability of the PEDI-G in a sample of children with and without activity limitations in Austria, Germany, and Switzerland. Investigate whether the reliability of the PEDI-G with added items shows reliability different from that of the translated PEDI without added items. Determine the smallest detectable difference.
III&IV	Evidence based on internal structure	<i>Quantitative approach</i> <ul style="list-style-type: none"> • Exploratory factor analysis (EFA) • Differential item functioning (DIF) • Differential test functioning (DTF) • Goodness of fit 	Evaluate the internal structure at item and domain level of the PEDI-G in a sample of children with and without activity limitations. A secondary aim of study four was to investigate whether the items of the PEDI-G were placed differently across Austria, Germany, and Switzerland and to ensure that possible DIF did not have an impact on the individual test function (DTF).

³ American Educational Research Association, American Psychological Association (APA) and National Council on Measurement in Education (NCME), 2014

OVERVIEW OF THE TRANSLATION PROCEDURE OF THE PEDI-G

Before explaining the methods used in the four studies in this thesis, a short description of the translation procedure will be given. The PEDI was translated and adapted into German following a translation procedure in accordance with international guidelines (Beaton et al., 2002; Beaton, Bombardier, Guillemin, & Ferraz, 2000; Dorer, 2012; Geisinger, 1994; Guillemin et al., 1993; Herdman et al., 1997; WHO, 2012; Wild et al., 2005).

Figure 1 presents an overview of the translation procedure of the PEDI-G. Three German translations with a focus on the clinical perspective - one translation from each country - were compared and synthesized by myself in collaboration with the 'clinical' translators. Two professional translators from Switzerland independently translated the PEDI into German. Their translations were also compared and synthesized by the responsible 'clinical' translators and myself. Finally, two synthesized versions were again reviewed and combined by the two professional translators and myself into a single version. Differences were discussed and suggestions, questions, and comments were documented (see Figure 1).

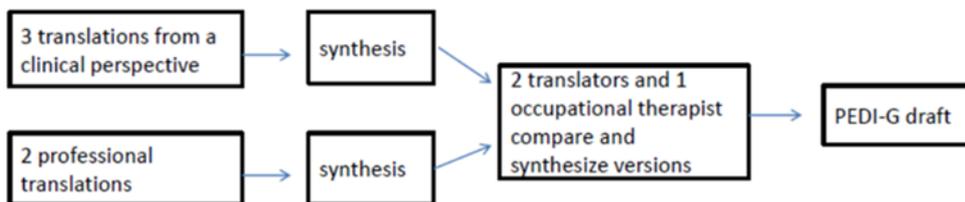


Figure 1: Overview of the translation procedure of the PEDI-G

During the translation and adaptation of the PEDI into German, it was determined that altogether nine items had to be added to the Functional Skills Scale in order for it to include ADL that are considered relevant to children living in a new cultural context, viz. the three German-speaking countries. Eight of these items reflect transfer into a stroller and using a three-wheeler in the mobility domain (see Table 5); and one item, “eating with fork and knife”, was added to the self-care domain.

Table 5: Item description of the added German items in the mobility domain of the PEDI

N	Vehicles (driving vehicle with wheels as a driver, not only as a passenger, pushing the vehicles not included. Not included are safety issues)
60.	Moves with Bobbycar, crawlers or skateboard (the vehicles are low and are driven indoors)
61.	Rides on tricycle, pedal tractor or Tretbulldog (the vehicles are driven outside, demands for balance, distance and speed are low)
62.	Rides kids bike / bicycle with training wheels, like a bike or therapy bike (used outside demands for balance, distance and speed a little higher)
63.	Rides kids bike / bicycle without training wheels (used outside demand for balance, distance and speed a little higher)
O	Transfer vehicles
64.	Helps to keep body parts in position during getting in and out of a child seat on the bike
65.	Helps to keep body parts in position during getting in and out of a bicycle trailer
66.	Gets in and out of a bike trailer independently
67.	Closes and opens belts

INVESTIGATION OF TEST CONTENT (STUDY I)

The expert committee members were recruited to represent a range of competencies, knowledge, and experience, including child rehabilitation and development; and various professional backgrounds such as medicine, therapy, psychology, translation, as well as mastery of moderation techniques. Participants from all three countries were included (see Table 6).

Table 6: Description of expert committee members in study I

Profession	Years of Experience	Country	Involved in translation process / additional notes	Gender
Developmental Psychologist	>3	CH		Female
Medical Doctor in Paediatrics	>15	G	German translation	Male
Medical Doctor in Paediatrics	>15	CH	Initiator of project	Male
Social Scientist	>10	CH	Moderator ⁴	Female
Translator	>10	CH	Professional translation	Female
Occupational Therapist	>10	A	Austrian translation	Female
Occupational Therapist	>10	G	Research assistant ⁵	Female

Note: CH= Switzerland; G= Germany; A = Austria

⁴ The moderator was one of the co-supervisor of this thesis.

⁵ I myself was involved in Study I.

Data collection procedures

The expert review committee meetings were held once the translation and synthesis of the PEDI in German had been completed. The PEDI-G draft was distributed together with the PEDI to the expert committee members in preparation for the meeting. Key questions were: “Can you identify anything that is unclear, irrelevant, or difficult to understand? Are you missing any aspects with regard to...?” To obtain the raw data, all audiotaped sessions were transcribed by myself into German. The meeting was set up as a one-day workshop divided into four sessions of approximately 1-2 hours’ duration each, and all the sessions were audiotaped and transcribed. This resulted in a total of 96 pages of transcripts.

Data analysis

Data was analyzed using qualitative content analysis based on the guidelines by Graneheim & Lundman (2004), with a view to identifying and resolving any issue related to the translation and localization of the PEDI in the German context. First, the content of the text was investigated by reading it several times and listened to the recordings, conferring regularly with the German-speaking supervisor of this thesis⁶. Second, searching for significant statements, units of meaning were identified and coded by going through the transcript line by line and paragraph by paragraph. Next, the codes were sorted into categories in conformity with the aim of the study; and differences and similarities between the original PEDI and the PEDI-G were identified and discussed. A first analysis and summary of the results was composed in English. This first draft helped establish an overview for all authors involved.

To obtain a deeper understanding we raised follow-up questions such as: What was the discussion about? What reasons and arguments triggered the debate? Was it difficult to resolve issues; and if so, why? What influenced the decision-making process? And how were problems solved? During the analysis, we all

⁶ The third person plural (we, our) is used to show that the research has been performed in collaboration with co-authors.

worked together and in the last step of the analysis differences, similarities, and connections between the categories were defined and, finally, the findings were formulated into two main categories with subcategories. An iterative approach was adopted in all stages of the analysis, going back and forth and comparing transcript, contents, argumentation, and interpretation of the data.

INVESTIGATION OF PRECISION OF MEASUREMENT (STUDY II)

Participants

The participants in this study include children with a known activity limitation as well as children without known activity limitations. In practice the PEDI-G will be used to evaluate ADL of children with and without activity limitations. Therefore, we wanted to find out whether the PEDI-G was applicable with children with and without known limitations. Children were included if: (i) they were aged between 6 months and 7.5 years, (ii) had an activity limitations (e.g. a cerebral palsy diagnosis), and (iii) their primary caregivers gave informed consent and understood and spoke German. Children with cerebral palsy, Gross Motor Function Classification System (GMFCS) (Palisano, Rosenbaum, Bartlett, & Livingston, 2008) level V, and children with autism were excluded from the study. Participation was voluntary, and the parents were informed that they could withdraw from the study at any stage. Due to the young age of the children included in the studies and because of their activity limitations, we performed the PEDI-G interviews with their parents or caregivers. This procedure is also recommended in the PEDI manual (Haley et al., 1992). Although the parents were interviewed the children and not the parents were considered to be the participants as the focus of interest was the children's performance, not the parents or the caregiver. In total, 117 children were included, 64 children with and 53 without activity limitations. Details about the sample characteristics are provided in Table 7.

Table 7: Description of the sample characteristics by country for Study II

	Austria (n=36)	Germany (n=32)	Switzerland (n=49)	Stat. test	p- value
Sex;					
female, n (%)	17(47.2)	14 (43.8)	17 (34.7)	$\chi^2=1.48$	0.48
male, n (%)	19 (56.2)	18 (56.2)	32 (65.3)	df=2	
Age, mean in months (SD)	37.9 (20.9)	31.41 (16.1)	32 (19.6)	$F=1.31$ df=(2,11.4)	0.27
Activity limitations					
without n (%)	16 (44.4)	20 (62.5)	28 (57.1)	$\chi^2=2.43$	0.29
with, n (%)	20 (55.6)	12 (37.5)	21 (42.9)	df=2	

Note: Chi-squared test for comparison among countries for gender and activity limitations and one-way ANOVA for age; df = Degree of freedom; F = Variance of the group means

Data collection procedures

Convenience sampling procedures were used for data collection. The PEDI-G was administered by 11 occupational therapists from Austria, Germany, and Switzerland during interviews with primary caregivers. All therapists had experience from working with young children, knowledge of child development, and were able to help the parent understand and judge the PEDI-G items according to the criteria, as recommended in the manual (Haley, et al. 1992). Before starting the data collection, all occupational therapists intensively studied the PEDI manual and received a minimum of 3 hours of training in administering the PEDI-G from the first author. Each therapist conducted at least two pilot PEDI-G interviews before initiation of the data collection. Data from these pilot interviews were not retained in the data set. During the data collection, any additional questions and misunderstandings were resolved among the therapists who performed the interviews and me either at regular meetings, through Skype sessions, or by email correspondence.

To examine test-retest reliability, the children's parents were interviewed twice by the same therapist. The second PEDI-G interview was carried out within a maximum of 2 weeks after the first interview. The mean time between test and

retest was 10.5 days with a minimum of 7 days and a maximum of 14 days. Parents could choose where the PEDI-G interviews took place, either in the occupational therapy department or at their homes. The duration of the interviews ranged from 20 minutes to 90 minutes.

To examine inter-rater reliability, parents were asked whether they agreed to have the interview audiotaped. For this purpose, two occupational therapists worked as a pair, with one scoring items during the parent interview and the other therapist scoring items after having listened to the audiotape. All audiotapes were rated by the same occupational therapist. In sum, 117 interviews were performed twice, and the first interview was audiotaped and then rated by the second therapist.

Data analysis

The PEDI is based on Item Response Theory (IRT). This implies that a child with higher ADL ability will probably receive a higher score within each of the items than a child with lower ADL ability. This also implies that there should be a logical pattern of item responses, meaning that a child should not have a high score on a hard item and at the same time a low score on an easier item, i.e. a high score indicating high level of ability to perform the particular ADL, and vice versa. The sum score of a subscale can vary from 0-6. Likewise, the items of the three domains of the Functional Skills Scale were summed. This sum score could vary from 0 to 74 for the self-care domain, from 0 to 67 for the mobility domain, and from 0 to 65 in the social function domain of the Functional Skills Scale (see Table 1).

For the Caregiver Assistance Scale, the items were also summed up to attain a sum score for each domain. The sum score could vary from 0 to 40 in the self-care domain, from 0 to 35 in the mobility domain, and from 0 to 25 in the social function domain of the Caregiver Assistance Scale. Even though not all studies evaluated the reliability of the Modification Scale we decided to include the Modification Scale in our analysis, and calculated test-retest and inter-rater agreement of all 20 items (see Table 2).

The subscales and item sum scores of the six PEDI-G domains were used to investigate reliability. To examine the test-retest and inter-rater reliability, intraclass correlation coefficients (ICC) were calculated from a one-way ANOVA as $(BMS - WMS) / (BMS + (r - 1) WMS)$, where BMS and WMS are the between subject and the within subject mean square, respectively, and r the number of trials. ICCs are used as a measure of association when studying the reliability of test scores over time and among raters (test-retest and inter-rater) (Streiner & Norman, 2008). Specifically, the ICC₁ (one-way random effect model) is used when each subject is rated by a different rater. An ICC value > 0.75 is considered excellent, 0.60-0.75 good, 0.40-0.59 moderate, and < 0.40 poor (Shrout & Fleiss, 1979). Test-retest and inter-rater reliability were examined using the ICC₁ for all 43 subscales for the Functional Skill Scale and the 20 items of the Caregiver Assistance Scale of the PEDI-G. Confidence intervals for ICCs were adjusted by the Bonferroni correction when multiple comparisons were made.

Cohen's weighted Kappa was used to calculate test-retest and inter-rater agreement for the Modification Scale. Cohen's Kappa is based on frequencies for nominal or ordinal data and is a measure of agreement that is adjusted for the agreement that can be expected by chance (Cohen, 1968). Kappa values range from -1 to 1 and are interpreted according to Landis & Koch, (1977) as follows: values less than 0.20 imply *poor* agreement; values between 0.21 and 0.40 indicate *fair* agreement; values between 0.41 and 0.6 indicate *moderate* agreement; values between 0.61 and 0.80 indicate *good* agreement; and *very good* agreement is indicated by values between 0.81 and 1.00.

In addition to the reliability calculation for PEDI-G subscales and items, and in order to further calculate the precision of the PEDI-G measurement, the absolute reliability of scorings was determined using the standard error of measurement (SEM) and the smallest detectable difference (SDD) for the three domains of the Functional Skills Scale and the Caregiver Assistance Scale. All statistical analyses were conducted using the R Core Team (2013) software packages.

INVESTIGATION OF INTERNAL STRUCTURE (STUDY III & IV)

Participants

The sample used in Study III and Study IV comprised children with and without activity limitations recruited in Austria, Germany, and Switzerland. Children with activity limitations spanned a wide range of diagnosis such as cerebral palsy, developmental coordination disorder, trisomy 21, and others. These samples were chosen to capture the full contents of the PEDI as it was assumed that these children would have varying physical and cognitive limitations. In total, 262 parents were interviewed; 64 (24.4%) parents of children without and 198 (75.6%) parents of children with an activity limitation. The diagnoses of the children with activity limitations were cerebral palsy (CP) $n=91$ (34.7%); developmental coordination disorder (DCD) $n=35$ (13.4%), pre-term delivery $n=13$ (5.0%), trisomy 21 $n=9$ (3.4%), epilepsy $n=8$ (3.0%), spina bifida $n=6$ (3.0%), acquired brain injury (ABI) $n=4$ (1.5%), and other syndromes not classified $n=32$ (12.2%). For more information concerning gender and age of the children, see Table 8.

Table 8: Description of the sample characteristics by country for Study III and IV

	Austria (n=86)	Germany (n=76)	Switzerland (n=100)	Stat. test	p- value
Sex; female, n (%)	42(49)	39(51)	37(37)	$F=0.92$	0.12
male, n (%)	44(51)	37(49)	63(63)	$df=(2,11.4)$	
Age, mean in months (SD)	4.1 (1.98)	3.7 (1.83)	4.1 (1.92)	$\chi^2=4.30$ $df=2$	0.40
Activity limitations					
without n (%)	70(81)	56(74)	72(72)	$\chi^2=2.43$	0.29
with, n (%)	16(19)	20(26)	28(28)	$df=2$	

Note: Chi-squared test for comparison among countries for gender and diagnosis and one-way ANOVA for age; df = Degree of freedom; F = Variance of the group means

Data collection procedures

Seven occupational and six physical therapists working in seven rehabilitation centres in Austria, Germany, and Switzerland recruited children between September 2011 and December 2012. Parents could choose whether the interview should be performed in the clinic or at their homes. Overall, almost half the interviews ($n=119$) were performed at the parents' homes. The duration of the interviews ranged from 20 minutes to 90 minutes with a mean duration of 45 minutes. In this cross-sectional study, the PEDI-G was administered as an interview by occupational or physical therapists with a primary caregiver or parent. In order to train the interviewer as much and as meticulously as possible, the administration of the PEDI-G was trained several times and during many meetings, skype, phone, or email. Several visits in which experiences with the PEDI-G were collected were performed to ensure that the standardized procedure described in the PEDI manual was followed by all test takers. We therefore assumed that the interviewers were overall confident and comfortable performing the PEDI-G.

Data analysis

Internal structure of the PEDI-G with explorative factor analysis (EFA)

In Study III, internal structure was investigated using CTT through the use of exploratory factor analysis (EFA) on the subscales of the Functional Skills Scale and the items of the Caregiver Assistance Scale (there are no subscales in the Caregiver Assistance Scale) of the six PEDI domains.

The Kaiser Meyer Olkin measure (KMO) was used to investigate the adequacy of the data for factor analysis. The KMO measure for all 43 subscales of the Functional Skills Scale across the three domains ranged from 0.92 to 0.98, and was thus marvelous (Kaiser, 1970). The KMO measure of all 20 items across the three domains of the Caregiver Assistance Scale ranged from 0.84 to 0.95, i.e. meritorious to marvelous. To investigate the internal consistency of the subscales of all six PEDI-G domains Cronbach`s alpha were assessed.

The four classic extraction methods were used to identify latent factors: the *Kaiser-criterion*, *parallel analysis* (pa), *Cattell`s scree test* or *acceleration factor* (af), and *optimal coordinates* (oc). As disagreement exists concerning which is the best method for extracting latent factors, the guidelines from Norman & Streiner (2014) was followed who suggested running the factor analysis a number of times with different methods and selecting the results of the analysis that makes the most clinical sense.

Finally, the last measure for deciding how many factors to retain was the ratio between the first and the second factor. If a PEDI domain is unidimensional, then all items should load highly on the first factor; and the amount of variance accounted for by the first factor`s eigenvalue should be considerably higher than the eigenvalues of the succeeding factors (Streiner & Norman, 2008). To specify “considerably” in this study, a conservative approach (to decide for a factor model) was used, i.e. a ratio of the first and second factor eigenvalues was used. Lumsden (1957) proposed that the first eigenvalue should be about 10 times higher than the eigenvalue of the second factor. Items with loadings below .05 indicate that the item does not belong to a group. After the identification of the number of factors to retain for each domain, a non-orthogonal rotation (oblique) by optimizing the oblimin criterion was performed (Norman & Streiner, 2014).

Internal structure of PEDI-G with Rasch analysis

In Study IV, aspects of internal structure using MTT in relation to goodness of fit, differential item functioning (DIF), and differential test functioning (DTF) were investigated. The individual items of the Functional Skills Scale and the Caregiver Assistance Scale were evaluated by the use of Rasch analysis. The WINSTEPS analysis software program was used to conduct all analyses (Linacre, 2014a).

Within the Rasch analysis, the person's raw scores on the PEDI items are converted into a linear measure expressed in logits (log odds probability units) based upon the pattern of responses across all items. Simultaneously, the PEDI item's raw scores are converted into a linear measure (item difficulty calibration) based upon the pattern of responses on that item across all persons, also expressed in logits (Wilson, 2005).

A basic assumption when performing a Rasch analysis is that the responses on items should demonstrate acceptable goodness-of-fit to the Rasch model in order to generate valid measures on items and persons. A commonly used criterion for goodness-of-fit for dichotomous scales is that the generated item infit and outfit mean square values (MnSq) should not exceed 1.3 (Linacre, 2013, 2014b; Linacre, Heinemann, Wright, Granger, & Hamilton, 1994; Wright & Linacre, 1994). As the infit statistics are more informative when exploring internal scale validity, we chose the infit statistics for evaluation of goodness-of-fit across items, which is also in accordance with other empirical studies (Bond & Fox, 2015; Nilsson & Fisher, 2006; Patomella, Tham, Johansson, & Kottorp, 2010; Patomella, Tham, & Kottorp, 2006). It is generally accepted that 5% of the items may by chance not demonstrate acceptable goodness-of-fit without this being a serious threat to person–response validity (Bonsaksen, Kottorp, Gay, Fagermoen, & Lerdal, 2013; Lerdal, Fagermoen, Bonsaksen, Gay, & Kottorp, 2014).

Rasch analysis is used to determine whether items are placed similarly across groups of persons from different regions. In this study, the magnitude of DIF was evaluated in two different ways reflecting the fairly small sample size: 1) statistically by using the Mantel-Hanzl statistic in the WINSTEPS program, and

2) clinically by monitoring the actual DIF size (Linacre, 2014a, 2014b). In the present study, we used a Bonferroni correction yielding a 1% alpha for statistical DIF (Tennant & Conaghan, 2007). An item DIF size exceeding 13 units on a 50-centered scale was considered a clinical DIF (Linacre, 2013; Tristan, 2006).

The DIF is considered to be trivial if the person estimates remains largely unchanged when comparing measures from two iterations (Tennant & Pallant, 2007). Differential test functioning (DTF) was therefore used to determine the potential impact of any identified DIF on the individually generated measures (Wright & Panchapakesan, 1969). DTF was also evaluated in two different ways: 1) statistically and 2) clinically. For the statistical analysis, standardized z-comparisons were used to evaluate any statistical difference between a person's measures obtained from the country-specific (Austria, Germany and Switzerland) item hierarchies and the standard error (SE) was calculated for each individual PEDI-G measure. The difference between the paired measures should fall within the 95% confidence bands based on the SEs of the estimated person measures (Linacre, 2014b). A significant DTF is detected if more than 5% of the paired measures are located outside the 95% confidence bands and in this case test bias would be present (Patomella et al., 2010). A clinical DTF was here considered a difference in the person estimates from the two analyses of more than 5 logits (Tennant & Pallant, 2007; Wright & Panchapakesan, 1969). Again 5% of our sample was used as a cut-off for clinical DTF.

ETHICAL CONSIDERATIONS

The Ethics Research Committees of the participating hospitals and centers in Austria, Germany, and Switzerland approved the project of validating the PEDI for its use in German-speaking countries.

One critical ethical aspect of the data collection for Study II to IV could be that we interviewed the parents and not the children. The parents were asked to provide information about their children's ADL performance because the PEDI is intended as a parent-administered interview and not as an interview administered by the children themselves. It was considered neither whether the children should be informed that their parents were asked about their ADL performance, nor whether they should consent in writing to their parents'

participation. Hence, we do not know how children feel about their parents' participation or if they would allow their parents to share details about their everyday lives.

All parents and caregivers received written information that was later complemented with oral information during the interview sessions. The written material included information about the research project, the rights to end participation at any time, issues of confidentiality, and practical information such as duration and aim. All interviewees consented to participation in the study.

In order to enhance participation and maximize the participants' comfort, they were allowed to decide whether the interview should be undertaken during their child's therapy session or in the privacy of their own home. Parents who did not understand German or with insufficient command of German were excluded from the data collection process.

FINDINGS

This chapter presents the main findings of the thesis. Overall, all four studies provide evidence of the reliability and the validity of the PEDI-G for use in German-speaking countries. In line with the definitions of validity and reliability described by Aera et al. (2014) the findings are presented together under the following headings: evidence based on test content, evidence based on precision of the PEDI-G, evidence based on internal structure of the PEDI-G, and evidence of the validity of items added in the German version of the PEDI.

EVIDENCE BASED ON TEST CONTENT (STUDY I)

Study I reports issues in relation to test content and illustrates aspects influencing the adaptation of the PEDI as proposed by the review committee of the three German-speaking countries. The analysis of the discussion in the review committee identified two relevant categories. In the first category, “*adapting an assessment to work in new contexts*”, there were three subcategories: “dealing with different understandings of terms when naming items”, “ensuring applicability of included items for a new context”, and “struggling with underlying constructs”. In the second category, “*dynamics influencing the harmonisation process*”, there were two subcategories; “feeling of shared responsibility for the process” and “from order to chaos”.

Adapting an assessments to work in new contexts

The first category shows difficulties and challenges of finding a consensus during the adaptation process and illustrates the reasoning process of the review committee members. More specifically the first subcategory was about “dealing with different understandings of terms when phrasing items”. In order to make the PEDI-G fit for its intended context, the members of the review committee prioritized the use of appropriate and understandable terms to designate the items in the PEDI-G.

Another subcategory “struggling with underlying constructs” illustrates the difficulty of finding a consensus in a review committee, too. The members of the committee, found themselves faced with the challenge of critically appraising the underlying constructs of the assessment; i.e. whether the constructs aimed at assessing the children's performance or their actual ability to perform ADL. The committee members were not unanimous as to how the underlying construct should be phrased in the German version of the items. Some preferred the phrasing “*the child can sit*” and “*the child cannot stand*” in order to convey that the child's ability to sit / stand was being measured. Others wanted to focus on the child's actual performance rather than its potential to perform the activity, and they suggested phrasings like “*sits*” or “*stands*” because these phrasings do not evoke the notion of ability through the use of the verb “*can*”. The discussion brought to light that members of the review committee with clinical experience were aware of a difference between the construct of capability and the construct of performance; and depending on the phrasing of items – the children might achieve higher or lower scores; hence, a risk of being misled when measuring ADL would arise. To solve this problem and to reach agreement, the participants consulted the PEDI manual and validity studies of other translated PEDI versions to benefit from the experience already reaped by other groups having translated the PEDI into other languages before. This process revealed that the concepts in the manual were not consistently defined, and this gave rise to even more questions.

The third subcategory was about “ensuring applicability of included items for a new context”. There the expert committee members discussed whether the items of the original PEDI had to be changed, added, or even deleted in order to map culturally applicable activities in the new context. All members agreed that in order to be useful an assessments evaluating ADL should embrace culturally applicable activities of the target context as well as more generic activities.

Dynamics influencing the harmonization process

This category contained issues that gave rise to discussion about the translation and the adaptation process. The first subcategory covers issues related to “feeling a shared responsibility” among the review committee members. The members demonstrated shared responsibility which shaped the way they discussed individual items to achieve agreement. In order to reach consensus, the members asked one other questions to find solutions that would take into account as many culturally and professionally relevant aspects as possible. This consensus-seeking process of deciding whether to keep, add, or change items in PEDI-G evolved into an iterative process, which implied, among others, that consensus already reached would occasionally need to be reconsidered later in the process in the light of new understandings; and this could result in rejection of previous decisions.

Another subcategory “from order to chaos” relates to the framing and planning of the work of the review committee (e.g. in terms of necessary preparation and its schedule) which also impacted the decision-making process. It became apparent that in some instances, the planned structure and setup of the committee did not suit the committee’s purpose. A schedule with four sessions was planned, each session focusing on a specific part of the assessment. However, this plan and the contents of the sessions did not always go together, and a more flexible handling of the program became necessary in order to ensure due progress in the committee’s work.

EVIDENCE BASED ON PRECISION OF PEDI-G (STUDY II)

The results of Study II confirm that the PEDI-G produces reliable measures for the domains of the Functional Skills Scale and the Caregiver Assistance Scale. The reliability of the Modification Scale indicates overall sufficient reliability for some, but not for all items. The findings related to the Functional Skills Scale and the Caregiver Assistance scale are summed together under one heading, and the findings related to the Modification Scale are presented subsequently.

Evidence based on test-retest reliability

The ICC for each domain (self-care, mobility, and social function) of the Functional Skills Scale and the Caregiver Assistance Scale indicate excellent test-retest reliability with ICC values above 0.95 (see Table 9). Specifically, the test-retest reliability of the 43 subscales of the Functional Skills Scale also show excellent reliability with ICCs above 0.75. All the 95% CIs were very narrow, ranging from 0.96 to 0.98 (see Table 9). The SEM and the SDD for each domain of the Functional Skills Scale and the Caregiver Assistance Scale are presented in Table 9. The SEM varied from 1.46 to 2.01.

Table 9: Test-retest reliability for the Functional Skills Scale and the Caregiver Assistance Scale

Test-Retest	Functional Skills Scale			Caregiver Assistance Scale			
	Domain	ICC (95%CI)	SEM	SDD	ICC (95%CI)	SEM	SDD
	<i>Self-care</i>	0.97 (0.96-1.00)	1.95	5.42	0.97 (0.95-0.99)	1.98	5.48
	<i>Mobility</i>	0.98 (0.97-0.99)	2.01	5.58	0.97 (0.94–0.98)	1.67	4.63
	<i>Social Function</i>	0.96 (0.96-0.99)	1.81	5.02	0.96 (0.92–0.98)	1.46	4.06

Note: Intraclass correlation coefficients (ICC), Standard error of measurement (SEM), Smallest detectable difference (SDD)

Likewise, the SDD varied from 4.06 for the social function domain of the Caregiver Assistance Scale to 5.58 for the mobility domain of the Functional Skill Scale. The results of the SDD show that the estimation of the SDD between two PEDI scores (at different time points) must exceed 4 to 6 raw scores in both scales to ensure that the measurement difference is true and not attributable to measurement error.

Evidence based on inter-rater reliability

The consistency of the PEDI-G interviews across raters was also excellent; with ICCs exceeding 0.95 for all six domains (see Table 10). The inter-rater reliability of the 43 subscales indicated excellent reliability with ICCs above 0.75 for the Functional Skills Scale and with ICCs above 0.97 for the items of the Caregiver Assistance Scale. For the six domains, all the 95% CIs of the ICCs were even narrower than for the test-retest reliability with values ranging from 0.97 to 1.00 (see Table 10). Additionally, the precision of the measurement was demonstrated for both measurement scales with SEM ranging from 0.30 to 0.66 for both scales and SDD ranging from 0.83 to 1.84.

Table 10: Inter-rater reliability for the Functional Skills Scale and the Caregiver Assistance Scale

Inter-Rater	Functional Skills Scale			Caregiver Assistance Scale			
	Domain	ICC (95%CI)	SEM	SDD	ICC (95%CI)	SEM	SDD
	<i>Self-care</i>	0.98 (0.99-1.00)	0.66	1.84	0.98 (0.97-0.99)	0.61	1.7
	<i>Mobility</i>	0.98 (0.99-1.00)	0.66	1.82	0.99 (0.98-1.00)	0.37	1.00
	<i>Social Function</i>	0.99 (0.99-1.00)	0.54	1.50	0.97 (0.97-0.99)	0.30	0.83

Note: Intraclass correlation coefficients (ICC), Standard error of measurement (SEM), Smallest detectable difference (SDD)

Evidence based on test-retest and inter-rater reliability of the Modification Scale

The test-retest reliability of the Modification Scale with overall 20 items in three domains varied from *fair agreement* to *very good agreement*. The investigation of the test-retest agreement revealed overall 15 items demonstrating *fair or moderate agreement* (values below 0.60). These items were five items from the self-care domain (Bathing, Grooming, Dressing upper body, Dressing lower body and Bowel management), six items from the mobility domain (Chair/toilet transfer, Car transfer, Bed mobility/transfer and Tub transfer, Indoor locomotion, Stairs), and four items from the social function domain (Functional expression, Peer play and Safety, Joint agreement). The remaining five items demonstrated *good to very good agreement* for the test-retest reliability (Eating, Bladder and Bowel management from the self-care domain and Outdoor locomotion from the mobility domain and Functional comprehension from the social function domain).

The values for the inter-rater reliability of the Modification Scale ranged from 0.60 to 1.00, indicating *good to very good agreement*.

EVIDENCE BASED ON INTERNAL STRUCTURE (STUDY III & IV)

In the following section, the findings concerning the internal structure of Study III and IV are summarized and presented. The mean scores of the six PEDI-G domains are presented in Table 11. One-way ANOVA was used to compare the mean scores for the six PEDI-G domains of the three countries (Austria, Switzerland and Germany). There were no statistically significant differences between the countries with respect to the six domain scores.

Table 11: The mean scores of the six domain scores of the three countries (Austria, Germany and Switzerland)

Functional Skills Scale	Austria (A) (n=86)	Germany (G) (n=76)	Switzerland (CH) (n=100)	Stat. test	p-value test
<i>Self-care</i>					
Mean (SD)	41.26 (20.40)	1.43 (22.94)	45.14 (21.53)	$F=0.925$	0.39
Range	4-74	6-74	4-74	$df=(2,11.4)$	
<i>Mobility</i>					
Mean (SD)	41.71 (21.00)	42.41 (22.98)	47.83 (18.37)	$F=2.203$	0.11
Range	1-67	2-67	4-67	$df=(2,11.4)$	
<i>Social Function</i>					
Mean (SD)	43.67 (17.68)	38.91(20.49)	43.06 (17.75)	$F=1.564$	0.21
Range	5-65	5-65	3-65	$df=(2,11.4)$	
Caregiver Assistance Scale					
<i>Self-care</i>					
Mean (SD)	18.44 (11.76)	17.71 (13.86)	18.58 (12.99)	$F=0.105$	0.90
Range	0-40	0-40	0-39	$df=(2,11.4)$	
<i>Mobility</i>					
Mean (SD)	22.21 (11.28)	21.42 (12.66)	24.68 (11.15)	$F=1.907$	0.15
Range	0-40	0-35	0-35	$df=(2,11.4)$	
<i>Social Function</i>					
Mean (SD)	16.2 (7.84)	13.75 (8.93)	16.16 (7.19)	$F=2.474$	0.09
Range	0-25	0-25	0-25	$df=(2,11.4)$	

Note: For comparisons among countries, one-way ANOVA, df =Degree of freedom; F =Variance of the group means, SD =Standard deviation

Evidence based on exploratory factor analysis (EFA)

The results of the four different methods (*Kaiser-criterion*, *parallel analysis* (pa), *Catell's scree test / acceleration factor* (af), or *optimal coordinates* (of)) suggest that each method produces a different number of relevant factor solutions. Table 11 outlines the suggested number of factors that should be retained, ranging from one factor to three factor solutions for the Functional Skills Scale and the Caregiver Assistance Scale. The methods suggesting most factors were parallel analysis and the optimal coordinate analysis, whereas the *Kaiser-criterion* and the af analysis suggested fewer factor solutions. The ratio of the first and second eigenvalue for the mobility domain of the Caregiver Assistance Scale is the lowest with 13 points. All ratios of the first and second eigenvalues for all six domains lie between 13 and 82. As Lumsden (1957) suggested that for stating unidimensionality, the first eigenvalue should be 10 times higher than the eigenvalues of the second factor, the results of this EFA suggest a one-factor model for all six PEDI-G domains as the first factor is in clear distance from the second. The internal consistency of the 43 subscales of all six PEDI-G domains was assessed with Cronbach's alpha values. All values lie above 0.90 which indicates excellent internal consistency.

Table 12: Number of factor to retain

Functional Skills Scale	Ratio first to second eigenvalue	N <i>Kaiser-criterion</i>	N pa	N af	N oc
<i>Self-care</i>	16	1	3	1	3
<i>Mobility</i>	15	1	2	2	2
<i>Social Function</i>	23	1	2	1	2
Caregiver Assistance Scale					
<i>Self-care</i>	15	1	3	1	3
<i>Mobility</i>	13	1	2	1	2
<i>Social Function</i>	82	1	1	1	1

Note: N *Kaiser-criterion*: Number of factors with the *Kaiser-criterion* rule; N pa: Number of factor with the *parallel analysis* method; N af: Number of factors with the *scree test*/or the *acceleration factor* method; N oc: Number of factors with the *optimal coordinates* method

Evidence based on Rasch analysis

Study IV offers evidence of the internal structure of the PEDI-G at item level among Austrian, German, and Swiss children with and without activity limitations. Overall, 24 (11.6 %) out of the 206 items of the Functional Skills Scale did not demonstrate acceptable goodness-of-fit according to the Rasch model. More precisely, eight items (10.8 %) out of 74 items of the self-care domain, ten items (14.9 %) out of 67 items of the mobility domain, and six items (9.2 %) out of 65 items of the social function domain of the Functional Skills Scale did not fit the expectations of the Rasch model.

In conclusion, there was a higher than expected level of misfit for all three domains of the Functional Skills Scale. Older age of the children was associated with a higher number of unexpected responses among the items demonstrating misfit with the Rasch model.

One item (5 %) out of the 20 items from the Caregiver Assistance Scale did not demonstrate acceptable goodness-of-fit. The misfitting item, “Eating”, belonged to one of the eight items of the self-care domain. All items of the mobility and social function domain of the Caregiver Assistance Scale demonstrated acceptable goodness-of-fit according to the Rasch model.

Differential item functioning (DIF)

Overall, 172 (83.5 %) items of the Functional Skills Scale showed no statistical or clinical DIF by country, which means that 34 (16.5 %) items out of the 206 items demonstrated statistical or clinical DIF by country. Among these 34 DIF items, 12 (35.3%) items were from the self-care domain of the Functional Skills Scale, 11 (32.4%) items were from the mobility domain and 11 (32.4%) items from the social function domain demonstrated either clinical or statistical DIF.

When monitoring the statistical DIF of the Functional Skills Scale, six items out of 74 items (8.1%) from the self-care domain, two items out of the 65 items (2.9%) from the mobility domain and six items out of the 65 items (9.2%) from the social function domain demonstrated DIF. When monitoring the clinical DIF of the Functional Skills Scale, 11 (14.9 %) items out the self-care domain, 10 (14.9 %) items out the mobility, and 11 (16.9 %) items out of the social function domain demonstrated clinical DIF.

The DIF analyses of the Caregiver Assistance Scale items identified no statistical or clinical differences in item functioning between Austria, Germany, and Switzerland, which indicates that the 20 Caregiver Assistance Scale items were functioning in a similar manner across the three countries.

A comparison of the findings from the item goodness-of-fit analysis and the DIF analysis showed that altogether 11 items demonstrating misfit also demonstrated DIF (46%), viz. two items from the self-care domain (1. Eats pureed/blended/strained foods; and 65. Helps to keep body parts in position while getting in and out of a child seat on the bike), six items from the mobility domain (6. Sits if supported by equipment or caregiver; 20. Sits if supported by equipment or caregiver in a tub or sink; 28. Moves within a room but with difficulty (falls; slow for age); 60. Moves with bobby-car, crawlers, or skateboard (vehicles are low to the ground and are driven indoors); 64. Helps to keep body parts in position while getting in and out of a child seat on the bike; 67. Closes and opens belts or seat devices), and three items from the social function domain (21. Tries to show you the problem or communicate what is needed to help the problem; 22. If upset because of a problem, child must be helped immediately or behavior deteriorates; 61. Child may play safely at home without being watched constantly). So, we concluded that the higher-than-expected item misfit on the PEDI-G scales may be associated with DIF by country.

Differential test functioning (DTF)

None of the paired measures from 1) the context-specific item hierarchies, and 2) the generic item hierarchies generated from all three countries demonstrated a z-value exceeding +/- 1.96 (z-values ranged from -.54 to .65 in the self-care domain, from -.34 to 1.57 in the mobility domain, and from -.34 to .43 in the social function domain of the Functional Skills Scale). Furthermore, none of the participants changed their measures across the six domains more than 5 logits. Thus, the identified DIF items of the PEDI-G across the German-speaking countries did not have a statistical or clinical impact on any of the generated individual PEDI-G measures.

EVIDENCE BASED ON VALIDITY OF THE ADDED GERMAN ITEMS

During the review committee's process of adapting the PEDI, it was discussed whether the original items had to be changed or even deleted or if new items had to be added in order to tailor the assessment to its new context (Study 1). All review committee members were unanimous that culture-specific German items had to be added to make the assessment applicable in the target context. Nine German items were added to the PEDI-G; two subscales each with four items (see Table 5) in the mobility domain of the Functional Skills Scale and one item in the self-care domain of the Functional Skills Scale.

In Study II and III the added German items of the Functional Skills Scale were included in the data analysis. In study II reliability was assessed for all 43 subscales and for the original 41 subscales separately. The comparison between the reliability with and without added subscales and items did not have an impact on both test-retest and inter-rater reliability on domain level.

In Study III the added German subscales in the mobility domain of the Functional Skills Scale showed no differences in regard to mean, standard deviation and factor loadings compared to the other subscales.

In Study IV, it was shown that the Rasch model expectations regarding item goodness-of-fit were not met for the following added items: 60. Moves with Bobby car, crawlers or skateboard; 64. Helps to keep body parts in position while getting in and out of a child seat on the bike; 65. Helps to keep body parts in position while getting in and out of a bicycle trailer; 67 Closes and opens belts or seat devices) (45%). The added items therefore demonstrated a higher than expected misfit than the PEDI-US items. Additionally, the investigation of DIF across the three countries revealed that three of the added items (60, 64, 67) (33%) demonstrated either clinical or both statistical and clinical DIF by country.

GENERAL DISCUSSION

The following section first discusses the challenges and difficulties encountered in the process of validating the PEDI-G. It then summarizes and discusses the most important findings in relation to the validity and the reliability of the PEDI-G used to measure ADL in children with and without activity limitations in German-speaking countries in light of existing literature on the PEDI. The present thesis has contributed with evidence of validity and reliability of the PEDI-G for use in German-speaking countries; and it offers knowledge, methodological reflections, and suggestions relevant to central aspects of adaptation and validation of assessments in general.

TEST CONTENT (STUDY I)

Illustrating issues relevant to the translation and adaptation of the PEDI, the results of Study I show that the process was characterized by complex group processes where a balance was struck between (i) adhering to the original assessment while adapting the translation to the new context, (ii) making decisions and reaching consensus, (iii) bending to group dynamics without jeopardizing the review committee's schedule. Another main finding was that the review committee's work was challenged in particular by differences in underlying ADL concepts related to capacity and performance.

The discussion related to whether the wording of the items should refer to the child's capacity or the child's performance of an ADL brought to light that review committee members with clinical experience were aware of a difference between the constructs of capacity and performance which was less known to members without clinical experience. According to the ICF, capacity reflects what a child can do in a standardized environment or situation, whereas performance reflects what a child does in everyday settings (WHO, 2001, 2007). The discussions testified to the necessity to clearly distinguish the used concepts from one another; and although these concepts may be easy to define and to distinguish theoretically, their application in practice and their use in assessments may be much more difficult. Even more difficult was the formulation of the

items and concepts in consistency with their use and intention in the original scoring form and manual of the PEDI, where the concepts were not always defined consistently (Haley et al., 1992). From the results from Study I and the discussions of the review committee, it also appeared that the concept of ADL performance was somewhat ‘slippery’; viz. a child’s ADL performance is not necessarily attributed only to the skills he/she demonstrated while performing the task; it is also related to the context and the families’ values (Fisher, 1998; Holsbeeke et al., 2009; Law & Letts, 1989; Malheiros et al., 2013; Young et al., 1996). For example, for the ADL item, “eats all textures of table food”; imagine that the caregivers answer, “yes, he/she does eat all textures”, then it may be possible that the answer refers to at least three different situations. In the first situation, the caregivers may refer not to the child’s actual ability, but to their own expectations. Maybe they do not expect the child to eat the whole bread (with the corners). In a second situation, the caregivers may say “yes” because they cut the corners of the bread for some reasons, so it is not a part of the performed task. In a third situation, the caregivers may say “yes” because the child does eat all textures of food including the corners of the bread. If those scores are used to draw conclusions, e.g., to compare the child with his/her peers, and are changed into normative scores, a risk of being misled when measuring may occur. In the first two situations, the child will get a score of “1” without actually showing his/her performance because of the circumstances of the context, i.e. the caregiver’s expectations or the context-specific task.

On first sight, ADL concepts seem to be easily transferred into different contexts, but as indicated by the findings of Study I and partly of Study IV, contextual variables are many, including, among others, parental attitudes, different environments, and social expectations (Case-Smith & O’Brien, 2014; Fisher, 2009; James et al., 2014; Magalhães, Fisher, Bernspång, & Linacre, 1996; Prado, Magalhães, & Wilson, 2009). In other words, children’s ADL are influenced by the family’s habits and routines and it often takes several years from the time a child is able to skillfully perform a task occasionally until the child can perform the task on a habitual, everyday basis in relevant situations (Bourke-Taylor & Hudson, 2005; Coster, 1998; Coster, 2006; Kielhofner, 2008; Lynch & Hanson, 2011; Rodger & Ziviani, 2006).

Many researchers have faced the challenge of developing an assessment to measure children's ADL (Eakin, 1989; Fisher, 2012; Grimby et al., 2005; Law, 2003; Law, Baum, & Dunn, 2005). Altogether three literature reviews were found in which 13 assessments were described that all assess ADL in children (Debusse & Brace, 2011; Harvey, Robin, Morris, Graham, & Baker, 2008; James et al., 2014). The results of the literature review illustrate that different ADL assessments are recommended for specific clinical questions, populations, and purposes. Among researcher the construct of activity is used differently (Berg, 2008; Dolva, Coster, & Lilja, 2004; Fisher, 2012; Kottorp, 2003; Ostensjø, 2003; WHO, 2007). For example, according to the reviews, the Assessment of Motor and Process Skills (AMPS) evaluates the quality of ADL motor and process performances in activities, whereas other assessments, for example the Functional Assessment Questionnaire (FAQ), evaluates only walking ability and whether this activity is difficult or easy (Debusse & Brace, 2011; Harvey et al., 2008; James et al., 2014). Furthermore, in some of the assessments, the concept of activity is defined in relation to its frequency, i.e. how often is the activity performed, or its importance, i.e. how important is the activity for the child. These examples illustrate some of the subtler aspects related to the definition of the constructs of ADL and they underscore two key findings of Study I; first, the importance of distinguishing between the ADL concepts measured (e.g. capacity versus performance); and, second, the need for translating and adapting these concepts when the PEDI is used in other contexts. The findings of Study I illustrated the need for a further, more detailed analysis of the ADL construct of the PEDI, and these issues were accordingly further investigated in Study III and IV.

The validity of the added German items will be further discussed below under the heading "Validity of added German items".

PRECISION (STUDY II)

The results of Study II demonstrate that the PEDI-G produces reliable results across raters and across two time points. The test-retest and inter-rater reliability study had excellent ICC for the three domains of the two scales (Functional Skills Scale and Caregiver Assistance Scale). While the inter-rater agreement on

the Modification Scale overall was for some items insufficient, the test-retest agreement varied from *poor* to *very good agreement* across all three domains

Test-retest reliability of the Functional Skills Scale and Caregiver Assistance Scale

The findings of test-retest reliability in this study are comparable with the test-retest reliability findings reported, e.g., in the PEDI and the Dutch PEDI version (Haley et al., 1992; Knox & Usen, 2000; McCarthy et al., 2002; Rich, Rigby, & Wright, 2013; Worth et al., 2014; Wassenberg et al., 2003; Wright & Broschen, 1993). Nichols & Case-Smith (2003) examined test-retest reliability in 23 children with activity limitations aged 1 to 6 years and found ICCs between 0.70 and 0.98, too. Correspondingly, e.g. Wassenberg-Severijnen et al. (2003) found ICCs between 0.91 and 0.98 in 53 children with activity limitations and in 63 children without a known activity limitation aged 6 months to 7 years and 6 months.

Although the ICC obtained in our study was very good, there was slightly more variation in the test-retest reliability than in the inter-rater reliability. This could be explained by a test effect because caregivers might be more conscious of the children's ADL after the first administration which was also concluded for the test-retest reliability of the Norwegian and Dutch PEDI (Berg et al., 2004; Wassenberg-Severijnen et al., 2003).

Inter-rater reliability of the Functional Skills Scale and Caregiver Assistance Scale

Overall, the present findings are comparable with the results of previous studies of the inter-rater reliability of translated PEDI versions despite small differences in samples and in languages, e.g. Chinese (Chen et al., 2009), Danish (Stahlhut et al., 2010), Dutch (Wassenberg-Severijnen et al., 2003), Norwegian (Berg et al., 2004) and Turkish (Erkin et al., 2007).

Test-retest and inter-rater reliability of the Modification Scale

While the inter-rater agreement on the Modification Scale overall was *good to very good*, the test-retest agreement varied from *poor to very good agreement* across all three domains. Unfortunately many studies investigating the reliability did not report or discussed results of the Modification Scale (Nichols & Case-Smith; 1996; Stahlhut et al., 2010; Wassenberg et al.; 2003). For the Chinese PEDI (PEDI-C) test-retest and the inter-rater agreement of the Modification Scale has been investigated (Chen et al., 2009; Chen et al., 2010).

Chen et al (2010) investigated test-retest reliability of the Modification Scale on item and domain level of the PEDI-C in 58 children with cerebral palsy. However, in contrast to our findings, they found *excellent* reliability for the self-care and social function domain with Kappa values above 0.88 and *good* test-retest reliability for the mobility domain with a Kappa value of 0.78 and higher. Unfortunately, their findings cannot be directly compared with our findings as their samples either consisted of only children without activity limitations or only children with cerebral palsy. As the test-retest reliability of the Modification Scale in the present study is overall insufficient, this scale may not be useful for measuring change in its current form.

Testing the inter-rater reliability of the PEDI-C, Chen et al. (2009) analyzed 494 children without a known activity limitation and found agreement between raters for the Modification Scale with Kappa values ranging from 0.96-1.00 for each item. Although the findings for the PEDI-C overall indicate better inter-rater reliability, with values above 0.96 compared with our values which ranged from 0.60 to 1.00 indicating *good to very good agreement*, the inter-rater agreement of this scale remains sufficient.

INTERNAL STRUCTURE (STUDY III & IV)

Studies III and IV were both complementary validity studies investigating the internal structure of the PEDI-G. Study III was performed to evaluate the factor structure of the six PEDI-G domains, specifically the unidimensional character. The aim of Study IV was to assess whether the scales of the Functional Skills Scale and Caregiver Assistance Scale of the PEDI-G meet the criteria of a valid

and generic tool to evaluate ADL abilities and performances in children with and without activity limitations. While Study III used exploratory factor analysis (EFA) on the subscales of each of the six domains, Study IV conducted Rasch analysis to analyze goodness-of-fit, DIF and DTF at item level of the six PEDI-G domains. Overall, both studies indicate that the two scales (six domains) and the majority of items of the two scales can be used as a valid tool to assess the ADL of children with and without activity limitations. Hence, the two studies also provide somehow mixed evidence concerning specific aspects of the internal structure of the PEDI-G. In the following, the results of the two studies will be discussed separately and afterwards the results of the two studies will be related to one another.

Internal structure using EFA

The findings of Study III confirm that the six PEDI-G domains of the Functional Skills Scale and the Caregiver Assistance Scale can be used as unidimensional domains in research and practice to evaluate ADL of children with and without activity limitations in German-speaking countries. Although the number of factors to be retained differed depending on the method used (see Table 11), a one-factor model was chosen because all ratios between the first and the second eigenvalues were higher than the suggested 10 ratio for all six domains, thus pointing to a one-factor solution. Additionally, all subscales of the Functional Skills Scale and all items of the Caregiver Assistance Scale but one loaded high on the first factor and accounted for more than 70% of the variance. Only the subscale *Food texture* from the self-care domain of the Functional Skills Scale showed a relatively low loading (0.51) on the first factor. A detailed description of the contents and meaning of all items and subscales of the PEDI is provided in the PEDI manual (Haley et al., 1992). Here, it says that the items from the subscale *Food texture* give an indication of the intactness of the child's swallowing mechanism and degree of oral motor control. In contrast, the other subscales of the self-care domain focus more on actual self-care activities. It is possible that the subscale *Food texture* captures specific body functions more than other subscales and that it focuses less on activity than the other subscales. However, according to Norman and Streiner (2008), factor loadings above 0.5

can be used, and we therefore suggest retaining the subscale *food texture* and to still consider this domain as unidimensional.

The results of this EFA are in line with the results of previous studies investigating the unidimensionality of the PEDI (Chen et al., 2009; Gannotti & Cruz, 2001). Although the authors of the study with the Spanish PEDI version did not use an EFA, but a confirmatory factor analysis without further explanation for their methodological choice, results of their validity study can to some extent be compared with our findings. They also used the ratio of the first and second eigenvalue as a criterion to conclude on unidimensionality. The ratios of the six domains of the Spanish PEDI version ranged from 3.0 to 4.5 for the Functional Skills Scale and from 4.53 to 6.70 for the Caregiver Assistance Scale. Compared with their results, the ratio between the first and the second eigenvalue of the six PEDI-G domains in our study are much higher and therefore support a one-factor (unidimensional) solution for the six PEDI-G domains, too (Gannotti & Cruz, 2001).

Internal structure using Rasch analysis

The findings of Study IV indicate that most of the items of the Functional Skills Scale and the Caregiver Assistance Scale remain stable and meet the criteria of goodness-of-fit. Wright & Panchapakesan (1969) suggested that 95% or more of the items of an assessment should meet the criteria of goodness-of-fit. For Study IV, this would mean that we could accept 10 or 11 items of the Functional Skills Scale lying outside the criterion. However, in our study, 24 items did not meet the criterion of goodness-of-fit. To understand the possible sources for the misfit of the items, the mentioned domains (self-care, mobility, and social function of the Functional Skills Scale) were further evaluated. For example, four misfitting items (1, 6, 20, and 28) from the mobility domain were all relatively easily performed activity items in the PEDI. This is also illustrated in the social function domain where four of the six misfitting items are also relatively easy items. An easy item demonstrating misfit is probably the result of a few individuals with overall higher ability that unexpectedly did not manage this item in combination with a few individuals with overall very low ability who unexpectedly did manage this item.

A study we found investigating goodness-of-fit was conducted in Turkey. In that study, Erkin et al. (2007) found 37 items out of the Functional Skills Scale to be misfitting. Although the PEDI-G consists of 206 items instead of 197 (no items were added to the Turkish PEDI version), the Turkish version had 13 items more demonstrating misfit than the PEDI-G did. Even if the studies are different in many ways, there are some similarities and aspects that may be compared and discussed. In both of the PEDI studies, the same nine items were misfitting (self-care: 1, 2, 4; mobility 1, 14, 20; social function 22, 46, 56, 61). Although these findings indicate that the level of misfit is slightly less outspoken when the PEDI-G is used than when the Turkish adaptation is used, the misfitting items still exceed the set criterion. Thus, in accordance with Erkin et al. (2007), the present study confirms the influence of social and cultural differences and family approaches on children's ADL abilities and performance, as also discussed in relation to Study I and the above discussion.

Study IV is the first study evaluating DIF for the PEDI used across countries. The PEDI-G demonstrated higher than expected levels of both clinical and statistical item DIF. The findings regarding item DIF varied across the scales, but they had no impact on the individual measures generated from the PEDI-G, which is clinically important for comparing pediatric rehabilitation outcomes across countries. In the present study, 34 items of the Functional Skills Scale were placed differently across the three countries where only minimal variations were expected. Not all of these 34 items were demonstrating DIF across all three countries; 18 items (52.9 %) demonstrated DIF only between two countries. As the evaluation of DIF items had minimal impact on the generated individual PEDI-G measure of the two PEDI-G scales, these scales can still be used as valid tools across the three countries.

Identifying potential systematic reasons for DIF is desirable because resolving them could ultimately lead to a more valid assessment across countries and diagnostic groups. Many studies have explored possible systematic reasons or sources of DIF, albeit unfortunately none investigating the PEDI (Dallmeijer et al., 2005; Hambleton, 2006; Linacre et al., 1994; Lundgren-Nilsson et al., 2005; Rogers, Gwinn, & Holm, 2001; Scott et al., 2010; Wright, Linacre, Smith, Heinemann, & Granger, 1997). If, for example, the DIF is due to difficulties in understanding and interpreting specific items in a similar way across countries,

additional PEDI-G workshops could be performed to train interviewers on how to score the items similarly. Another solution could be to use item-split techniques for specific countries, as demonstrated with other assessments (Lundgren-Nilsson et al., 2005; Pallant & Tennant, 2007; Tennant & Pallant, 2007; Scott et al., 2010).

Overall, 10 items demonstrated both DIF by country and misfit (self-care: 1 and 65, mobility 6, 20, 28, 60, 64 and 67 and social function 21 and 22). As we found a relation between item demonstrating misfit and DIF in the present study, the actions suggested above to minimize DIF could also potentially produce a lower degree of item misfit in the present study. However, no pattern in how the DIF items were placed across the three countries, e.g. whether the German items were mostly either harder or easier items could be identified.

No clear reasons for the identified DIF for the items could be identified, which is, unfortunately, not unusual. Other authors report difficulties in identifying the causes of DIF, too (AERA et al., 2014; Gantschnig, Page, & Fisher, 2012; Osterlind & Everson, 2009). It is also possible that the DIF and misfit items are due to the ADL construct which might be influenced by context and by habits within individual families. As we had very limited information regarding family structure and context in our data, in-depth analysis to explore systematic patterns was unfortunately not possible. Future studies should therefore seek to obtain more demographic information about family and context in order to be able to explore systematic patterns in relation to DIF and item misfit of the PEDI-G.

While the results of Study III clearly point to a one-factor solution that can be interpreted as evidence supporting the unidimensionality of the six PEDI-G domains, the findings of Study IV show that more items than expected demonstrate misfit and DIF. This means that depending on the methodological approach used and the level (item or domain) that is being evaluated, the analysis of the internal structure of the PEDI-G somehow yield inconsistent results. This is not surprising since the two studies used different analyses to illuminate specific aspects of the internal structure of the PEDI-G, and since they are based upon different assumptions regarding data. A more detailed discussion concerning strengths and weaknesses of the analysis used for investigating the internal structure is given below.

VALIDITY OF ADDED GERMAN ITEMS

One item in the self-care domain (eating with knife and fork) and two additional subscales in the mobility domain which aimed to describe transfer abilities into stroller and bikes were considered to be important for the German version of the PEDI as discussed in the review in study I and mentioned above (see Table 5). Therefore, nine items were added in the PEDI-G.

While in Study I, the necessity to add items specific for German-speaking countries was highlighted, the findings of study IV might question the valid use of these nine added German country-specific items. The values of Cronbach's alpha did not decrease when the two added subscales in the mobility domain of the Functional Skills Scale were dropped; however, they also do not seem to add new information as the Cronbach's alpha remained above 0.97. Moreover, the investigation of differential item functioning in Study IV may raise some concern regarding the added German items as they demonstrate a higher than expected misfit than the PEDI items. More specifically, four items of the nine added German items were demonstrating misfit and three of those misfit items (60, 64 and 67) were also demonstrating DIF by country.

Although the added items demonstrate evidence of validity based on test content from the translation and adaptation process described in Study I, they were not necessarily confirmed with evidence relating to internal structure in Study III and IV. Similar findings have been reported recently (Malinowsky et al., 2015), which indicates that items that can be viewed as highly relevant in a new context/country do not necessarily fit the same underlying construct when the unidimensionality of a scale is evaluated (Tennant et al., 2004). It may therefore be stated that adding country-specific items to ensure that the assessment contains relevant ADL may be necessary and, indeed, recommended from a content validity perspective. But adding items to assessments where sum scores are used to interpret the results may be a mistake because this will impede cross-cultural comparability, e.g. of norm values (note: for PEDI-G no norm values exist and the American ones are used). Thus, differences in a measure modified to improve cultural validity need to be taken into consideration when comparing findings across countries. Added items may therefore not be included in the scoring sum scores, but could be used to obtain additional information.

While some countries, e.g. China and Turkey did not add any country specific items in their PEDI translation in the Dutch version items were added. In those countries where items were added country-specific norm values for the PEDI were developed without evaluating DIF. According to Tennant et al. (2004) assessing and adjusting for cross-cultural validity through DIF is of crucial importance for the sound use of assessments in practice and research and therefore the evaluation of DIF may be important for further research with (translated) assessments used cross countries.

In summary it can be concluded that adding items during the translation and adaption process is a balancing act and its decision for adding items should include a consideration concerning the use and purpose of the assessment and may include the original developer of the assessment. Assessments which have a bigger item pools such as the PEDI-CAT and the AMPS are not dependent on summary scores of all items and may therefore be more applicable when country specific items need to be added (Dumas, Fragala-Pinkham, Feng, & Haley, 2012; Goldman & Fisher, 1997; Munkholm, Berg, Löfgren, & Fisher, 2010). This enables the test taker to extract those items that are relevant for the child's ADL performance. However whether the results gained from assessments with an item pool provide also results that are valid and cross-cultural comparable e.g. to evaluate ADL performance remains questionable.

METHODOLOGICAL CONSIDERATIONS

The following section discusses the methodological considerations in relation to different approaches and methods used for adapting and validating the PEDI for its use in German-speaking countries. The section is structured into four parts a) the first reflects on the choices we made concerning the combination of approaches at different PEDI-G levels (item, subscale, and domain); b) the second part considers the sample across all four studies; c) the third part discusses the sample size in Study III and IV; and d) the fourth part discusses language differences and how we dealt with this challenge to ensure that there no meaning would be lost.

Investigation of reliability and validity on item, subscale, and domain level with different approaches

One important assumption underlying the statistical methods used in classic test theory (CTT) is that the data are interval-based and not ordinal data in order to allow more sophisticated kinds of mathematical analysis such as EFA. As the popular assumption that ordinal data can be treated as interval data has repeatedly been questioned (Bond & Fox, 2015; Fisher, 1993; Merbitz, Morris, & Grip, 1989; Wright, 1982), detailed empirical analyses with a combination of different methods to evaluate the psychometric properties such as the reliability and the internal structure of PEDI-G items, subscales, and domains were performed. It was also decided in Study II to use a complementary approach (in addition to the ICC) for the investigation of reliability on the Functional Skills Scale and the Caregiver Assistance Scale, i.e. to use Cohen's Kappa statistics (not assuming interval data) on item, subscale, and domain level (data not shown). To further justify the use of CTT methods such as ICC and EFA was made to ensure that the underlying assumptions of those approaches were not violated. Data adequacy was appraised by calculating the Kaiser Meyer Olkin measure and by evaluating the internal consistency using Cronbach's alpha. Although these additional approaches do not solve the issue of treating ordinal data as interval data, they support a more comprehensive view of the findings than would a single methodological approach alone. Based on the findings of Study III and IV, it is obvious that depending on which approach and which scale level is used for evaluation of the internal structure, some inconsistencies occurred. It is important to acknowledge these inconsistencies when reviewing the overall complex findings on validity and reliability issues of the PEDI-G.

Since the thesis comprises a combination of multiple approaches and methods (qualitative and quantitative; CTT and MTT) this is considered a strength. The knowledge provided in Study I and its critical reflections in regard to the translation and adaptation of assessments in particular may assist other researchers who are faced with the task of translating and validating an assessment tool for cross-cultural use. Moreover as illustrated in Study II-IV, CTT and MTT approaches may complement each other. More specifically they seem to be useful for relating the findings gained from each method to each other and for discussing how the findings complement or contradict one another.

Therefore it should be acknowledged that the combination of approaches in this thesis generated a more complex but also inconsistent view of the internal structure of the PEDI-G when combined and put together. Even though it is common that studies evaluating psychometric properties of clinical assessments use only one approach - either MTT or CTT - to evaluate aspects of validity, the variety of findings across methodological approaches and various demographics illustrate consequences for the overall judgement of evidence of validity and the reliability of each assessment. Many studies investigating the validity of assessments e.g. for the Functional Independence Measure (FIM) (Cotter, Burgio, Stevens, Roth, & Gitlin, 2002; Dorans & Kulick, 2006; Kielhofner, 2006; Ravaud, Delcey, & Yelnik, 1999; Wright et al., 1997) or the Mini Mental (Dickson & Kohler, 1995; Ramirez, Teresi, Holmes, Gurland, & Lantigua, 2006) using either MTT or CTT exist. If those findings are put together it can be seen that they provide somehow inconsistent or even contradictory results concerning conclusions of the degree of validity or specific aspects of validity too. This, to some extent reflects also our results. Therefore, to ensure that the results from our studies are not merely due to errors in design or methodological approaches, Aera, et al. (2014), Hambleton (2006) and Pae & Park (2006) provided suggestions for the performance of “good validity” studies e.g. for the identification for DIF which were followed in this thesis.

Sample size and demographics in Study I-IV

The review committee members of Study I were recruited to represent a range of competencies, knowledge, and experiences. The PEDI may be used as an interdisciplinary assessment because its contents cover many issues within health-related and education-related disciplines. Unfortunately, the review committee did not include representatives from all the professional groups that might use the PEDI-G; nor did it include the parents or children with activity limitations. It is therefore possible that a more broadly composed review committee would have been able to address a richer variety of aspects. However, the group size was selected carefully according to the criteria mentioned above; and those criteria were based on Krueger and Casey's (2009) suggestions for focus groups. Having a bigger group in the review committee would also have

introduced issues of management, especially as far as the process of translating the PEDI and establishing consensus are concerned.

In Study II to IV, children with and without activity limitations were included in conformity with the use of the PEDI in practice. In accordance with other research, the findings of this thesis pointed out that cultural/contextual differences are important to consider when evaluating the reliability and the validity of assessments in a new context (Custers et al., 2000; Gannotti & Handwerker, 2002; Groleger et al., 2005; Hammell, 2001; Nordmark et al., 1999). Unfortunately, we lacked additional information on the parents, e.g., gender, education, age, context, and their beliefs and views on ADL and activity limitations. Although the PEDI with its measurement scales recognizes the importance of context for understanding and working with children and their ADL, especially if they face activity limitations, the child's individual life was not investigated in this thesis either. Therefore the potential influence of caregivers and/or contextual factors (e.g. siblings, teachers, therapists, and friends) on the child's characteristics was not examined, which must be considered a limitation of the present thesis in relation to study II-IV. While everyday life apparently is individual and might not be standardized in general, there are other situations where the need for measuring is both relevant and necessary.

The fact that the four studies included review committee members (Study I) and participants (children) (Study II-IV) from three German-speaking countries can be interpreted as strength and at the same time as a limitation. Owing to the design of the present thesis, it was possible to perform a DIF study to evaluate item functioning across three German-speaking countries and also to achieve a certain cost-effectiveness in this process because it was possible to stage a concerted effort to translate and adapt a version in German for three countries at the same time; something which is rarely possible. However, the fairly small sample size, especially for Study IV, where group comparisons of three countries were used could be seen as a limitation, and this needs to be considered in the interpretation of the results. A larger sample would have allowed us to conduct more in-depth analysis of the three country subgroups. Although there are no established guidelines on the sample size required for DIF analyses (the number of respondents will depend on the type of methodological approach used to

evaluate DIF) (Linacre, 2013; Scott et al., 2010; Tristan, 2006) the sample size in this study may be considered as a limitation due to its small number. To generate stable statistical findings, we complemented our statistical analysis by also monitoring the size of the DIF, as has been suggested in the literature (Tennant et al., 2004). As a small sample size could result in underestimation of the DIF, supplementary evaluation of the clinical DIF could therefore compensate for this shortcoming (Tennant & Pallant, 2007; Scott et al.; 2010). This approach confirmed that several, but not all of the DIF noticed statistically were also demonstrated clinically. This analysis also indicated that a higher than expected number of items demonstrates clinical DIF without demonstrating statistical DIF, which confirms the risk of underestimation of statistical DIF with small sample sizes.

Due to the limited sample size, it remains speculative whether the existing DIF by country could also be explained by other variables such as gender, age, disability, or other factors. Future studies using larger samples, especially within different age groups, will likely inform any conclusions concerning the use of the PEDI-G.

Language issues

The use of language is another issue that requires some consideration. This project was performed in three German-speaking countries, and the studies are all here written in English. Furthermore, communication between the supervisors and me was in English, albeit it is nobody's first language. Similarly, the findings have been disseminated in English to ensure international reach and to allow comparison with extant research. However, any rendition of meaning in another language than one's mother tongue in general and any translation in qualitative studies in particular invariably come with some loss of meaning as discussed by Haak, Himmelsbach, Granbom, & Löfqvist (2013) and van Nes, Abma, Jonsson, & Deeg (2010). To reduce this loss of meaning, we followed the recommendations given in the literature and conducted as many meetings as possible as face-to-face meetings, using video calls, mails in every step of the research phase and software like atlas.ti to facilitate the analysis within the native languages for as long as possible (ATLAS.ti, 1999).

CLINICAL IMPLICATIONS

This chapter will illustrate how the PEDI-G may be used in child rehabilitation given that its validity and reliability has been duly established. As validation proceeds and new evidence is produced regarding the interpretations afforded by the assessment scores, continuous revisions of the assessment and its concepts may be needed. Overall, the results of this thesis have implications for health professionals such as occupational therapists, physical therapists, and speech therapists. The PEDI-G with its psychometric properties provides health professionals with a standardized, reliable, and valid assessment for describing and measuring ADL of children with and without activity limitations. Moreover, within the use of the PEDI-G health professionals in the German speaking countries are featured with an assessment that allow an evidence based approach and enable them to evaluate the effectiveness of their interventions. The PEDI-G domains can be used by different raters; for instance, this means that the self-care domain can be appraised by other health professionals than for example the mobility and social function domain. Although some differences across Austria, Germany, and Switzerland exist at item level, this does not impact the generated individual PEDI-G measure of the PEDI-G domains. The domains can still be used for valid and stable assessment.

While the German added items in the PEDI-G seem to be important from a contents validity perspectiv, as some of them also demonstrate misfit and DIF, the added items may not be included in the scoring sum scores, but could be used for additional information. Occupational therapists in German-speaking countries have so far not been regular users of standardized assessments; yet, the findings of this thesis may enhance and support the use of a specific ADL assessment for children in these countries. Following publication of the PEDI-G manual in 2014 (Haley, Coster, Ludlow, Haltiwanger, & Andrellos, 2014), workshops on the administration of the PEDI are regularly being held in all three countries. This is experienced as strength since, on the one hand, this brings the health professionals across the three countries closer together and, on the other hand, the items demonstrating DIF can be studied and their utility discussed. Additionally, the interviewers are trained to score the items correctly, and in those workshops it is possible to emphasize and clarify the construct of the PEDI being measured.

Several studies describe how the PEDI can be used in practice (Dolva et al., 2004; Engelbert et al., 1997; Haley & Fragala-Pinkham, 2006; Kothari, Haley, Gill-Body, & Dumas, 2003; Rich et al., 2013). Overall these studies illustrated that the PEDI may be used for several, different purposes and showed how the discussion of scores inspired a strategy that therapists can use to support collaborative relationships with parents in setting therapy goals that are meaningful, contextually relevant and built on the child's de facto abilities. Moreover the studies demonstrated that completing the PEDI increased parental awareness of their child's progress in relation to typical skill acquisition patterns and level of assistance needed to undertake daily activities. The validation process of PEDI-G evaluated in this thesis can now also support pediatric practitioners in the German-speaking countries to apply and benefit from these findings as well, by using PEDI-G in their clinical practice for multiple purposes.

CONCLUSIONS

Overall, the findings of the four studies provide evidence supporting that the PEDI-G is valid and reliable in clinical practice in Austria, Germany and Switzerland.

- The adaptation of the PEDI for use in a new context within a review committee was a balancing act that involved (i) staying close to the original assessment while adapting the translation to the new context, (ii) making decisions and reaching consensus, (iii) weighing group dynamics against versus the planned agenda of the review committee.
- The PEDI-G demonstrates high levels of precision across raters and across two time points for the six domains; however, the Modification Scale may not be suitable for measuring change.
- The six PEDI-G domains can be used as unidimensional domains to measure ADL of children with and without activity limitations in German-speaking countries. Although some differences exist at item level, this does not impact the generated individual PEDI-G measure of the PEDI-G domains. The domains can still be used as valid tools across the three countries.
- The German added items in the PEDI-G seem to be important from a content validity perspective, but some of them also demonstrate misfit and DIF. Therefore, the added items may not be included in the scoring sum scores, but could be used for additional information.
- The combination of qualitative and quantitative approaches was supportive to demonstrate the complex process and outcomes in validating the PEDI-G for children with and without activity limitations.
- It also remains to further examine the psychometric properties of the PEDI-G within a larger sample and possibly with inclusion of more details about the parents and the caregivers. Furthermore, evidence of the PEDI-G related to fairness in testing and the relationship with the PEDI-G measures and other variables and outcomes (evidence of validity in relation to other variables) is needed to further demonstrate the validity of the PEDI-G in the new context.

FUTURE RESEARCH

This chapter gives recommendations for future research based on the use of the PEDI-G. Although the four studies overall support the use of the PEDI-G in German-speaking countries, the validation process of the PEDI-G across these countries may be further strengthened and more research is needed.

- As the findings from the studies indicate several country-specific DIF items, further exploration of DIF by country within larger datasets is warranted. As the results of study IV also indicate that older children may have more unexpected scores on specific PEDI-G items, DIF by age groups may be another factor that needs to be further explored with larger datasets.
- As the sources of DIF could not be identified, additional studies (qualitative as well as quantitative) must be conducted to enhance the understanding and scoring of potential systematic sources of DIF on single items.
- The results of these studies allow no decision as to whether the normative values provided for the PEDI from the US are applicable or not in German speaking countries. Evidence concerning the applicability of the PEDI norms in other contexts such as the Netherlands and Norway do exist, and this evidence supports that country-specific normative values should be developed. Studies evaluating DIF between German countries and the United States could be a starting point for investigating whether the scales work in the same way between those countries in relation to norms. Further research on norms for children in German-speaking countries may be needed as the norms provided for the US-PEDI may not fit all contexts. Conclusions related to whether children living in German-speaking countries develop ADL earlier or later than children living in other European countries or other parts of the world are still pending.
- For therapists, outcome measures such as the PEDI-G serve to objectively identify areas of difficulty, and they enable therapists to monitor change over time. Sensitivity of the PEDI-G to measure change related to an intervention is another aspect that could be evaluated by

using it in an intervention program as this would procure evidence for the usefulness of the PEDI-G for evaluating change over time.

- Qualitative data investigating the impact of context on ADL across German speaking countries would enhance the understanding of ADL performance of children with and without activity limitations.

ACKNOWLEDGEMENTS

Writing this thesis was for me like learning to cook a sophisticated course to be served at a four-star restaurant. It has been an educational journey and I now know some magic receipts and special spices which will support and encourage me to continue my scientific development and endeavors. I therefore wish to deeply thank all those who have been supporting me and in various ways taught me the ‘tricks of cooking’ and contributed to the present thesis. First and foremost, I wish to extend my warmest gratitude to my three supervisors.

First, *Anders Kottorp*, my main supervisor, my profound gratitude goes to you for giving me the possibility to become a PhD student at Karolinska Institutet and for giving me my first cooking lessons. Thank you for sharing with me your profound knowledge, expertise, and pragmatic reasoning. You guided me through the long process with its ups and downs which I have now come to accept with gratitude. Especially, I want to thank you for the last challenging months and for your never failing confidence in my ability to accomplish the task before me. Thank you for being always available when I needed it and for your patience. Our valuable discussions gave me the possibility to develop and extend my scientific reasoning and to develop the courage to finish this thesis.

I also owe my sincerest thanks to *Julie Page*, my co-supervisor, for her caring and whole-hearted engagement and supervision throughout this process. You helped me embark on this educational journey, letting me know that the PEDI project could develop into a PhD project; and after this you never failed to support and help me, not only in professional matters, but also by showing concern for me as a person. You have played a vital role in the completion of this thesis, both with your expertise and with your supportive coaching. Thank also for being able to find some pockets with money for financing various tasks and for covering travel expenses.

Margareta Lilja, my other co-supervisor, thank you for sharing your enormous experience and knowledge of occupational therapy and qualitative study designs. You have been a tremendous support in guiding me into the world of research. Your commitment to develop occupational therapy practice, and research has been truly inspiring. I will miss our discussion and critical reflections.

I am also grateful to *André Meichtry* who gave me advice and guided me through the world of statistics with so much knowledge and expertise. You have often added additional perspectives to our discussions by posing difficult questions and suggesting statistical alternatives.

Renate Fink, thank you for being my wise mentor and for taking interest in my work. Your humor, support, and critical reflections have been invaluable. It felt so good to be on this journey with you.

My warm thanks also go to the children and their parents for their participation, and to the health professionals (physical and occupational therapists) for collecting the data. Thanks are also due to the following persons who contributed substantially to this thesis: *Martin Häussler, Christoph Künzle, Ruth Amann, Barbara Brändli, and Andrea Espei*.

I also wish to thank my colleagues from the School of Health Professions in Winterthur and in particular my colleagues from Research and Development who in various ways supported me during this time. A special thank you to *Yvonne Treusch*, my research colleague and friend for spending so many hours, weekends and Sundays in the office with me. It was a relief to share the office with you, and thank you for trying to answer all my questions. I also wish to extend my gratitude to *Beate Krieger*. Without your interest, engagement and generous support, this thesis would not have been finished. I also want to thank the following persons and friends from the Institute of Occupational Therapy: *Andrea Citrini-Hunger, Christiane Mentrup, Christine Mosimann, Kim Roos and Uta Jakobs*. Thank you for your interest and company on this long way, you have supported and stimulated me in various ways.

My appreciation goes to my PhD colleagues at Karolinska Institutet. Although I was not a part of the research school, you always welcomed me warmly. Thank you, *Louise Nygard*, for Kreativa Konditoriet. It has been wonderful to be part of the group, and you have been the skilled leader of inspiring scientific discussions. Thank you for inviting me.

Special thanks to *Thomas Baumann* from the ZKSK, my working place in Solothurn. Thank you for your believe in me and for our valuable discussions during the last years. *Christine Bigolin* chief of the ZKSK for your generosity and support on a daily basis and to all my colleagues, *Andrea*, *Franziska*, and *Sabina* in particular, thanks a lot for taking interest in my work.

A warm thanks to *Katrin Grüntzig* for her skillful painting of the front cover. I really feel privileged to have this painting from you and appreciate it so much.

Finally my family and friends are profoundly thanked for giving me the time and space for working during visits and for supporting me where possible.

This work was supported by grants from the Foundation of Cerebral Palsy and the Commission for Innovation (KTI) in Switzerland.

REFERENCES

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. Washington, DC USA: American Educational Research Association.
- AOTA. (2014). Occupational Therapy Practice Framework (OTPF): Domain & Process, 3rd Edition. *American Journal of Occupational Therapy*, 1-48.
- Árnadóttir, G. (2010). *Measuring the Impact of Body Functions on Occupational Performance: Validation of the ADL-focused Occupation-based Neurobehavioral Evaluation (A-ONE)*. (Doctoral dissertation), Umeå university. Retrieved from <http://www.diva-portal.org/smash/get/diva2:300916/FULLTEXT01.pdf>.
- ATLAS.ti (Version 4.2.). (1999). Berlin: Scientific Software Development.
- Beaton, D., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2002). Recommendations for the Cross-Cultural Adaptation of Health Status Measures. from <http://www.dash.iwh.on.ca/assets/images/pdfs/xculture2002.pdf>.
- Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, 25(24), 3186-3191.
- Berg, M. (2008). *Norwegian validation of the Pediatric Evaluation of Disability Inventory (PEDI)*. (Doctoral dissertation), Oslo, Norway. Retrieved from <https://www.duo.uio.no/bitstream/handle/10852/29998/Berg-publ-avh.pdf?sequence=2> .
- Berg, M., Aamodt, G., Stanghelle, J., Krumlinde-Sundholm, L., & Hussain, A. (2008). Cross-cultural validation of the Pediatric Evaluation of Disability Inventory (PEDI) norms in a randomized Norwegian population. *Scandinavian Journal of Occupational Therapy*, 3(1), 1 - 10. doi:10.1080/11038120802022011
- Berg, M., Fräslie, K. F., & Hussain, A. (2003). Applicability of Pediatric Evaluation of Disability Inventory in Norway. *Scandinavian Journal of Occupational Therapy*, 10(3), 118 - 126. doi: 10.1080/11038120310013330
- Berg, M., Jahnsen, R., Frosli, K. F., & Hussain, A. (2004). Reliability of the Pediatric Evaluation of Disability Inventory (PEDI). *Physical and Occupational Therapy in Pediatrics*, 24(3), 61-77. doi: 10.1300/J006v24n03_05
- Bonsaksen, T., Kottorp, A., Gay, C., Fagermoen, M. S., & Lerdal, A. (2013). Rasch analysis of the General Self-Efficacy Scale in a sample of persons with morbid obesity. *Health and Quality of Life Outcomes*, 11, 202. doi: 10.1186/1477-7525-11-202
- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (Third ed.). New York: Routledge.
- Bonder, B. R., Martin, L., & Miracle, A. W. (2004). Culture Emergent in Occupation. *American Journal of Occupational Therapy*, 58(2), 159-168. doi:10.5014/ajot.58.2.159

- Bourke-Taylor, H., & Hudson, D. (2005). Cultural differences: The experience of establishing an occupational therapy service in a developing community. *Australian Occupational Therapy Journal*, 52(3), 188-198. doi: 10.1111/j.1440-1630.2005.00493.x
- Brislin, R. W. (1970). Back-Translation for Cross-Cultural Research. *Journal of Cross-Cultural Psychology*, 1(3), 185-216. doi: 10.1177/135910457000100301.
- Bronfenbrenner, U. (1979). *The ecology of human development*. Cambridge, MA: Harvard University Press.
- Brown, L., Schwarz, P., Udvari-Solner, A., Kampschroer, E. F., Johnson, F., Jorgensen, J., & Gruenewald, L. (1991). How Much Time Should Students with Severe Intellectual Disabilities Spend in Regular Education Classrooms and Elsewhere? *Research and Practice for Persons with Severe Disabilities*, 16(1), 39-47. doi: 10.1177/154079699101600105.
- Bundy, A. C. (1995). Assessment and Intervention in School-Based Practice. *Physical & Occupational Therapy in Pediatrics*, 15(2), 69-88. doi:10.1080/J006v15n02_05.
- Cairney, J., Hay, J. A., Faught, B. E., Wade, T. J., Corna, L., & Flouris, A. (2005). Developmental coordination disorder, generalized self-efficacy toward physical activity, and participation in organized and free play activities. *The Journal of Pediatrics*, 147(4), 515-520. doi: 10.1016/j.jpeds.2005.05.013.
- Case-Smith, J., & O'Brien, J. (2014). *Occupational Therapy for Children and Adolescents* (7th Edition ed.). St Louis, USA: Mosby.
- Chen, K.-L., Hsieh, C.-L., Sheu, C.-F., Hu, F.-C., & Tseng, M.-H. (2009). Reliability and validity of a Chinese version of the pediatric evaluation of disability inventory in children with cerebral palsy. *Journal of Rehabilitation Medicine*, 41(4), 273-278. doi: 10.2340/16501977-0319.
- Chen, K.-L., Tseng, M.-H., Hu, F.-C., & Koh, C.-L. (2010). Pediatric Evaluation of Disability Inventory: A cross-cultural comparison of daily function between Taiwanese and American children. *Research in Developmental Disabilities*, 31(6), 1590-1600. doi: 10.1016/j.ridd.2010.05.002.
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*.
from <http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=yrovfta&AN=00006823-196810000-00001>.
- Cordier, R., Bundy, A., Hocking, C., & Einfeld, S. (2009). A model for play-based intervention for children with ADHD. *Australian Occupational Therapy Journal*, 56(5), 332-340. doi: 10.1111/j.1440-1630.2009.00796.x.
- Costa, U., & Schönthaler, E. (2011). Pediatric Occupational Therapy in Austria. *Journal of Occupational Therapy, Schools, & Early Intervention*, 4(3-4), 176-185. doi: 10.1080/19411243.2011.629547.
- Coster, W. J. (1998). Occupation-Centered Assessment of Children. *American Journal of Occupational Therapy*, 52(5), 337-344. doi: 10.5014/ajot.52.5.337.
- Coster, W. J. (2006). Evaluating the Use of Assessments in Practice and Research. In G. Kielhofner (Ed.), *Research in Occupational Therapy* (pp. 201-212). Philadelphia: Davies Company.

- Cotter, E., Burgio, L., Stevens, A., Roth, D., & Gitlin, L. (2002). Correspondence of the functional independence measure (FIM) self-care subscale with real-time observations of dementia patients' ADL performance in the home. *Clinical Rehabilitation*, 16(1), 36 - 45.
- Custers, J., Hoijsink, H., van der Net, J., & Helders, P. (2000). Cultural differences in functional status measurement: Analyses of person fit according to the Rasch model *Quality of Life Research*, 9(5), 571-578.
- Custers, J. W. H., Wassenberg-Severijnen, J. E., Hoijsink, H., Vermeer, A., & Helders, P. J. M. (2003). Dutch adaptation of the Pediatric Evaluation of Disability Inventory (PEDI). *Nederlands Tijdschrift Voor Fysiotherapie*, 33-36.
- Dallmeijer, A., Dekker, J., Roorda, L., Knol, D., van Baalen, B., de Groot, V., . . . Lankhorst, G. (2005). Differential item functioning of the functional independence measure in higher performing neurological patients. *Journal of Rehabilitation Medicine*, 37(6), 346 - 352. doi: 10.1080/16501970510038284
- Darragh, A. R., Sample, P. L., & Fisher, A. G. (1998). Environment effect on functional task performance in adults with acquired brain injuries: Use of the assessment of motor and process skills. *Archives of Physical Medicine and Rehabilitation*, 79(4), 418-423. doi: 10.1016/S0003-9993(98)90143-5.
- Debusse, D., & Brace, H. (2011). Outcome Measures of Activity for Children With Cerebral Palsy: A Systematic Review. *Pediatric Physical Therapy*, 23(3), 221-231. doi:10.1097/PEP.0b013e318227bbc6
- Dekker, J., Dallmeijer, A. J., & Lankhorst, G. J. (2005). Clinimetrics in rehabilitation medicine: current issues in developing and applying measurement instruments *Journal of Rehabilitation Medicine*, 37(4), 193-201. doi: 10.1080/16501970510035179.
- DeVellis, R. F. (2006). Classical Test Theory. *Medical Care*, 44(11), p50-59. doi: 10.1097/01.mlr.0000245426.10853.30.
- Dickson, H., & Kohler, F. (1995). The functional independence measure: A comparative validity and reliability study. *Disability and Rehabilitation*, 17(8), 456.
- Dolva, A.-S., Coster, W., & Lilja, M. (2004). Functional Performance in Children With Down Syndrome. *The American Journal of Occupational Therapy*, 58(6), 621-629. doi: 10.5014/ajot.58.6.621.
- Dorans, N. J., & Kulick, E. (2006). Differential Item Functioning on the Mini-Mental State Examination: An Application of the Mantel-Haenszel and Standardization Procedures. *Medical Care*, 44(11), S107-S114. doi: 10.1097/01.mlr.0000245182.36914.4a.
- Dorer, B. (2012). ESS Round 6 Translation Guidelines. from http://www.europeansocialsurvey.org/docs/round6/methods/ESS6_translation_guidelines.pdf.
- Dumas, H. M., Fragala-Pinkham, M. A., Feng, T., & Haley, S. M. (2012). A preliminary evaluation of the PEDI-CAT Mobility item bank for children using walking aids and wheelchairs. *Journal of Pediatric Rehabilitation Medicine*, 5(1), 29-35. doi: 10.3233/prm-2011-0184.

- Dumas, H. M., Fragala-Pinkham, M. A., Haley, S. M., Ni, P., Coster, W., Kramer, J. M., . . . Ludlow, L. H. (2012). Computer adaptive test performance in children with and without disabilities: prospective field study of the PEDI-CAT. *Disability & Rehabilitation, 34*(5), 393-401. doi: 10.3109/09638288.2011.607217.
- Duncan, B., Shen, K., Zou, L.-P., Han, T.-L., Lu, Z.-L., Zheng, H., . . . Caspi, O. (2012). Evaluating Intense Rehabilitative Therapies With and Without Acupuncture for Children With Cerebral Palsy: A Randomized Controlled Trial. *Archives of Physical Medicine and Rehabilitation, 93*(5), 808-815. doi: 10.1016/j.apmr.2011.12.009.
- Dunn, L., Magalhaes, L. C., & Mancini, M. C. (2014). Internal structure of the Children Helping Out: Responsibilities, Expectations, and Supports (CHORES) measure. *American Journal of Occupational Therapy, 68*(3), 286-295. doi: 10.5014/ajot.2014.010454.
- Dunn, W. (1993). Measurement of Function: Actions for the Future. *American Journal of Occupational Therapy, 47*(4), 357-359. doi: 10.5014/ajot.47.4.357.
- Dunn, W., Brown, C., & McGuigan, A. (1994). The Ecology of Human Performance: A Framework for Considering the Effect of Context. *American Journal of Occupational Therapy, 48*(7), 595-607. doi: 10.5014/ajot.48.7.595.
- Eakin, P. (1989). Assessments of Activities of Daily Living: A Critical Review. *The British Journal of Occupational Therapy, 52*(1), 11-15. doi: 10.1177/030802268905200104.
- Elad, D., Barak, S., Eisenstein, E., Bar, O., Herzberg, O., & Brezner, A. (2012). Reliability and validity of Hebrew Pediatric Evaluation of Disability Inventory (PEDI) in children with cerebral palsy -- health care professionals vs. mothers. *Journal of Pediatric Rehabilitation Medicine, 5*(2), 107-115. doi: 10.3233/PRM-2012-0202.
- Engelbert, R. H. H., Custers, J. W. H., van der Net, J., van der Graaf, Y., Beemer, F. A., & Helders, P. J. M. (1997). Functional Outcome in Osteogenesis Imperfecta: Disability Profiles Using the PEDI. *Pediatric Physical Therapy, 9*(1), 18-22. doi: 10.1542/peds.99.2.e3
- Erkin, G., Elhan, A., Aybay, C., Sirzai, H., & Ozel, S. (2007). Validity and reliability of the Turkish translation of the Pediatric Evaluation of Disability Inventory (PEDI). *Disability & Rehabilitation, 29*(16), 1271-1279. doi:10.1080/09638280600964307
- Fawcett, A. L. (2007). *Principles of Assessment and Outcome Measurement for Occupational Therapists and Physiotherapists: Theory, Skills and Application*. New York: Wiley.
- Feldman, A. B., Haley, S. M., & Coryell, J. (1990). Concurrent and construct validity of the Pediatric Evaluation of Disability Inventory. *Physical Therapy, 70*(10), 602-610.
- Fisher, A. G. (1998). Uniting practice and theory in an occupational framework. 1998 Eleanor Clarke Slagle Lecture. *American Journal of Occupational Therapy, 52*(7), 509-521. doi:10.5014/ajot.52.7.509

- Fisher, A. G. (2009). *Occupational Therapy Intervention Process Model: A Model for Planning and Implementing Top-down, Client-centered, and Occupation-based Interventions*: Three Star Press Incorporated.
- Fisher, A. G. (2012). *Assessment of Motor and Process Skills: User Manual* (seventh Ed.). Ft. Collins: Three Star Press.
- Fisher, J. W. P. (1993). Measurement-Related Problems in Functional Assessment. *American Journal of Occupational Therapy*, 47(4), 331-338. doi: 10.5014/ajot.47.4.331.
- Flaherty, J. A., Gaviria, F. M., Pathak, D., Mitchell, T., Wintrob, R., Richman, J. A., & Birz, S. (1988). Developing instruments for cross-cultural psychiatric research. *Journal of Nervous & Mental Disease*, 176(5), 257-263. doi:10.1097/00005053-198805000-00001
- Gannotti, M. E. (1998). *The validity and reliability of the Pediatric Evaluation of Disability Inventory for children living in Puerto Rico*. (Doctorial dissertation), University of Connecticut. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=cin20&AN=2005043419&site=ehost-live>.
- Gannotti, M. E., & Cruz, C. (2001). Content and construct validity of a Spanish translation of the Pediatric Evaluation of Disability Inventory for children living in Puerto Rico. *Physical and Occupational Therapy in Pediatrics*, 20(4), 7-24. doi: 10.1080/J006v20n04_02
- Gannotti, M. E., & Handwerker, W. P. (2002). Puerto Rican understandings of child disability: methods for the cultural validation of standardized measures of child health. *Social Science & Medicine*, 55(12), 2093-2105. doi:10.1016/S0277-9536(01)00354-9.
- Gantschnig, B. E. (2014). *Occupation-based and Occupation-focused Evaluation and Intervention with Children: A Validation Study of the Assessment of Motor and Process Skills (AMPS)*. (Doctoral dissertation), Umea, Sweden. Retrieved from <http://www.diva-portal.org/smash/get/diva2:679436/FULLTEXT02.pdf>.
- Gantschnig, B. E., Page, J., & Fisher, A. G. (2012). Cross-regional validity of the assessment of motor and process skills for use in middle Europe. *Journal of Rehabilitation Medicine*, 44(2), 151-157. doi: 10.2340/16501977-0915
- Geisinger, K. F. (1994). Cross-Cultural Normative Assessment: Translation and Adaptation Issues Influencing the Normative Interpretation of Assessment Instruments. *Psychological Assessment*, 6(4), 304-312. doi:10.1037/1040-3590.6.4.304
- Goldman, S. L., & Fisher, A. G. (1997). Cross-Cultural Validation of the Assessment of Motor and Process Skills (AMPS). *The British Journal of Occupational Therapy*, 60(2), 77-85. doi: 10.1177/030802269706000210.
- Graneheim, U. H., & Lundman, B. (2004). Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse Education Today*, 24(2), 105-112. doi: 10.1016/j.nedt.2003.10.001.

- Grimby, G., Ekholm, J., Fisher, A. G., & Sunnerhagen, K. S. (2005). Measurement and evaluation of outcomes in rehabilitation. *Journal Rehabilitation Medicine*, 37, 1-2. doi:10.1080/16501970510027060
- Groleger, K., Vidmar, G., & Zupan, A. (2005). Applicability of the Pediatric Evaluation of Disability Inventory in Slovenia. *Journal of Child Neurology*, 20(5), 411-416. doi: 10.1177/08830738050200050301
- Guillemin, F., Bombardier, C., & Beaton, D. (1993). Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *Journal of Clinical Epidemiology*, 46(12), 1417-1432. doi:10.1016/0895-4356(93)90142-N
- Haak, M., Himmelsbach, I., Granbom, M., & Löfqvist, C. (2013). Cross-National and Multi-Language Qualitative Research: Challenges and Recommendations. *The British Journal of Occupational Therapy*, 76(7), 333-336. doi: 10.4276/030802213x13729279115059.
- Haley, S. M., Coster, W. J., & Binda-Sundberg, K. (1994). Measuring physical disablement: the contextual challenge. *Physical Therapy*, 74(5), 443-451.
- Haley, S. M., Coster, W. J., Dumas, H. M., Fragala-Pinkham, M. A., Kramer, J., Ni, P., . . . Ludlow, L. H. (2011). Accuracy and precision of the Pediatric Evaluation of Disability Inventory computer-adaptive tests (PEDI-CAT). *Developmental Medicine & Child Neurology*, 53(12), 1100-1106. doi: 10.1111/j.1469-8749.2011.04107.x.
- Haley, S. M., Coster, W. J., Ludlow, L. H., Haltiwanger, J. T., & Andrellos, P. J. (1992). *Pediatric Evaluation of Disability Inventory: Development, standardization, and administration manual*. Boston, MA: New England Medical Centre Inc. and PEDI Research Group.
- Haley, S. M., Coster, W. J., Ludlow, L. H., Haltiwanger, J., & Andrellos, P. (2014). *PEDI-D Pediatric Evaluation of Disability Inventory: Assessment zur Erfassung von Aktivitäten des täglichen Lebens bei Kindern mit und ohne Beeinträchtigung*: Schulz-Kirchner Verlag GmbH.
- Haley, S. M., & Fragala-Pinkham, M. A. (2006). Interpreting Change Scores of Tests and Measures Used in Physical Therapy. *Physical Therapy*, 86(5), 735-743.
- Hambleton, R. K. (2001). The next generation of the ITC test translation and adaptation guidelines. *European Journal of Psychological Assessment*, 17(3), 164-172.
- Hambleton, R. K. (2006). Good Practices for Identifying Differential Item Functioning. *Medical Care*, 44(11), S182-S188. doi: 10.1097/01.mlr.0000245443.86671.c4.
- Hammell, K. W. (2001). Using Qualitative Research to Inform the Client-Centred Evidence-Based Practice of Occupational Therapy. *The British Journal of Occupational Therapy*, 64, 228-234. doi: 10.1177/030802260106400504
- Harvey, A., Robin, J., Morris, M. E., Graham, H. K., & Baker, R. (2008). A systematic review of measures of activity limitation for children with cerebral palsy. *Developmental Medicine & Child Neurology*, 50(3), 190-198. doi: 10.1111/j.1469-8749.2008.02027.x.

- Herdman, M., Fox-Rushby, J., & Badia, X. (1997). "Equivalence" and the translation and adaptation of health-related quality of life questionnaires. *Quality of Life Research*, 6, 237-247. doi: 10.1023/A:1026410721664
- Hocking, C. (2001). Implementing occupation-based assessment. *The American Journal of Occupational Therapy: official publication of the American Occupational Therapy Association*, 55(4), 463-469.
- Hollenweger, J., & Kraus, O. E. (2011). *ICF-CY: Internationale Klassifikation der Funktionsfähigkeit, Behinderung und Gesundheit bei Kindern und Jugendlichen*. Bern, Switzerland: Hans Huber Verlag Bern.
- Holsbeeke, L., Ketelaar, M., Schoemaker, M. M., & Gorter, J. W. (2009). Capacity, Capability, and Performance: Different Constructs or Three of a Kind? *Archives of Physical Medicine and Rehabilitation*, 90(5), 849-855. doi: org/10.1016/j.apmr.2008.11.015.
- Jasmin, E., Couture, M., McKinley, P., Reid, G., Fombonne, E., & Gisel, E. (2009). Sensori-motor and daily living skills of preschool children with autism spectrum disorders. *Journal of Autism and Developmental Disorder*, 39(2), 231-241. doi: 10.1007/s10803-008-0617-z.
- James, S., Ziviani, J., & Boyd, R. (2014). A systematic review of activities of daily living measures for children and adolescents with cerebral palsy. *Developmental Medicine & Child Neurology*, 56(3), 233-244. doi: 10.1111/dmcn.12226.
- Jerosch-Herold, C. (2005). An Evidence-Based Approach to Choosing Outcome Measures: a Checklist for the Critical Appraisal of Validity, Reliability and Responsiveness Studies. *British Journal of Occupational Therapy*, 68(8), 347-353. doi: 10.1177/030802260506800803
- Jette, A. M., & Haley, S. M. (2005). Contemporary measurement techniques for rehabilitation outcomes assessment. *Journal of Rehabilitation Medicine*, 37(6), 339-345. doi: 10.1080/16501970500302793.
- Kaiser, H. (1970). A second generation little jiffy. *Psychometrika*, 35(4), 401-415. doi: 10.1007/BF02291817.
- Ketelaar, M., Vermeer, A., & Helders, P. J. (1998). Functional motor abilities of children with cerebral palsy: a systematic literature review of assessment measures. *Clinical Rehabilitation*, 12(5), 369-380. doi: 10.1191/026921598673571117.
- Ketelaar, M., & Wassenberg-Severijnen, J. (2010). Developments in Measuring Functional Activities: Where Do We Go with the PEDI-CAT? *Physical & Occupational Therapy in Pediatrics*, 30(3), 185-189. doi: 10.3109/01942638.2010.486352.
- Kielhofner, G. (2006). *Research in occupational therapy: methods of inquiry for enhancing practice* Philadelphia F.A. Davis.
- Kielhofner, G. (2008). *Model of Human Occupation: Theory and Application*. Baltimore: Lippincott Williams & Wilkins.
- Knox, V., & Usen, Y. (2000). Clinical Review of the Pediatric Evaluation of Disability Inventory. *The British Journal of Occupational Therapy*, 63, 29-32. doi: 10.1177/030802260006300106

- Kothari, D. H., Haley, S. M., Gill-Body, K. M., & Dumas, H. M. (2003). Measuring functional change in children with acquired brain injury: Comparison of generic and ABI-specific scales using the PEDI. *Physical Therapy*, 83, 776-785.
- Kottorp, A. (2003). *Occupation-based evaluation and intervention. Validity of the Assessment of Motor and Process Skills when used with persons with mental retardation*. (Doctoral dissertation), Umea, Sweden. Retrieved from <http://www.diva-portal.org/smash/get/diva2:763461/FULLTEXT01.pdf>.
- Kramer, J. M., Coster, W. J., Kao, Y.-C., Snow, A., & Orsmond, G. I. (2012). A new approach to the measurement of adaptive behavior: Development of the PEDI-CAT for children and youth with autism spectrum disorders. *Physical & Occupational Therapy in Pediatrics*, 32(1), 34-47. doi: 10.3109/01942638.2011.606260.
- Krueger, R. A., & Casey, M. A. (2009). *Focus groups: A practical guide for applied research*. Thousand Oaks: Sage.
- Landis, R., & Koch, G. (1977). The Measurement of observer Agreement for Categorical Data. *International Biometrics Society*, 33(1), 159-174.
- Law, M. (2003). Outcome Measurement in Pediatric Rehabilitation. *Physical & Occupational Therapy in Pediatrics*, 23(2), 1-4. doi: doi:10.1080/J006v23n02_01.
- Law, M., Baum, C., & Dunn, W. (Eds.). (2005). *Measuring Occupational Performance: Supporting Best Practice in Occupational Therapy*. Thorofare: Slack Incorporated.
- Law, M., Cooper, B. A., Strong, S., Stewart, D., Rigby, P., & Letts, L. (1996). The Person-Environment-Occupational Model: a transactive approach to occupational performance. *Canadian Journal of Occupational Therapy*, 63(1), 9-23. doi:10.1177/000841749606300103
- Law, M., & Letts, L. (1989). A Critical Review of Scales of Activities of Daily Living. *American Journal of Occupational Therapy*, 43(8), 522-528. doi: 10.5014/ajot.43.8.522.
- Law, M., Polataiko, H., Baptiste, W., & Townsend, E. (1997). *Enabling occupation: An occupational therapy perspective*. Ottawa: Canadian Association of Occupational Therapists.
- Law, M. C., Darrah, J., Pollock, N., Wilson, B., Russell, D. J., Walter, S. D., . . . Galuppi, B. (2011). Focus on function: a cluster, randomized controlled trial comparing child- versus context-focused intervention for young children with cerebral palsy. *Developmental Medicine & Child Neurology*, 53(7), 621-629. doi: 10.1111/j.1469-8749.2011.03962.x.
- Lee, C.C., Li, D., Arai, S., & Puntillo, K. (2009). Ensuring Cross-Cultural Equivalence in Translation of Research Consents and Clinical Documents. *Journal of Transcultural Nursing*, 20(1), 77-82. doi: 10.1177/1043659608325852.
- Lerdal, A., Fagermoen, M., Bonsaksen, T., Gay, C., & Kottorp, A. (2014). Rasch analysis of the sense of coherence scale in a sample of people with morbid

- obesity – a cross-sectional study. *BioMedCentral Psychology*, 2(1), 1-10. doi: 10.1186/2050-7283-2-1
- Linacre, J. (2013). Differential Item Functioning DIF Sample Size Nomogram. *Rasch Measurement Transactions*, 26(4), 131.
- Linacre, J. (2014a). Winsteps Rasch measurement computer program. Beaverton, Oregon: Winsteps.com.
- Linacre, J. (2014b). *Winsteps Rasch measurement computer program User`s Guide*. Beaverton, Oregon: Winsteps.com.
- Linacre, J., Heinemann, A., Wright, B., Granger, C., & Hamilton, B. (1994). The structure and stability of the functional independence measure. *Archives of physical medicine and rehabilitation*, 75(2), 127 - 132.
- Linderman, T. M., & Stewart, K. B. (1999). Sensory integrative-based occupational therapy and functional outcomes in young children with pervasive developmental disorders: a single-subject study. *American Journal Occupational Therapy*, 53(2), 207-213. doi: 10.5014/ajot.53.2.207
- Lollar, D. J., & Simeonsson, R. J. (2005). Diagnosis to function: classification for children and youths. *Journal of Developmental and Behavioral Pediatrics*, 26(4), 323-330.
- Lumsden, J. (1957). A factorial approach to unidimensionality. *Australian Journal of Psychology*, 9(2), 105-111. doi: 10.1080/00049535708256147.
- Lundgren-Nilsson, A., Grimby, G., Ring, H., Tesio, L., Lawton, G., Slade, A., . . . Tennant, A. (2005). Cross-cultural validity of functional independence measure items in stroke: a study using Rasch analysis. *Journal Rehabilitation Medicine*, 37(1), 23-31. doi: 10.1080/16501970410032696.
- Lynch, E. W., & Hanson, M. J. (2011). *Developing Cross-cultural Competence: A Guide for Working with Children and Their Families*. Baltimore: Paul H. Brookes Pub.
- Magalhães, L. C., Fisher, A. G., Bernspång, B., & Linacre, J. M. (1996). Cross-Cultural Assessment of Functional Ability. *OTJR: Occupation, Participation and Health*, 16(1), 45-63. doi: 10.1177/153944929601600103.
- Majnemer, A. (2010). Benefits of Using Outcome Measures in Pediatric Rehabilitation. *Physical & Occupational Therapy in Pediatrics*, 30(3), 165-167. doi: 10.3109/01942638.2010.484353.
- Malheiros, S., Monteiro, C. B. d. M., da Silva, T. D., Torriani-Pasin, C., de Andrade, M., Valenti, V., . . . de Abreu, L. C. (2013). Functional capacity and assistance from the caregiver during daily activities in Brazilian children with cerebral palsy. *International Archives of Medicine*, 6(1), 1. doi: 10.1186/1755-7682-6-1
- Malinowsky, C., Kottorp, A., Tanemura, R., Asaba, E., Nagao, T., Noda, K., . . . Nygard, L. (2015). *Validation of the Everyday Technology Use Questionnaire in a Japanese context*. Hong Kong Journal of Occupational Therapy, (accepted for publication).
- Mancini, M. (2005). Inventory Pediatric Evaluation of Disability (PEDI): Brazilian adapted version of the manual. *Belo Horizonte UFMG*, 6, 55 - 133.

- Mandich, A. D., Polatajko, H. J., & Rodger, S. (2003). Rites of passage: understanding participation of children with developmental coordination disorder. *Human Movement Science*, 22(4-5), 583-595. doi:10.1016/j.humov.2003.09.011
- McCarthy, M. L., Silberstein, C., Atkins, E., Harryman, S., Sponseller, P., & Hadley-Miller, N. (2002). Comparing reliability and validity of pediatric instruments for measuring health and well-being of children with spastic cerebral palsy. *Developmental Medicine & Child Neurology*, 44, 468 - 476. doi:10.1111/j.1469-8749.2009.03382.x
- Merbitz, C., Morris, J., & Grip, J. C. (1989). Ordinal scales and foundations of misinference. *Archives of Physical Medicine & Rehabilitation*, 70, 308-312.
- Munkholm, M., Berg, B., Löfgren, B., & Fisher, A. G. (2010). Cross-Regional Validation of the School Version of the Assessment of Motor and Process Skills. *American Journal of Occupational Therapy*, 64(5), 768-775. doi: 10.5014/ajot.2010.09041.
- Nichols, D. S., & Case-Smith, J. (1996). Reliability and validity of the Pediatric Evaluation of Disability Inventory. *Pediatrics of the American Physical Therapy Association*, 8, 15-24.
- Nilsson, I., & Fisher, A. G. (2006). Evaluating leisure activities in the oldest old. *Scandinavian Journal of Occupational Therapy*, 13(1), 31-37. doi: doi:10.1080/11038120500360655.
- Nordmark, E., Orban, K., Hägglund, G., & Jarnlo, G. B. (1999). The American Paediatric Evaluation of Disability Inventory (PEDI). Applicability of PEDI in Sweden for children aged 2.0-6.9 years. *Scandinavian Journal of Rehabilitation Medicine*, 31(2), 95-100.
- Norman, G. R., & Streiner, D. L. (2014). *Biostatistics: The Bare Essentials*. USA: People`s Medical Publishing House.
- Ostensjø S, C. E., Vøllestad NK. (2003). Everyday functioning in young children with cerebral palsy: functional skills, caregiver assistance, and modifications of the environment. *Developmental Medicine and Child Neurology*, 45(9), 603-612. doi: 10.1017/S0012162203001105
- Ostensjø, S., Bjorbaekmo, W., Carlsberg, E. B., & Vollestad, N. (2006). Assessment of everyday functioning in young children with disabilities: An ICF-based analysis of concepts and content of the Pediatric Evaluation of Disability Inventory (PEDI). *Disability and Rehabilitation*, 28:8, , 489 - 504. doi:10.1080/09638280500212013
- Osterlind, S. J., & Everson, H. T. (2009). *Differential item functioning* (second ed.). Los Angeles: Sage Publications, Inc.
- Ottenbacher, K. J., Msall, M. E., Yon, N., Duffy, L. C., Granger, C. V., & Braun, S. (1999). Measuring developmental and functional status in children with disabilities. *Developmental Medicine and Child Neurology*, 41, 186- 194. doi: 10.1111/j.1469-8749.1999.tb00578.x
- Pae, T.-I., & Park, G.-P. (2006). Examining the relationship between differential item functioning and differential test functioning. *Language Testing*, 23(4), 475-496. doi: 10.1191/0265532206lt338oa.

- Page, J., Roos, K., Bänziger, A., Margot-Cattin, I., Agustoni, S., Rossini, E., . . . Meyer, S. (2015). Formulating goals in occupational therapy: State of the art in Switzerland. *Scandinavian Journal of Occupational Therapy*, early online, posted on 19 June, 1-13. doi:10.3109/11038128.2015.1049548.
- Palisano, R. J., Rosenbaum, P., Bartlett, D., & Livingston, M. H. (2008). Content validity of the expanded and revised Gross Motor Function Classification System. *Developmental Medicine & Child Neurology*, 50(10), 744-750. doi: 10.1111/j.1469-8749.2008.03089.x
- Pallant, J. F., & Tennant, A. (2007). An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). *British Journal of Clinical Psychology*, 46(1), 1-18. doi: 10.1348/014466506X96931.
- Patomella, A.-H., Tham, K., Johansson, K., & Kottorp, A. (2010). P-drive on-road: internal scale validity and reliability of an assessment of on-road driving performance in people with neurological disorders. *Scandinavian Journal of Occupational Therapy*, 17(1), 86-93. doi: 10.1080/11038120903071776
- Patomella, A.-H., Tham, K., & Kottorp, A. (2006). P-drive: assessment of driving performance after stroke. *Journal Rehabilitation Medicine*, 9(5), 273-279. doi: 10.1080/16501970600632594
- Ponte-Allan, M., & Giles, G. M. (1999). Goal setting and functional outcomes in rehabilitation. *American Journal of Occupational Therapy*, 53(6), 646-649. doi:10.5014/ajot.53.6.646
- Prado, M., Magalhães, L., & Wilson, B. (2009). Cross-cultural adaptation of the Developmental Coordination Disorder Questionnaire for brazilian children. *Revista Brasileira de Fisioterapia*, 13, 236-243.
- Projektgruppe ergotherapeutische Befundinstrumente in der Pädiatrie, D. (Ed.). (2004). *Befundinstrumente in der pädiatrischen Ergotherapie*. Idstein: Schulz-Kirchner Verlag.
- R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Ramirez, M., Teresi, J. A., Holmes, D., Gurland, B., & Lantigua, R. (2006). Differential Item Functioning (DIF) and the Mini-Mental State Examination (MMSE): Overview, Sample, and Issues of Translation. *Medical Care*, 44(11), 95-106. doi: 10.1097/01.mlr.0000245181.96133.db.
- Ravaud, J., Delcey, M., & Yelnik, A. (1999). Construct validity of the functional independence measure (FIM): Questioning the unidimensionality of the scale and the "value" of FIM scores. *Scandinavian Journal of Rehabilitation Medicine*, 31(1), 31 - 41. doi: 10.1080/16501977313141
- Reid, D. T., Boschen, K., & Wright, V. (1994). Critique of the Pediatric Evaluation of Disability Inventory (PEDI). *Physical & Occupational Therapy in Pediatrics*, 13(4), 57-87. doi:10.1080/J006v13n04_04
- Rennie, D. J., Attfield, S. F., Morton, R. E., Polak, F. J., & Nicholson, J. (2000). An evaluation of lycra garments in the lower limb using 3-D gait analysis and

- functional assessment (PEDI). *Gait & Posture*, 12(1), 1-6. doi:10.1016/S0966-6362(00)00066-7.
- Rich, D., Rigby, P., & Wright, V. (2013). Mothers' Experiences with the Pediatric Evaluation of Disability Inventory (PEDI). *Physical & Occupational Therapy in Pediatrics*. doi: 10.3109/01942638.2013.823473.
- Rodger, S., & Ziviani, J. (2006). *Occupational Therapy with Children: Understanding Children's Occupations and Enabling Participation*. Oxford, United Kingdom: Wiley - Blackwell.
- Rogers, C., Gwinn, S., & Holm, M. (2001). Comparing activities of daily living assessment instruments: FIMTM, MDS, OASIS, MDS-PAC. *Physical & Occupational Therapy in Geriatrics*, 18(3), 1 - 25. doi:10.1080/J148v18n03_01
- Romain, E. (2004). Ist die Verbesserung von Feinmotorik ein ergotherapeutisches Therapieziel? (Is improvement of fine motor skills a therapeutic goal in occupational therapy?). *Ergotherapie & Rehabilitation*, 43, 5-13.
- Schädler, S., Kool, J., Lüthi, H., Marks, D., Oesch, P., Pfeffer, A., & Wirz, M. (Eds.). (2005). *Assessments in der Neurorehabilitation*. Bern: Hans Huber.
- Schum, T. R., Kolb, T. M., McAuliffe, T. L., Simms, M. D., Underhill, R. L., & Lewis, M. (2002). Sequential Acquisition of Toilet-Training Skills: A Descriptive Study of Gender and Age Differences in Normal Children. *Pediatrics*, 109(3), e48. doi: 10.1542/peds.109.3.
- Scott, N., Fayers, P., Aaronson, N., Bottomley, A., de Graeff, A., Groenvold, M., . . . Group, t. Q. o. L. C.-C. M.-A. (2010). Differential item functioning (DIF) analyses of health-related quality of life instruments using logistic regression. *Health and Quality of Life Outcomes*, 8(1), 81. doi: 10.1186/1477-7525-8-81
- Shepard, K., Procter, S., & Coley, I. L. (1996). Self-care and Adaptations for independent living. In J. Case-Smith (Ed.), *Occupational Therapy for Children* (third ed., pp. 461-503). St. Louis: Mosby.
- Shrout, P. E., & Fleiss, J. L. (1979). *Intraclass correlations: Uses in assessing rater reliability*: *Psychological Bulletin*;86(2):420-428.doi:10.1037/0033-2909.86.2.420
- Simeonsson, R. J., Leonardi, M., Lollar, D., Bjorck-Akesson, E., Hollenweger, J., & Martinuzzi, A. (2003). Applying the International Classification of Functioning, Disability and Health (ICF) to measure childhood disability. *Disability & Rehabilitation*, 25(11-12), 602-610. doi: 10.1080/0963828031000137117.
- Smits, D., Gorter, J., Ketelaar, M., Van Schie, P., Dallmeijer, A., Lindeman, E., & Jongmans, M. (2010). Relationship between gross motor capacity and daily-life mobility in children with cerebral palsy. *Developmental Medicine & Child Neurology*, 52, e60 - e66. doi:10.1111/j.1469-8749.2009.03525.x
- Smyth, M. M., & Anderson, H. I. (2000). Coping with clumsiness in the school playground: Social and physical play in children with coordination impairments. *British Journal of Developmental Psychology*, 18(3), 389-413. doi: 10.1348/026151000165760.
- Stahlhut, M., Christensen, J., & Aadahl, M. (2010). Applicability and Intraresponder Reliability of the Pediatric Evaluation of Disability Inventory in a Random

- Danish Sample. *Pediatric Physical Therapy*, 22(2), 161-169.
doi:10.1097/PEP.0b013e3181dbf965
- Stahlhut, M., Gard, G., Aadahl, M., & Christensen, J. (2011). Discriminative Validity of the Danish Version of the Pediatric Evaluation of Disability Inventory (PEDI). *Physical & Occupational Therapy in Pediatrics*, 31(1), 78-89.
doi:10.3109/01942638.2010.486824.
- Steultjens, E. M. J., Dekker, J., Bouter, L. M., van de Nes, J. C. M., Lambregts, B. L. M., & van den Ende, C. H. M. (2004). Occupational therapy for children with cerebral palsy: a systematic review. *Clinical Rehabilitation*, 18(1), 1-14. doi: 10.1191/0269215504cr697oa
- Streiner, D. (1993). A checklist for evaluating the usefulness of rating scales. *Canadian Journal of Psychiatry*, 38, 140 - 148.
- Streiner, D. L., & Norman, G. R. (2008). *Health Measurement Scales: A Practical Guide to Their Development and Use*: Oxford University Press.
- Summers, J., Larkin, D., & Dewey, D. (2008). Activities of daily living in children with developmental coordination disorder: Dressing, personal hygiene, and eating skills. *Human Movement Science*, 27(2), 215-229. doi: org/10.1016/j.humov.2008.02.002.
- Spector, W. D., Katz, S., Murphy, J. B., & Fulton, J. P. (1987). The hierarchical relationship between activities of daily living and instrumental activities of daily living. *Journal of Chronic Diseases* 40(6), 481-489. doi:10.1016/0021-9681(87)90004-X
- Tennant, A., & Conaghan, P. G. (2007). 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 Care & Research*, 57(8), 1358-1362. doi: 10.1002/art.23108.
- Tennant, A., & Pallant, J. F. (2007). DIF matters: A practical approach to test if Differential Item Functioning makes a difference. *Rasch measurement transactions*, 20(4), 1082-1084.
- Tennant, A., Penta, M., Tesio, L., Grimby, G., JL, T., Slade, A., . . . S, P. (2004). Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the rasch model: the PRO-ESOR project. *Medical Care*, 42(1), 137-148. doi: 10.1097/01.mlr.0000103529.6313277
- Tristan, A. (2006). An Adjustment for Sample Size in DIF Analysis. *Rasch Measurement Transactions*, 20(3), 1070-1071.
- Vacha-Haase, T. (1998). Reliability Generalization: Exploring Variance in Measurement Error Affecting Score Reliability Across Studies. *Educational and Psychological Measurement*, 58(1), 6-20. doi: 10.1177/0013164498058001002.
- van Nes, F., Abma, T., Jonsson, H., & Deeg, D. (2010). Language differences in qualitative research: is meaning lost in translation? *European Journal of Ageing*, 7(4), 313-316. doi: 10.1007/s10433-010-0168-y.

- Vos-Vromans, D. C. W. M., Ketelaar, M., & Gorter, J. W. (2005). Responsiveness of evaluative measures for children with cerebral palsy: The Gross Motor Function Measure and the Pediatric Evaluation of Disability Inventory. *Disability and Rehabilitation*, 27(20), 1245-1252. doi:10.1080/09638280500076178.
- Wassenberg-Severijnen, J. E. (2005). *Pediatric Evaluation of Disability Inventory (PEDI): Calibrating the Dutch version*. (PhD Doctoral thesis), Universiteit Utrecht.
- Wassenberg-Severijnen, J. E., Custers, J. W., Hox, J. J., Vermeer, A., & Helders, P. J. (2003). Reliability of the Dutch Pediatric Evaluation of Disability Inventory (PEDI). *Clinical Rehabilitation*, 17(4), 457-462. doi: 10.1191/0269215503cr634oa.
- Weisner, T. (1984). Ecocultural niches of middle childhood: A cross-cultural perspective. *Panel to review the status of basic research on school-age children. Development during middle childhood: The years from 6 to 12*. (pp. 335-369). New York: National Academic Press.
- WFOT. (2012). Definition of occupation. Retrieved from <http://www.wfot.org/aboutus/aboutoccupationaltherapy/definitionofoccupationaltherapy.aspx>.
- WHO. (2001). *International Classification of Functioning, Disability and Health (ICF)*. Geneva, Switzerland: World Health Organization.
- WHO. (2007). *International Classification of Functioning, Disability, and Health: Children & Youth Version : ICF-CY*: World Health Organization.
- WHO. (2012). Process of translation and adaptation of instruments. Retrieved from http://www.who.int/substance_abuse/research_tools/translation/en/.
- Wilcock, A., & Townsend, E. (2014). Occupational justice. In B. A. Boyt Schell, G. Gillen & M. Scaffa (Eds.), *Willard and Spackman's occupational therapy 12th ed.* (pp. 541-552). Philadelphia: Lippincott Williams & Wilkins.
- Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., & Erikson, P. (2005). Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value in Health*, 8(2), 94-104. doi: 10.1111/j.1524-4733.2005.04054.x.
- Wiley, S., Meinzen-Derr, J., Grether, S., Choo, D. I., & Hughes, M. L. (2012). Longitudinal functional performance among children with cochlear implants and disabilities: a prospective study using the Pediatric Evaluation of Disability Inventory. *International Journal of Pediatric Otorhinolaryngology*, 76(5), 693-697. doi: 10.1016/j.ijporl.2012.02.022.
- Wilson, M. (Ed.). (2005). *Constructing Measures - An Item Response Modeling Approach*. New York: Taylor & Francis Group.
- Worth, M., Darrah, J., Magill-Evans, J., Wiart, L., & Law, M. (2014). Stability of parent report on Mobility and Self-Care item scores of the Pediatric Evaluation of Disability Inventory. *Physical and Occupational Therapy in Pediatrics*, 34(3), 289-299. doi: 10.3109/01942638.2013.823475.

- Wright, & Broschen. (1993). The Pediatric Evaluation of disability inventory: validation of a new functional assessment outcome instrument. *Canadian Journal of Rehabilitation*, 7, 41-42.
- Wright, B., & Linacre, J. (1994). Reasonable mean-square fit values. *Rasch measurement transactions*, 8, 370.
- Wright, B., Linacre, J., Smith, R., Heinemann, A., & Granger, C. (1997). FIM measurement properties and rasch model details. *Scandinavian Journal of Rehabilitation Medicine*, 29(4), 267 - 272.
- Wright, B. D., & Panchapakesan, N. (1969). A procedure for sample-free item analysis. *Educational and Psychological Measurement*, 29, 23-48. doi:10.1177/001316446902900102
- Wright, B. D. M., Geofferey. (1982). *Rating Scale Analysis. Rasch Measurement*. Chicago: MESA Press.
- Young, N. L., Williams, J. I., Yoshida, K. K., Bombardier, C., & Wright, J. G. (1996). The context of measuring disability: Does it matter whether capability or performance is measured? *Journal of Clinical Epidemiology*, 49(10), 1097-1101. doi:10.1016/0895-4356(96)00214-4.
- Ziviani, J., Ottenbacher, K. J., Shepard, K., Foreman, S., Astbury, W., & Ireland, P. (2001). Concurrent validity of the Functional Independence Measure for Children (Wee-FIM) and the Pediatric Evaluation of Disability Inventory for children with developmental disabilities and acquired brain injury. *Physical and Occupational Therapy in Pediatrics*, 21, 91- 101. doi:10.1080/J006v21n02_08

APPENDIX

Table 13: Evidence of Validity and Reliability of the PEDI

First Author / year	Method /design	Country
Berg (2003)	Translation and adaptation of PEDI for its use in Norway	Norway
Berg (2003)	Applicability of normative reference values of PEDI	Norway
Berg (2004)	Inter-rater, inter-respondent and intra- rater Reliability with intraclass correlation coefficients (ICC)	Norway
Chen (2009)	Test-retest reliability, concurrent validity and discriminant validity, internal consistency and intraclass correlation coefficients, correlation between PEDI and Wee-FIM, logistic regression analysis	China
Chen (2010)	Discriminant validity of children with developmental disabilities and cross-cultural comparison of Taiwanese and American children, internal consistency and inter-rater reliability, unidimensionality by Mokken analysis	China
Custers (2000)	Cross-cultural validation process using a Rasch model	Netherlands
Custers (2002)	Discriminative validity with sensitivity and specificity of the PEDI, predictions of group membership (children with versus without a disability)	Netherlands
Custers (2003)	Dutch adaptation and content validity	Netherlands
Dolva (2004)	Functional performance in children with Down syndrome	Denmark
Elad (2011)	Inter-rater reliability with internal consistency using Cronbach`s alpha and absolute agreement with intraclass correlation coefficients (ICC), discriminative validity using collapsed strata of the Gross Motor Functional Classification System (GMFCS)	Israel
Engelbert (1997)	The purpose of this study was to determine if the severity of osteogenesis imperfecta in childhood might have any influence on the	Netherlands

	functional outcome	
Erkin (2007)	Test-retest reliability with intraclass correlation coefficient) and Cronbach's alpha, Spearman's correlation coefficients, internal construct validity by using Rasch model	Turkey
Feldman (1990)	Concurrent validity with Pearson product moment correlations between Batelle Developmental Inventory screening test and PEDI. Construct validity by significant differences between groups disabled and nondisabled children	USA
Gannotti (1998)	Content validity is determined by comparison of PEDI, analysis of structured questionnaires, and expert content review, factor analysis and reliability by testing the items of the PEDI further to evaluate the construct validity, logistic regression tests the ability of the PEDI to discriminate between children with disabilities from those without	USA Dissertation
Gannotti (2001)	Content and construct validity of a Spanish translation of the Pediatric Evaluation of Disability Inventory (PEDI) for use with children living in Puerto Rico	USA
Gannotti (2002)	Methods utilize principal components analysis to establish the validity of cultural concepts and multiple regression analysis to identify intracultural variation.	USA
Grolegger (2005)	Applicability of PEDI for use in Slovenia; whether the influence of gender, parent's education community size and presence of siblings on children's performance is measured with the PEDI	Slovenia
Haley manual (1992)	Chapter 3 Standardization Chapter 5 Psychometric Properties (reliability, validity)	/ USA
Haley (2006)	Clinically significant change	USA
Harvey (2008)	Reliability responsiveness clinical utility compared with Child Health Questionnaire (CHQ), Functional Assessment Questionnaire, Functional Mobility Scale (FMS), Pediatric Outcomes Data Collection Instrument (PODCI) and PEDI	Australia
Knox (2000)	Clinical Review of the Pediatric Evaluation of Disability Inventory (PEDI)	United Kingdom

Iyer, (2003)	Establishing clinically important differences	USA
Kothari (2003)	To determine whether a hierarchical subscale developed in this study for children with acquired brain injuries corresponds to the generic PEDI subscales, and, if not, whether condition-specific (ABI-specific) PEDI subscales are more sensitive for measuring change	USA
Mc Carthy (2002)	Reliability and validity of PEDI and the GMFM) and the Pediatric Outcomes Data Collection Instrument (PODCI) and the Child Health Questionnaire (CHQ) with internal consistency	USA
Nichols (1996)	Inter-rater, inter-respondent reliability and concurrent validity with the Peabody Developmental Motor Scales (PDMS) and the PEDI	USA
Nordmark (2000)	Compared Gross Motor Function Measure and PEDI over time in children with cerebral palsy; measuring sensitive to change	Sweden
Nordmark (1999)	Compared results using the Swedish PEDI and the American normative values; Pearson correlation analysis	Sweden
Ostensjo (2003)	Everyday functioning in young children	Sweden
Stahlhut (2010)	Applicability and intra-respondent of the PEDI, linear regression models and comparisons of item difficulty, intraclass correlation coefficients in intra-respondent	Denmark
Stahlhut (2011)	Discriminative validity of children without disability and cerebral palsy and juvenile idiopathic arthritis; receiver operating characteristics	Denmark
Reid (1993)	Review and critique of PEDI on purpose, content and scale development, standardization, psychometric characteristics, administration, scoring and examiner training	Canada
Rigby (2013)	Mother`s experiences with the Pediatric Evaluation of Disability Inventory	Canada
Vos-Vromans (2005)	Responsiveness of the Gross Motor Function Measure (GMFM) and the Pediatric Evaluation of Disability Inventory (PEDI) in a group of children with cerebral palsy	Netherlands
Wassenberg (2003)	Inter-interviewer reliability, test-retest reliability inter-respondent on item level	Netherlands

	percentage identical scores were computed and on scale level intraclass correlation coefficients (ICC) and Cronbach`s alpha	
Worth (2013)	Stability of PEDI item across time	Canada
Wright (1993)	Validation of the PEDI	Canada
Ziviani (2001)	Concurrent validity for children with developmental disability and acquired brain injury	Australia
