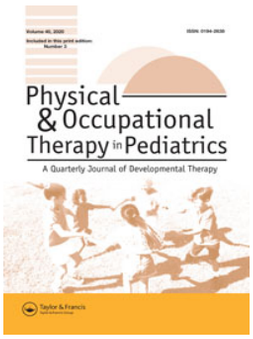

Taylor & Francis
Taylor & Francis Group

Physical & Occupational Therapy In Pediatrics

ISSN: 0194-2638 (Print) 1541-3144 (Online) Journal homepage: https://www.tandfonline.com/loi/ipop20

Administration of the German Pediatric Evaluation of Disability Inventory (PEDI-G) Using the Mode of Observation in Children Undergoing Inpatient Rehabilitation: A Reliability and Validity Study

Ulrike C. Ryll, Caroline H.G. Bastiaenen, Nicole Iten & Hubertus J.A. van Hedel

To cite this article: Ulrike C. Ryll, Caroline H.G. Bastiaenen, Nicole Iten & Hubertus J.A. van Hedel (2020) Administration of the German Pediatric Evaluation of Disability Inventory (PEDI-G) Using the Mode of Observation in Children Undergoing Inpatient Rehabilitation: A Reliability and Validity Study, Physical & Occupational Therapy In Pediatrics, 40:3, 345-359, DOI: 10.1080/01942638.2019.1695698

To link to this article: https://doi.org/10.1080/01942638.2019.1695698

Published online: 05 Dec 2019.

PHYSICAL & OCCUPATIONAL THERAPY IN PEDIATRICS
2020, VOL. 40, NO. 3, 345–359
https://doi.org/10.1080/01942638.2019.1695698

Taylor & Francis
Taylor & Francis Group

Administration of the German Pediatric Evaluation of Disability Inventory (PEDI-G) Using the Mode of Observation in Children Undergoing Inpatient Rehabilitation: A Reliability and Validity Study

Ulrike C. Ryll[a,b,c], Caroline H.G. Bastiaenen[b,d], Nicole Iten[a,e], and Hubertus J.A. van Hedel[a,e]

[a]Pediatric Rehab Research Group, Rehabilitation Center Affoltern am Albis, University Children's Hospital Zurich, Switzerland; [b]Caphri Research Institute, Program Functioning and Rehabilitation, Department of Epidemiology, Maastricht University, The Netherlands; [c]Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden; [d]School of Health Professions, Department of Health, Zurich University of Applied Sciences, Switzerland; [e]Children's Research Center, University Children's Hospital Zurich, Switzerland

ABSTRACT

Aims: To investigate inter-rater reliability, concurrent validity, and feasibility of the German Pediatric Evaluation of Disability Inventory (PEDI-G) using the mode of observation in a Swiss inpatient rehabilitation setting with the Functional Independence Measure for Children (WeeFIM®) as criterion.

Methods: Cross-sectional clinimetric study including 36 children and adolescents with median age 10.8 (quartiles 8.7, 13.0) years with neurological/neuro-orthopedic disorders. Data were collected by healthcare professionals through observation. Analyses were performed using intraclass correlation coefficients ($ICC_{2,1}$), standard error of measurement ($SEM_{Agreement}$), Bland-Altman plots, Cohen's Kappa κ, percentage agreement, and correlations.

Results: Excellent inter-rater reliability ($ICCs_{2,1} \geq 0.97$), small SEMs and acceptable limits of agreement for the Functional Skills Scale (FSS) and Caregiver Assistance Scale (CAS) were found. No systematic differences between raters existed. Cohen's Kappa for inter-rater agreement of the Modifications Scale (MS) ranged from poor to strong ($-0.06 \leq \kappa \leq 0.85$). Excellent concurrent validity for FSS and CAS with the WeeFIM® ($\rho \geq 0.96$), and excellent correlations of FSS and CAS with each other ($\rho \geq 0.98$) were identified.

Conclusion: The German PEDI-G seems to be a reliable and valid, but time-consuming tool when applied in an inpatient setting using observation.

ARTICLE HISTORY
Received 14 February 2018
Accepted 18 November 2019

KEYWORDS
Inpatient setting; observational checklist; psychometric investigation; rehabilitation

Diagnostic tools and evaluation measures in rehabilitation are essential to assess and document the health condition of patients, to plan therapeutic interventions, and to

CONTACT Ulrike C. Ryll ulrike.ryll@ki.se Neuropediatric Unit, Q2:07, Department of Women's and Children's Health, Astrid Lindgren Children's Hospital, 171 76 Stockholm, Sweden.

Supplemental data for this article is available online at https://doi.org/10.1080/01942638.2019.1695698.

evaluate their effectiveness. The Pediatric Evaluation of Disability Inventory (PEDI™) (Haley, Coster, Ludlow, Haltiwanger, & Andrellos, 1992) and the Functional Independence Measure for Children (WeeFIM®) (Uniform Data System for Medical Rehabilitation, 2006) are two well-known instruments to measure functional ability and determine levels of functional independence in children with disabilities. Function, here, relates to performance and reflects the execution of a task in real-life situations (Haley et al., 1992; Uniform Data System for Medical Rehabilitation, 2006; World Health Organization, 2002). The theoretical constructs in both measures are rather similar, with particularly the PEDI Caregiver Assistance Scale resembling the WeeFIM®, and both are used to assess the functional status of children aged 0.6–7 or 7.5 years on daily life activities (ICF domain activities and participation).

The PEDI can be administered as parent report, structured interview with parents, and as observation or professional judgment by healthcare professionals, or a combination (Berg, Jahnsen, Froslie, & Hussain, 2004; Haley et al., 1992). The German version of the PEDI (PEDI-G) was translated and cross-culturally investigated in collaboration with Swiss, Austrian, and German centers. Measurement properties were investigated for PEDI-G scores from structured interviews with caregivers of children and adolescents in outpatient centers (Schulze, Page, Lilja, & Kottorp, 2017; Schulze, Kottorp, Meichtry, Lilja, & Page, 2015; Schulze, Meichtry, Page, & Kottorp, 2019; Schulze & Page, 2010). As the improvement of functional independence in daily life situations is also an important goal in the outcome of inpatient rehabilitation settings, the PEDI-G could be valuable to monitor functional changes in patients in such a setting. However, as the inpatient rehabilitation environment is adjusted to meet the requirements of patients with different levels of impairments, it would be of great interest to determine validity and reliability in such an adapted environment. Most commonly the PEDI is performed as structured interview with the child's parent(s)/caregivers. However, one alternative is of particular interest in an inpatient setting and has not been investigated earlier, that is, to administer the PEDI as an observational checklist in an inpatient rehabilitation setting (Dumas et al., 2010). Inpatient rehabilitation centers are common in the countries involved in the translation of the PEDI-G, and in such a setting, one cannot solely rely on parent-reported measures as parents/guardians often are not able to stay in the center together with their children and, therefore, have limited insight into their rehabilitation progress.

The similarity in constructs motivates the choice of the WeeFIM® as comparator to the PEDI™ as does the fact that it has been used as part of the standard clinical assessment in the investigating rehab center for several years (personal communication). However, in the quest for evidence-based rehabilitation, newly available evaluation instruments are always taken into account.

The purpose of this research was first to determine the inter-rater reliability and concurrent validity of the PEDI-G applied as an observational checklist in children with neurological and/or orthopedic disorders in an inpatient rehabilitation setting in the German-speaking part of Switzerland. For the concurrent validity, the PEDI-G was compared with German version of the WeeFIM® as external criterion.

We hypothesized that the inter-rater reliability of the PEDI-G subscales, Functional Skills Scale (FSS) and Caregiver Assistance Scale (CAS), will be at least good (Intraclass

correlation coefficient $ICC_{2,1} > 0.75$) and expected the FSS and CAS to show a positive and strong (≥ 0.80) linear relationship with the WeeFIM® total score.

Further, we evaluated aspects of feasibility of the measure by (1) assessing the time needed to observe the children performing the PEDI items and (2) gathering qualitative information about the use of the German version of the PEDI in an inpatient rehabilitation setting. To simplify reading, we will refer to 'children' rather than 'children and adolescents' throughout this paper.

Methods

Participants

This cross-sectional clinimetric study included 36 children aged 1.1–17.5 years with congenital or acquired neurological and/or orthopedic disorders (Table 1) who were admitted for an intense inpatient rehabilitation program at the Rehabilitation Center for Children and Adolescents in Affoltern am Albis, Switzerland. Most children had a neurological ($n = 24$) or neuro-orthopedic ($n = 9$) condition. Excluded were participants that exceeded the typical development of a 7-year-old (Uniform Data System for Medical Rehabilitation, 2006). This criterion was chosen to meet the requirements for the WeeFIM® and the PEDI. For the WeeFIM®, normal development is defined as reaching a total score of 126 points. Every eligible child that agreed to participate (consecutive sampling) was recruited through a research nurse (NI) from the inpatient rehabilitation center, over a period of four months.

This study was approved by the Ethics Committee of the Canton of Zurich. Informed assent was obtained from children under 18 years of age as well as informed consent from their parents/guardians, whose data were evaluated and analyzed for this study. Children and parents were informed about the study through a phone call or personal conversation and a participant information letter.

Measures

The Pediatric Evaluation of Disability Inventory (PEDI™) is a clinical instrument to assess the capability and performance of children aged 0.6–7.5 years on daily life

Table 1. Characteristics of participants.

Characteristic	
n	36
Age in years, *median* (25th, 75th percentiles)	10.8 (8.7, 13.0)
Gender girls/boys, *n* (%)	12 (33.3)/24 (66.7)
Origin, *n* (%)	
Albania	3 (8.3)
Kosovo	1 (2.8)
Switzerland	31 (86.1)
Turkey	1 (2.8)
Diagnosis	
Neurological, *n* (%)	24 (66.7)
Congenital, *n* (%)	19 (79.2)
Acquired, *n* (%)	5 (20.8)
Orthopedic, *n* (%)	3 (8.3)
Both, *n* (%)	9 (25)

activities (Haley et al., 1992). It can also be administered to children beyond that age, whose cognitive and motor abilities do not exceed those of a normally developed 7.5-year-old child. Three domains, self-care, mobility, and social function, are assessed by three scales. The Functional Skills Scale (FSS, 197 items) directly measures the capability of a child to execute an item (1 – 'capable', 0 – 'not capable'). The Caregiver Assistance Scale (CAS, 20 items) measures the performance of a child indirectly through the assistance (0 – total assistance to 5 – independent), while the Modifications Scale (MS, 20 items) assesses the modifications (none, children-specific, rehabilitation-specific, extensive) the child requires to execute tasks (Haley et al., 1992; Kothari, Haley, Gill-Body, & Dumas, 2003). The administration time varies, according to the method, age, and level of disability of the child, from 20 to 30 min (observational checklist) to 45–60 min (structured interview) (Berg, Frooslie, & Hussain, 2003; Berg et al., 2004; Haley et al., 1992; Reid, Boschen, & Wright, 1994; Wassenberg-Severijnen, Custers, Hox, Vermeer, & Helders, 2003). Each PEDI™ scale or domain can be used independently or in combination with one other (Erkin, Elhan, Aybay, Sirzai, & Ozel, 2007; Haley et al., 1992).

The German version, PEDI-G, was translated and cross-culturally adapted by Schulze and Page (Schulze et al., 2017; Schulze & Page, 2010) from the Zurich University of Applied Sciences, Switzerland, in cooperation with clinical partners in Austria, Germany, and Switzerland, and was later calibrated and norm-referenced scores determined in a German-speaking population. One item in the FSS self-care domain (eating with knife and fork) and eight items in the FSS mobility domain (four items regarding the use of a three-wheeler or bicycle, four items regarding the transfer into a stroller/bicycle carrier or child bike seat) were added resulting in a total of 246 items for the PEDI-G.

The Functional Independence Measure for Children (WeeFIM®) was designed in a similar conceptual format as the Functional Independence Measure (FIM™) for adults. The WeeFIM® can be administered in about 15–20 min to children aged 0.6–7 years, but is also used beyond this age when individuals' cognitive and motor abilities do not exceed those of a typically developed 7-year-old child (Uniform Data System for Medical Rehabilitation, 2006). The WeeFIM® includes 18 items and measures the performance of daily activities through the amount of assistance a child with disabilities requires to execute these activities (1 – 'total assistance' to 7 – 'complete independence'. Each item score of the WeeFIM® corresponds to a percentage reflecting the degree of independent task completion. The measure has six self-care, two sphincter control, three transfer, two locomotion, two communication, and three social and cognition items. For this study, we used the German WeeFIM® translation that was approved by the company, which was internally validated by center staff and approved by the Uniform Data System for Medical Rehabilitation (UDS) for clinical use in the center. Appendix 1 summarizes an outline of both instruments.

Data Collection Procedures

Two healthcare professionals (NI, UCR) collected PEDI-G data independently by observing each child on the same occasion. NI is a trained research nurse, with nine years of nursing experience in different fields of pediatrics. UCR is a physiotherapist

with three years of experience and a researcher in the field of neuropediatrics. Both raters were native speakers of the German language and studied a preliminary version of the PEDI-G manual that was available at the time and consisted the PEDI-G form with the description of all items (Schulze & Page, 2010). They received three hours of training. In the training, the person who translated the PEDI-G (Christina Schulze) went through the handbook and the evaluation form with the raters, explained the items, and demonstrated how to use the PEDI-G. They went through examples of activities and more elaborately discussed the newly added items. Based on their prior experiences, the two raters asked questions and extensively discussed the MS in the context of a rehabilitation setting. For the assessment, the raters primarily relied on this preliminary version and only in some cases additionally applied the original English manual next to the German version for clarification of some items.

Information about specific items that could not be assessed by observation was obtained through short interviews with the nurse responsible for the child, or answered directly by the adolescent. UCR conducted the structured interview with the nurses following the PEDI-G questionnaire form in presence of the other assessor. In 14 cases (38.9%), the additional short interview with the responsible nurse was necessary for the PEDI-G, and five adolescents (14%) were asked about items regarding getting into a car or using a bicycle or three-wheeler. The raters recorded the time needed to observe the children and documented field notes of issues with observing and scoring items of the PEDI-G in the inpatient setting after the scoring of each child was completed.

The WeeFIM® was completed as an observational checklist by several nurses from the inpatient unit who were different from those raters administering the PEDI-G. All nurses were trained and certified, and observed the children during their routine clinical practice. The time-point for data collection could vary for both assessments as could the order of their administration due to practical constraints related to the clinical flow of the nursing department, but each of the assessments was performed in a one-day period. Mean time between PEDI-G and WeeFIM® assessment was 1.1 days (*SD* 1.8).

Data Analysis

Raw scores were used for all statistical analyses because no normative data of German-speaking children were available at the time of this research. Data for scale and domain scores of PEDI-G FSS, CAS, and WeeFIM® were not normally distributed. Histograms showed that our sample constituted of children with relative low or relative high function. No missing values or outliers existed. For the analysis of concurrent validity, the PEDI-G data obtained by NI was used due to her primary role as healthcare professional in the rehabilitation center. PASW 19 was used to perform all analyses, except for the computation of the Kappa values, for which we used MATLAB® version 2010 b. Field notes were studied by UCR searching for similar words and topics in order to compile the aspects in which both raters agreed into keywords.

To analyze *inter-rater reliability* for the FSS and CAS, we used a two-way random effects model to calculate intraclass correlation coefficients type 2,1/Agreement ($ICC_{2,1}$) on scale and domain score level and derived the standard error of measurement for agreement (SEM) from the error variance. The SEM presents an absolute value for the

error of the measurement; it does not indicate an actual difference or change in scores (de Vet, Bouter, Bezemer, & Beurskens, 2001). The smaller the SEM, the less variability there is around the mean of both raters and the greater the reliability of the instrument (Bruton, Conway, & Holgate, 2000). A priori sample size calculation was based on an ICC of 0.8 and a 95% confidence interval ± 0.15, resulting in a minimum of 22 participants. (de Vet et al., 2001; Giraudeau & Mary, 2001). ICC values of 0.8 were considered good reliability and values below 0.75 poor to moderate reliability (Portney & Watkins, 2008).

Absolute reliability and systematic error between raters were estimated by limits of agreement (Bland & Altman, 1986). Systematic differences between raters were analyzed with the Wilcoxon signed-rank test, because data were not normally distributed. As the MS is of nominal measurement level, inter-rater agreement was calculated using Cohen's Kappa (κ) for all items. A Kappa lower than 0.00 is considered poor agreement, 0.00 to 0.20 slight, 0.21 to 0.40 fair, 0.41 to 0.60 moderate, 0.61 to 0.80 strong and 0.81 to 1.00 almost perfect agreement (Landis & Koch, 1977). Complementing Kappa, we computed percentage agreement for each MS item. Four categories can be scored on the MS, so that chance agreement equates to 25%. Confidence intervals (95%) are presented for $ICC_{2,1}$, mean difference d, and Cohen's Kappa, but are of limited meaning since only two raters were involved.

To investigate *concurrent validity*, Spearman correlation coefficients were used with the German WeeFIM® as criterion. We hypothesized that the PEDI-G would show good agreement with the WeeFIM® and specified *a priori* a minimum correlation ≥ 0.8 between the WeeFIM® and PEDI-G FFS and CAS. The level of agreement was calculated considering the average reliability (X) of the PEDI and preliminary reliability of the German WeeFIM® (Y). Thus, the expected correlation cannot exceed $\sqrt{(rel[X] \times rel[Y])}$ (Lord & Novick, 1968). Furthermore, we investigated how the two main scales of the PEDI-G, FSS measuring capability, and CAS assessing performance, relate to each other. We hypothesized that overall capabilities to execute tasks (high FSS score) will be highly positively correlated (≥ 0.8) to a proficient performance of tasks requiring a low amount of caregiver assistance (high CAS score). Analyses were performed on scale and domain level.

Results

Results of the *inter-rater reliability* of the PEDI-G are presented in Table 2. Relative reliability was excellent with very high $ICCs_{2,1}$ for scale and domain scores ($ICCs_{2,1} \geq 0.97$). The largest SEM was obtained for the FSS (5.21 out of 206 points) indicating an absolute difference of approximately five points between raters. The Wilcoxon signed-rank test showed no systematic difference between medians of both raters for scale and domain scores of the FSS and CAS ($p > 0.05$). Indeed, Bland-Altman plots (Figure 1) showed that both raters scored alike with a systematic mean difference (d) close to zero. The largest mean difference between raters was found in the FSS domain score of social function ($d = -1.50$, 95% CI: -2.74 to -0.26). Likewise, the FSS sum score ($d = -1.08$, 95% CI: -3.50 to 1.33) showed a larger but non-significant difference between raters. The random error seemed large as illustrated by wide limits of agreements for FSS and CAS sum scores, but also for FSS domain scores self-care and mobility. The

Table 2. Inter-rater reliability, mean differences between raters including 95% confidence intervals (CI) and limits of agreements for the PEDI-G FSS and CAS for the total sample ($n = 36$) on total score and domain level.

Scale/Domain	Relative reliability,$ICC_{2,1}$ (95% CI)	Absolute reliability, SEM (range of scores)	*d* (95% CI)	Limits of agreement of *d*
FSS	1.00 (0.99 to 1.00)	5.21 (0–206)	−1.08 (−3.50 to 1.33)	−15.57 to 13.40
CAS	0.99 (0.98 to 0.99)	3.94 (0–100)	0.53 (−1.31 to 2.37)	−10.50 to 11.56
FSS self-care	0.99 (0.98 to 1.00)	2.62 (0–74)	0.03 (−1.20 to 1.26)	−7.34 to 7.40
FSS mobility	0.98 (0.96 to 0.99)	3.28(0–67)	−0.39 (−1.14 to 1.92)	−8.80 to 9.57
FSS social function	0.99 (0.98 to 1.00)	2.85 (0–65)	−1.50 (−2.74 to −0.26)*	−8.95 to 5.95
CAS self-care	0.99 (0.97 to 0.99)	1.77 (0–40)	0.44 (−0.37 to 1.26)	−4.46 to 5.35
CAS mobility	0.98 (0.96 to 0.99)	2.12 (0–35)	0.44 (−0.54 to 1.42)	−5.43 to 6.32
CAS social function	0.97 (0.95 to 0.99)	1.85 (0–25)	−0.36 (−1.22 to 0.49)	−5.50 to 4.77

*Significant at $p < 0.05$. FSS, Functional Skills Scale; CAS, Caregiver Assistance Scale; $ICC_{2,1}$, Intraclass correlation coefficient (Agreement); SEM, standard error of measurement (Agreement); *d*, mean difference between raters.

limits of agreement implied that an improvement of more than 15 points for the FSS, and more than 11 points for the CAS are required to detect a real change in a child's performance status.

Inter-rater agreement for MS items ranged from -0.06 $\leq \kappa \leq$ 0.85, i.e., poor to almost perfect agreement (Table 3). Almost perfect agreement could only be shown for "Bladder management" (self-care item G). Strong agreement, for example, was found for "Bowel management", "Locomotion indoors" and "Locomotion outdoors" (self-care item H, mobility items E and F), and poor agreement was observed for "Transfer into cars" (mobility item B). In cases where the observed ratings were equal to the agreement expected by chance, the numerator equals zero and no meaningful 'true' Kappa could be obtained due to the way Kappa is calculated. Marginal distributions indicated that raters systematically differed for "Transfer into tub/shower", "Locomotion Stairs" (mobility items D and G) and "Safety" (social function item E). Contingency tables can be found in Appendix 2. *Percentage agreement* for each MS item (Table 3) showed by how many percentage points both raters agreed to classify a child into the same category. Overall, agreement ranged from 58 to 92% within the self-care domain, from 53 to 89% within the mobility domain and from 47 to 97% within the social function domain.

Excellent *concurrent validity* ($\rho \geq 0.93$) was indicated for corresponding scales and domains of the PEDI-G and WeeFIM® (Table 4) exceeding *a priori* levels for sufficient agreement. The association between the FSS and CAS, i.e., whether a high functional score (FSS score) was not biased by receiving a lot of support from caregivers (CAS score), was very strong ($\rho \geq 0.94$) for scale and domain level (Table 4).

Keywords from field notes captured the *experiences of both raters* with the PEDI-G and can be summarized into three distinct areas: difficulties with the added items in the German version after cross-cultural validation, differences in interpretation of some items between the original English version and the German version because of the wording of some items, and aspects regarding the use of the PEDI-G in daily practice in an inpatient setting.

Problems after the cross-cultural validation: the simple addition of the aspect showering to FSS items covering bathing may be more appropriate with German-speaking

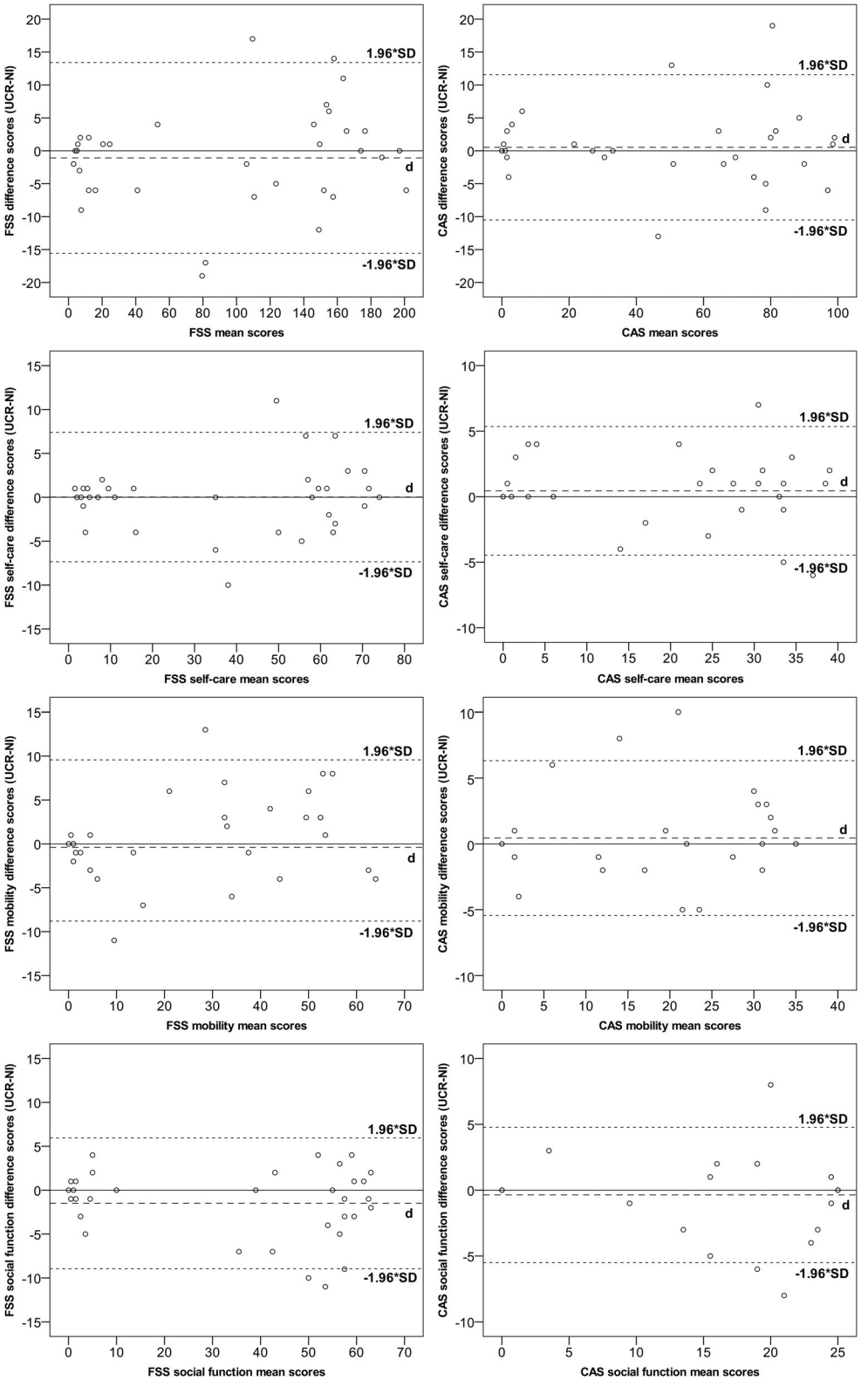

Figure 1. Bland-Altman plots on scale and domain level for FSS and CAS ($n = 36$).

Table 3. Inter-rater agreement (Cohen's Kappa (95% CI)) and percentage agreement (%) between raters for PEDI-G Modifications Scale (MS) for the total sample on item level ($n = 36$).

	Items							
MS domains	A	B	C	D	E	F	G	H
Self-care	Inter-rater agreement							
	0.51*	0.44*	0.38*	n.c.†	0.39	0.24	0.85*	0.71*
	(0.26 to 0.77)	(0.07 to 0.99)	(0.07 to 0.68)		(−0.17 to 0.95)	(−0.05 to 0.53)	(0.70 to 1.01)	(0.49 to 0.92)
	Moderate	Moderate	Fair		Fair	Fair	Almost perfect	Strong
	Percentage agreement							
	72%	81%	69%	92%	89%	58%	92%	83%
Mobility	Inter-rater agreement							
	0.04	−0.06	0.05	0.37	0.65*	0.67*	0.02	–
	(−0.40 to 0.48)	(−0.42 to 0.31)	(−0.49 to 0.59)	(0.06 to 0.68)	(0.32 to 0.97)	(0.36 to 0.97)	(−0.31 to 0.36)	
	Slight	Poor	Slight	Fair	Strong	Strong	Slight	
	Percentage agreement							
	67%	53%	75%	69%	89%	89%	53%	–
Social function	Inter-rater agreement							
	n.c.†	n.c.†	n.c.†	n.c.†	0.17	–	–	–
					(−0.08 to 0.43)			
					slight			
	Percentage agreement							
	92%	94%	97%	97%	47%	–	–	–

*Significant at $p < 0.05$. † Not calculated, undefined. No Kappa could be computed, because at least one variable of each two-way table was a constant. Levels of percentage agreement: < 25% poor, 26–50% fair, 51–75% moderate, 75–100% strong.

Table 4. Concurrent validity and association between PEDI-G FSS and CAS on scale and domain level ($n = 36$).

Concurrent validity	Spearman correlation coefficient, ρ
PEDI-G - WeeFIM®	
Scale level	
FSS - WeeFIM® total	0.98
CAS - WeeFIM® total	0.96
Domain level	
FSS self-care - WeeFIM® self-care	0.98
FSS mobility - WeeFIM® mobility	0.97
FSS social function - WeeFIM® cognition	0.92
CAS self-care - WeeFIM® self-care	0.97
CAS mobility - WeeFIM® mobility	0.95
CAS social function - WeeFIM® cognition	0.95
Association between PEDI-G FSS and CAS	
Scale scores	0.97
Self-care domain	0.96
Mobility domain	0.95
Social function domain	0.94

FSS, Functional Skills Scale; CAS, Caregiver Assistance Scale; WeeFIM®, Functional Independence Measure for Children.

standards; however, both options were included into one single item without differentiating which alternative is predominantly used. For example, getting in and out a bathtub is far more difficult than going under the shower. Newly added FSS items regarding transfer into vehicles such as bike trailers may not reflect an option that all families use, and may further not be applicable for daily transport of a five to seven-year-old child.

Differences in interpretation between the original items and the items translated into German: the raters reported that in the CAS, the item "Security" was not described clearly enough to know which situation should be considered a danger to a child at a certain age. There are no clear-cut limits to differentiate between the categories of the scale. Similar difficulties were experienced with the MS. Among the four categories, it is not clear at which age certain assistive devices such as a nappy or children's toothbrush should still be considered a child-specific rather than a rehabilitative modification. The translation of FSS items regarding food consistencies turned out to be incoherent with the description in the original English items and manual.

Performance of the PEDI-G in daily practice: the raters experienced a general problem with the PEDI™ regarding the child's ability to execute certain items, e.g., using a knife or crossing a road on its own, as they are dependent on the degree to which she/he is allowed to do these activities. Next, depending on their working field, assessors were specialized in different areas covering items of a specific domain. The nurse experienced slightly more challenges with mobility items regarding locomotion outdoors as well as speed and distance indoors, whereas the physiotherapist was challenged by self-care items such as eating, washing, and toileting. Both raters reported less confidence in assessing areas like problem-solving and complexity of communication.

Feasibility

PEDI-G raters observed the children on average for 3.4 hours (*SD* 1.5) and 3.6 hours (*SD* 1.2), respectively, on all items except those not frequently performed in a rehabilitation center, i.e., "Locomotion outdoors", "Vehicles" and "Transfer into vehicles" (FSS

mobility I, J, K, N and O), "Sense of danger" and "Involvement in a social environment" (FSS social function L and M).

Discussion

Inter-Rater Reliability

Excellent ICCs and small SEMs indicate that the FSS and CAS are very reliable when applied by two different healthcare professionals. Though not investigating the mode of observation, excellent inter-interviewer reliability was reported on scale (Berg et al., 2003, 2004; Kothari et al., 2003) and domain level (Kothari et al., 2003; Wassenberg-Severijnen et al., 2003), implying that inter-rater reliability may hold across different administration procedures and populations, however, wide limits of agreement suggest large random error. Maybe, the numerous items, the different design of the scales and the level of experience in scoring the PEDI-G may have contributed to the random error. For example, the social function domain leaves much room for subjective judgment since social abilities are difficult to observe, especially in this study with its relatively short period of observation and its specific inpatient rehabilitation setting.

Schulze et al. (2015) reported much lower values for the SEM than found in our study. Their slightly different choice for the $SEM_{consistency}$ and its calculation based on the $ICC_{1,1}$ may only have had a minor influence on the magnitude of the SEM (de Vet, Terwee, Mokkink, & Knol, 2011), since the systematic error in our study was rather low and ICCs are shown to be nearly equal. Their approach to investigate inter-rater reliability using audiotaped interviews, however, is more likely to have reduced the variance in measurements to a large extent, and thus the size of the SEMs. Two ratings based on one single interview leave much less room for variation than two ratings based on observation. Therefore, a smaller random error can be expected in their ratings.

For some items, two of the four categories of the MS (rehabilitative and extensive modifications) did not occur in our sample, and therefore Cohen's Kappa could not be calculated. Kappa can be influenced by systematic differences between raters, the number of categories, and the marginal distributions (de Vet et al., 2001), as well as the prevalence of findings under consideration (Viera & Garrett, 2005). In these cases, Kappa may not be an appropriate measure for inter-rater agreement. The majority of marginal distributions in every domain was rather skewed indicating a homogeneous study population regarding modifications. This is known to result in higher chance agreement, leaving less room for real agreement between raters (de Vet et al., 2001). Regarding inter-rater agreement, percentage agreement was not always in line with the results of Cohen's Kappa since it only takes into account observed agreement and not chance agreement (e.g., items "Transfer into cars" and "Transfer into/mobility in bed", mobility item B and C). In contrast to our investigation, Schulze et al. (2015) used Cohen's weighted Kappa, considering the MS as being of ordinal measurement level and reported moderate to very good agreement. This again may be a result of their choice to analyze audiotaped interviews instead of ratings collected in a clinically more relevant observational setting as well as their decision to weight the responses to calculate Kappa. Further, the difficulty in differentiating between MS categories in an inpatient setting especially in young children may have contributed to differences in results.

Concurrent Validity

High correlation coefficients between the PEDI-G and the German WeeFIM® indicate excellent concurrent validity and support the assumption that both measure the same underlying construct. Other studies reported a strong relationship, but also a slightly wider range of correlation coefficients, between the corresponding domains of the PEDI™, administered as a structured interview, and the WeeFIM® ($\rho = 0.59$–0.96 (Schultz, 1992), $r = 0.80$–0.90 (Ziviani et al., 2001), $\rho = 0.68$–0.99 (Chen, Hsieh, Sheu, Hu, & Tseng, 2009)). The association between the FSS and CAS implied strong relationships on scale and domain levels. This supports our hypothesis that higher FSS scores, representing better overall capabilities, are strongly associated with higher CAS scores, indicating a minor need for assistance. This does not show, however, which modifications a child requires achieving this level of performance.

Clinical Feasibility

The concurrent use of the comprehensive manual alongside the PEDI-G assessment was rather time-consuming and limits its feasibility in daily practice, although we assume that increased experience would decrease these difficulties to a certain extent. An advantage in using the PEDI-G is its availability free of charge and without obligatory verification of assessors (for comparison, institutions have to pay yearly to use the WeeFIM® and the assessors need to be recertified every two years). In addition, the PEDI scales can be used separately of each other. Further, the PEDI-G may be the preferred tool for a more detailed, but infrequent assessment as it captures more in-depth information on the performance of activities, which may be useful for intervention as well as discharge planning or when justification for an extension of the rehabilitation stay is required by healthcare providers.

As this study investigated the validity of the PEDI-G using the mode of observation, we noticed during data collection that some items were not applicable to children in an inpatient rehabilitation setting. Since this was our target population, this aspect cannot strictly be considered a weakness, it rather points out items (Locomotion outdoors, Vehicles and Transfer into vehicles) that are not feasible for the evaluation through observation in this setting. Therefore, we decided to collect those data by interviews. That could be seen as a minor limitation of the applicability of the PEDI-G in this particular setting. Furthermore, usefulness of the data gathered by the MS in a rehabilitation setting seems questionable since it is difficult to distinguish between modifications specifically required by the child and those provided by the rehabilitation setting as standard equipment, for example a handicapped accessible room or bathroom (i.e., grab bars and adapted handles, wheel-in shower, nonskid surface/mat, raised toilet seat).

The extensive information requested by the PEDI-G was experienced as a challenge for accurate assessment by the observers. An alternative to administer the PEDI in an inpatient setting may be a team approach, that is nurses or occupational therapists could assess the self-care, physiotherapists the mobility, and psychologists or teachers the social function domain (Haley et al., 1992). This, however, may not necessarily result in a decrease of administration time in total. Moreover, it requires training of several professional groups that may require many resources for some centers. Last, the aspects of costs are of interest: applying the PEDI-G for dozens of children per year at

multiple time-points during rehabilitation might produce high personal costs due to a long application time.

Methodological Considerations

A sample of 36 participants may be considered small for psychometric investigations since 30 participants are considered to be fair but a minimum of 50 is required to be considered good (de Vet et al., 2011). This was not possible due to the fluctuating number of children in the center and limited admissions during the study period. Inter-rater reliability analyses involved two PEDI-G raters, which may limit the generalizability of findings to some extent. Further, ICC estimates are known to be excellent in heterogeneous populations, due to the large between-subject variance. Therefore, parameters of absolute agreement may be more meaningful for clinicians. To determine inter-rater agreement for the MS, the Cohen's Kappa seemed the most appropriate, but still limited choice due to skewed marginal distributions of categories in this sample. A further limitation of this study is the small percentage of children that provided self-reports on their abilities on two items although the PEDI-G has not been validated as a self-report measure. Ideally, items should be scored as to the child's current status demanding that if children are unable to move outdoors due to medical or physical limitations, items regarding transfers into a car of cycling abilities should be scored as unable to represent their function at the present time.

Conclusions

Our results support the PEDI-G as a reliable and valid tool when applied by trained healthcare professionals using the clinically relevant mode of observation in an inpatient rehabilitation setting. However, observation of social abilities is challenging (social function domain) as is the distinction between modifications required by the child and those provided as standard equipment (MS).

The clinical feasibility of the PEDI-G was experienced as limited to a certain extent because of the long administration time and the comprehensive manual. The advantages of the PEDI-G are the low initial costs, no burden of the verification of assessors as well as the possibility to use each scale separately.

In general, the PEDI-G will be of interest in a German speaking inpatient rehabilitation setting, when a more detailed, but rather infrequent, assessment is required or open access is a crucial factor. For the use of the PEDI-G, we firmly recommend scoring in strict accordance with the manual.

Acknowledgments

We are grateful to all children and parents, who participated in this project, and all nurses from the rehab center in Affoltern, who helped to collect the data. Many thanks to Christina Schulze and Julie Page, from the ZHAW Winterthur, for their cooperation and provision of the preliminary version of the PEDI-G.

Disclosure statement

The authors have no conflicts of interest to disclose.

Funding

This work was partly sponsored by a grant from the Mäxi-Foundation, Zurich.

About the Authors

Ulrike C. Ryll, MSc, PT, PhD candidate in neuropediatric and clinical epidemiologist at the Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden.

Caroline H.G. Bastiaenen, PhD, PT, is an Associate Professor, Department of Epidemiology, Research line Functioning, Participation & Rehabilitation CAPHRI, Maastricht University, Maastricht, the Netherlands.

Nicole Iten, MSc in Disability and Participation, is a Research Nurse at the Pediatric Rehab Research Department of the University Children's Hospital Zurich, Affoltern am Albis, Switzerland.

Hubertus J.A. van Hedel, PhD, PT, is a Profssor for Neurorehabilitation and Head of the Pediatric Rehab Research Department of the University Children's Hospital Zurich, Affoltern am Albis, Switzerland and Honorary Professor at the Queen Margaret University, Edinburgh, UK.

ORCID

Hubertus J.A. van Hedel http://orcid.org/0000-0002-9577-5049

References

Berg, M., Frooslie, K. F., & Hussain, A. (2003). Applicability of Pediatric Evaluation of Disability Inventory in Norway. *Scandinavian Journal of Occupational Therapy*, *10*(3), 118–126. https://doi.org/10.1080/11038120310013330 doi:10.1080/11038120310013330

Berg, M., Jahnsen, R., Froslie, K., & Hussain, A. (2004). Reliability of the Pediatric Evaluation of Disability Inventory (PEDI^{TM}). *Physical & Occupational Therapy in Pediatrics*, *24*(3), 61–77. doi:10.1300/J006v24n03_05

Bland, J., & Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, *1*(8476), 307–310. https://doi.org/10.1016/S0140-6736(86)90837-8 doi:10.1016/S0140-6736(86)90837-8

Bruton, A., Conway, J., & Holgate, S. (2000). Reliability: What is it, and how is it measured? *Physiotherapy*, *86*(2), 94–99. doi:10.1016/S0031-9406(05)61211-4

Chen, K., Hsieh, C., Sheu, C., Hu, F., & Tseng, M. (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. https://doi.org/10.2340/16501977-0319 doi:10.2340/16501977-0319

de Vet, H., Bouter, L., Bezemer, P., & Beurskens, A. (2001). Reproducibility and responsiveness of evaluative outcome measures. Theoretical considerations illustrated by an empirical example. *International Journal of Technology Assessment in Health Care*, *17*(4), 479–487. doi:10.1017/S0266462301107038

de Vet, H., Terwee, C., Mokkink, L., & Knol, D. (2011). *Measurement in medicine: A practical guide* (1st ed.). Cambridge: Cambridge University Press.

Dumas, H., Fragala-Pinkham, M., Haley, S., Coster, W., Kramer, J., Kao, Y. C., & Moed, R. (2010). Item bank development for a revised pediatric evaluation of disability inventory

(PEDI). *Physical & Occupational Therapy in Pediatrics*, *30*(3), 168–184. doi:10.3109/01942631003640493

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^{TM}). *Disability and Rehabilitation*, *29*(16), 1271–1279. doi:10.1080/09638280600964307

Giraudeau, B., & Mary, J. (2001). Planning a reproducibility study: How many subjects and how many replicates per subject for an expected width of the 95 percent confidence interval of the intraclass coefficient. *Statistics in Medicine*, *20*(21), 3205–3214. doi:10.1002/sim.935

Haley, S., Coster, W., Ludlow, L., Haltiwanger, J., & Andrellos, P. (1992). *Pediatric Evaluation Disability Inventory (PEDITM): Development, Standardization and Administration Manual*. Boston, MA: New England Medical Centre Hospitals.

Kothari, D., Haley, S., Gill-Body, K., & Dumas, H. (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. doi:10.1093/ptj/83.9.776

Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

Lord, F., & Novick, M. (1968). *Statistical theory of mental test scores*. Oxford, England: Addison-Wesley Publishing Company, Inc.

Portney, L., & Watkins, M. (2008). *Foundation of clinical research: Applications to practice* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Reid, D., Boschen, K., & Wright, V. (1994). Critique of the Pediatric Evaluation of Disability Inventory (PEDI^{TM}). *Physical & Occupational Therapy in Pediatrics*, *13*(4), 57–87. doi:10.1080/J006v13n04_04

Schultz, C. (1992). *Concurrent validity of the Pediatric Evaluation of Disability Inventory*. Medford, MA: Tufts University.

Schulze, C., Kottorp, A., Meichtry, A., Lilja, M., & Page, J. (2015). Inter-Rater and Test–Retest Reliability of the German Pediatric Evaluation of Disability Inventory (PEDI-G). *Physical & Occupational Therapy in Pediatrics*, *35*(3), 296–310. https://doi.org/10.3109/01942638.2014.975311 doi:10.3109/01942638.2014.975311

Schulze, C., Meichtry, A., Page, J., & Kottorp, A. (2019). Psychometric properties of the German Version of the Pediatric Evaluation of Disability Inventory (PEDI-G): A factor analysis. *Scandinavian Journal of Occupational Therapy*, 1–10. https://doi.org/10.1080/11038128.2019.1618392 doi:10.1080/11038128.2019.1618392. [Epub ahead of print]

Schulze, C., & Page, J. (2010). *Deutscher Beurteilungsbogen der Pediatric Evaluation Disability Inventory - Vorversion Stand Oktober 2010*. Winterthur: Zürcher Hochschule Für Angewandte Wissenschaften, Departement Gesundheit, Institut Für Ergotherapie.

Schulze, C., Page, J., Lilja, M., & Kottorp, A. (2017). Cross-cultural validity of the German version of the Pediatric Evaluation of Disability Inventory (PEDI-G)-a Rasch model application. *Child: Care, Health and Development*, *43*(1), 48–58. https://doi.org/10.1111/cch.12401 doi:10.1111/cch.12401

Uniform Data System for Medical Rehabilitation. (2006). *WeeFIM IITM Clinical Guide, Version 6.0*. Buffalo: UDS MR.

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine*, *37*(5), 360–364.

Wassenberg-Severijnen, J., Custers, J., Hox, J., Vermeer, A., & Helders, P. (2003). Reliability of the Dutch Pediatric Evaluation of Disability Inventory (PEDI^{TM}). *Clinical Rehabilitation*, *17*(4), 457–462. doi:10.1191/0269215503cr634oa

World Health Organization. (2002). *Towards a common language for functioning, disability and health*. Geneva: ICF.

Ziviani, J., Ottenbacher, K., Shephard, K., Foreman, S., Astbury, W., & Ireland, P. (2001). Concurrent validity of the Functional Independence Measure for Children (WeeFIM®) and the Pediatric Evaluation of Disabilities Inventory in children with developmental disabilities and acquired brain injuries. *Physical & Occupational Therapy in Pediatrics*, *21*(2–3), 91–101. doi:10.1080/J006v21n02_08