



ELSEVIER

Contents lists available at ScienceDirect

Research in Developmental Disabilities

journal homepage: www.elsevier.com/locate/redevdis

Development, psychometrics and feasibility of the School Participation Questionnaire: A teacher measure of participation related constructs

Donald Maciver^{a,*}, Vaibhav Tyagi^a, Jessica M. Kramer^b, Janet Richmond^c, Liliya Todorova^d, Dulce Romero-Ayuso^e, Hiromi Nakamura-Thomas^f, Margo van Hartingsveldt^g, Lorna Johnston^{a,h}, Anne O'Hareⁱ, Kirsty Forsyth^a^a School of Health Sciences, Queen Margaret University, Edinburgh, Scotland EH21 6UU, UK^b Department of Occupational Therapy, College of Public Health and Health Professions, University of Florida, United States^c School of Medical and Health Sciences, Edith Cowan University, Joondalup, Australia^d Occupational Therapy, Faculty of Public Health and Health Care, University of Ruse, Ruse, Bulgaria^e Department of Physical Therapy, Occupational Therapy Division, Faculty of Health Sciences, University of Granada, Granada, Spain^f Saitama Prefectural University, Graduate School of Health, Medicine and Welfare, Saitama, Japan^g Amsterdam University of Applied Sciences, Faculty of Health, School of Occupational Therapy, Amsterdam, The Netherlands^h Additional Support for Learning Service, Communities and Families, City of Edinburgh Council, Edinburgh, Scotland, UKⁱ Child Life and Health, SMC Research Centre, Edinburgh University, Edinburgh, Scotland, UK

ARTICLE INFO

Keywords:

Assessment
School
Validity
Rasch analysis
Participation

ABSTRACT

Background: We report development of the SPQ (School Participation Questionnaire) a teacher-completed measure of participation related constructs. The SPQ was developed to support participation-related assessment, interventions, and research in the inclusive school context.

Methods: Several iterative steps were undertaken. An international panel of experts reviewed content validity. A 66-item pilot questionnaire was administered in schools. Mokken and Rasch model analysis were applied. Internal consistency was assessed using Cronbach's alpha. Analyses were conducted on associations with teacher and child demographic variables. Feedback was sourced from users. Participants were teachers of 101 children (5–12 years old) with a range of disabilities, including intellectual disability, autism spectrum disorder and learning difficulties.

Results: Four participation-related dimensions of the SPQ were confirmed. Rasch person and item reliability were good, and 2–4 strata were confirmed per scale. Internal consistency was good (all scales, Cronbach $\alpha > 0.8$). Mean administration time was 11.7 min. Mean SPQ scores were independent of teacher characteristics. A significant effect of school support level, eligibility for free school meals and gender was found. Through synthesising analytic results and feedback, a new 46-item tool was obtained.

Conclusion: The results of this study provide evidence of acceptability, practicality and validity. The SPQ is the first tool developed to assess participation related constructs in schools, and it contains novel information not given by other assessments. The SPQ may be used by practitioners and researchers to understand and improve the participation of children with a range of disabilities in schools.

* Corresponding author.

E-mail address: dmaciver@qmu.ac.uk (D. Maciver).<https://doi.org/10.1016/j.ridd.2020.103766>

Received 23 September 2019; Received in revised form 8 July 2020; Accepted 12 August 2020

Available online 19 September 2020

0891-4222/© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license

<http://creativecommons.org/licenses/by-nc-nd/4.0/>.

What this paper adds

- This study describes assessment of participation related constructs for children aged 5–12 years with a range of disabilities in order to support participation-related interventions and research in the inclusive school context.
- The study describes the rationale for considering participation as an important intervention and assessment focus in education.
- The study identified that teachers can effectively assess participation related constructs using the SPQ.

1. Introduction

Participation, often defined as “involvement in a life situation” (WHO, 2007) is important to children, parents and professionals alike (Anaby et al., 2019; James Lind Alliance, 2016). Childhood onset conditions which result in reduced participation in school are increasingly common, including autism spectrum disorder, learning difficulties, and intellectual disability (Zablotsky et al., 2019). Restrictions in school participation have lifetime consequences for learning, development and well-being (Izadi-Najafabadi, Ryan, Ghafooripoor, Gill, & Zwicker, 2019; John-Akinola & Nic-Gabhainn, 2014; Peny-Dahlstrand, Krumlinde-Sundholm, & Gosman-Hedstrom, 2013). Historically there has been a lack of measures to support assessment of participation in schools (Adair et al., 2018).

In this study, we aimed to develop a teacher report measure of participation related constructs, titled the School Participation Questionnaire (SPQ). In recent years, participation has become of increasing interest to educators. Reduction of reliance on impairment focused or medical models of disability is a key theme (Norwich, 2008). Impairment focussed approaches have limited relevance to education planning, and there has also been recognition of stigma and labelling issues associated with the “medical” model (Norwich, 2016). Participation-focussed models are an alternative and advantageous framework, due to their focus on the “social” model of disability, or more recently “biopsychosocial” model, combining aspects of the “social” and “medical” models (WHO, 2007).

Conducting assessments of participation in the child’s natural environment, such as the classroom is beneficial (Anaby et al., 2019). However, there are several general barriers to using measures in schools. These include demands on time (Ingram, Louis, & Schroeder, 2004), training or knowledge (Chen, Heritage, & Lee, 2005), and whether a particular measure has face validity (Sturgis, Hughes, & Smith, 2006). Poor return for data collection efforts are common, and teachers may express concerns about measures not being necessary or relevant (Sturgis et al., 2006). For some educationalists, any “classification” is associated with a deficit model, and therefore to be criticized (Norwich, 2016). Teacher capacity may also depend upon experience or knowledge (van der Veen, Smeets, & Derriks, 2010) or self-perception of capabilities with regards to education of children with additional needs (Wilson, Woolfson, Durkin, & Elliott, 2016).

Considering these complexities, we developed a new measure. Several iterative steps were taken in collaboration with teachers, school leadership, paediatric therapists and parents. First, items were developed and subjected to content validity assessment by experts. Second, a pilot measure was deployed in schools, followed by psychometric analysis. We assessed practicality and acceptability for users, and produced an improved version of the tool for further use.

1.1. Theoretical background

Middle childhood (approximately 5–12 years) was the target age range. During middle childhood a child’s mastery of developmental challenges is strongly influenced by school experiences. Children are moving from nursery/kindergarten provision to increasingly formal education settings, but have not yet entered the adolescent phase where other challenges appear.

Regarding broader perspectives, the idea that a participation-focussed model could be useful in education has been a topic of debate for some time, and there are similarities with the more traditionally understood concept of “inclusion” (Hollenweger, 2011; Norwich, 2008, 2016). Historically, the “inclusion” perspective has been concerned with placement of children with additional needs in non-segregated environments (Göransson & Nilholm, 2014). The field has since moved to debating how to “do” inclusion and ensure quality, rather than placement alone. However, identifying what counts as good practice and good outcomes remains complex (Ruijs & Peetsma, 2009). Key principles for reforms to facilitate inclusion are available, but teachers and schools still have difficulty operationalizing these (Florian, 2012). This is understandable, particularly when schools tend to be rated on academic achievement, rather than on how inclusive they are (Glazzard, 2013). Whilst the inclusion perspective has been driven by educationalists, participation research is primarily driven by health professionals. Such discrepancies have led to different emphases, but both perspectives represent a social model of disability, stressing that impairments interact with environments to influence outcomes. In attempting to meld perspectives, educational views on curriculum, instruction and inclusion are complemented by health perspectives on disability, function and participation. A focus on participation may provide a common language and beneficial shared framework. Through focussing on participation, educators’ knowledge of children with developmental and other disabilities, and the factors which influence daily life activities, may be increased. Participation of children in the daily life of the school is also, arguably, a core indicator of the quality of placement, and therefore the “effectiveness” of inclusion.

1.2. SPQ conceptual model

Instead of measuring the amount or type of participation engaged in by children, SPQ development focussed on producing a measure to identify reasons for suboptimal participation. Participation in school, as defined in the contemporary literature, includes everyday life situations or activities which are required or desired to fulfil the role of the school pupil within or around the school

context (Maciver et al., 2019). School participation therefore includes a range of possible roles and activities, including classroom activity, school work, events, trips, teams, clubs, and relationships (Maciver et al., 2019). School participation can be measured in terms of how much, how often and what the child does (attendance), as well as their subjective experience (involvement) (Imms et al., 2016). However, knowing the amount that a child is participating, or that a child is participating poorly does not explain why this is the case, or how to help. Instead, it was intended that the SPQ would provide information about environmental and child determinants of participation. It was important to include a range of possible influences. A multidimensional assessment would support identification of the multiple factors leading to restricted participation, allowing for development and implementation of targeted interventions. Such assessment, inclusive of a biopsychosocial concept of child and environment fit, is congruent with current conceptualisations of participation (Imms et al., 2016) and approaches to intervention (Adair et al., 2015).

A pre-existing approach to assessing the determinants of participation is to ask teachers or others to assess the extent to which determinants are restricting (or facilitating) the child's participation. However, this is problematic for several reasons. Firstly, determining the extent to which each determinant is restricting participation requires identifying the determinant, and then observing its impact. Secondly, users may not be able to identify independent effects of determinants. Thirdly, perceiving a factor as restricting participation is not the same as the factor actually restricting participation (Whiteneck & Dijkers, 2009). Instead, the SPQ was designed to map onto a recent school participation conceptual framework, derived from a systematic literature review of key determinants of participation (Maciver et al., 2019). This review was the first work to comprehensively describe participation determinants and their comprising dimensions in the school setting. It was intended that the SPQ would provide information about the extent to which the child and environment contained features identified as important for participation in this review.

The conceptual framework comprised four components: identity, competence, experience of mind and body (symptoms), and environment (Maciver et al., 2019). In the framework, 'identity' or 'being' included sense of efficacy, what children found interesting and valuable to do, a sense of routines of the school day, perceptions around role, and perceptions around inclusion. 'Competence' or 'doing' focussed on what the child actually did in school. Competence induced such aspects as fulfilling expectations, standards of performance, maintaining a routine, discharging responsibilities, and skills. 'Experience of mind and body' (symptoms) included issues commonly experienced by children with disabilities in the school context: pain, anxiety, mood, and fatigue. Environment was operationalised as the physical and social attributes of the school.

1.3. Initial development of the SPQ

A multi-professional, international group developed the SPQ items (i.e. the authors). This group had expertise in teaching, psychology, community paediatrics, neurodisability, community health sciences, occupational therapy, participation, rehabilitation and global health. Items were designed to represent the hypothesized causal model of influences on participation as described above. Input was also sought from teachers, therapists and parents to ensure relevance. Additionally, we completed a review of existing instruments (additional file 1). These processes reinforced gaps, identifying that a tool and a common language to describe and measure participation related constructs in schools was lacking.

1.4. Items and scales

Four scales were included in the SPQ, reflecting the originating conceptual framework (Maciver et al., 2019). The 'competence' scale included items on making choices, persistence, meeting expectations, performing roles and skills, with higher scores being indicative of favourable characteristics. Example item: "The child does what is expected of them." Items in the 'symptoms' scale were targeted to measure the extent of disability-related symptoms: lack of energy, tiredness/sleep, pain, low mood and anxiety. Higher scores would indicate a favourable symptom status. Example item: "The child seems well-slept when they arrive for school." Items in the 'identity' scale aimed to assess the child's thoughts, feelings, knowledge, preferences, self-perceptions and role perceptions, with higher scores being indicative of favourable characteristics. Example item: "The child understands school routines." The 'environment' scale was designed to measure the physical (spaces and the objects) and social (peers, teachers, routines and typical or expected ways of doing things) environment of the school. Higher scores are indicative of a more facilitative environment. Example item: "The school building is fully accessible to the child."

In consultation with teachers a 4-point scale was considered optimal, using the categories: 1 = Disagree; 2 = Somewhat Disagree; 3 = Somewhat Agree; and 4 = Agree. Teachers reported that items should be positively worded to reflect the ethos of education. As well as scores for each individual item, responses may be averaged to generate a summary score ranging from 1 to 4 for each scale. It was envisaged that a total score (across scales) would be unhelpful due to differential functioning.

To reflect real world practices, there were no restrictions on how information for completing the questionnaire was gained, including direct experience with the child, paperwork and communication with others. Teachers were advised to not think only about good days or bad days, rather an overall assessment of the previous two-week period. They were asked to consider the child with normal supports in place (e.g. if a child habitually wore glasses then consider that child when they are wearing glasses). Throughout, the focus was on school. Therefore, when a question addressed activities the child took part in, or roles, we were only concerned with those that took place at school. In terms of the environment, teachers were advised to be reflective and adopt a critical, probing, problem-seeking approach. A "walk through" was advised.

2. Methods

This study draws on best practice for development and evaluation of instruments via the COSMIN framework (Mokkink et al., 2010) and the Medical Research Council's guidance on development of complex interventions (Craig et al., 2008). We had the following objectives. Firstly, to undertake a content validity procedure with an expert group, and winnow items accordingly. Secondly, to assess psychometric properties of a pilot version of the SPQ. Thirdly, to explore the relationship of the SPQ to child and teacher characteristics. Such analyses were aimed at hypothesis testing (Mokkink et al., 2010) in terms of the expected behavior of the measure. We hypothesized that SPQ scores would be associated with characteristics of children, but that they would be independent of teacher characteristics. Finally, a focus was to produce a short effective tool that was practical and acceptable to the target community (teachers). We therefore gathered feedback from users, and examined the acceptability and time burden. To finalise the SPQ, changes were made according to the statistical results and in collaboration with users.

2.1. Participants and procedures

2.1.1. Content validity

Eight independent experts were recruited through professional networks. Experts included parent/family advocacy organisations ($n = 2$), education/psychology services ($n = 3$), and health professionals ($n = 3$). Experts had at least masters education, four held PhDs, and two held professorships. Average experience was 23 years. One hundred and eight items were presented for review. Data were collected through electronic survey.

2.1.2. Rasch model analysis

Prior to starting, researchers determined that the sample should comprise at least 100 SPQs as this would provide 95 % confidence in calibrations within ± 0.5 logits for the Rasch model analysis (Linacre, 1994). One hundred and one children were recruited from inclusive primary schools, in a large urban local authority in Scotland. Four schools were selected randomly using Probability Proportional to Size (PPS) methods (Organisation for Economic Co-operation & Development, 2012). Schools selected children using a lottery method (aiming to provide 28 children each, with 4 children selected per primary stage). Children were 5–12 years old. Senior teachers identified children for randomisation. Criteria for having disability reflected those used nationally: children having a physical, developmental, behavioural, or emotional condition and who also required health and/or related services of a type or amount beyond that required by children generally. Categories of need were recorded as follows: learning disability; dyslexia; other specific learning difficulty; other moderate learning difficulty; visual impairment; hearing impairment; deafblind; physical or motor impairment; language or speech disorder; autistic spectrum disorder; social, emotional and behavioral difficulty; physical health problem; or mental health problem (Scottish Government, 2019). Children could have one or more condition/need. Paper SPQs and other forms were distributed to schools, with return requested in a three week timeframe. SPQs were completed by teachers.

2.1.3. Exploratory analysis

Children's demographic characteristics included were level of school support, free school meals entitlement, age and gender. The post-completion questionnaire completed by each participating teacher included qualification status and years of experience. Five items capturing teacher self-efficacy concerning education of children with additional needs was also included. During analysis we tested for reliability and found Cronbach's alpha to be $> .8$ for this teacher self-efficacy measure, indicating reliability for the purposes of this study.

2.1.4. Acceptability and time burden

The post-completion questionnaire completed by each participating teacher also included free-text feedback, and eight items on the qualities of the SPQ. Each school's senior teacher(s) was debriefed, with feedback gathered. Time taken to complete each SPQ was recorded to the nearest minute.

2.1.5. Finalisation

Our analytic results were based primarily on quantitative statistical techniques, and this may lead to items of conceptual and clinical interest being removed, hence involvement of users and experts is appropriate in finalising a tool (Chen et al., 2014). Decisions on changes were therefore made in a series of meetings including teachers, school leadership, and experienced paediatric therapists.

2.2. Data analysis

2.2.1. Content validity

Content validity is judged by experts with the proportion of expert agreement captured in the content validity index (CVI). Experts rated draft SPQ items covering five aspects: relevance to concept, relevance to population; relevance to purpose; ease of reading; and clarity. Item-level content validity index (I-CVI) and modified kappa statistic (k^*) were calculated (Polit, Beck, & Owen, 2007). Where an item received $I-CVI < .78$ (on any criteria) it was considered for revision/exclusion (Polit et al., 2007). For modified kappa statistic, content validity for items is described as excellent for $k^* > .74$ (Cicchetti & Sparrow, 1981).

2.2.2. Rasch model analysis

A Rasch model analysis for polytomous item scales was conducted using a Partial Credit Model (PCM). Checks were conducted for monotonicity, local independence, Invariant Item Ordering (IIO) and unidimensionality using Mokken techniques (Na, Gross, & West, 2015). Checks for unidimensionality indicate that items measure a unidimensional trait (Kreiner, 2007). Item scalability coefficients $>.5$ indicate strong evidence for unidimensionality. Monotonicity indicates that the probability of receiving a higher score on an item will increase monotonically (or stay the same but will not decrease) as the score on the latent trait increases (Kreiner, 2007). Local independence ensures that responses on each item are independent of responses on other items (Kreiner, 2007). It is tested through largest associations (correlation $>.5$) in the standardised residual (Kreiner, 2007). Assumptions around IIO are that item difficulty remains the same across respondents (Ligtvoet, Van der Ark, te Marvelde, & Sijtsma, 2010). It is beneficial that the items have the same order with respect to difficulty as such an ordering facilitates interpretation (Ligtvoet et al., 2010). IIO is interpreted using plots of item response functions, and visual examination to identify problematic items.

How well items conformed to the Rasch model was assessed through fit statistics. Fit statistics include the item locations (in logits), standard errors, residuals and fit to the model. For our study we focussed on Infit MnSq values, which are less sensitive to extreme responses that outfit (Bond & Fox, 2013). Infit MnSq statistics should range from .5 to 1.5 to be productive for measurement (Bond & Fox, 2013). A cut-off of Infit MnSq >1.4 with standardised z score (ZStd) >2 may be used to identify particularly problematic items (Bond & Fox, 2013). Items and persons were also examined using a person-item map. Examinations were made for ceiling and floor effects, and gaps. Using this technique, in cases where there were multiple items representing the same level of difficulty, and the items reflected similar concepts, items may be omitted to shorten questionnaires (Boone, 2016). Separation was calculated. Separation gives an estimate of the spread of items or individuals along the continuum of ability and reflects the number of strata into which the sample can be divided. This is an indicator of quality as it evaluates in terms of whether items can separate individuals meaningfully. The separation ratio may be transformed into strata index describing the number of significantly different levels (Bond & Fox, 2013). Related is the reliability of these indices, providing the degree of confidence in estimates. For reliability, coefficients of .67–.80 are fair, .81–.90 are good, .91–.94 are very good and >0.94 are excellent (Fisher, 2007). For strata <2 strata are considered poor, 2–3 are fair, 3–4 are good, 4–5 are very good and >5 strata are excellent (Fisher, 2007). Cronbach's alpha, was also produced for each scale, with values $>.8$ considered good (Tavakol and Dennick, 2011).

2.2.3. Exploratory analysis

To assess the relationship between SPQ and child and teacher characteristics Bayesian pairwise correlations were conducted. One way analyses of variance were also conducted to assess effects of child and teacher characteristics on SPQ scores. Statistical significance was set to $p < .05$ for ANOVAs and statistical support for correlations was set to $BF_{10} > 3$ for Bayesian analyses.

2.2.4. Acceptability and time burden

Time-burden was assessed by identifying the average time taken to complete the SPQ. Qualitative feedback was transcribed and analysed using content analysis. Categorical questionnaire data were expressed as frequencies and percentages.

2.2.5. Finalisation

In order to produce the next version of the SPQ, items identified as problematic in the preceding analyses, inclusive of the Rasch model analysis, Mokken analysis, feedback, and teacher post-completion questionnaire, were reviewed.

Analyses were completed in R (v3.5.0), R Studio (v1.1.453), and SPSS V.22.0 (IBM Corp, Armonk, New York, USA). Mokken checks were conducted through the 'Moken' package (v2.8.11) in R. Rasch modelling was conducted via joint maximum likelihood estimation in WINSTEPS (v3.91.2).

2.3. Ethics

The investigation was carried out following the Declaration of Helsinki. Ethical approval was provided by the institutional committee of Queen Margaret University ("School participation feasibility study" August 2017) and Local Authority Research Access Service ("CIRCLE project" July 2017). Written consent was obtained from each school's head teacher and implied consent secured from teachers (completion of questionnaire and feedback indicating consent). Teachers were given the opportunity to opt out at any time and without giving a reason. Parental consent was not sought, as teachers completed SPQs based on professional knowledge and school-held records. Children were not directly involved in the research and all data was fully anonymised before release to the research team.

3. Results

3.1. Content validity

For 108 items, on review of k^* , $>90\%$ of items had $k^* > .74$, indicating good content validity. However, twenty items had an I-CVI $< .78$ (across any criteria). On review, 15 items were removed and five were reworded. Qualitative feedback led to removal of four items. In consultation with end-users, it was identified that the instrument should contain as few items as possible (as respondents could be required to evaluate several children). Items were therefore ranked for relevance by the development team to winnow the pool. Rankings were aggregated using the Cross-Entropy Monte Carlo algorithm, applied to Spearman footrule distances, implemented

in the RankAggreg package for R. The item reduction steps resulted in a pool of 66 items (see additional file 2 for summary of all item reduction steps).

3.2. Rasch model analysis

Six schools were approached, with four participating. After randomisation, there were 108 children and 44 teachers eligible. Two teachers declined, leading to a loss of seven eligible children. SPQs were therefore completed by 42 teachers (95.45 % response) for 101 children (93.5 % response). Characteristics of children are in Table 1, teachers in Table 2. Checks of the sample against the national population were completed for gender, free school meals, and needs. This approach was used to infer that the sample was representative. The comparator dataset, a yearly government-completed school census, was secured. This dataset included all government sponsored primary schools in Scotland, and data on children with additional needs aggregated at the school level. Summaries of the characteristics of children were developed and compared. Differences in proportions and means were inspected between the sample and local/national data. Data and analyses are presented in additional file 3. The overall conclusion is that the current sample shows an acceptable level of representativeness.

In line with the underlying conceptual framework, each scale of the SPQ was modeled separately. As per recommendations (Linacre, 2002) we collapsed categories after an unsatisfactory evaluation of the response categories. We collapsed responses 1 and 2 in

Table 1
Children (n = 101).

Children		
Age	Mean (SD; Range)	
Years, Months	8y 2mo (1y 10mo; 5y-12y 1mo)	
Characteristics	N	%
Gender		
Female	31	30.69
Male	67	66.33
Missing	3	2.97
Ethnicity		
White - British/Scottish	81	80.19
African	3	2.97
Asian - British/Scottish	3	2.97
Mixed/multiple/other	4	3.96
Missing	10	9.99
Language		
English	87	86.13
Other	11	10.89
Missing	3	2.97
Classifications^a		
Autistic spectrum disorder	16	15.84
Communication support need	10	9.9
Deafblind	0	0
Dyslexia	22	21.78
Hearing impairment	3	2.97
Language or speech disorder	20	19.8
Learning disability	10	9.99
Mental health problem	6	5.94
Other moderate learning difficulty	6	5.94
Other specific learning difficulty (e.g. Numeric)	8	7.92
Physical health problem	5	4.95
Physical or motor impairment	9	8.91
Social emotional and behavioural difficulty	19	18.81
Visual impairment	2	1.98
Missing	14	13.86
School support level^b		
I	27	26.73
II	24	23.76
III	43	42.57
Missing	7	6.93

^a Children may be in multiple categories. Individual categories may not sum to 100 %. 'Learning disability' matches internationally used definition 'Intellectual disability'. Moderate and Specific Learning Difficulties are umbrella terms for often co-occurring difficulties (Dyslexia, Dyspraxia, Dyscalculia, Attention Deficit Hyperactivity Disorder). A child may be diagnosed with a Learning Difficulty where there is a lack of achievement for age/ability, or a discrepancy between achievement and ability. Dyslexia is recorded separately due to national practice, and impact on education. 'Communication support need' represents children who experience difficulties communicating and/or understanding others and is used in place of more specific diagnosis.

^b Level I: child's needs are managed by the class teacher; Level II: child's needs are managed with help from specialist or more senior teachers within the school; Level III: child's needs are managed with support from partnership services or agencies (e.g. therapists or psychologists).

Table 2
Teachers (n = 42).

Teachers		
Experience	Mean (SD: Range)	
Years	9.44 (9.36: .25–32)	
Characteristics	N	%
Role ^a		
Class Teacher	31	73.81
SfL	2	4.76
Head Teacher	1	2.38
Head Teacher + SfL	1	2.38
Missing	7	16.67
Full time		
Yes	30	71.43
No	3	7.14
Job share	2	4.76
Missing	7	16.67
Probationary ^b		
Yes	6	17.14

^a SfL: Support for learning teacher is experienced teacher with school wide responsibility for additional support of children. Head teacher matches internationally used definition of “school principal”.

^b Probationary teacher in their first year post qualification.

the Identity scale, the Competence scale and Symptoms scale. In the Environment scale, ratings 1, 2 and 3 were collapsed. After these adjustments, thresholds and category measures increased as appropriate, indicating that the categories capture increasing levels of the latent trait.

3.2.1. Unidimensionality, monotonicity, local independence, invariant item ordering

In reviewing unidimensionality, coefficients for each scale showed moderate-strong support for unidimensionality. Homogeneity coefficients for Identity scale ranged from $H_i = .44-.65$ with an overall scalability coefficient of $H_s = .56$ ($SE = .049$). Homogeneity coefficients for the Competence scale ranged from $H_i = .36-.7$ with $H_s = .58$ ($SE = .04$). Homogeneity coefficients for the Environment scale ranged from $H_i = .39-.66$ with $H_s = .53$ ($SE = .04$). Homogeneity coefficients for the Symptoms scale ranged from $H_i = .6-.67$ with $H_s = .64$ ($SE = .05$). Automated item selection procedure with increasing homogeneity thresholds indicated that at levels of $.05-.35$, the items in each scale measured a unidimensional latent trait. There were no violations of monotonicity. For local independence, nine items in the Environment scale and one item in the Identity scale were identified as problematic. For invariant item ordering, two items in the Competence scale and three in the Environment scale were identified as problematic. Items with issues at this stage were removed from further analysis, and tagged for review, rewording or removal.

3.2.2. Overall performance of items

Fifty-one items were included in the Rasch model analysis (see additional file 4 for a full overview of items, and Table 3 for a summary). Across each scale items were ordered according to difficulty as clinically and developmentally expected, supporting validity. In general, Infit values were within range, showing acceptable fit with the Rasch model. One item each in the Identity, Competence and Environment scales was Infit MnSq >1.4 . Subsequent removal of items showed little improvement. With one item misfitting per scale the decision was taken to proceed retaining those items. The separation index was acceptable for persons and items (all scales). Reliability was good-excellent, except for the Symptoms scale which was fair (.79). Across the scales there were 2–4 person strata, indicating acceptable precision. Cronbach’s alphas were $>.8$ for all scales, indicating good internal consistency.

Table 3
Summary table of indicators.

Criterion ^{a,b}	Target	Environment	Identity	Competence	Symptoms
Cronbach’s alpha	$>.8$.94	.89	.94	.84
Item reliability	$>.8$.94	.91	.95	.93
Item separation	–	3.89	3.23	4.28	3.59
Item strata [*]	>2	5.52	4.64	6.04	5.12
Person reliability	$>.8$.86	.83	.9	.79
Person separation	–	2.48	2.24	2.97	1.92
Person strata [*]	>2	3.64	3.32	4.29	2.89

^a Recommended values for Rasch are based on Fisher (2007).

^b Cronbach’s alpha values from Tavakol and Dennick (2011).

^{*} Calculated as ‘ $H = (4G + 1)/3$ ’, where $G = \text{item/person separation}$.

3.3. Exploratory analysis

Full details are in additional file 5. In summary, mean SPQ scores were unaffected by teachers' characteristics – their experience, self-assessed self-efficacy, satisfaction with SPQ, probationary status and full or part time working. For children, eligibility for free school meals was a negative predictor of mean scores on three SPQ scales (Identity – $t(74) = -2$, $p = .049$; Symptoms – $t(74) = -2.18$, $p = .032$; Environment – $t(74) = 2.79$, $p = .007$). A significant effect of school support level was also found for mean scores on two SPQ scales (Identity – $F = 3.147$, $p = .048$; Environment – $F = 4.914$, $p = .009$) with children receiving support level III scoring lower than those on level I. Gender was a significant predictor for mean SPQ score on the Competence scale ($t(95) = -2.36$, $p = .02$) with girls scoring higher ($M_{diff} = .325$). Children's age was not related to mean SPQ scores.

3.4. Acceptability and time burden

SPQ mean administration time was 11.7 min. The post completion feedback questionnaire is presented in Table 4. Users reported the SPQ was an appropriate length, with clear instructions, and that they would use it again. However, feedback also indicated some issues, including items which were irrelevant, misleading, or superfluous and items which were unnecessary or repetitive. Qualitative analysis also identified factors that best facilitate successful liaison with schools (see additional file 6).

3.5. Finalisation

Post completion feedback from teachers was considered in tandem with the outcomes of the psychometric analysis. Items showing local dependency were examined first. Although item dependency can arise due to assistance or fatigue (Yen, 1993), we ruled out these since the SPQ was completed by skilled professionals. Instead, we hypothesised that dependence was arising due to common or overlapping concepts (Marais & Andrich, 2008). There was a pattern. Seven locally dependant items were related to peers. We hypothesised that issues were caused by items being perceived to be targeting the same idea, and/or by inadequately differentiated wording. The items were reviewed, with three removed and four reworded. Two additional items were discarded on the basis of local dependency, with alterations made to one item. Invariant item ordering (IIO) was examined. Requirement for IIO is stringent standard (Sijtsma & Hemker, 1998), and is particularly suitable for tests which may be used to guide clinical decision making. Questionnaires with IIO are increasingly seen as desirable. In the case of the SPQ, four items not demonstrating IIO were removed, and one item was rewritten. Concerning misfit, one item was removed. The two remaining misfitting items were rewritten as we hypothesised that the misfit was due to ambiguous or otherwise poorly constructed wording. With regards to item difficulty, the environment scale contained items which were easier to endorse highly. A group of high scoring items were related to 'adults' (teachers and support staff) within the school. As teachers were rating, this phenomena was interpreted as a halo effect. To address this, in consultation with teachers, we retained items with a focus on supports/strategies (e.g. 'strategies to foster involvement and interaction with peers are regularly applied') and removed four items relating to teacher self-assessment (e.g. 'adults are responsive to the needs of the child'). In terms of the person-item map, six items representing the same level of difficulty, and similar concepts, were omitted to shorten the questionnaire.

Minimal data existed for the lower response categories for the environment items (i.e. Disagree, Somewhat Disagree and Somewhat Agree). The collapsing of categories was considered necessary for the purposes of the Rasch model analysis. A consideration of how these items were operating and the latent trait which they were eliciting was made in light of the responses obtained. A practical rationale for collapsing the responses is that the rubric of SPQ regards any score of less than 4 as indicative of an issue or problem for that item. The scores of 1, 2, and 3 therefore indicate a problem or issue, whilst 4 would indicate no issue. Such collapsing was required for the environment scale, which equated to a dichotomous response option for the environment items (Disagree + Disagree Somewhat + Agree Somewhat, versus Agree). The remaining three scales required collapsing of responses 1 and 2 only. This led to scales differentiating three levels (Disagree + Disagree Somewhat, versus Agree Somewhat, versus Agree). At least in the analysis phase, such collapsing could replace the 4-point scales. When combined with a Rasch model approach, which provides more precise estimates compared to mean scores, reduced response options may function well. However, regarding practical use, collapsed responses might not serve as well. Although teachers in the current study did not favour the lower response options, it was strongly supported by end-users that a range of scores should be available to reflect the variability of children/environments. We therefore retained the 4-point rating scale.

Table 4
Post-completion evaluation (n = 33 teachers).

Qualities of the SPQ	Yes (%)	Partially (%)	No (%)
Instructions appropriate?	30 (90.9)	1 (3.0)	2 (6.1)
Appropriate length?	28 (84.8)	1 (3.0)	4 (12.1)
Happy to use in future?	25 (75.8)	5 (15.2)	3 (9.1)
Most teachers know the answers for most children?	22 (66.7)	3 (9.1)	8 (24.2)
Questions irrelevant/misleading/superfluous?	9 (27.3)	14 (42.4)	10 (30.3)
Items unnecessary/repetitive?	8 (24.2)	19 (57.6)	6 (18.2)
Difficulties completing?	12 (36.4)	14 (42.4)	7 (21.2)
Changes required?	12 (36.4)	11 (33.3)	10 (30.3)

All steps resulted in a questionnaire of 46 items (Identity: 9 items, Competence: 11 items, Experience of mind and body (symptoms): 5 items, Environment: 21 items). A summary of all item reduction steps is in additional file 2.

4. Discussion

In developing the SPQ, a goal was to produce an instrument that could be used for assessment of participation related constructs. In this study, we present information for teachers, rehabilitation professionals and other clinicians on a new measure for 5–12 year old children with additional needs and disabilities in the school setting. We have advanced the conceptual development work of [Maciver et al. \(2019\)](#) and consulted with experts and practitioners. We tested the SPQ with 101 children, evaluated measurement properties, and produced a new version.

Evidence suggests the SPQ is an instrument that captures four participation related dimensions. The SPQ is based on the ideas that teachers are important in developing participation-fostering practices ([Norwich, 2016](#)) and that to do this requires a contemporary model that sees participation as shaped by environment and child factors ([Maciver et al., 2019](#)). Evaluating measurement properties of new instruments is of key importance. There is recognition of the value of raising the current standard of assessment and particularly developing assessments that generate ‘true’ measures.

We found internal consistency to be good. Cronbach’s alphas were $>.8$ for all scales. Rasch person and item reliability were also good ($>.8$) except for the Symptoms scale which was fair (.79). Separation was also good (>2 strata) for all scales. However it is notable that the Symptoms scale reflected a somewhat lower separation index, indicating that this section was distinguishing two strata only. This reduced differentiation is less desirable. Overall however, there is evidence that scores from the SPQ may be considered reliable, and are able to distinguish among higher and lower performing groups of children.

SPQ responses were independent of teacher characteristics, a beneficial feature for tool reliability. Enhancing teacher ability to work with children is a key task, considering that initial teacher training may not fully prepare teachers for working with children who have complex additional needs ([Florian, 2012](#)). Additionally, measures with good reliability are needed in schools, where several people may be involved with each child.

In terms of child characteristics, results indicated no significant differences in scores regarding child age. Lower SPQ scores were reported for children receiving Free School Meals (FSM), from children receiving the highest level of school support (school support level III) and in the case of one scale, for boys. Children eligible for FSM are a socioeconomically disadvantaged group, and children experiencing poverty experience more participation restrictions ([Arakelyan, Maciver, Rush, O’hare, & Forsyth, 2019](#)). Regarding findings on school support, SPQ scores were lower for children receiving school support level III (children receiving the highest level of support). These children may have been receiving input from psychologists, occupational therapists, and/or other specialists and may therefore be considered to have higher need. The findings on lower scores for boys on one scale require investigation in future studies. However, reflecting other research, boys are more likely to be identified by teachers as having greater levels of need ([EASNIE, 2018](#)).

The SPQ was designed to be concise. A shorter and quicker to complete form is generally preferred by practitioners ([Dillman, Sinclair, & Clark, 1993](#)). Although it was reported to be of a suitable length, teachers also identified issues with the pilot measure, which we have taken steps to remediate. Teachers did complete the instrument within acceptable timescales (<12 min per child), and it is likely this will be reduced due to the shortening of the tool as a result of this study.

The study provides insight into theoretical perspectives. Assessment of participation is synonymous the things that people do, for example activities or satisfaction with activities ([Adair et al., 2018; Imms et al., 2016](#)). Yet it is also important to understand other elements when attempting to address barriers and issues to participation. Research related to the environment, choices, preferences, and contemporary approaches to intervention all reference the simultaneous contributions of several interrelated factors, including those related to the child and the environment ([Anaby, Law, Feldman, Majnemer, & Avery, 2018; Imms et al., 2016](#)). This thinking moves away from deficit, instead focussing on biopsychosocial explanations – and by encouraging a focus on the environment and the role of practitioners (rather than what a child can and cannot do) decentralizes children’s personal limitations and disabilities ([Maciver et al., 2018](#)). Building application of such theories into practice are important for teacher development in inclusive education ([Florian, 2012](#)).

We believe the SPQ may find use internationally. There is potential for generalisability because of the shared context of increasingly inclusive education systems. The SPQ is intended for use by teachers (and related services personnel), with an English language version available, which has been developed with international collaboration. In Scotland, where the research was carried out, the presumption of general education for almost all children has been national policy for several years ([Scottish Government, 2019](#)). Significant numbers of learners with disabilities are in general schools. Improving teacher and school practices are a key focus. Scotland is not unique in this respect, and analogous trends are apparent internationally ([Riddell, 2012](#)). The demographic and cultural similarities of Scotland’s education system to others around the world means that the applicability of the SPQ to a range of contexts is increased. Further research supporting use in the different contexts, especially non-English speaking contexts, remains necessary.

Overall, we believe that the measure may find application for the population of children with additional needs, developmental and other disabilities in schools. The SPQ is intended to support fast and efficient child and classroom assessment, as well as audit and research. Across the world teachers are being challenged to develop systems for identifying, monitoring and improving outcomes. Reliable, valid tools are required. The SPQ is a structured rating scale designed to complement practice with new participation related indicators. It will be of interest and use to classroom teachers, specialised teachers and related services personnel. The SPQ could be used for individual learners in audit of support needs, identifying current support and potential gaps, as well as resourcing and staff development. For example, advance knowledge of low SPQ environment scores, could help ensure that further assessment and supports focus on the classroom environment. Results could also support planning by school leadership and to inform and support strategic

decision making around resourcing. Examining child and the environment should be the first step of any support. The SPQ allows the user to identify where problems lie, and develop interventions, or components for interventions that will target them in a participation-focussed framework. These ideas are congruent with evidence that environment-focussed adaptations and individual coaching/mentoring may improve participation of children with disabilities (Adair, Ullenhag, Keen, Granlund, & Imms, 2015). Participation related constructs are also a key topic of interest for researchers focusing on school participation. The SPQ can measure four participation related areas, and can serve as a foundation for studies on key mechanisms for school participation outcomes.

4.1. Limitations

The study did not evaluate the effects of rewording/removal of some items. Recall bias may occur with teachers reporting on events and children within larger cohorts. Some information was not immediately available or known to teachers. Another potential form of bias was social desirability, which arises when teachers offer ‘better’ answers, believe some may be ‘wrong’ or do not wish to present a negative picture of the school. A repeat of the Rasch analysis with the 46 item post-pilot measure with a larger data set is warranted, as the robustness and stability of the findings will be enhanced if a larger sample is used. Further research is also needed to assess convergent validity, test-retest reliability and inter-rater reliability. Using the SPQ alongside a tool which measures frequency or involvement in participation would provide the opportunity to test our hypotheses on the relationships between environment, child, and participation.

5. Conclusions

The study has demonstrated a method to address participation-related assessment through establishment of low-burden data collection using teachers as lead respondents. An additional advantage of the SPQ is its combination of three scales relating to child factors, and one environment scale, into a comprehensive instrument, which sets it apart from the other available measures.

CRedit authorship contribution statement

Donald Maciver: Conceptualization, Methodology, Validation, Formal analysis, Writing - review & editing, Investigation, Resources, Data curation, Writing - original draft, Supervision, Project administration, Funding acquisition. **Vaibhav Tyagi:** Conceptualization, Methodology, Software, Validation, Formal analysis, Writing - review & editing, Investigation, Resources, Data curation, Writing - original draft, Visualization, Project administration. **Jessica M. Kramer:** Conceptualization, Methodology, Validation, Formal analysis, Writing - review & editing. **Janet Richmond:** Conceptualization, Methodology, Validation, Formal analysis, Writing - review & editing. **Liliya Todorova:** Conceptualization, Methodology, Validation, Formal analysis, Writing - review & editing. **Dulce Romero-Ayuso:** Conceptualization, Methodology, Validation, Formal analysis, Writing - review & editing. **Hiroki Nakamura-Thomas:** Conceptualization, Methodology, Validation, Formal analysis, Writing - review & editing. **Margo van Hartingsveldt:** Conceptualization, Methodology, Validation, Formal analysis, Writing - review & editing. **Lorna Johnston:** Conceptualization, Methodology, Validation, Formal analysis, Writing - review & editing, Investigation, Resources, Project administration. **Anne O’Hare:** Conceptualization, Methodology, Validation, Formal analysis, Writing - review & editing, Writing - original draft, Supervision, Project administration, Funding acquisition. **Kirsty Forsyth:** Conceptualization, Methodology, Validation, Formal analysis, Writing - review & editing, Writing - original draft, Supervision, Project administration, Funding acquisition.

Acknowledgements

The study was led by a partnership between the CIRCLE Collaboration (Child Inclusion: Research into Curriculum, Learning and Education) research team at Queen Margaret University, NHS Lothian, and the City of Edinburgh Council. Thanks to the schools, teachers and parents who have supported this research and to the international collaborators. This research was funded by NHS Lothian, City of Edinburgh Council & Scottish Government. The SPQ has been developed for use by schools, teachers, researchers and clinicians and can be used for the benefit of improving the participation of 5-12 year old children with special educational needs or disabilities. To use the scale please contact the corresponding author, and cite in any subsequent write up.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.ridd.2020.103766>.

References

Adair, B., Ullenhag, A., Keen, D., Granlund, M., & Imms, C. (2015). The effect of interventions aimed at improving participation outcomes for children with disabilities: A systematic review. *Developmental Medicine & Child Neurology*, 57(12), 1093–1104.

- Adair, B., Ullenhag, A., Rosenbaum, P., Granlund, M., Keen, D., & Imms, C. (2018). Measures used to quantify participation in childhood disability and their alignment with the family of participation-related constructs: A systematic review. *Developmental Medicine & Child Neurology*, 60(11), 1101–1116.
- Anaby, D. R., Campbell, W. N., Missiuna, C., Shaw, S. R., Bennett, S., Khan, S., ... GOLDs (Group for Optimizing Leadership and Delivering Services). (2019). Recommended practices to organize and deliver school-based services for children with disabilities: A scoping review. *Child: Care, Health and Development*, 45(1), 15–27.
- Anaby, D. R., Law, M., Feldman, D., Majnemer, A., & Avery, L. (2018). The effectiveness of the pathways and resources for engagement and participation (PREP) intervention: Improving participation of adolescents with physical disabilities. *Developmental Medicine & Child Neurology*, 60(5), 513–519.
- Arakelyan, S., Maciver, D., Rush, R., O'hare, A., & Forsyth, K. (2019). Family factors associated with participation of children with disabilities: A systematic review. *Developmental Medicine & Child Neurology*, 61(5), 514–522.
- Bond, T. G., & Fox, C. M. (2013). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*.
- Boone, W. J. (2016). Rasch analysis for instrument development: why, when, and how? *CBE—Life Sciences Education*, 15(4), rm4.
- Chen, W. H., Lenderking, W., Jin, Y., Wyrwich, K. W., Gelhorn, H., & Revicki, D. A. (2014). Is Rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An example using PROMIS pain behavior item bank data. *Quality of Life Research*, 23(2), 485–493.
- Chen, E., Heritage, M., & Lee, J. (2005). Identifying and monitoring students' learning needs with technology. *Journal of Education for Students Placed at Risk*, 10(3), 309–332.
- Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*.
- Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: The new Medical Research Council guidance. *BMJ*, 337, a1655. <https://doi.org/10.1136/bmj.a1655>.
- Dillman, D. A., Sinclair, M. D., & Clark, J. R. (1993). Effects of questionnaire length, respondent-friendly design, and a difficult question on response rates for occupant-addressed census mail surveys. *Public Opinion Quarterly*, 57(3), 289–304.
- European Agency for Special Needs and Inclusive Education. (2018). In J. Ramberg, A. Lénárt, & A. Watkins (Eds.), *European agency statistics on inclusive education: 2016 dataset cross-country report*. Odense, Denmark.
- Fisher, W. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transactions*, 21, 1. Available from: <https://www.rasch.org/rmt/rmt211.pdf> (Accessed 10 July 2019).
- Florian, L. (2012). Preparing teachers to work in inclusive classrooms: Key lessons for the professional development of teacher educators from Scotland's inclusive practice project. *Journal of Teacher Education*, 63(4), 275–285.
- Glazzard, J. (2013). A critical interrogation of the contemporary discourses associated with inclusive education in England. *Journal of Research in Special Educational Needs*, 13(3), 182–188.
- Göransson, K., & Nilholm, C. (2014). Conceptual diversities and empirical shortcomings—a critical analysis of research on inclusive education. *European Journal of Special Needs Education*, 29(3), 265–280.
- Hollenweger, J. (2011). Development of an ICF-based eligibility procedure for education in Switzerland. In *BMC public health* (Vol. 11, No. 4). BioMed Central. p. S7.
- Imms, C., Adair, B., Keen, D., Ullenhag, A., Rosenbaum, P., & Granlund, M. (2016). 'Participation': a systematic review of language, definitions, and constructs used in intervention research with children with disabilities. *Developmental Medicine & Child Neurology*, 58(1), 29–38.
- Ingram, D., Louis, K. S., & Schroeder, R. G. (2004). Accountability policies and teacher decision making: Barriers to the use of data to improve practice. *Teachers College Record*, 106(6), 1258–1287.
- Izadi-Najafabadi, S., Ryan, N., Ghafooripoor, G., Gill, K., & Zwicker, J. G. (2019). Participation of children with developmental coordination disorder. *Research in Developmental Disabilities*, 84, 75–84.
- James Lind Alliance. (2016). *Learning Difficulties (Scotland) Top 10*. Available from: <http://www.jla.nihr.ac.uk/top-10-priorities>.
- John-Akinola, Y. O., & Nic-Gabhainn, S. (2014). Children's participation in school: A cross-sectional study of the relationship between school environments, participation and health and well-being outcomes. *BMC Public Health*, 14(1), 964.
- Kreiner, S. (2007). Validity and objectivity: Reflections on the role and nature of Rasch models. *Nordic Psychology*, 59(3), 268–298.
- Ligtvoet, R., Van der Ark, L. A., te Marvelde, J. M., & Sijtsma, K. (2010). Investigating an invariant item ordering for polytomously scored items. *Educational and Psychological Measurement*, 70(4), 578–595.
- Linacre, J. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3(1), 85–106.
- Linacre, J. (1994). Sample size and item calibration stability. *Rasch Mes Trans.*, 7, 328.
- Maciver, D., Hunter, C., Adamson, A., Grayson, Z., Forsyth, K., & McLeod, I. (2018). Supporting successful inclusive practices for learners with disabilities in high schools: A multisite, mixed method collective case study. *Disability and Rehabilitation*, 40(14), 1708–1717.
- Maciver, D., Rutherford, M., Arakelyan, S., Kramer, J. M., Richmond, J., Todorova, L., ... O'Hare, A. (2019). Participation of children with disabilities in school: A realist systematic review of psychosocial and environmental factors. *PLoS One*, 14(1), Article e0210511.
- Marais, I., & Andrich, D. (2008). Formalizing dimension and response violations of local independence in the unidimensional Rasch model. *Journal of Applied Measurement*, 9(3), 200–215.
- Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, 63(7), 737–745.
- Na, M., Gross, A. L., & West, K. P. (2015). Validation of the food access survey tool to assess household food insecurity in rural Bangladesh. *BMC Public Health*, 15(1), 863.
- Norwich, B. (2008). *Perspectives and Purposes of Disability Classification Systems*. *Disability Classification in Education: Issues and perspectives* (p. 131).
- Norwich, B. (2016). Conceptualizing special educational needs using a biopsychosocial model in England: the prospects and challenges of using the International Classification of Functioning framework. *Frontiers in Education*, 1(December), 5. Frontiers.
- Organisation for Economic Co-operation and Development. (2012). *PISA 2012 Technical Report*. Available from: <https://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf>.
- Peny-Dahlstrand, M., Krumlinde-Sundholm, L., & Gosman-Hedstrom, G. (2013). Patterns of participation in school-related activities and settings in children with spina bifida. *Disability and Rehabilitation*, 35(21), 1821–1827.
- Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health*, 30(4), 459–467.
- Riddell, S. (2012). *Education and disability/special needs, policies and practices in education, training and employment for students with disabilities and special educational needs in the EU* (p. 201). Brussels: NESSE Network.
- Ruijs, N. M., & Peetsma, T. T. (2009). Effects of inclusion on students with and without special educational needs reviewed. *Educational Research Review*, 4, 67–79.
- Scottish Government. (2019). *Guidance on the presumption to provide education in a mainstream setting*. https://www.parliament.scot/S5_Education/Inquiries/20190326Guidance_on_the_presumption_to_mainstreaming.pdf.
- Sijtsma, K., & Hemker, B. T. (1998). Nonparametric polytomous IRT models for invariant item ordering, with results for parametric models. *Psychometrika*, 63(2), 183–200.
- Sturgis, P., Hughes, G., & Smith, P. (2006). *A study of suitable methods for raising response rates in school surveys*. Department for Education and Skills.
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55.
- van der Veen, I., Smeets, E., & Derricks, M. (2010). Children with special educational needs in the Netherlands: Number, characteristics and school career. *Educational Research*, 52(1), 15–43.
- 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, S22–S35.

- World Health Organization. (2007). *International classification of functioning, disability, and health: Children & youth version*. Geneva: WHO.
- Wilson, C., Woolfson, L. M., Durkin, K., & Elliott, M. A. (2016). The impact of social cognitive and personality factors on teachers' reported inclusive behaviour. *British Journal of Educational Psychology, 86*(3), 461–480.
- Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement, 30*(3), 187–213.
- Zablotsky, B., Black, L. I., Maenner, M. J., Schieve, L. A., Danielson, M. L., Bitsko, R. H., Blumberg, S. J., Kogan, M. D., & Boyle, C. A. (2019). Prevalence and trends of developmental disabilities among children in the United States: 2009–2017. *Pediatrics, 144*(4), Article e20190811.