

Reliability, Construct Validity and Completeness of the “Winterthur Observation Protocol” Applied to Early Intervention in Pediatric Physiotherapy Using Coaching

Schirin Akhbari Ziegler^a  and Ximena Suárez^{b,c}

^aSchool of Health Sciences, Institute of Physiotherapy, Zurich University of Applied Sciences ZHAW, Winterthur, Switzerland; ^bFacultad de Medicina, Universidad Nacional Autónoma de México Mexico City, Mexico; ^cMedicine faculty, Universidad de Salamanca, Salamanca, Spain

ABSTRACT

Aim: To assess intra- and inter-rater reliability, construct validity and completeness of the “Winterthur Observation Protocol” (WOP).

Methods: Intra-rater and inter-rater reliability were calculated with the interclass correlation coefficient (ICC) with 95% confidence intervals and Cohens’s kappa coefficient based on 18 5-min sequences from six video-recorded early intervention sessions rated by two assessors. Construct validity was based on an increase of coaching over time (three video-recorded sessions) in six therapists during a course on coaching techniques. WOP’s completeness – the capacity to cover at least 95% of the specific coaching strategies – was assessed based on 18 video-recorded sessions.

Results: Intra-rater reliability was excellent: ICCs of continuous variables varied from .87 to 1.0; the kappa-value of categorical variables was 0.90. Inter-rater reliability was sufficient to excellent: ICCs of continuous variables were .47–1.0; the kappa-value of categorical variables was .89, indicating excellent intra-rater reliability. Increasing application of coaching behavior during the course provided evidence of construct validity. More than 95% of the specific coaching strategies could be scored with the WOP.

Conclusion: The WOP is a promising instrument to quantify the contents of early intervention approaches in pediatric physical therapy using coaching. Results must be interpreted with caution due to limited generalizability.



ARTICLE HISTORY


Received 13 September 2022
Accepted 10 February 2023

KEYWORDS

Coaching approaches; infants; pediatric physical therapy; psychometric properties; Winterthur Observation Protocol

Pediatric physical therapists use a variety of approaches and actions to support motor development of infants at risk of developmental motor disorders. As a result, pediatric physical therapy in early intervention is characterized by heterogeneity (Dirks et al., 2011). This heterogeneity, e.g., in the type and dosing of interventions, hampers the understanding of which elements are effective and which are not (Hadders-Algra et al., 2017). Hence, there is a need to characterize and quantify pediatric physical therapy interventions for reliable implementation of a specific intervention, replication of

CONTACT Schirin Akhbari Ziegler  akhb@zhaw.ch  School of Health Sciences, Institute of Physiotherapy, Zurich University of Applied Sciences ZHAW, Katharina-Sulzer-Platz 9, Winterthur 8401, Switzerland

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/01942638.2023.2181724>.

© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

research findings (Moher et al., 2010), and to bridge the research-to-practice caregiver coaching gap (Romano & Schnurr, 2022).

Guidelines for reporting parallel group randomized trials like the Consolidated Standards of Reporting Trials (CONSORT 2020) (Moher et al., 2010) mandate the reporting of details of intervention elements. Currently various instruments to describe interventions in detail are available, such as the “Template for Intervention Description and Replication” (TIDieR) (Hoffmann et al., 2014), the “Pediatric Rehabilitation Observational measure of Fidelity” (PROF) (Rezze et al., 2013), the Triadic Implementation Evaluation Rating Scale (TIERS) (Basu et al., 2010), and the Groningen Observation Protocol version 2 (GOP 2.0) (Hielkema et al., 2018; Sonderer et al., 2017). TIDieR and PROF allow us to classify, describe interventions, and measure fidelity. TIERS is a method for comparing parent and provider interactions in early intervention approaches using coaching. However, in contrast to the GOP 2.0, TIDieR, PROF, and TIERS do not quantify the contents of interventions. Quantification of the contents of interventions provides researchers and practitioners an in-depth description of the characteristics of an intervention (Blauw-Hospers et al., 2010). Elements of the intervention become concrete and quantifiable.

The GOP 2.0 has been used to quantify contents of infant physical therapy, to evaluate changes of contents over time, to compare intervention programs (Akhbari Ziegler et al., 2019; Dirks et al., 2011), and to conduct process evaluations, consisting of an evaluation of the associations between intervention strategies and developmental outcome (Akhbari Ziegler et al., 2021; Blauw-Hospers et al., 2011; Hielkema et al., 2020). The GOP 2.0 quantifies therapeutic actions (behaviors) in terms of continuous variables providing information on their relative duration during a pediatric physical therapy-session (percentage of total treatment time). These therapeutic behaviors are grouped in five main categories, *neuromotor actions*, *educational actions toward caregiver*, *communication*, *position of the infant* and *situation*. Some of the behaviors in the main categories may be further specified by subcategories. Further, GOP 2.0 also allows for the recording of additional categorical variables, such as the role of the caregiver and family member involvement (Hielkema et al., 2018; Sonderer et al., 2017). However, the GOP 2.0 does not differentiate in sufficient detail the various coaching skills described by Rush et al. (2011) and applied in the family-centered early intervention program “Coping with and Caring for Infants with special Needs” (COPCA) (Akhbari Ziegler et al., 2019; Dirks et al., 2011).

Family-centered practices, which recognize the importance of including the family in the child’s care, have become practice-of-choice in pediatric rehabilitation and in early childhood intervention programs (Dunst et al., 2007). The caregivers are regarded as the key decision-makers in the child’s life (Rosenbaum et al., 1998). Understanding caregivers’ needs and supporting caregivers in making informed decisions enhances the well-being and participation of the whole family, including the child with a disability. Systematic review on early intervention indicated that caregiver involvement is associated with better outcome for the infant and family (Vanderveen et al., 2009). It is becoming increasingly clear that the manner in which support is provided to families has an influence on the outcome of the intervention (Rush et al., 2011). Literature suggests that coaching strategies form an essential means of family support (Kemp & Turnbull, 2014; Schwellnus et al., 2015; Ward et al., 2020).

However, coaching is not a uniform method. Different practices with different pre-suppositions regarding mechanisms of effects exist (Ives, 2008). Many ambiguities persist that hinder the application of coaching in relationship-directed forms of interventions based on family-centered practice (Akhbari Ziegler & Hadders-Algra, 2020). We define caregiver coaching as follows: “The therapist, in the role of a coach, supports family members in the process of decision making on functional activity and participation in daily life with the aim of family empowerment and optimization of child development. Therefore, the coach applies the coaching strategies like “shared observation,” relevant information exchange,” “provision of feedback and the opportunity to reflect,” provision of strategies to promote development (hints and suggestions) and “illustration with example.” Generally, caregiver coaching takes place during daily care situations.”

Currently, we do not know which components of coaching are responsible for the positive results described in literature (Akhbari Ziegler & Hadders-Algra, 2020; Kemp & Turnbull, 2014). Thus, our understanding of the application and effect of different forms of coaching is insufficient. One possibility to overcome this uncertainty, would be to examine associations between coaching strategies and defined child and family outcomes, i.e., to undertake a process evaluation. To perform such process evaluation, a reliable and valid assessment instrument to characterize and quantify physical therapy interventions using coaching strategies is requested. At present we lack such an instrument. We, therefore, developed a novel evaluation tool based on the principles of the GOP 2.0, i.e., the Winterthur Observation Protocol (WOP).

The purpose of this study was to assess the intra- and inter-rater reliability of the continuous and categorical variables of the WOP and to evaluate the WOP’s construct validity and completeness.

Methods

Study design

In this study, the psychometric properties, i.e., intra-, and inter-rater reliability of the continuous and categorical variables of the WOP, and WOP’s construct validity and completeness, were examined.

Before the study onset, we informed the ethics committee of Kanton Zurich of the study and asked them whether ethical approval was needed. The committee decided that the study did not need such approval, as the study does not come under the scope of human research law (BASEC-Nr. Req-2019-00588). Informed consent was obtained as recommended by the ethics committee, implying that participants were fully informed about the details of the study. This means that we informed the potential participants, why they were asked to participate in the study, about the aim of the study, we provided general information about the trial and the study process, informed them, that participation is voluntary, and we provided information on benefits, risk of the study and about the confidentiality of the data.

Development and composition of the WOP

The WOP was developed in collaboration with one of the developers of the GOP 2.0. We were guided by the principles of the GOP 2.0, our experiences in the COPCA[®] course which provided us with ample experience on coaching strategies, and own experiences as COPCA coaches. The coaching strategies described in the WOP are based on “The Early Childhood Coaching Handbook” by Rush et al. (2011) and the coaching skills used in COPCA[®] (Akhbari Ziegler et al., 2019). In COPCA[®] caregiver coaching means that the therapist, in the role of a coach, supports family members in the process of decision making on functional activity and participation in daily life with the aim of family empowerment and optimization of child development.

The WOP includes continuous and categorical variables. The continuous variables are classified into five main categories: “*Situation of the intervention session*,” “*Educational actions toward caregivers*,” “*Coaching strategies*,” “*Communication*” and “*Neuromotor actions*.” Within each category specific therapeutic intervention elements, and so-called behaviors are defined. For instance, Coaching Strategies include the behaviors “observation/shared observation,” “relevant information exchange,” “provision of feedback and the opportunity to reflect” and “provision of strategies to promote development.” Some behaviors are further specified with so-called modifiers; for example, the behavior “provision of strategies to promote development” has two modifiers “provision of strict instructions” and “provision of hints and suggestions.” The duration of each behavior during an intervention session is quantified.

There are 11 categorical variables, i.e., variables summarizing overall behavior during a video sequence. Examples are “location,” “role of the caregiver,” “role of the therapist,” or “goal orientation.” Each categorical variable has between two and four response options. The rater selects the response option that summarizes best the events occurring during the entire intervention session. For details on the WOP see the [Supplementary Appendix](#).

Differences between the GOP 2.0 and the WOP are described in [Table 1](#). The category “coaching skills” has been added to the WOP whereas the category “position” from the GOP 2.0 had been removed. The WOP categories “situation of the session” and “neuromotor actions” include fewer behaviors than the GOP 2.0. In addition, some behaviors that came under category “communication” in GOP 2.0 had been transferred to WOP’s category “coaching strategies.” Generally, the WOP is easier to apply than the GOP 2.0 as it comprises fewer continuous variables, which saves time and reduces complexity. The differences in the categorical variables of the GOP 2.0 and WOP are shown in [Table 2](#). GOP 2.0 has five categorical variables, the WOP 11. Only three items: “type of session,” “involvement of family members,” and “role of the caregivers” are present in both protocols. The eight new categorical items on the WOP represent current treatment characteristics, and they include items as “role of the physiotherapist,” “goal orientation” or “joint planning.”

Similar to all previous GOP-analyses (Akhbari Ziegler et al., 2021; Blauw-Hospers et al., 2011; Dirks et al., 2011; Hielkema et al., 2018, 2020), the duration of the behaviors and modifiers is quantified from videos of early intervention sessions using the software program, *The Observer XT* (Version 11.5; Noldus, Wageningen, the Netherlands) (Blauw-Hospers et al., 2010). In *The Observer*, onset and end of each

Table 1. Differences in categories and behaviors (continuous variables) between the WOP and the GOP 2.0.

WOP 1.0	GOP 2.0
A.1 Situation of the intervention session	A5. Situation of treatment session
1.1 Daily care situation	5.1. Motor activity/play
1.2 No daily care situation	5.2. Feeding
	5.3. Bathing
	5.4. Dressing/ undressing
	5.5. Changing Diapers
	5.6. Carrying
	5.7. Not specified situation
A.2 Educational actions toward caregivers	A2. Educational actions
2.1. Caregiver training	2A Educational actions toward the infant;
2.2. Caregiver coaching	a. PT/Caregiver interferes with activities of infant point event
2.3. Not specified, i.e., neither training nor coaching	b. Not specified
	2B Educational actions toward caregiver
	2.1. Caregiver training
	2.2. Caregiver coaching
	2.3. Not specified
A.3 Coaching strategies	
3.1. Observation/shared observation	
3.2. Relevant information exchange	
3.3. Provision of feedback and the opportunity to reflect	
3.4. Provision of strategies to promote development	
3.5. Illustration with examples	
3.6. None of the above specified coaching strategies	
A.4 Communication	A3. Communication
4.1. Communication related to the session	3.1. Information exchange
4.2. Communication not related to the session	3.2. Instruct
4.3. No communication	3.3. Provide feedback
	3.4. Not specified, e.g. com. with the infant
	3.5. No communication
	A4. Position
	4.1. Supine
	4.2. Prone
	4.3. Side
	4.4. Sitting
	4.5. Standing
	4.6. Walking
	4.7. Transition
	4.8. Not specified position
A.5 Neuromotor actions	A1. Neuromotor actions
5.1. Transition	1.1. Facilitation techniques
5.1. Hands-on techniques	1.2. Reflex Locomotion
5.2. Self-produced motor behavior (SPMB)	1.3. Sensory experience; state event
5.3. Challenging to SPMB	1.4. Sensory experience; point event
5.4. Not specified	1.5. Passive motor experience
	1.6. Self-produced motor behavior (SPMB)
	1.7. Self-produced motor behavior (SPMB) in combination with constraint of one upper limb, no interference of caregiver/PT
	1.8. Challenged to SPMB (CSPMB), infant is allowed to continue activity by him/herself
	1.9. CSPMB in combination with constraint of upper limb, infant is allowed to continue activity by him/herself
	1.10. CSPMB, activity flows over into facilitation, sensory or passive experience
	1.11. Craniosacral therapy
	1.12. Not specified

CSPMB: Challenged to self-produced motor behavior; GOP: Groningen Observation Protocol; PT: physiotherapist; SPMB: self-produced motor behavior; WOP: Winterthur Observation Protocol.

Table 2. Differences in the categorical variables of the WOP and the GOP 2.0.

WOP	GOP 2.0
B.1 Type of session	B.1 Type of session
B.2 Location	
B.3 Degree of involvement of family members	B.2 Dressing
B.4 Role of the caregiver	B.3 Family member involved in the treatment session
B.5 Role of the physiotherapist	B.4 Role of the caregivers
	B.5 Presence of twins
B.6 Goal orientation	
B.7 Positions in correspondence with the goal	
B.8 Postural support in correspondence with the goal	
B.9 Activities in correspondence with the goal	
B.10 Joint planning	
B.11 Involvement of the infant/child	

GOP: Groningen Observation Protocol; WOP: Winterthur Observation Protocol.



Figure 1. Illustration of the WOP-analysis of a 5-min sequence of an intervention video. In bold the 5 WOP-categories are denoted and below each category, the applied behaviors and modifiers are listed. The onset of a bar indicates the onset of a behavior/modifier, the end of the bar the end of the behavior/modifier, implying that the length of the bar indicates the duration (time spent with) of that behavior/modifier. Addition of all duration values results in the total time spent with a specific behavior/modifier, which is expressed as a proportional value, i.e., as the percentage of total intervention time. In this example 100% of the time was spent with “caregiver coaching,” approximately 20% of the time with “observation/shared observation,” and approximately 36% with “provision of feedback and the opportunity to reflect.”

behavior and modifier are recorded continuously and simultaneously, allowing *The Observer* to calculate the relative duration, i.e., the percentage of total intervention time spent on each specific behavior (see Fig. 1). The Figure illustrates that one 5-min sequence involves many decisions on the beginning and the end of behaviors, with the number of decisions of the sequences used in the present study varying from 42 to 92 (median number 73). The scores of the modifiers represent relative percentages of time spent with this modifying behavior. In other words, the outcomes of *Observer* are proportions of intervention time spent with specific actions. For example, in the sequence shown in Figure 1, approximately 36% of the intervention time was spent with “provision of feedback and the opportunity to reflect.”

Participants

Participants included the pediatric physical therapists carrying out the intervention, caregivers, infants receiving intervention, and two WOP-assessors. The authors (SA + XS) are the two WOP-assessors. Inclusion criterion for the pediatric physical therapists carrying out the video-recorded early intervention sessions were participation in a COPCA[®] course 2019 held in Winterthur, Switzerland. The pediatric physical therapists informed eligible families, i.e., families with an infant or child with an indication for physical therapy due to a developmental disorder, about the course and its associated study and selected one family with a young child with special needs. Informed consent from the therapists and the family was an additional inclusion criterion. Six pediatric physical therapists and family dyads (mother and infant) fulfilled these criteria and participated in the COPCA course and the study. Video recordings of six families with children aged 10–36 months corrected age (median 22 months) with delayed motor development based on neurological dysfunction or a genetic syndrome and not able to stand and walk independently were included in the study. Background information on the therapists, caregivers, children, and the two WOP-assessors can be found in [Table 3](#).

A selection of six videos was used for the assessment of reliability. Three videos were from sessions of a Swiss pediatric physical therapist with 15 years of professional experience in treating young children. She worked with a girl of three years with Rett-Syndrome and her mother originating from the USA. The other three videos were from sessions of a Lithuanian pediatric physical therapist with three years professional experience in treating young children. She worked with a girl born preterm of 10–12 months corrected age who was at high risk of developmental disorders along with her Lithuanian mother.

Procedures

The data for this study were collected just prior to and during the COPCA[®] course. The course started in May 2019 with two days of contact teaching (Part 1), followed by two days of contact teaching in August and December 2019 (Part 2 + 3). COPCA[®] courses aim to educate pediatric physical therapists to become a certified COPCA-coach. Each participating physical therapist recorded three of their own intervention sessions. Instructions for capturing the required data included videorecording an entire intervention session (video and audio signal) with a video camera on a tripod placed in such a way that it allowed for adequate visibility of all events and participants. The duration of the 18 recorded intervention sessions varied from 18 to 54 min (median 32 min).

The first author (SA) has extensive experience with the quantification of behaviors during early intervention sessions, especially using the GOP 2.0 (Sonderer et al., 2017) and she developed the WOP. The second author (XS) received two days of training on the WOP by first author, whereafter she practiced the WOP by independently assessing four training videos prior to data analysis for this study. For the inter- and intra-rater reliability of the WOP, three sequences each lasting 5 min, were selected from six treatment sessions. The number of video sequences was based on previous studies assessing

Table 3. Characteristics of study participants.

Characteristics of the WOP-assessors ($n = 2$)	
Sex: male/female, n	0/2
Country of basic physical therapy education	
Switzerland, n	1
Mexico, n	1
Level of education	
PhD, n	1
PhD cand., n	1
Professional experience physical therapy with infants in years	
31, n	1
23, n	1
Professional experience in research in years	
10, n	1
4, n	
Professional experience in quantifying early intervention by means of GOP or WOP in years	
9, n	1
1, n	1
Characteristics of the therapists ($n = 6$)	
Sex: male/female, n	0/6
Country of basic physical therapy education	
Switzerland, n	4
Lithuania, n	1
Austria, n	1
Level of education	
Bachelor of sciences, n (%)	100
Professional experience physical therapy with infants in years	
<1, n	1
1–10, n	2
>10, n	3
Characteristics of the infants/children ($n = 6$)	
Sex: male/female, n	3/3
Preterm birth, n	2
Age at first video recording	
≤1 year, n	2
>1 year, n	4
Severity of disorder ^a	
Mild-to-moderate disorder, n	4
Severe disorder, n	2
Characteristics of the care givers ($n = 6$)	
Sex: male/female, n	0/6
Maternal age (years): median range	34.5 (26–40)
Maternal education, ^b low/middle/high, n (%)	2 (33.3)/2 (33.3)/2 (33.3)

^aBased on the Gross Motor Function Classification System (Gorter et al., 2009; Palisano et al., 1997) Infants/children classified as GMFCS levels I–III were classified as having a mild-to-moderate disorder; those classified as GMFCS IV–V as having a severe mobility impairment. ^bLevels of education: low = primary education/junior vocational training, middle = secondary education/senior vocational training, high = university education/vocational colleges.

reliability of quantification of behaviors during early intervention sessions using *the Observer* (Blauw-Hospers et al., 2010; Sonderer et al., 2017). These sequences were sampled from three video recorded COPCA sessions of two COPCA[®] course participants. Sampling of the six videos occurred at fixed time periods of the videos, i.e., from 2 to 6 min, from 15 to 19 min, and from 26 to 30 min after video onset. The videos were selected based on the language spoken since English was a requirement of the assessors.

To calculate the intra-rater reliability, the first author (SAZ) of the study assessed all 18 sequences twice: the second time occurring three weeks after the first assessment. To determine inter-rater reliability, the two authors independently assessed the 18 sequences. The interclass correlation coefficient (ICC), as an expression of inter-assessor

agreement of the two assessors' evaluations of these training videos, was between 0.75 and 0.80 for the behaviors of the continuous variables, which was sufficient to allow the current reliability study.

Construct validity was assessed by determining the changes in the prevalence of coaching behaviors over time in all six COPCA[®] course participant and family dyads who were filmed three times: (1) just prior to the course (T0), (2) one month after the first part of the course (T1), and (3) two months after the second part (T2). Considering that coaching is COPCA's basic communication strategy it is expected (as has been demonstrated previously (Akhbari Ziegler et al., 2018) that the proportion of time spent with coaching increases between T0 and T2. Construct validity of the WOP is supported when such an increase is demonstrated. Completeness was assessed based on these 18 (3 × 6) video-recorded physical therapy sessions. WOP's completeness was considered to be present when at least 95% of the specific coaching strategies could be covered with the WOP. This was operationalized as follows: the sum of time spent in the coaching strategies described under A.3 (A.3.1–A.3.5), except “provision of strict instructions” should be approximately similar to the time spent with the educational action of “caregiver coaching.”

Data analyses

Characteristics of the WOP-assessors, the pediatric physical therapists, caregivers, and children were quantified by descriptive statistics. To determine the reliability of the continuous variables of the WOP, the relative durations of the behaviors were compared. The interclass correlation coefficient (ICC) with 95% confidence intervals (CIs) was used to quantify reliability. The intra-rater reliability was calculated with the ICC1 (1-way random model), the inter-rater reliability with the ICC2 (2-way random model). Values <0.4 were interpreted as insufficient agreement, values between 0.4 and 0.75 as sufficient to good and values >0.75–0.9 as excellent (Shrout & Fleiss, 1979).

Time spent with coaching behaviors was quantified with medians and ranges of relative duration of treatment sessions. The Wilcoxon-sign rank test was used to analyze changes in these behaviors over time. Median changes and corresponding non-parametric confidence intervals were computed using the Hodges-Lehmann estimator (HL) and the associated CIs (Bauer, 1972).

To evaluate the completeness of the WOP, we compared time spent on the educational action “caregiver coaching” with the sum of the time spent on the coaching strategies “observation,” “relevant information exchange,” “provision of reflective feedback and the opportunity to reflect,” the modifier “provided strategy: hints” of the coaching strategy “provision of strategies to promote development” and “illustration with examples” in all 18 video-recorded COPCA sessions. Completeness was defined when at least 95% of the specific coaching strategies could be scored with the WOP. These statistical analyses were done with IBM SPSS 28 Software (IBM Corp, Armonk, New York).

To determine the intra- and inter-rater reliability of the categorical variables of the WOP, we used Cohens's kappa coefficient (κ ; using Microsoft[®] Excel[®] version 12.3.6). Values between 0.60 and 0.79 were interpreted as moderate, values between 0.80 and 0.90 as strong, and values >0.90 as almost perfect (McHugh, 2012).

Table 4. Intra-rater and inter-rater agreement on continuous items assessing behaviors and modifiers of the WOP.

Intervention	Intra-rater agreement (ICC 95% CI)	Inter-rater agreement (ICC 95% CI)
Situation of the session		
Daily care situation	Not calculated	Not calculated
Educational actions toward the caregivers		
Caregiver coaching	0.99 (0.98–0.99)	1.00 (0.99–1.00)
No educational actions	0.99 (0.98–0.99)	1.00 (0.99–1.00)
Coaching strategies		
Observation	0.98 (0.96–0.99)	0.66 (0.29–0.86)
Relevant Information exchange	0.99 (0.98–0.99)	0.92 (0.79–0.97)
Provision of feedback and the opportunity to reflect	0.97 (0.93–0.99)	0.82 (0.58–0.93)
Provision strategies to promote development	0.89 (0.74–0.96)	0.47 (0.02–0.76)
Provided strategy “hints”	0.89 (0.74–0.96)	0.47 (0.02–0.76)
Illustration with examples	0.99 (0.97–0.99)	0.54 (0.11–0.8)
None of the above specified coaching strategies	0.97 (0.92–0.99)	0.60 (0.12–0.83)
Communication		
Communication related to the session	1.00 (1.00–1.00)	0.68 (0.33–0.87)
No communication	0.99 (0.99–1.00)	0.68 (0.33–0.87)
Neuromotor actions		
Hands-on techniques	0.87 (0.69–0.95)	0.95 (0.87–0.98)
SPMB	0.98 (0.96–0.99)	0.78 (0.50–0.91)
SPMB exploration	0.98 (0.93–0.99)	0.72 (0.41–0.89)
SPMB no exploration	0.97 (0.92–0.99)	0.81 (0.55–0.92)
CSPMB	0.89 (0.74–0.96)	0.72 (0.40–0.89)
CSPMB adequate	0.98 (0.95–0.99)	0.51 (0.07–0.78)
CSPMB not interested	0.74 (0.44–0.89)	0.36 (–0.11–0.77)
CSPMB too easy	0.92 (0.80–0.97)	0.56 (0.14–0.80)
Not specified	0.99 (0.99–1.00)	0.60 (0.20–0.83)
Transition	0.99 (0.99–1.00)	0.79 (0.52–0.91)

CI: confidence interval; CSPMB: challenged to self-produced motor behavior; ICC: interclass correlation coefficient; SPMB: self-produced motor behavior; WOP: Winterthur Observation Protocol.

Results

Intra-class correlations for intra-rater and inter-rater agreement of the continuous variables of the WOP and their associated 95% CIs are presented in Table 4. ICCs could not be calculated for the behavior “daily care situation,” as both raters scored a consistent duration of 100% in all 18 video sequences.

For intra-rater reliability, the ICC of all 15 behaviors (“caregiver coaching,” “no educational action,” “observation,” “relevant information exchange,” “provision of feedback and the opportunity to reflect,” “provision of strategies to promote development,” “illustration with examples,” “none of the above specified coaching strategies,” “communication related to the session,” “no communication,” “hands-on techniques,” “self-produced motor behavior,” “challenged to self-produced motor behavior,” “not specified neuromotor action” and “transition”) varied from .87 to 1.0, indicating excellent intra-rater reliability. The ICC of five modifiers (“provided strategy: hints,” “self-produced motor behavior exploration and no exploration,” “challenged to self-produced motor behavior adequate and too easy”) varied from .89 to .98 indicating excellent intra-rater reliability. The ICC for (“challenged to self-produced motor behavior not interested) was .74 indicating sufficient to good intra-rater reliability.

For inter-rater reliability, the ICC of seven of the 15 behaviors (“caregiver coaching,” “no educational action,” “relevant information exchange,” “provision of feedback and

the opportunity to reflect,” “hand-on techniques,” “self-produced motor behavior and “transition”) varied from .78 to 1.0 indicating excellent inter-rater reliability. For the remaining eight behaviors (“observation,” “provision of strategies to promote development,” “illustration with examples,” “none of the above specified coaching strategies,” “communication related to the session,” “no communication,” “challenged to self-produced motor behavior” and “not specified neuromotor action”) the ICC varied from .47 to .72 indicating was sufficient to good inter-rater reliability. The ICC of five modifiers showed sufficient to good inter-rater reliability (“Provided strategy: hints,” “self-produced motor behavior exploration and no exploration,” “challenged to self-produced motor behavior adequate and too easy”) varied from .47 to .81 indicating sufficient to good inter-rater reliability. The ICC of “challenged to self-produced motor behavior not interested” was .36 indicating a poor inter-rater reliability. Intra-rater and inter-rater reliability of the categorical variables of the WOP, measured with Cohens’s kappa coefficient was strong ($\kappa = 0.89, 0.90$ respectively).

Table 5 shows the change in coaching behavior over time, as an expression of the construct validity of the WOP. “Caregiver coaching” increased significantly between T0 and T2 to 57.3% relative duration. Also, the coaching strategies “observation” and “provision of reflective feedback and opportunity to reflect,” showed a significant increase in relative duration between T0 and T2, with 36.5% and 5.4%, respectively. Time spent on the coaching strategies “relevant information exchange” and “provided strategy: hints” increased 3.2% and 5.3%, however these increases were not statistically significant.

The evaluation of the WOP’s completeness revealed similarity as time spent on the educational action “caregiver coaching” and the sum of the time spent on the various coaching strategies ranged from 0.00% to 4.05% (median 0.00%). Consequently, in all 18 video recorded COPCA sessions, more than 95% of the specific coaching strategies used to apply coaching during physical therapy intervention could be scored with the WOP.

Discussion

The results indicate, based on 18 video fragments from six videos of two pediatric physical therapists, that coaching strategies used during pediatric physical therapy sessions can be assessed reliably with the WOP. Most items had excellent intra- and inter-rater agreement values. WOP’s construct validity was demonstrated by increasing the application of coaching behavior during a COPCA[®] course. WOP covered more than 95% of the coaching strategies used during coaching illustrating its completeness.

Our results are consistent with the results of the psychometric properties of GOP 2.0 (Sonderer et al., 2017). The GOP 2.0 is a useful observation protocol, to quantify the contents of more traditional infant physiotherapy approaches, to evaluate changes of these approaches over time, and to compare different intervention programs. However, to quantify the content of intervention approaches using coaching, we recommend using the WOP since it allows to quantify coaching skills in a differentiated and reliable way. In comparison to the TIERS (Basu et al., 2010) which uses a Likert scale to

Table 5. Coaching behaviors over time.

Coaching behavior	T0	T1	Median change T0–T1		T2	Median change T0–T2	
	Median (range)	Median (range)	HL-estimate (99% CI) lower	upper	Median (range)	HL-estimate (99% CI) lower	upper
Educational actions toward caregiver							
Caregiver coaching	21.5 (0.0–56.0)	74.3 (0.0–97.9)	43.7 (–25.4;97.9)		83.5 (43.6–99.3)	57.3 (22.8;99.3)	
Coaching strategies							
Observation/shared observation	6.0 (0.0–21.3)	33.8 (0.0–76.6)	23.2 (–3.8;76.6)		45.1 (23.9–66.9)	36.5 (14.4;58.6)	
Relevant information exchange	6.1 (0.0–9.6)	14.3 (0.0–22.4)	6.4 (–9.5;17.2)		8.8 (4.4–17.6)	3.2 (–2.4;17.6)	
Provision of reflective feedback and opportunity to reflect	2.5 (0.0–6.0)	3.4 (0.0–9.2)	0.5 (–2.6;9.1)		8.0 (3.5–12.3)	5.4 (1.0;12.3)	
Provision of strategies to promote development	4.8 (0.0–20.8)	7.5 (5.6–18.4)	1.6 (–15.2;18.4)		6.9 (6.0–11.3)	1.5 (–9.7;6.4)	
Hints/suggestions	0.3 (0.0–8.7)	6.2 (0.0–10.0)	4.3 (–3.1;10.0)		6.7 (4.8–11.3)	5.3 (–1.4;10.6)	

CI: confidence interval; HL: Hodges-Lehmann estimator; T0: 1 month before the start of the COPCA course; T1: 1 months after the first part of the COPCA course; T2: 2 months after the second part of the COPCA course. The numbers represent the relative duration of coaching behaviors, in percentage of the total treatment time. Results in bold represent significant differences ($p \leq .01$). Note that the coaching strategy preceded by cubed numeration is a modifier. The modifiers represent **relative** percentages of time spent in the behavior mentioned above the modifier.

examine approaches using coaching, the WOP quantifies the content of the intervention in percentage of total intervention time.

The modifier “not interested” of the behavior “challenged to self-produced motor behavior” is an exception to the good to excellent inter-rater reliability. This modifier had also the lowest value in the intra-rater reliability, although it is still sufficient to good. An explanation for these results could be that the child in half of the video sequences was diagnosed with Rett-Syndrome, which made it difficult to decide whether the child was not interested in the challenge or was not able to express her interest the challenge.

The application of coaching behaviors steadily increased during the COPCA[®] course. This suggests that the WOP is a valid tool to quantify physiotherapy interventions using coaching strategies. Five months after the start of the COPCA[®] course (T2), the participating physical therapists spent more than 80% of treatment time with caregiver coaching. The most frequently applied coaching strategy at T2 was “observation/shared observation” (during 45.1% of treatment time), followed by “relevant information exchange” (8.8% of time) and “provision of reflective feedback and the opportunity to reflect” (8.0% of time).

The completeness of the WOP was very good, more than 95% of the specific coaching strategies could be scored with the WOP. The completeness of the WOP matches that of the GOP 2.0 (Sonderer et al., 2017), even though the completeness of the GOP 2.0 focused on the neuromotor actions.

Thus, the WOP is a new observation protocol which allows a detailed quantification of coaching skills and a characterization of pediatric physical therapy approaches using coaching in a reliable way. This offers the possibility to explore potential associations between coaching skills and defined child and family outcomes. Process analyses like this have the potential to increase understanding of which coaching components might be responsible for the positive results of coaching, e.g., developmental gains for children, reduction of caregiver stress, and improved caregiving capacities (Akhbari Ziegler & Hadders-Algra, 2020; Kemp & Turnbull, 2014). As family-centered practices currently

are the practice-of-choice and coaching strategies are intended to provide family support, it is important to know which components of coaching are associated with beneficial effects of intervention and which components do not contribute.

A strength of this study was that inter- and intra-rater reliability of the continuous as well as the categorical variables of the WOP were examined. Another strength was that the study also addressed and confirmed WOP's construct validity and completeness. The study had several limitations. One of the study's limitations was that for the reliability analyses only two infant-mother and therapist combinations could be assessed, which reduces the generalizability of the study. Two additional limitations are (1) the use of only two raters for the reliability evaluation, and (2) the fact that one of the assessors is a co-developer of the WOP. To increase understanding for the effectiveness of different coaching strategies, we recommend studies examining potential associations between coaching strategies and defined child and family outcomes, i.e., to undertake studies with process evaluation.

Conclusion

The purpose of this study was to assess the intra- and inter-rater reliability of the continuous and categorical variables of the WOP and to evaluate the WOP's construct validity and completeness. WOP's reliability was based on 18 video sequences of six videos of two infant-mother and therapist combinations. Intra-rater reliability of continuous and categorical variables was excellent. Inter-rater reliability of continuous variables was excellent in 36%, sufficient in 54.5% of items, whereas it was excellent for categorical items. WOP's construct validity and completeness were good. As these results are based on limited data they must be interpreted with caution. The results indicate that the WOP is a promising instrument to quantify contents of early intervention approaches in pediatric physical therapy using coaching.

Acknowledgements

The authors greatly acknowledge the constructive feedback of Mijna Hadders-Algra, Michèle Brühlhart and Partizia Sonderer on a previous draft of the manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

Notes on contributors

Schirin Akhbari Ziegler, physiotherapist, received her PhD 2020 from the University of Groningen on the topic "Implementation of the family-centred early intervention program COPCA". She works as a lecturer and researcher at the Zurich University of Applied Sciences and as a freelance paediatric physiotherapist in the field of family-centred early intervention.

Ximena Suárez is a Paediatric Physical Therapist. She has been a professor at different Universities in México, she is part of the research team for Spinal Muscular Atrophy at Hospital Infantil de México Federico Gómez in México City. Actually she is a PhD student at Universidad de Salamanca Spain.

ORCID

Schirin Akhbari Ziegler  <http://orcid.org/0000-0001-8640-0774>

References

- Akhbari Ziegler, S., Dirks, T., Reinders-Messelink, H. A., Meichtry, A., & Hadders-Algra, M. (2018). Changes in therapist actions during a novel pediatric physical therapy program: Successes and challenges. *Pediatric Physical Therapy*, 30(3), 223–230. <https://doi.org/10.1097/PEP.0000000000000509>
- Akhbari Ziegler, S., Dirks, T., & Hadders-Algra, M. (2019). Coaching in early physical therapy intervention: The COPCA program as an example of translation of theory into practice. *Disability and Rehabilitation*, 41(15), 1846–1854. <https://doi.org/10.1080/09638288.2018.1448468>
- Akhbari Ziegler, S., & Hadders-Algra, M. (2020). Coaching approaches in early intervention and paediatric rehabilitation. *Developmental Medicine and Child Neurology*, 62(5), 569–574. <https://doi.org/10.1111/dmcn.14493>
- Akhbari Ziegler, S., von Rhein, M., Meichtry, A., Wirz, M., Hielkema, T., & Hadders-Algra, M. (2021). The coping with and caring for infants with special needs intervention was associated with improved motor development in preterm infants. *Acta Paediatrica*, 110(4), 1189–1200. <https://doi.org/10.1111/apa.15619>
- Basu, S., Salisbury, C. L., & Thorkildsen, T. A. (2010). Measuring collaborative consultation practices in natural environments. *Journal of Early Intervention*, 32(2), 127–150. <https://doi.org/10.1177/1053815110362991>
- Bauer, D. F. (1972). Constructing confidence sets using rank statistics. *Journal of the American Statistical Association*, 67(339), 687–690. <https://doi.org/10.1080/01621459.1972.10481279>
- Blauw-Hospers, C. H., Dirks, T., Hulshof, L. J., Bos, A. F., & Hadders-Algra, M. (2011). Pediatric physical therapy in infancy: From nightmare to dream? A two-arm randomized trial. *Physical Therapy*, 91(9), 1323–1338. <https://doi.org/10.2522/ptj.20100205>
- Blauw-Hospers, C. H., Dirks, T., Hulshof, L. J., & Hadders-Algra, M. (2010). Development of a quantitative tool to assess the content of physical therapy for infants. *Pediatric Physical Therapy*, 22(2), 189–197. <https://doi.org/10.1097/PEP.0b013e3181dbd5f1>
- Dirks, T., Blauw-Hospers, C. H., Hulshof, L. J., & Hadders-Algra, M. (2011). Differences between the family-centered “COPCA” program and traditional infant physical therapy based on neuro-developmental treatment principles. *Physical Therapy*, 91(9), 1303–1322. <https://doi.org/10.2522/ptj.20100207>
- Dunst, C. J., Trivette, C. M., & Hamby, D. W. (2007). Meta-analysis of family-centered helping practices research. *Mental Retardation and Developmental Disabilities Research Reviews*, 13(4), 370–378. <https://doi.org/10.1002/mrdd.20176>
- Gorter, J. W., Ketelaar, M., Rosenbaum, P., Helders, P. J. M., & Palisano, R. (2009). Use of the GMFCS in infants with CP: The need for reclassification at age 2 years or older. *Developmental Medicine and Child Neurology*, 51(1), 46–52. <https://doi.org/10.1111/j.1469-8749.2008.03117.x>
- Hadders-Algra, M., Boxum, A. G., Hielkema, T., & Hamer, E. G. (2017). Effect of early intervention in infants at very high risk of cerebral palsy: A systematic review. *Developmental Medicine and Child Neurology*, 59(3), 246–258. <https://doi.org/10.1111/dmcn.13331>

- Hielkema, T., Hamer, E. G., Boxum, A. G., Gemert, S. L. B.-V., Dirks, T., Reinders-Messelink, H. A., Maathuis, C. G. B., Verheijden, J., Geertzen, J. H. B., & Hadders-Algra, M. (2020). LEARN2MOVE 0–2 years, a randomized early intervention trial for infants at very high risk of cerebral palsy: Neuromotor, cognitive, and behavioral outcome. *Disability and Rehabilitation*, 42(26), 3752–3761. <https://doi.org/10.1080/09638288.2019.1610508>
- Hielkema, T., Toonen, R. F., Hooijsma, S. J., Dirks, T., Reinders-Messelink, H. A., Maathuis, C. G. B., Geertzen, J. H. B., & Hadders-Algra, M. (2018). Changes in the content of pediatric physical therapy for infants: A quantitative, observational study. *Physical & Occupational Therapy In Pediatrics*, 38(5), 457–488. <https://doi.org/10.1080/01942638.2017.1405863>
- Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D., Altman, D. G., Barbour, V., Macdonald, H., Johnston, M., Lamb, S. E., Dixon-Woods, M., McCulloch, P., Wyatt, J. C., Chan, A.-W., & Michie, S. (2014). Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide. *BMJ (Clinical Research Ed.)*, 348, g1687. <https://doi.org/10.1136/bmj.g1687>
- Ives, Y. (2008). What is “coaching”? An exploration of conflicting paradigms. *International Journal of Evidence Based Coaching and Mentoring*, 6(2), 100–113.
- Kemp, P., & Turnbull, A. P. (2014). Coaching with parents in early intervention: An interdisciplinary research synthesis. *Infants & Young Children*, 27(4), 305–324. <https://doi.org/10.1097/IYC.0000000000000018>
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3), 276–282. <https://doi.org/10.11613/BM.2012.031>
- Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., Elbourne, D., Egger, M., & Altman, D. G. (2010). CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *BMJ (Clinical Research Ed.)*, 340, c869. <https://doi.org/10.1136/bmj.c869>
- Palisano, R., Rosenbaum, P., Walter, S., Russell, D., Wood, E., & Galuppi, B. (1997). Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Developmental Medicine and Child Neurology*, 39(4), 214–223. <https://doi.org/10.1111/j.1469-8749.1997.tb07414.x>
- Rezze, B. D., Law, M., Eva, K., Pollock, N., & Gorter, J. W. (2013). Development of a generic fidelity measure for rehabilitation intervention research for children with physical disabilities. *Developmental Medicine and Child Neurology*, 55(8), 737–744. <https://doi.org/10.1111/dmcn.12114>
- Romano, M., & Schnurr, M. (2022). Mind the gap: Strategies to bridge the research-to-practice divide in early intervention caregiver coaching practices. *Topics in Early Childhood Special Education*, 42(1), 64–76. <https://doi.org/10.1177/0271121419899163>
- Rosenbaum, P., King, S., Law, M., King, G., & Evans, J. (1998). Family-centred service. *Physical & Occupational Therapy In Pediatrics*, 18(1), 1–20. https://doi.org/10.1080/J006v18n01_01
- Rush, D., Shelden, M., & Dunn, W. (2011). *The early childhood coaching handbook* (Tion). BROOKES PUB.
- Schwellnus, H., King, G., & Thompson, L. (2015). Client-centred coaching in the paediatric health professions: A critical scoping review. *Disability and Rehabilitation*, 37(15), 1305–1315. <https://doi.org/10.3109/09638288.2014.962105>
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428. <https://doi.org/10.1037/0033-2909.86.2.420>
- Sonderer, P., Akhbari Ziegler, S., Gressbach Oertle, B., Meichtry, A., & Hadders-Algra, M. (2017). Psychometric properties of a standardized observation protocol to quantify pediatric physical therapy actions. *Pediatric Physical Therapy*, 29(3), 244–250. <https://doi.org/10.1097/PEP.0000000000000398>
- Vanderveen, J. A., Bassler, D., Robertson, C. M. T., & Kirpalani, H. (2009). Early interventions involving parents to improve neurodevelopmental outcomes of premature infants: A meta-analysis. *Journal of Perinatology*, 29(5), 343–351. <https://doi.org/10.1038/jp.2008.229>
- Ward, R., Reynolds, J. E., Pieterse, B., Elliott, C., Boyd, R., & Miller, L. (2020). Utilisation of coaching practices in early interventions in children at risk of developmental disability/delay: A systematic review. *Disability and Rehabilitation*, 42(20), 2846–2867. <https://doi.org/10.1080/09638288.2019.1581846>