

Measurement properties of questionnaires assessing participation in children and adolescents with a disability: a systematic review

Linda Rainey · Ruth van Nispen · Carlijn van der Zee · Ger van Rens

Accepted: 16 June 2014 / Published online: 27 June 2014
© Springer International Publishing Switzerland 2014

Abstract

Purpose To critically appraise the measurement properties of questionnaires measuring participation in children and adolescents (0–18 years) with a disability.

Methods Bibliographic databases were searched for studies evaluating the measurement properties of self-report or parent-report questionnaires measuring participation in children and adolescents (0–18 years) with a disability. The methodological quality of the included studies and the results of the measurement properties were evaluated using a checklist developed on consensus-based standards.

Results The search strategy identified 3,977 unique publications, of which 22 were selected; these articles evaluated the development and measurement properties of eight different questionnaires. The Child and Adolescent Scale of Participation was evaluated most extensively, generally showing moderate positive results on content validity, internal consistency, reliability and construct validity. The

remaining questionnaires also demonstrated positive results. However, at least 50 % of the measurement properties per questionnaire were not (or only poorly) assessed.

Conclusions Studies of high methodological quality, using modern statistical methods, are needed to accurately assess the measurement properties of currently available questionnaires. Moreover, consensus is required on the definition of the construct ‘participation’ to determine content validity and to enable meaningful interpretation of outcomes.

Keywords Disability · Children · Adolescents · Validation studies · Psychometrics · Participation

Introduction

Participation is one of the most important outcomes of rehabilitation for children and adolescents with a disability. Successful participation can lead to increased emotional and psychological well-being and ultimately improved quality of life (QoL) [1–3]. Participation is essential for the development of skill competencies, socialisation, exploring personal interests and the enjoyment of life [4]. For example, without the opportunity to participate in leisure activities ‘people are unable to explore their social, intellectual, emotional, communicative and physical potential and are less able to grow as individuals’ [5]. Although several measures of participation are available for children and adolescents with a disability, there is no general consensus on the definition of ‘participation’. A clear definition of the construct is essential to determine the validity of questionnaires, and for meaningful selection and usage of these instruments [6].

Since the development of the International Classification of Functioning, Disability and Health for Children and

Electronic supplementary material The online version of this article (doi:10.1007/s11136-014-0743-3) contains supplementary material, which is available to authorized users.

L. Rainey (✉) · R. van Nispen · G. van Rens
Department of Ophthalmology, EMGO + Institute for Health and Care Research, VU University Medical Centre,
PO Box 7057, 1007 MB Amsterdam, The Netherlands
e-mail: l.rainey@vumc.nl

C. van der Zee
Department of Epidemiology and Biostatistics, EMGO +
Institute for Health and Care Research, VU University Medical
Centre, PO Box 7057, 1007 MB Amsterdam, The Netherlands

G. van Rens
Department of Ophthalmology, Elkerliek Hospital,
Wesselmanlaan 25, 5707 HA Helmond, The Netherlands

Youth (ICF-CY), several instruments have been modelled on the World Health Organisation's definition of participation [7]. The ICF-CY defines participation as 'a person's involvement in a life situation'. This indicates that participation represents the societal perspective of functioning and includes the concept of involvement, which is further defined by the ICF-CY as 'taking part, being included or engaged in an area of life, being accepted or having access to needed resources'. The authors add that 'the concept of involvement should also be distinguished from the subjective experience of involvement (a sense of belonging or satisfaction with the extent of one's involvement)'. Participation is part of the 'Activities and Participation' category of the ICF-CY. Nine domains are included in its assessment: (1) 'learning and applying knowledge', (2) 'general tasks and demands', (3) 'communication', (4) 'mobility', (5) 'self-care', (6) 'domestic life', (7) 'interpersonal interactions and relationships', (8) 'major life areas', and (9) 'community, social and civic life'. However, by combining the two terms, it is often unclear what constitutes 'activity' and what 'participation'. Where participation refers to the involvement in a life situation, activity is regarded as 'the execution of a task' [7]. Researchers have tried to further clarify this distinction [8–15]. According to Whiteneck and Dijkers [8], 'activities' are tasks performed individually. Activities focus on functional performance of an individual that can be done in solitude, while participation is the social role performance as a member of society with or for others. Activities tend to be straightforward and unambiguous, while participation tends to be more abstruse and will generally involve the performance of several activities. Generally, agreeing with this operationalisation, Eyssen et al. [9] defined participation as 'performing roles in the domains of social functioning, family, home, financial, work/education, or in a general domain'. They distinguished activities from participation by stating that participation requires a social context, including both environmental factors and other people. However, McConachie et al. [10] stated that participation might be defined or operationalised differently when children are involved. They argue that the following life situations must be covered by an instrument that aims to measure participation in children and adolescents: 'participation essential for survival', 'participation in relation to child development', 'discretionary participation' and 'educational participation'. The distinction that participation is more likely to involve other people and to be more environmentally dependent may not be beneficial for children. When performing activities, children tend to be in interaction with others (caregivers) or supported in their performance by environmental supports and modifications [5]. McConachie et al. [10] agree with this notion and add: 'it may not be practical to place a clear boundary around

the child when describing their participation; survey instruments should encompass the notion that for some purposes the child participates as part of a family rather than as an individual'. In accordance with previous research, using the framework of the ICF-CY and developing distinct sets of activity domains and participation domains by dividing the current classification into two mutually exclusive lists (without overlap) seems to be a useful strategy [8].

To contribute to evidence-based instrument selection, the measurement properties of questionnaires evaluating participation in children and adults with a disability have to be evaluated. Several reviews have been conducted [16–19]; most of these reviews have focused on instruments applied in specific populations (e.g. acquired brain injury, cerebral palsy). Morris et al. [16] demonstrated that all selected measures have sound psychometric properties; however, when selecting instruments, no distinction was made between the constructs of 'activity' and 'participation'. Moreover, four of the included instruments pertained to measure QoL or general health status. Sakzewski et al. [17] showed that most participation instruments included in their study had adequate reliability and validity, but limited data were available to determine the responsiveness of several questionnaires; they also stated that 'a combination of assessments is required to capture participation of children in home, school and community environments'. In their review, Ziviani et al. [18] noted that there is a paucity of information on the psychometric properties of participation instruments; moreover, each systematic review used a different definition of participation for instrument selection.

Besides a lack of consensus on the definition of participation, previous reviews of participation measures have also lacked the application of an adequate tool to critically appraise the methodological quality of the included studies [16–19]. This is required in order to (a) draw valid and reliable conclusions about the methodology applied, (b) to be able to reliably infer quality assessments and (c) to recommend specific participation questionnaires. Several authors have proposed guidelines for how measurement properties of health status questionnaires should be measured and criteria for what constitutes good measurement properties [20–23]. The Scientific Advisory Committee (SAC) defined a set of 8 attributes together with some criteria to perform instrument assessments (SACMOS) [20]. The attributes encompassed properties such as 'conceptual and measurement model', 'reliability', 'validity', 'responsiveness' and 'cultural and language adaptations'. The criteria mainly focussed on information the author should provide in an article. In addition, a few criteria are offered on acceptable reliability coefficients and standard error of measurement (SEM). Another tool for the

standardised assessment of patient-reported outcome measures (PROs) is Evaluating the Measurement of Patient-Reported Outcomes (EMPRO) [21]. It also addresses 8 attributes, e.g. ‘reliability’, ‘validity’, ‘responsiveness’, ‘interpretability’ and ‘administration burden’. Similar to the SACMOS, the criteria corresponding to these attributes mainly consist of information authors need to report and some suggestions on acceptable values for reliability measures. Andresen reported another standard, offering additional criteria to contribute a grade to the quality of the measurement properties (A: high standard; B: adequate standard; C: low or inadequate standard) [22]. All these guidelines offer some insight into the standardised assessment and appraisal of measurement properties; however, they all lack comprehensive, detailed and consensus-based descriptions of what constitutes an adequate measurement property. Although some recommendations were made about acceptable values for Cronbach’s alpha, intraclass correlation coefficient (ICC) or the SEM, no relevant specifics were established for attributing factors e.g. sample size. To assess construct validity, hypotheses testing is often recommended by the standards. However, no information is provided about the number of hypotheses that need to be drawn up and to what extent these hypotheses should be confirmed.

The international ‘COnsensus-based Standards for the selection of health status Measurement Instruments (COSMIN) checklist’ fills this gap. It was developed by more than 50 international professionals and can be applied to evaluate the methodological quality of studies on measurement properties [23]. It offers a standardised assessment method to comprehensively appraise the validity, reliability, responsiveness, generalisability and interpretability and assign a score from poor to excellent to each measurement property. The checklist aims to improve evidence-based instrument selection.

This is the first study to (1) select measures of participation developed for children and adolescents (aged 0–18 years) with a disability according to a clearly pre-defined operationalisation and (2) critically appraise the measurement properties of these measures using the standardised approach of the COSMIN checklist.

The study could contribute to instrument selection, improving the assessment of participation in clinical practice.

Methods

Construct definition

We adhered to the following definition of participation: ‘Social role performance in the domains of interpersonal

relations (e.g. with family or friends), education/employment, recreation and leisure and community life as a member of society in interaction with others’. This definition is in concordance with the method of Whiteneck and Dijkers [8] who also divided the Activity and Participation subscales of the ICF-CY in two mutually exclusive lists. By removing the ICF-CY items covered by ‘(d660) assisting others’ from domain (6): ‘domestic life’, domain six now consists solely out of activities. In order not to discard this subsection of items, it will be added as a new ‘major life area’ to domain (8) entitled: ‘care giving and assisting others’. The final three domains of the ICF-CY: (7) ‘interpersonal interactions and relationships’, (8) ‘major life areas’, and (9) ‘community, social and civic life’ are considered as ‘participation’.

Search strategy

The following computerised databases were searched to obtain a comprehensive result: Medline (1966–Oct 2013), EMBase (1974–Oct 2013) and PsycINFO (1806–Oct 2013). Index terms such as: ‘participation’, ‘personal autonomy’ and ‘role functioning’ were combined with terms identifying minors, for example: ‘infant’, ‘child’, ‘adolescent’ and ‘schoolchild’. To identify articles assessing measurement properties, an adaptation of the sensitive search filter for measurement properties was used, which included terms as ‘reliability’, ‘validity’, ‘generalisability’ and ‘psychometric’ [24]. The full search strategy is available upon request. Reference lists of included articles were screened to identify additional relevant studies.

Selection criteria

Articles were selected for inclusion when it concerned a full-text original article (e.g. not a review or manual), aiming to develop a self-report or parent proxy-report measure of participation in disabled children and adolescents (0–18 years old) or to evaluate one or more of the measurement properties of such a measure. The measure had to meet the definition of participation stated under the heading ‘construct definition’. Disability was defined according to the definition of the WHO: ‘any restriction or lack (resulting from any impairment) of ability to perform an activity in the manner or within the range considered normal for a human being’ [7]. Articles had to be written in English (although the developed/evaluated questionnaires can be in a different language). Articles were excluded when the measurement instrument consisted predominantly of activity items, according to our preset definition. In addition, because QoL and ‘adaptive behaviour’ were not considered to measure the same underlying construct as participation, measures assessing these constructs were

Table 1 COSMIN definitions and operationalisation of measurement properties (adapted from www.cosmin.com)

| Term | | | Definition [28] + operationalisation [27] |
|-----------------------------------|----------------------|------------------------------|--|
| Domain | Measurement property | Measurement property aspects | |
| Reliability | | | The degree to which the measure is free from measurement error |
| Reliability (extended definition) | | | The extent to which scores for patients who have not changed are the same for repeated measurement under several conditions: e.g. using different sets of items from the same health-related patient-reported outcomes (HR-PRO) (internal consistency); over time (test–retest); by different persons on the same occasion (inter-rater); or by the same persons (i.e. raters or responders) on different occasions (intrarater) |
| | Internal consistency | | The degree of interrelatedness among the items, calculated with Cronbach's α or the Kuder–Richardson Formula 20 (KR-20). A positive rating is obtained when the subscale is unidimensional and Cronbach's $\alpha \geq 0.70$ |
| | Reliability | | The proportion of the total variance in the measurements which is due to 'true' differences between patients. For continuous scores, it is expressed by the intraclass correlation coefficient (ICC) and for dichotomous, nominal and ordinal scores by Kappa A positive score is acquired when the ICC or the weighted Kappa is ≥ 0.70 or when Pearson's is $r \geq 0.80$ |
| | Measurement error | | The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured. It can be reported using the standard error of measurement (SEM), which can be converted into the smallest detectable change (SDC). Limits of agreement (LoA) can also be calculated. In order to allow for meaningful interpretation, information regarding the minimal important change (MIC) needs to be provided. This enables to relate the measurement error (noise) to the 'signal' that one wants to assess. A positive score is obtained when the MIC is larger than the SDC, or when the MIC is outside the LoA |
| Validity | | | The degree to which an HR-PRO instrument measures the construct(s) it purports to measure |
| | Content validity | | The degree to which the content of an HR-PRO instrument is an adequate reflection of the construct to be measured. Items need to be relevant to the construct, the study population and the purpose of the measurement instrument. A positive score is acquired when the target population considers all items in the questionnaire to be relevant and considers the questionnaire to be complete |
| | Construct validity | | The degree to which the scores of an HR-PRO instrument are consistent with hypotheses (for instance with regard to internal relationships, relationships to scores of other instruments, or differences between relevant groups) based on the assumption that the HR-PRO instrument validly measures the construct to be measured |
| | | Structural validity | The degree to which the scores of an HR-PRO instrument are an adequate reflection of the dimensionality of the construct to be measured. This can be determined by performing confirmatory or exploratory factor analysis or an IRT test. When the identified factors explain at least 50 % of the variance, a positive score can be assigned |
| | | Hypotheses testing | The degree to which the scores of an instrument are consistent with hypotheses based on the assumption that the instrument validly measures the construct to be measured. Hypotheses regarding correlation or mean differences (preferably including a prediction of the direction and magnitude) need to be formulated a priori. A positive score can be assigned when: (1) correlations ≥ 0.60 are obtained with an instrument measuring a similar construct; or (2) when at least 75 % of the results are in accordance with the formulated hypotheses; and (3) when correlations with related constructs are higher than correlations with unrelated constructs |
| | | Criterion validity | The degree to which the scores of an instrument are an adequate reflection of a 'gold standard'. Because no 'gold standard' is not available in the field of health-related questionnaires, the criterion validity was not assessed. Instead, the methodological quality of the hypotheses formed in the articles regarding the participation questionnaire and questionnaires measuring a similar construct were evaluated |

Table 1 continued

| Term | | | Definition [28] + operationalisation [27] |
|------------------|----------------------|------------------------------|--|
| Domain | Measurement property | Measurement property aspects | |
| | | Cross-cultural validity | The degree to which the performance of the items on a translated or culturally adapted HR-PRO instrument are an adequate reflection of the performance of the items of the original version of the HR-PRO instrument. Since actual field tests (including confirmatory factor analysis) are rarely performed when adapting a questionnaire to another language, the methodological quality of the translation process will be evaluated instead |
| Responsiveness | | | The ability of an HR-PRO instrument to detect change over time in the construct to be measured. It can be assessed by calculating the area under the receiver operator curve (AUC) or by calculating correlations between change scores. For dichotomous scales sensitivity and specificity can be determined. Requirements necessary to obtain a positive score are similar to those of hypotheses testing, with an additional criterion of the $AUC \geq 0.70$ |
| Interpretability | | | The degree to which one can assign qualitative meaning—that is, clinical or commonly understood connotations—to an instrument’s quantitative scores or change in scores. Although this is not a measurement property, it can be seen as an important characteristic of a measurement instrument. Information about missing items and floor and ceiling effects can be reported here |

excluded. Two researchers (LR, RvN) independently screened the titles, abstracts and full-text articles. For the updated search strategy, one researcher (LR) and a research assistant screened the additional studies independently on title, abstract and full-text. A third researcher (GvR) was consulted when consensus could not be reached through discussion.

Methodological quality assessment and quality criteria

The COSMIN checklist (with a four-point scale) was used to evaluate and calculate overall methodological quality scores per study on a measurement property [25]. It assesses measurement properties on three dimensions: reliability, validity and responsiveness. In addition, general requirements for studies that applied Item Response Theory (IRT) models and evaluated interpretability and generalisability were also appraised when applicable. Items can be rated excellent, good, fair and poor. Two researchers (LR, CvdZ) independently rated the included articles per measurement property. An overall score was established by taking the lowest rating of the items in a box. A third researcher (RvN) helped reach consensus when necessary. Although the inter-rater agreement of the checklist is adequate, the item inter-rater agreement is low; this is thought to be due to the required subjective judgement and being accustomed to different standards [26]. Therefore, decisions were made a priori regarding the appraisal of items. Quality criteria by Terwee et al. [27] were used to rate the quality of the evaluated measurement properties. The COSMIN definitions of the

measurement properties and the quality criteria are described in Table 1.

Synthesis of results

The evidence on the measurement properties of the questionnaires was synthesised by combining the results, taking into consideration a) the number of studies, b) the methodological quality of the studies and c) the consistency of their results. The overall score can be rated ‘positive’, ‘negative’ or ‘indeterminate due to conflicting evidence’. Criteria developed by Terwee et al. (2007) were used to determine the overall score of the measurement properties per questionnaire [24].

Results

The search strategy identified 3,977 unique publications, of which 277 articles were selected after title and abstract screening. The full-text of these 277 articles was assessed, which resulted in the exclusion of 260 articles. Reference checking identified five additional clinimetric studies. In total 22 articles, evaluating the measurement properties of eight different questionnaires were included in the present study (Fig. 1). No other articles, evaluating the measurement properties of the eight included participation questionnaires, were available. The search strategy identified another 53 measurement instruments that pertained to measure participation or contained several items that could be considered participation items; these instruments were

not included in the review, because the questionnaires did not comprehensively evaluate participation according to the definition used in this review. However, because they do provide insight into our operationalisation of the construct ‘participation’, these questionnaires are presented in Appendix A of ESM: ‘Characteristics of the excluded measurement instruments: construct clarification’. The general characteristics of the included studies and participation measures are presented in Tables 2 and 3, respectively. The methodological quality of each study per measurement property is presented in Table 4. The synthesis of results for each questionnaire and the quality of its measurement properties is presented in Table 5 (+++ or --- Strong evidence positive/negative result). The results for each measure are discussed, separately, below.

Assessment of Preschool Children’s Participation (APCP)

No studies were found examining the measurement error, content validity and structural validity of the APCP. The evaluation of the internal consistency and the responsiveness was of poor methodological quality. Hypothesis testing for the English version of the APCP showed significant differences in diversity and intensity scores for gender, age and income [29, 30]. Moderate positive correlations were found between baseline diversity and intensity scores on the APCP and the PEDI, the WeeFIM and the GMFM-66 ($r = 0.51$ – 0.78 ; $r = 0.46$ – 0.82 ; $r = 0.51$ – 0.77 , respectively) [29, 30]. With regard to the interpretability of the English questionnaire, estimates were provided for the minimal detectable change (MDC) and minimal clinically important difference (MCID) [30]. The MDC_{95} values for diversity scales of the APCP were play (PA): 5.1 %, skill development (SD): 2.5 %, active physical recreation (AP): 7.8 %, social activities (SA): 16.7 %, and total score: 3.8 %. The MDC_{95} values for intensity scales were PA: 0.6, SD: 0.1 AP: 0.5, SA: 0.7, and total score: 0.2. The MCID was estimated based on anchor- and distribution-based approaches. The anchor-based MCID values for diversity scales were PA: 16.7 %, SD: 19.4 %, AP: 11.0 %, SA: 16.5 %, and total score: 16.3 %. The anchor-based MCID values for intensity scales were PA: 1.1, SD: 1.2, AP: 0.8, SA: 0.9, and total score: 1.0. The distribution-based MCID values for diversity scales were PA: 11.7 %, SD: 11.4 %, AP: 11.0 %, SA: 9.6 %, and total score: 10.1 %. The distribution-based MCID values for intensity scales were PA: 0.7, SD: 0.7, AP: 0.6, SA: 0.4, and total score: 0.6. The Dutch APCP was translated from English using one forward and one backward translation [31]. The Dutch version was not pretested. The test–retest reliability of the overall diversity and intensity scores for children with and without physical disabilities is acceptable

(ICC = 0.83–0.91). Two out of five subscales had an ICC below 0.70 [31]. The ICC cannot be evaluated for the group of participants with a disability due to small sample size ($N = 24$). There is moderate positive evidence for hypothesis testing, showing significant differences in diversity and intensity scores for gender, disability and age [31]. Floor and ceiling effects were not reported.

Children’s Assessment of Participation and Enjoyment (CAPE)

No articles were found evaluating the structural validity and responsiveness of the CAPE. The Greek translation provided limited evidence of inadequate internal consistency with Cronbach’s α for the five subscales ranging from 0.08 to 0.64 [32]. The internal consistency of the Swedish [33], Norwegian [34] and Spanish [35, 36] versions could not be evaluated due to assessments of poor methodological quality. The test–retest reliability of the five domains of the Dutch translation was shown to be adequate (ICC = 0.61–0.78) [37]. Limited positive evidence is provided for the intrarater reliability of the Dutch translation (ICC = 0.65–0.83) [37]. There is limited evidence of adequate reliability of the Norwegian translation for children with typical development (ICC_{Total} = 0.66) [34]. The sample of children with a disability was too small to make any meaningful conclusions. The reliability of an English version of the CAPE specifically adapted for children and adolescents with high functioning autism (HFA) was also assessed, but the study was of poor quality due to small sample size ($N = 14$) [38]. The measurement error cannot be formally determined, as the three studies evaluating the measurement error, do not provide an estimate of the minimal important change (MIC). Therefore, a comparison between the smallest detectable change (SDC) or limits of agreement (LoA) and the MIC cannot be made, and thus, the measurement error cannot be assessed. Bult et al. [37] reported SDC values ranging from 0.89 to 1.91 for the inter-rater reliability and from 1.14 to 1.86 for test–retest reliability. Hypotheses testing showed a positive correlation between Dutch CAPE activity scores and scores on instruments measuring family environment ($r = 0.26$ – 0.34), adaptive behaviour ($r = 0.31$ – 0.51) and picture vocabulary ($r = 0.24$) [37]. In addition, positive correlations between English CAPE activity scores and measures of athletic competence ($r = 0.29$) and physical functioning ($r = 0.15$ – 0.42) were found [39, 40]. As well as negative correlations between the CAPE and environmental factors (perceptions of barriers in the physical structural environment; $r = -0.17$) and financial constraints ($r = -0.13$ to -0.21) [39]. Children from families with a lower income participated less often in active physical activities. Activity scores on the Spanish version

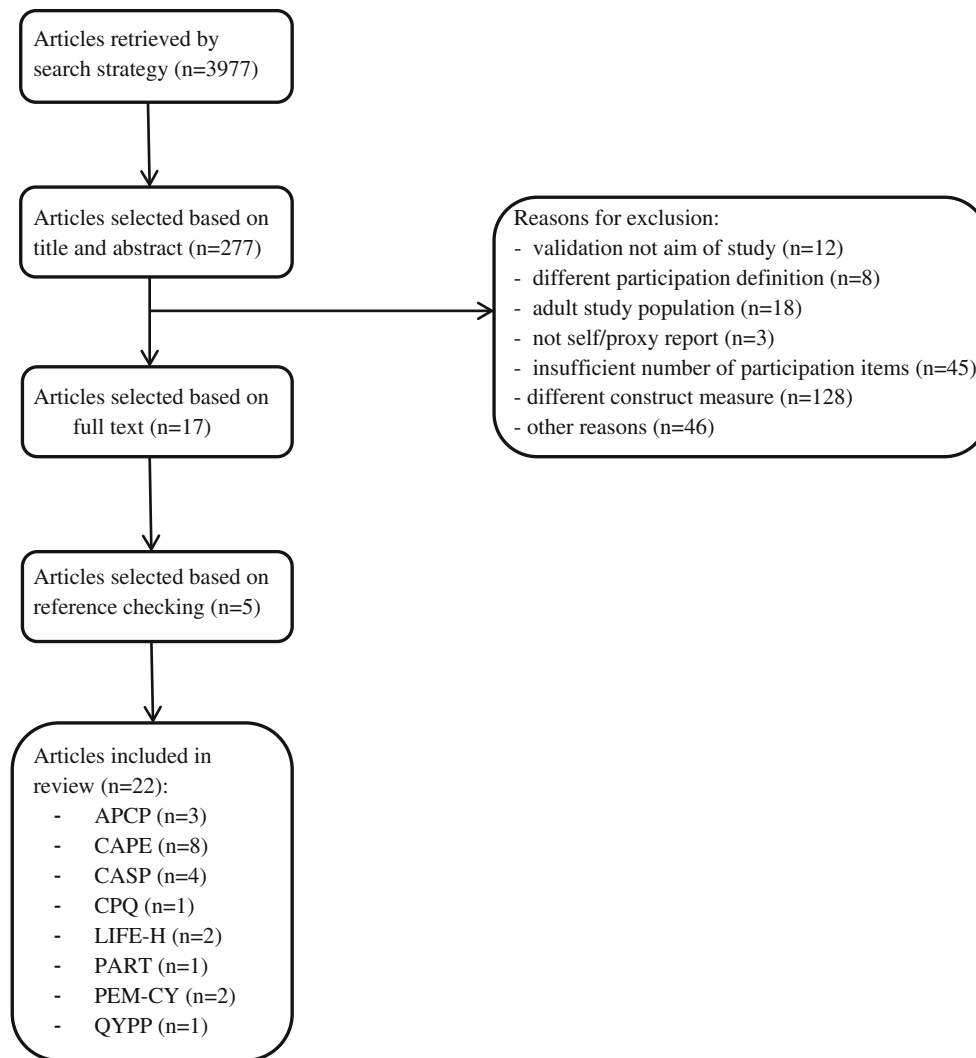


Fig. 1 Flowchart search strategy and article selection

of the CAPE were both positively and negatively correlated with a QoL measure. For example, CAPE diversity scores were positively correlated with ‘social support and peers’ ($r = 0.41$) and negatively correlated with ‘self-perception’ ($r = -0.29$) [36]. This last result was considered to be due to cultural differences. When parents judged their children’s self-perception to be low, they would motivate them to participate in more activities, due to the importance of family interaction and involvement. Differences in scores between subgroups (children vs. adolescents vs. adults; male vs. female; disability vs. no disability) have also been reported [32, 35–37]. No information was available on floor and ceiling effects.

Child and Adolescent Scale of Participation (CASP)

No methodological studies were found evaluating measurement error and responsiveness of the CASP. There is

limited positive evidence for content validity of the Chinese adaptation [41]. The methodological quality of the translation process is fair. Exploratory factor analysis showed limited evidence for a two-factor solution on the parent-report version [41], but there is also limited evidence for a three-factor solution [42, 43]. For the youth version, exploratory factor analyses showed limited evidence for a three-factor solution [43]. Evidence of good internal consistency is demonstrated with regard to the total score of the parent-report measure (Cronbach’s $\alpha = 0.96$) and the youth-report measure (Cronbach’s $\alpha = 0.87$) [41, 43]. The four subscales of the CASP parent report show acceptable internal consistency (Cronbach’s $\alpha = 0.88$ – 0.90) [41]. One study assessed the internal consistency of the subscores on the three-factor solution model identified for both the youth and parent report, resulting in a Cronbach’s α of 0.67–0.90 [43]. Three studies performing Rasch analysis on the parent-report questionnaire showed moderate evidence for a

Table 2 Characteristics of the included studies

| Study | Population | Total <i>n</i> | Gender % male | Age Mean (SD) | Country | Language instrument | Setting |
|-------------------|--|-------------------|---------------------|-------------------------|--------------------------------|------------------------|-----------------------------------|
| Anastasiadi [32] | CP, mental retardation, ADHD, Down syndrome | 302 | 45.7 | n/a | Greece | Greek | School and rehabilitation setting |
| Bedell [44] | Acquired brain injury | 60 | 49.3 | 13.2 (5.2) | USA | English | Home based |
| Bedell [42] | Acquired brain injury, developmental disability | 313 | 55 | 12.8 (4.6) | Australia, Canada, Israel, USA | English | Occupational therapist |
| Bult [37] | CP, Spina bifida, Arthritis | 232 | 47.4 | 12.0 (3.4) ^a | The Netherlands | Dutch | School and home based |
| Bult [31] | Central nervous system and musculoskeletal disabilities | 126 | 67 | 2.9 (1.8) ^a | The Netherlands | Dutch | Email based/home based |
| Chen [30] | CP | 82 | 63.4 | 3.9 (1.4) | Taiwan | English | Physical therapist |
| Colon [35] | Developmental disability | 249 | 50.6 | n/a | Puerto Rico | Spanish | Occupational therapist |
| Coster [51] | Developmental delay, orthopaedic-, emotional impairment | 576 | 54.0 | 11.2 (3.1) | Canada/USA | English | Web based |
| Fougeyrollas [47] | Myelodysplasia | 24 | 33.3 | 10.9 (4.7) | Canada | English | Not specified |
| Hwang [41] | Intellectual disability, physical disability, autism spectrum disorder | 231 | 58 | 11.6 (3.5) | China | Chinese | Hospital |
| Kemps [49] | Congenital deformities of the extremities, spina bifida, CP | 325 | 63.0 | 3.8 (0.9) ^a | The Netherlands | Dutch | Home based |
| King [39] | Central nervous system disorders and musculoskeletal disorders | 426 | 53.6 | n/a | Canada | English | Home based |
| Law [29] | CP | 120 | 59.2 | 4.1 (1.1) | Canada | English | Not specified |
| Longo [36] | CP | 398 | 56.8 | 12.12 (3.02) | Spain | Spanish | Rehabilitation centre |
| McDougall [43] | CP, acquired brain injury, communication/cleft lip-palate | 430 | 55.0 | 14.0 (2.2) | Canada | English | Treatment centre or home based |
| Nordtorp [34] | CP, muscular disease, metabolic disorder | 141 | 47.1 | 14.2 (2.3) ^a | Norway | Norwegian | Home based, rehabilitation centre |
| Noreau [48] | CP, myelomeningocele, traumatic brain injury | 94 | 39.6 | 8.10 (2.4) ^a | Canada | English | Not specified |
| Potvin [38] | High functioning autism | 61 | 86.9 | n/a | Canada | English | Not specified |
| Rosenberg [45] | Developmental disability | 480 | 81.3 | 5.16 (0.66) | Israel | English | Home based |
| Tuffrey [52] | CP | 107 | 60 | 16.6 (n/a) | United Kingdom | English | Home based |
| Ullenhag [33] | Normal development | 337 | 51.0 | 12.0 (2.0) | Sweden | Swedish | School |

CP Cerebral palsy, ADHD attention deficit and hyperactivity disorder, n/a not available

^a Only the mean age of the population with a disability was reported, not for the total sample

unidimensional construct [41, 42, 44]. The average CASP item difficulty order ranged from 1.46 logits to -1.51 logits and from 1.36 logits to -1.97 logits [42, 44]. All three

studies identified items pertaining to community participation as most challenging, whereas items regarding skills learned at a younger age, such as mobility, communication

Table 3 General characteristics of the selected participation measures

| Questionnaire(publication date) | Framework | Target population | Dimensions/subscales | Mode of administration | Number of items, response options | Administration time | Levels of scoring | Available in |
|---------------------------------|---|---|---|-------------------------------|--|---------------------|---|--|
| APCP (2012) | ICF-CY | 2–5 years; with and without disabilities | Play; skill development; active physical recreation; social activities | Parent report | 45 Items, yes/no; 7-point ordinal scale | 30–40 min | Two scores: diversity and Intensity | English, Dutch |
| CAPE (2004) | ICF | 6–21 years; with and without disabilities | Diversity; Intensity; Enjoyment; With whom; Where | Self-report | 55 Items, 7-point scale | 30–45 min | Overall score, domain score (direct and indirect activities), activity type score (recreational, physical, social, skill-based, self-improvement) | English, Greek, Dutch, Spanish, Norwegian, Swedish |
| CASP (2004) | ICF | >5 years; brain injury | Home; School; community; home and community living activities | Parent report/ self-report | 20 Items, 4-point scale | 10–15 min | Overall score and four subscale scores | English, Chinese |
| CPQ (2010) | Occupational Therapy Practice Framework | 4–6 years; with and without a disability | Activities of daily living (ADL); Instrumental activities of daily living (IADL); Play; leisure; Social participation; Education | Parent report | 45 items, 6-point Likert scale | n/a | Five total scores: diversity, intensity, independence, enjoyment, parent satisfaction. Six subscale scores | English |
| LIFE-H (1998) | Disability Creation Process | 5–13 years; with disabilities | Nutrition; Fitness; Communication; Mobility; Responsibility; Education; Recreation; Personal care; Housing; Community life; Relationships | Parent report | 64 and 197 items (short and long version), 9-point scales; 5-point scale | n/a | Two concepts : Functional skills and Caregiver assistance; A total score and 11 subscale scores | English |
| PART (2011) | ICF-CY | 2–5 years; with and without disabilities | Interpersonal interactions and relationships; Major life areas; Community, social and civic life | Parent report | 44 items, 4-point scale | n/a | Total score and 3 subscale scores | Dutch |
| PEM-CY (2011) | ICF-CY | 5–17 years; with and without disabilities | Home; School; Community | Parent report | 58 items, 4-point scales; 5-point scales; 8-point scale | n/a | Six scores: Frequency, Involvement, Desire for change, Environment: supportiveness, Environment: supports, Environment: resources | English |

Table 3 continued

| Questionnaire(publication date) | Framework | Target population | Dimensions/subscales | Mode of administration | Number of items, response options | Administration time | Levels of scoring | Available in |
|---------------------------------|-----------|-----------------------------|--|-------------------------------|-----------------------------------|---------------------|-------------------------------|--------------|
| QYPP (2013) | ICF | 14–21 years; cerebral palsy | Home life; Getting on with other people; School or college life; Work life; Recreation and leisure; Autonomy; Preparing for the future | Parent report/ self-report | 45 items, 7-point likert scale | 20–30 min | A total score and 7 subscores | English |

APCP Assessment of Preschool Children's Participation, CAPE Children's Assessment of Participation and Enjoyment, CASP Child and Adolescent Scale of Participation, CPQ Children Participation Questionnaire, LIFE-H Assessment of Life Habits, PEM-CY Participation and Environment Measure for Children and Youth, QYPP Questionnaire of Young People's Participation

and self-care were identified as least challenging. Three items are identified as potential misfits or deviant to the Rasch measurement model. Inter-rater reliability examining agreement between parent and youth report showed moderate agreement on the total score (ICC = 0.63) [43]. On the subscales of the three-factor solution, there is limited evidence that the reliability is inadequate, showing limited to moderate agreement (ICC = 0.51–0.70). Hypotheses testing showed that the CASP has a positive correlation with an instrument measuring functional skills ($r = 0.51–0.72$) and a negative correlation with instruments measuring extent of impairment ($r = -0.58$ to -0.66) and environmental barriers ($r = -0.43$ to -0.57) [42, 44]. It was also found that children without a disability have significantly higher and less variable CASP scores than children with a disability ($p < 0.001$) [41]. Floor and ceiling effects have been reported [41–44]. No information was available on the MIC.

Children Participation Questionnaire (CPQ)

One methodological study evaluated the following measurement properties of the CPQ: internal consistency, reliability, content validity, and hypotheses testing [45]. No studies were available on the measurement error, structural validity and responsiveness of the questionnaire. There is limited positive evidence for content validity. The evaluation of the internal consistency was of poor quality (unidimensionality of the scale was not checked) and therefore yields little information. There is limited positive evidence for adequate test–retest reliability with the ICC for the total scores ranging from 0.84 to 0.89 and the ICC for subscores ranging from 0.72 to 1.00. Hypotheses testing showed that the questionnaire can distinguish between subgroups (age, socioeconomic status and disability vs. no disability). Floor and ceiling effects or the MIC have not been reported.

Assessment of Life Habits (LIFE-H)

There were no original methodological studies evaluating the internal consistency, measurement error, structural validity and responsiveness of the child version of the LIFE-H [46]. There is limited positive evidence for content validity [47, 48]. There is limited positive evidence for acceptable inter-rater reliability (ICC = 0.63–0.93) [48]. In addition, limited positive evidence is provided for satisfactory intrarater reliability for 10 of the 11 subscales of the short questionnaire (ICC > 0.78) [48]. Hypotheses testing showed positive correlations between the LIFE-H and questionnaires measuring functional capabilities ($r = 0.70–0.94$). The methodological quality of this assessment was rated 'poor', because the comparator

Table 4 Methodological quality of each study per questionnaire

| Study | Internal consistency | Reliability | Measurement error | Content validity | Structural validity | Hypotheses testing | Cross-cultural validity ^a | Responsiveness |
|--------------------------|----------------------|-------------|-------------------|------------------|---------------------|--------------------|--------------------------------------|----------------|
| <i>APCP</i> | | | | | | | | |
| Bult et al. [31] | Poor | Fair | | | | Fair | Poor | |
| Chen et al. [30] | | | | | | Fair | | Poor |
| Law et al. [29] | Poor | | | | | Fair | | |
| <i>CAPE</i> | | | | | | | | |
| Anastasiadi et al. [32] | Good | | | | | Good | Fair | |
| Bult et al. [37] | | Good | Good | | | Good | Poor | |
| Colon et al. [35] | Poor | | | | | Good | Poor | |
| King et al. [39] | | | | | | Good | | |
| Longo et al. [36] | | Poor | Poor | | | Good | Good | |
| Nordtorp et al. [34] | Poor | Poor | Poor | | | | Fair | |
| Potvin et al. [38] | | Poor | | Fair | | | | |
| Ullenhag et al. [33] | Poor | | | | | | Good | |
| <i>CASP</i> | | | | | | | | |
| Parent report | | | | | | | | |
| Bedell [44] | Poor | Poor | | | Poor | Fair | | |
| Bedell [42] | Poor | | | | Fair | Poor | | |
| Hwang et al. [41] | Good | | | Fair | Good | Fair | Fair | |
| Youth report | | | | | | | | |
| McDougall et al. [43] | Excellent | Good | | | Excellent | Poor | | |
| <i>CPQ</i> | | | | | | | | |
| Rosenberg et al. [45] | Poor | Fair | | Good | | Good | | |
| <i>LIFE-H</i> | | | | | | | | |
| Fougeyrollas et al. [47] | | Poor | | Fair | | | | |
| Noreau et al. [43] | | Good | | Fair | | Poor | | |
| <i>PART</i> | | | | | | | | |
| Kemps et al. [49] | | Good | | | | Good | | |
| <i>PEM-CY</i> | | | | | | | | |
| Coster et al. [51] | Poor | Fair | | | | Fair | | |
| Coster et al. [50] | | | | Fair | | | | |
| <i>QYPP</i> | | | | | | | | |
| Tuffrey et al. [52] | Fair | Good | | Fair | Poor | Fair | | |

APCP Assessment of Preschool Children's Participation, *CAPE* Children's Assessment of Participation and Enjoyment, *CASP* Child and Adolescent Scale of Participation, *CPQ* Children Participation Questionnaire, *LIFE-H* Assessment of Life Habits, *PEM-CY* Participation and Environment Measure for Children and Youth, *QYPP* Questionnaire of Young People's Participation

^a Only the translation process was evaluated

instruments were not adequately described regarding construct and methodological properties. Differences in scores between subgroups (cerebral palsy, neuropathy, myelomeningocele) were noted [48]. Floor and ceiling effects or the MIC have not been reported.

PART

No methodological studies were found evaluating the internal consistency, measurement error, content validity, structural

validity and responsiveness of the PART. There is limited positive evidence for reliability for the total score (ICC = 0.92) and for each of the subscales (ICC = 0.84–0.89) [49]. Hypothesis testing showed positive correlations between PART scores and scores on a measure of functional skills ($r = 0.35$ – 0.62) [46]. Differences in scores between subgroups (mobility limitations vs. no mobility limitations and environmental barriers vs. no experienced environmental barriers) have been reported [49]. No information was available on floor or ceiling effects or on the MIC.

Table 5 Result synthesis of measurement properties per questionnaire

| Questionnaire | Internal consistency | Reliability | Measurement error | Content validity | Structural validity | Hypothesis testing | Responsiveness |
|---------------|----------------------|-------------|-------------------|------------------|---------------------|--------------------|----------------|
| APCP | ? | + | N/A | N/A | N/A | ++ | ? |
| CAPE | – | + | + | + | N/A | ++ | N/A |
| CASP | ++ | ± | N/A | + | ± | ++ | N/A |
| CPQ | ? | + | N/A | + | N/A | + | N/A |
| LIFE-H | N/A | + | N/A | + | N/A | ? | N/A |
| PART | N/A | + | N/A | N/A | N/A | + | N/A |
| PEM-CY | ? | + | N/A | + | N/A | + | N/A |
| QYPP | + | + | N/A | + | ? | + | N/A |

APCP Assessment of Preschool Children's Participation, *CAPE* Children's Assessment of Participation and Enjoyment, *CASP* Child and Adolescent Scale of Participation, *CPQ* Children Participation Questionnaire, *LIFE-H* Assessment of Life Habits, *PEM-CY* Participation and Environment Measure for Children and Youth, *QYPP* Questionnaire of Young People's Participation

++ or -- moderate evidence positive/negative result, + or – limited evidence positive/negative result, ± conflicting evidence, ? Only poor methodological studies, N/A information not available

Participation and environment measure for children and youth (PEM-CY)

No methodologically studies were found assessing the measurement error, structural validity and the responsiveness of the PEM-CY. The assessment of internal consistency was of poor quality and will therefore not be reported. There is limited positive evidence of content validity [50]. Limited positive evidence is provided for reliability on the total scores for 'participation frequency' in each of the three settings ($ICC_{\text{school}} = 0.58$, $ICC_{\text{home}} = 0.84$ and $ICC_{\text{community}} = 0.79$). The reliability for individual items within each setting varied within the same range ($ICC_{\text{home}} = 0.68\text{--}0.96$, $ICC_{\text{school}} = 0.73\text{--}0.91$ and $ICC_{\text{community}} = 0.73\text{--}0.93$). Reliability estimates for the other sections across settings were all moderate to good ($ICC = 0.66\text{--}0.96$) [51]. Hypotheses testing showed a significant effect of age group for 'participation involvement' in both the home settings ($df = 3,512\text{--}3,568$; $F = 7.17$) and school settings ($df = 3,485\text{--}3,506$; $F = 3.81$). A significant negative correlation between the 'desire for change' score and the 'environmental supportiveness' total score was found for each setting ($r_{\text{home}} = -0.42$, $r_{\text{school}} = -0.59$ and $r_{\text{community}} = -0.53$) [51]. No information was available on floor or ceiling effects or on the MIC.

Questionnaire of Young People's Participation (QYPP)

No methodologically sound studies were found evaluating structural validity of the QYPP. The measurement error and responsiveness of the questionnaire were not assessed. One methodological study evaluated the internal consistency, reliability, content validity and hypotheses testing [52]. There is limited positive evidence for content validity.

Due to small sample size in relation to the number of items in the questionnaire ($N = 107$), the assessment of internal consistency offered insufficient information (Cronbach's $\alpha = 0.61\text{--}0.86$). There is limited positive evidence for adequate test–retest reliability with ICCs ranging from 0.83 to 0.98. Hypotheses testing showed that the questionnaire could distinguish between people with CP and the general population (Mann–Whitney U test $P < 0.01$ for all domains). Floor effects have been reported. No information regarding the MIC was available.

Discussion

The methodological quality of studies evaluating the measurement properties of eight different questionnaires measuring participation in children and adolescents with disabilities was evaluated, using the COSMIN taxonomy. Overall, the CASP was evaluated most extensively, generally showing moderate positive results on the assessed measurement properties. Remarkably, very few studies evaluating the measurement properties of participation questionnaires were available. In addition, at least 50 % of the measurement properties per questionnaire were not (or only poorly) assessed. Therefore, no final conclusions can be made about the methodological quality of the questionnaires. For some questionnaires (i.e. QYPP, APCP), this is understandable, because the measures are relatively new. However, other participation measures could have been validated more comprehensively.

The content validity of several of the questionnaires was not assessed. Good content validity is a prerequisite for sound validity and reliability. The fact that there is no general consensus on the definition of the construct 'participation' highlights the importance of a well-substantiated

framework, on which the questionnaire should be built. The finding that the items within this framework are often not appraised with regard to the relevance to the construct, study population and the purpose of the measurement instrument, gives cause for concern. This is not an attempt to imply that the content validity of the included questionnaires is of poor quality. It is merely an observation that emphasises the necessity of analysing the content validity of these questionnaires in future studies. The notion that each participation questionnaire is developed using a different take on the definition of the construct was highlighted in Appendix A of ESM. Several questionnaires purported to measure the construct of participation [i.e. Children Helping Out—Responsibilities, Expectations and Support (CHORES), Pediatric Community Participation Questionnaire (PCPQ), and Rotterdam Handicap Scale (RHS)], but upon inspection, the items of the measures were often single tasks performed alone [53–55]. The definition used in the present review combined generally accepted aspects of previously developed descriptions (i.e. role performance, interaction and community life). General agreement is needed in order to attribute meaningful interpretations to the outcomes of the questionnaires.

Another measurement property that is underreported is the responsiveness. This is unexpected, as questionnaires measuring participation are often used in rehabilitation settings, where increased participation is a main treatment outcome. To make meaningful comments about patients' progress in participation, the responsiveness of the questionnaires needs to be evaluated in longitudinal studies of good methodological quality; this will have positive implications for both research and clinical practice. Angst has voiced his concern about the COSMIN rules used to examine responsiveness [56]. According to Angst, responsiveness aims to detect change over time in the construct of interest. He argues that this is not solely a question of longitudinal validity, but a matter of determining which instrument detects changes over time more accurately using a quantitative measure. According to the COSMIN taxonomy, these methods are considered inappropriate and deem a study to be of poor methodological quality. Whereas it is true that other methods can be (and often are) used to evaluate the responsiveness of the questionnaire (i.e. effect size), these methods provide less insight than previously thought. The COSMIN panel argued that the effect size can only be used as a measure of responsiveness if the effect of an intervention has been determined or assumed beforehand [56]. However, these methods need not be disregarded. Although the COSMIN taxonomy has been criticised for its adherence to optimal statistical methods, rather than generally accepted and commonly used methods, it remains a consensus-based tool, which offers a standardised way of assessing

measurement properties. It is crucial to use one's own methodological knowledge and insight and apply the COSMIN checklist as a framework, not as an absolute truth. For clinical practice, it is important that future research aims to study the responsiveness of these measures, to enable valid evaluation of clients' progress and to allow for (cost) effectiveness studies assessing current treatment techniques. Despite of this clinical relevance of responsiveness, it has to be noted that for an individual child, a higher score on a participation questionnaire does not necessarily equal preferred improvement. It can indicate that a child is capable to perform certain activities more easily, but these might not be considered important or relevant by the child or parents. A higher score can also indicate that the child has received more help, rather than independently performed the activities. Therefore, scores on a participation questionnaire always need to be interpreted qualitatively as well as quantitatively. When looking at the qualitative applicability of these questionnaires, the responsiveness is a less important measurement property.

Cross-cultural validity was difficult to assess for the included questionnaires. The adaptation of questionnaires for other cultures and languages requires a rigorous and integral process of expert translation (including multiple forward and backward translations) item revision and pretesting in a similar study population [28]. Particularly, the pretest is often disregarded, but remains an essential part of the validation process. A simple translation of the items is not sufficient as the meaning and significance of items might vary according to culture, setting and circumstance. A poorly executed cross-cultural adaptation could result in less optimal findings regarding the validity and reliability of the instrument. This could explain the inconsistent findings in the present study when comparing the reliability and validity of the original questionnaire and some of the translated versions.

The studies of poor methodological quality showed similar shortcomings: small sample size and omitting to execute (confirmatory) factor analysis. Performing factor analysis is an important method to evaluate the internal consistency, structural validity and cross-cultural validity of an instrument. By excluding this analysis, no valuable information is obtained about the (uni) dimensionality of the scales and the distribution of items. Several studies still proceed to determine the internal consistency of a scale, without looking at unidimensionality, or even when unidimensionality has been disproved [31, 35, 42, 44]. Adhering to general statistical requirements when evaluating measurement properties will improve the methodological quality of these studies.

This is the first systematic review in which the measurement properties of eight participation questionnaires were analysed using a standardised, consensus-based

taxonomy, preceded by a construct operationalisation process. Based on the results, it can be concluded that there is still a shortage of good quality information regarding the psychometric properties of questionnaires measuring participation in children and adolescents with a disability. Therefore, a recommendation for future research is to assess the psychometric properties of these identified questionnaires using good qualitative research methods. The COSMIN checklist can be consulted when determining which statistical methods are required and preferred when assessing these properties. IRT can be a valuable and useful tool to determine the quality of a measurement instrument [57–59].

Future research should pay special attention to the content validity and the responsiveness of the questionnaires. The development of new questionnaires measuring participation in a general population of children and adolescents is not considered a direct priority at this time. Recently, several new questionnaires have been developed (e.g. APCP, PART, PEM-CY, and QYPP). Therefore, it is recommended to evaluate existing questionnaires using studies of high methodological quality, preferably using IRT models, to contribute to the practical application of the instruments and to be able to accurately measure participation in children and adolescents with disabilities.

References

- Petrenchik, T. M., & King, G. A. (2011). Pathways to positive development: Childhood participation in everyday places and activities. In S. Bazyk (Ed.), *Mental health promotion, prevention, and intervention in children and youth: A guiding framework for occupational therapy* (pp. 71–94). Bethesda, MD: American Occupational Therapy Association.
- McManus, V., Corcoran, P., & Perry, I. J. (2008). Participation in everyday activities and quality of life in pre-teenage children living with cerebral palsy in south west Ireland. *BMC Pediatrics*, *8*(50), 1–10.
- Law, M. (2002). Participation in the occupations of everyday life. *American Journal of Occupational Therapy*, *56*, 640–649.
- Simpkins, S. D., Ripke, M., Huston, A. C., & Eccles, J. S. (2005). Predicting participation and outcomes in out-of-school activities: Similarities and differences across social ecologies. *New Directions for Youth Development*, *105*, 51–69.
- King, G., Law, M., King, S., Rosenbaum, P., Kertoy, M. K., & Young, N. L. (2003). A conceptual model of the factors affecting the recreation and leisure participation of children with disabilities. *Physical and Occupational Therapy in Pediatrics*, *23*, 63–90.
- Coster, W., & Khetani, M. A. (2008). Measuring participation of children with disabilities: Issues and challenges. *Disability and Rehabilitation*, *30*, 639–648.
- WHO. (2007). *International classification of functioning, disability and health for children and youth (ICF-CY)*. Geneva: World Health Organization.
- Whiteneck, G., & Dijkers, M. P. (2009). Difficult to measure constructs: Conceptual and methodological issues concerning participation and environmental factors. *Archives of Physical Medicine and Rehabilitation*, *90*(supplement), 1.
- Eyssen, I. C., Steultjens, M. P., Dekker, J., & Terwee, C. B. (2011). A systematic review of instruments assessing participation: Challenges in defining participation. *Archives of Physical Medicine and Rehabilitation*, *92*, 983–997.
- McConachie, H., Colver, A. F., Forsyth, R. J., Jarvis, S. N., & Parkinson, K. N. (2006). Participation of disabled children: How should it be characterised and measured? *Disability and Rehabilitation*, *28*, 1157–1164.
- Colver, A. (2005). A shared framework and language for childhood disability. *Developmental Medicine and Child Neurology*, *47*, 780–784.
- Hemmingsson, H., & Jonsson, H. (2005). An occupational perspective on the concept of participation in the International Classification of Functioning, Disability and Health—Some critical remarks. *American Journal of Occupational Therapy*, *59*, 569–576.
- Ueda, S., & Okawa, Y. (2003). The subjective dimension of functioning and disability: What is it and what is it for? *Disability and Rehabilitation*, *25*, 596–601.
- Forsyth, R., & Jarvis, S. (2002). Participation in childhood. *Child: Care, Health and Development*, *28*, 277–279.
- Wade, D. T., & Halligan, P. (2003). New wine in old bottles: The WHO ICF as an explanatory model of human behaviour. *Clinical Rehabilitation*, *17*, 349–354.
- Morris, C., Kurinczuk, J. J., & Fitzpatrick, R. (2005). Child or family assessed measures of activity performance and participation for children with cerebral palsy: A structured review. *Child: Care, Health and Development*, *31*, 397–407.
- Sakzewski, L., Boyd, R., & Ziviani, J. (2007). Clinimetric properties of participation measures for 5- to 13-year-old children with cerebral palsy: A systematic review. *Developmental Medicine and Child Neurology*, *49*, 232–240.
- Ziviani, J., Desha, L., Feeney, R., & Boyd, R. (2010). Measures of participation outcomes and environmental considerations for children with acquired brain injury: A systematic review. *Brain Impairment*, *11*, 93–112.
- Phillips, R. L., Olds, T., Boshoff, K., & Lane, A. E. (2013). Measuring activity and participation in children and adolescents with disabilities: A literature review of available instruments. *Australian Occupational Therapy Journal*, *60*, 288–300.
- Scientific Advisory Committee of the Medical Outcomes Trust. (2002). Assessing health status and quality-of-life instruments: Attributes and review criteria. *Quality of Life Research*, *11*, 193–205.
- Valderas, J. M., Ferrer, M., Mendivil, J., Garin, O., Rajmil, L., Herdman, M., et al. (2008). Development of EMPRO: A tool for the standardized assessment of patient-reported outcome measures. *Value Health*, *11*, 700–708.
- Andresen, E. M. (2000). Criteria for assessing the tools of disability outcomes research. *Archives of Physical Medicine and Rehabilitation*, *81*(2), S15–S20.
- Mokkink, L. B., Terwee, C. B., Knol, D. L., et al. (2009). The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Medical Research Methodology*, *10*, 1–8.
- Terwee, C. B., Jansma, E. P., Riphagen, I. I., & de Vet, H. C. W. (2009). Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Quality of Life Research*, *18*, 1115–1123.
- Terwee, C. B., Mokkink, L. B., Knol, D. L., Ostelo, R. W. J. G., Bouter, L. M., & de Vet, H. C. W. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Quality of Life Research*, *21*(4), 651–657.

26. Mokkink, L. B., Terwee, C. B., Gibbons, E., et al. (2010). Inter-rater agreement and reliability of the COSMIN (consensus-based standards for the selection of health status measurement instruments) checklist. *BMC Medical Research Methodology*, *10*, 1–11.
27. Terwee, C. B., Bot, S. D. M., de Boer, M. R., et al. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*, 34–42.
28. Mokkink, L. B., Terwee, C. B., Patrick, D. L., et al. (2010). International consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes: results of the COSMIN study. *Journal of Clinical Epidemiology*, *63*, 737–745.
29. Law, M., King, G., Petrenchik, T., Kertoy, M., & Anaby, D. (2012). The Assessment of Preschool Children's Participation: internal consistency and construct validity. *Physical and Occupational Therapy in Pediatrics*, *32*, 272–287.
30. Chen, C., Chen, C., Shen, I., Liu, L., Kang, L., & Wu, C. (2013). Clinimetric properties of the Assessment of Preschool Children's Participation in children with cerebral palsy. *Research in Developmental Disabilities*, *34*, 1528–1535.
31. Bult, M. K., Verschuren, O., Kertoy, M. K., Lindeman, E., Jongmans, M. J., & Ketelaar, M. (2013). Psychometric evaluation of the dutch version of the Assessment of Preschool Children's Participation (APCP): Construct validity and test-retest reliability. *Physical and Occupational Therapy in Pediatrics*, *33*, 372–383.
32. Anastasiadi, I., & Tzetzis, G. (2013). Construct validation of the children's assessment participation and enjoyment (CAPE) and preferences for activities for children (PAC). *Journal of Physical Activity and Health*, *10*, 523–532.
33. Ullenhag, A., Almqvist, L., Granlund, M., & Krumlinde-Sundholm, L. (2012). Cultural validity of the Children's Assessment of Participation and Enjoyment/preferences for activities of children (CAPE/PAC). *Scandinavian Journal of Occupational Therapy*, *19*, 428–438.
34. Nordtorp, H. L., Nyquist, A., Jahnsen, R., Moser, T., & Strand, L. I. (2013). Reliability of the Norwegian version of the Children's Assessment of Participation and Enjoyment (CAPE) and preferences for activities of children (PAC). *Physical and Occupational Therapy in Pediatrics*, *33*, 199–212.
35. Colon, W. I., Rodriguez, C., Ito, M., & Reed, C. N. (2008). Psychometric evaluation of the Spanish version of the Children's Assessment of Participation and Enjoyment and preferences for activities of children. *Occupational Therapy International*, *15*, 100–113.
36. Longo, E., Badia, M., Orgaz, B., & Verdugo, M. A. (2014). Cross-cultural validation of the Children's Assessment of Participation and Enjoyment (CAPE) in Spain. *Child: Care, Health And Development*, *40*, 231–241.
37. Bult, M. K., Verschuren, O., Gorter, J. W., Jongmans, M. J., Piskur, B., & Ketelaar, M. (2010). Cross-cultural validation and psychometric evaluation of the Dutch language version of the Children's Assessment of Participation and Enjoyment (CAPE) in children with and without physical disabilities. *Clinical Rehabilitation*, *24*, 843–853.
38. Potvin, M. C., Snider, L., Prelock, P., Kehayia, E., & Wood-Dauphinee, S. (2013). Children's Assessment of Participation and Enjoyment/preferences for activities of children: Psychometric properties in a population with high-functioning autism. *American Journal of Occupational Therapy*, *67*, 209–217.
39. King, G. A., Law, M., King, S., et al. (2006). Measuring children's participation in recreation and leisure activities: Construct validation of the CAPE and PAC. *Child: Care, Health and Development*, *33*, 28–39.
40. King, G., Law, M., King, S., Hurley, P., Rosenbaum, P., Hanna, S., et al. (2004). *Children's Assessment of Participation and Enjoyment (CAPE) and preferences for activities of children*. San Antonio, TX: PsychCorp.
41. Hwang, A., Liou, T., Bedell, G., et al. (2013). Psychometric properties of the Child and Adolescent Scale of Participation—Traditional Chinese version. *International Journal of Rehabilitation Research*, *36*, 211–220.
42. Bedell, G. (2009). Further validation of the Child and Adolescent Scale of Participation (CASP). *Developmental Neurorehabilitation*, *12*, 342–351.
43. McDougall, J., Bedell, G., & Wright, V. (2013). The youth report version of the Child and Adolescent Scale of Participation (CASP): Assessment of psychometric properties and comparison with parent report. *Child: Care, Health and Development*, *39*, 512–522.
44. Bedell, G. (2004). Developing a follow-up survey focused on participation of children and youth with acquired brain injuries after discharge from inpatient rehabilitation. *NeuroRehabilitation*, *19*, 191–205.
45. Rosenberg, L., Jarus, T., & Bart, O. (2010). Development and initial validation of the children's participation questionnaire (CPQ). *Disability and Rehabilitation*, *32*, 1633–1644.
46. Lepage, C., Noreau, L., Bernard, P., & Fougeyrollas, P. (1998). Profile of handicap situations in children with cerebral palsy. *Scandinavian Journal of Rehabilitation Medicine*, *30*, 263–272.
47. Fougeyrollas, P., Noreau, L., Bergeron, H., Cloutier, R., Dion, S. A., & St-Michel, G. (1998). Social consequences of long term impairments and disabilities: Conceptual approach and assessment of handicap. *International Journal of Rehabilitation Research*, *21*, 127–141.
48. Noreau, L., Lepage, C., & Boissiere, L. (2007). Measuring participation in children with disabilities using the Assessment of Life Habits. *Developmental Medicine and Child Neurology*, *49*, 666–671.
49. Kempers, R. J. J. K., Siebes, R. C., Gorter, J. W., Ketelaar, M., & Jongmans, M. J. (2011). Parental perceptions of participation of preschool children with and without mobility limitations: Validity and reliability of the PART. *Disability and Rehabilitation*, *33*, 1421–1432.
50. Coster, W., Law, M., Bedell, G., Khetani, M., Cousins, M., & Teplicky, R. (2012). Development of the participation and environment measure for children and youth: conceptual basis. *Disability and Rehabilitation*, *34*, 238–246.
51. Coster, W., Bedell, G., Law, M., et al. (2011). Psychometric evaluation of the participation and environment measure for children and youth. *Developmental Medicine and Child Neurology*, *53*(11), 1030–1037.
52. Tuffrey, C., Bateman, B. J., & Colver, A. C. (2013). The Questionnaire of Young People's Participation (QYPP): A new measure of participation frequency for disabled young people. *Child: Care, Health and Development*, *39*, 500–511.
53. Dunn, L. (2004). Validation of the CHORES: A measure of school-aged children's participation in household tasks. *Scandinavian Journal of Occupational Therapy*, *11*, 179–190.
54. Washington, L. A., Wilson, S., Engel, J. M., & Jensen, M. P. (2007). Development and preliminary evaluation of a pediatric measure of community integration: The Pediatric Community Participation Questionnaire (PCPQ). *Rehabilitation Psychology*, *52*, 241–245.
55. Merckies, I. S. J., Schmitz, P. I. M., van der Meché, F. G. A., Samijn, J. P. A., & van Doorn, P. A. (2002). Psychometric evaluation of a new handicap scale in immune-mediated polyneuropathies. *Muscle and Nerve*, *25*, 370–377.

56. Angst, F. (2011). The new COSMIN guidelines confront traditional concepts of responsiveness. *BMC Medical Research Methodology*, *11*, 1–6.
57. Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
58. Edwards, M. C. (2009). An introduction to item response theory using the need for cognition scale. *Social and Personality Psychology Compass*, *3*(4), 507–529.
59. Pesudovs, K. (2006). Patient-centered measurement in ophthalmology: a paradigm shift. *BMC Ophthalmology*, *6*, 1–4.