Article

Intra-rater test-retest reliability of self-reported child functioning module

Kwok Ng [1,2,*]**, Piritta Asunta** [3]**, Niko Leppä** [4] **and Pauli Rintala** [5]

[1] School of Educational Sciences and Psychology, University of Eastern Finland, Joensuu, Finland; kwok.ng@uef.fi

[2] Physical Activity for Health Research Cluster, Department of Physical Education and Sport Sciences, University of Limerick, Ireland; kwok.ng@ul.ie

[3] LIKES Physical Activity for Health Research Centre, Jyväskylä, Finland; piritta.asunta@likes.fi

[4] City of Jyväskylä, Jyväskylä, Finland; niko.leppa@jkl.fi

[5] Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland; pauli.rintala@jyu.fi

* Correspondence: kwok.ng@hbsc.org; KN

Abstract: Determining disability prevalence is an important area for population statistics, especially among young adolescents. The Washington Group on Disability Statistics is one source of reporting disabilities through functional difficulties. However, young adolescents self-reporting this measure is in its infancy. The purpose of this study was to carry out an intra-rater test-retest reliability study on a modified set of items for self-reporting functional difficulties. Young adolescents (n=74; boys=64%; age m=13.7, SD=1.8) completed a self-reported version of the child functioning module in a supervised classroom. The second administration took place two weeks later. Intraclass correlation coefficient (ICC) and Kappa (k) statistics were used to test reliability of the items, and interpretation through Landis & Koch, and Cohen, respectively. The majority of items had substantial or moderate agreement, although there was only fair agreement for self-care (ICC=0.59), concentration (ICC=0.50), and routine (ICC=0.54). Kappa statistics of behaviour were interpreted to be large (k=0.65), and seeing (k=0.49), walking (k=0.49), and speaking (k=0.49) difficulties were moderate. The majority of the items in the self-reported version of the child functioning module can be used in a scale format, although some caution may be required on items of self-care and concentration when used as a dichotomous variable.

Keywords: disability statistics; kappa; intraclass correlation coefficient; young adolescents; functional difficulties; special education; survey; health behaviour

1. Introduction

Based on the Salamanca Agreement on Inclusive Education, all children have the right to education, irrespective of individual difficulties [1]. Since then, the Finnish education system has been progressing towards more inclusion in schools by passing the Education Act in 2010, where families have the choice for children to attend a general school, special educational class, or special school [2]. Changes to the educational structures has seen a year on year increase in the number of children in comprehensive schools who require special or intensified support from 8% in 2010 to 20% in 2019 [3].

A multi-tiered framework explains this big rise. In Finland there is a three-tier support system, with the purpose of support learning at the earliest possibility for the child and to be within inclusive environments. The Basic Education Act [2] and the three-tier framework was officially implemented in August 2010 among every Finnish school [4]. The support system allows these pupils to become part of the general school, be in environments whereby they have access, and can participate in the same activities as their peers. This type of support is described as Tier 1 – general support. In Tier 1, the support level is offered for every pupil in the Finnish education system, Tier 2 is intensified support, and in Tier 3, pupils are given special support.

In addition to monitoring academic progress, schools form a good place to recruit children for important health checks as well as carry out health surveys.   Monitoring tools of health behaviours should also include children with support needs [5]. However, few instruments do this. The majority of surveys often exclude children with disabilities [6], which may lead to response bias when it comes to national reporting. Furthermore, completion of survey instruments may be inappropriate for children with support needs, and thus a gap in knowledge of health behaviours among children with disabilities exists.

1.2 Difficulties in measuring disabilities

Conceptually, measurement of disabilities has its difficulties [7]. There is often a stigma related to reporting of disabilities and short measures often lack detail to understand what features make a person feel like they have disabilities [8]. To address these previously reported issues, items based on the WHO International Classification of Functioning, Disability and Health (ICF) are used as indicators for disabilities [9,10]. Core functions that influence children's development, based on the Washington Group on Disability Statistics short set [11], have been created with the assistance of UNICEF [12], to present the child functioning module (CFM). Although there has been a number of studies that have tested the viability of the CFM [13], these were primarily based on the proxy version of the questionnaire set. It is not known if these instruments can be used in the context of children taking part in a self-reported survey.

Adolescents need to be able to self-report their own overall health (physical, mental and social) and such information is often referred to health-related quality of life [14]. It is not uncommon for adolescents with disabilities to report lower ratings of their own health-related quality of life [15]. Self-reporting of health-related quality of life is a predictor of temporal functioning however, details of specific fixed impairments are often neglected in research [14]. Therefore, it is essential other health-related data is collected. For health behaviour surveys, it is important to have reliable instruments as part of the validation process. Intra-rater reliability can be carried out through a test-retest mode, whereby participants carry out the test twice [16]. Completion of the test-retest can yield recency effects, whereby responses reflect on memory of responses rather than reporting actual behaviours [17]. However, too much time between survey completion may generate true changes in the responses due to behavioural changes and that would alter the test-retest scores [18]. Given the importance for accurately measuring disabilities among children with special support needs, the aim of this study was to carry out a test and retest reliability study on the self-report version of the CFM among children with supportive needs in schools.

2. Materials and Methods

The study had received approval by the blind institution ethical committee. According to the Finnish Ministry of Education school lists, there are 60 schools with special education status. The location of the schools was examined, and a convenience sample was selected based on schools clustered in region of Finland. A one-tail test with power at .80, alpha at 0.05 and 0.30 as the hypothesized level of correlation, specified the target sample size needs to be 67 [19].

2.1 Procedures

Schools in the allocated region (n=10) were contacted. A researcher (NL) described the procedures of the study and asked if it was possible to obtain permission to take part in the study. Schools who agreed (n=4) to take part in the study were rewarded by receiving equipment for adapted physical education and sports. School principals selected a class in the school with children in equivalent grades whereby the age ranged between 11y-15y old. This age range was chosen as other items in the questionnaire were appropriate for young adolescents in other national and international health behaviour in school-aged children surveys. Principals were asked to make a list of pupils who would be able to complete a survey independently (ability to read questions and enter a response on a computer by clicking a mouse) and then randomly selected the pupils.

Researchers visited the school site to administer the online surveys. The class teachers were given a short website address link to give to each of the pupils. There were different links depending on the age or ability of the pupils (more about the surveys later). Researchers were present to give instructions to the pupils, teachers, and teacher assistants before telling the pupils they can start the survey. Some of the children had personal assistants with them, and some other children shared the assistant. Pupils entered their responses on the computers by themselves. Students were permitted to ask teachers, assistants, and researchers to clarify on some items they did not seem to understand, at times the assistants may have read out aloud the question directly to the pupil. Some pupils needed specific clarification for abstract questions for example, Cantrill's life satisfaction ladder [20]. These items were not included in this intra-rater test-retest study.

Teachers were asked to allocate a choice of four surveys to the pupils based on the age and developmental stages of the individual. The surveys were; 1) Long survey (L) with 60 questions targeted at pupils aged 15y; 2) an easy-to-read modification of the long (L-er) survey targeted at pupils aged 15y but with basic language requirements; 3) Medium survey (M) with 40 questions targeted at pupils aged between 11-13y; and, 4) an easy-to-read modification of the medium (M-er) survey targeted at pupils aged between 11-13y but with basic language requirements. The reduction of items between the two age groups were based on the experiences of survey design from the WHO Collaborative Health Behaviour in School-aged Children (HBSC) study [21].

The L and M versions of the survey were sent to the Finnish Easy-to-Read service to make the changes to the question items. The items were then sent back to the research team for consideration. Modifications continued until there was agreement between the Easy-to-Read service and the researchers so there would be consistency with original and modified constructs. Although there were differences in the number of questions in M and L, the placement of the CFM was the same, both at the beginning of the surveys. Placement at the beginning of the survey means, the aims of the study are unaffected by which version of the survey was completed.

The pupils completed the survey independently on two occasions. Teachers and researchers were available to clarify any questions the pupils had when completing the survey, but were instructed in the protocol to avoid answering it for them. The time between surveys was two weeks. Surveys were completed through an online survey platform. However, for part of the first data collection date, there were server outages and for those participants (n=14), the survey was carried out by pen and paper (print out of the online survey) and coded in by the researchers. Subsequent surveys were completed through the online survey. There were further server outages during the data collection period, however, responses were refreshed in order for the data to be saved.

2.2 Measures in this study

Pupils entered their sex (boy or girl) their month and year of birth. A calculation was made based on the time of survey completion to create an age variable.

2.2.1 Child Functioning Module

The child functioning module (CFM) was derived from the joint work of the Washington Group on Disability statistics and UNICEF [22]. However, the original was modified in several ways that allow for cultural differences. The first modification was to transfer the content from proxy reporting (by parents) to self-report. For example, the original question would begin with "Does your child have difficulties in…" and the modified version became, "Do you have difficulties in…" The next modification was based on item reduction. The CFM has a layered approach to functioning, and the modified version was based on a single item per function. For example, the CFM has three items related to the seeing function. The first item is a screener for whether the child uses glasses or contact lenses, and then depending on the answer, there is a skip function to assess the difficulty in seeing. The modified version we used was a single item about "seeing difficulties, even if the child wears glasses or contact lenses." This type of modification has been used in the development of the Washington Group Short Set questions [11]. The third modification was to group the items together to give the impression the child was answering fewer questions. In the CFM, there are separate

questions for each functioning. In the modification, the same header was used, "Compared to children of the same age, do you have difficulties in…", and then the corresponding functions were listed. This was the presentation of the items in the L and M version. The entire sentence was included in the easy-to-read versions. The differences between the L and M versions and the L-er and M-er versions are in the Table 1.

Table 1. Items of the self-report Child Functioning Module in plain and easy-to-read English

	Plain Language	Easy to Read version
	Compared to children of the same age, do you have difficulties in…	Compare yourself to other teenagers of the same age. Which things are easy or difficult for you?
1	Seeing (even if you have to wear glasses or contact lens)?	Do you have difficulties in seeing (even if you have to wear glasses or contact lens)?
2	Hearing (even if you have a hearing aid)?	Do you have difficulties in hearing (even if you have a hearing aid)?
3	Walking 100m, for example a length of a football pitch (even if you use assistance)?	Do you have difficulties in walking 100m, for example the length of a football pitch (even if you use assistance)?
4	Self-care, for example eating or dressing up?	Do you have difficulties in self-care, for example eating or dressing up?
5	Being understood when speaking (outside of the home)?	Do you have difficulties in being understood when speaking to people outside of your home?
6	Learning things?	Do you have difficulties in learning things?
7	Remembering things?	Do you have difficulties in remembering things?
8	Concentrating on things you enjoy?	Do you have difficulties in concentrating on things you enjoy?
9	Making changes to your own routine?	Do you have difficulties in making changes to your own routine?
10	Controlling your own behaviours?	Do you have difficulties in controlling you own behaviours?
11	Getting friends?	Do you have difficulties in getting friends?

All items had a four-category response scale with the following options, "None", "Some", "A lot", and "Cannot do". Translations of the items were carried out with contextual back translations. Unlike direct back translations, contextual translations take into context of the local language during the translation process [23]. The translations were corrected until experts in disability and adolescence surveys (KN, PR, NL, PA) were satisfied the items in Finnish matched the original items. Moreover, in the translation process, visual representations of the response scales were used to help the respondents to understand the differences between the response options. They were colour coded from green for 'None', orange for 'some', red for 'a lot', and a cross for 'cannot do'.

One final modification was made to this self-report version of the CFM. The CFM has items related to mental functions [9]. One item is related to the functions surrounding being very anxious, nervous or worried, and the other item is related to being sad or depressed. The response scale in the CFM is different to the other functions, whereby questions were related to frequency of mental dysfunction. This is because corresponding responses for items on mental dysfunction would be difficult to comprehend. Whereas frequencies of recalling symptoms are reliable ways among populations who complete the survey [24]. Based on earlier research on a psychosomatic symptom checklist, the two items closest in relation to these two items were also included in the survey [25]. The items used were headed with the following, "How often have you had the following symptoms over the past 6 months? Tick one box for each symptom". Symptoms listed were, depression or

feeling low, and nervousness. The response scale included the following; "almost daily", "more than once a week", "approximately once a week", "approximately once a month", and "less or never". Due to the differences in the way the CFM was used in our study, these results were not reported.

2.3 Analyses

The survey data was combined between the test and retest surveys. A unique identifier was coded for each participant for each survey. Data from participants who completed both surveys were included in the final data sheet. The data were imported into IBM SPSS version 24.0 for statistical analyses. Reliability between test and retest was computed through the single measure of intraclass correlation coefficients (ICC). The two-way random model with absolute agreement type was performed, and test statistics set to 95% confidence intervals (CI). Acceptable reliability criteria were based on the Landis and Koch divisions of agreement (Landis and Koch, 1977). To interpret the categories, the following were used, less than 0.20; Slight or poor, 0.21-0.40; Fair, 0.41-0.60; Moderate, 0.61-0.80; Substantial, and over 0.80; Almost perfect.

Single functions were also dichotomised to test various cut points between a state of 'disability' versus 'no disability'. Two sets of cut off values for each functions were set to 1) at least "some", and 2) at least "A lot" as guided by previous research [11]. To test this, the Cohen's Kappa statistics were used to estimate the stability of each function. Cohen's Kappa can be interpreted with the following correlation values, greater than 0.5 being large, 0.3-0.5 moderate, 0.1-0.3 small, and less than 0.1 trivial [26].

3. Results

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation as well as the experimental conclusions that can be drawn.

3.1. Descriptive Results

The majority of the participants (n=74) completed the M-version of the survey (Table 2). According to the cut-off points of at least some difficulties, to indicate disabling functions, almost two thirds of the respondents would be considered to have disabilities.

Table 2. Description of the breakdown of participants in the study

	M-er	M	L-er	L	Total
Total	6	54	5	9	74
Boys	5	32	3	7	47
Girls	1	22	2	2	27
Disabilities (% some→)	4 (67%)	31 (56%)	5 (100%)	5 (55%)	57 (65%)
Age in years (SD)	14.8 (1.2)	12.3 (1.5)	16.1 (0.5)	15.6 (0.6)	13.2 (1.8)

The prevalence of disabilities in the study varies depending on which cut-point is used (Table 3). The most common functional limitations where the individual has some difficulties were in the cognitive domain, such as difficulties in learning (32.4%) or in remembering (31.1%). The most common function adolescents reported they could not do (most severe limitation), was the domain of getting friends (5%).

Table 3. Distribution of reporting difficulties from retest survey and prevalence using the cut-points based at some difficulties and a lot of difficulties

Difficulties in…	None	Some	A lot	Cannot do	Missing	Some+ %	A lot+ %
Seeing	66	6	0	2	0	10.8	2.7
Hearing	66	7	0	1	0	10.8	1.4

Walking	68	4	0	2	0	8.1	2.7
Self-Care	65	7	1	1	0	12.2	2.7
Speaking	63	8	2	1	0	14.9	4.1
Learning	50	19	3	2	0	32.4	6.8
Remembering	51	20	2	1	0	31.1	4.1
Concentration	58	14	0	1	1	20.5	1.4
Routine	56	15	1	1	1	23.3	2.7
Behaviour	56	14	1	2	1	23.3	4.1
Friends	58	10	2	3	1	20.5	6.8

3.2 Test-retest results

According to the interpretation by Landis & Koch [27], six of the 11 functions had substantial agreement after a two-week gap between completing the survey (Table 4). Difficulties in learning and difficulties in getting friends had moderate agreement. Three items (self-care, concentration, and maintaining routines) had fair agreement.

Table 4. Intra-rater reliability results of self-reported child functioning module

Function	ICC	LCI	UCI	Landis & Koch	Kappa1	p	Cohen	Kappa2	p	Cohen
Seeing	0.753	0.608	0.845	Substantial	0.801	<.001	Large	0.486	<.001	moderate
Hearing	0.63	0.411	0.768	Substantial	0.457	<.001	Moderate	n/a		
Walking	0.645	0.449	0.782	Substantial	0.509	<.001	Large	0.486	<.001	moderate
Self-Care	0.345	-0.027	0.585	Fair	0.251	.018	Small	0		poor
Speaking	0.76	0.619	0.849	Substantial	0.472	<.001	Moderate	0.490	<.001	moderate
Learn	0.59	0.35	0.742	Moderate	0.410	<.001	Moderate	0.210	.061	small
Remembering	0.625	0.405	0.764	Substantial	0.563	<.001	Large	-0.210	.836	small
Concentration	0.204	-0.274	0.502	Fair	0.28	.016	Small	-0.019	.866	poor
Routine	0.277	-0.145	0.544	Fair	0.341	.003	Moderate	-0.034	.767	poor
Behaviour	0.732	0.573	0.832	Substantial	0.548	<.001	Large	0.652	<.001	large
Friends	0.57	0.314	0.73	Moderate	0.531	<.001	Large	0.256	.015	small

ICC – Intraclass correlation coefficient, LCI – lower confidence interval, UCI – Upper confidence interval, Kappa1 – Kappa based on cut-point of 'at least some difficulties', Kappa2 – Kappa based on cut-point of 'at least a lot of difficulties', n/a – not available

Kappa was tested on two cut-off points, at least some difficulties (Kappa1), and at least a lot of difficulties (Kappa2). According to the interpretation by Cohen & Cohen [16], five out of 11 functions had large (seeing, walking, remembering, behaviour, friends) Kappa1 values. There were four moderate (hearing, speaking, learn, routine) and two small (self-care, concentration) Kappa1 values. There was one large Kappa2 value for difficulties in "behaviour". In addition, three other difficulties had moderate (seeing, walking, speaking) Kappa2, three with small (learn, remembering, friends) Kappa2, and three with poor (self-care, concentration, routine) Kappa2 values.

Difficulties in seeing, walking, remembering and controlling behaviours performed consistently as an entire scale and as cut-off points used to determine disability classification. There was not enough test-retest data for pupils who reported difficulties in hearing to determine how well the test-retest performed. In other words, none of the individuals, who reported at least a lot of difficulties in hearing during the test survey, completed the retest survey.

Difficulties in remembering and making friends functions had inconsistent results. There was substantial agreement across the scale of remembering difficulties, large agreement for Kappa1, but small Kappa2 values. Difficulties in making friends had large Kappa1 values, small Kappa2 values, and moderate agreement across the scale. Other subtle differences across the results were noted. Difficulties in self-care and concentration are items that have fair agreement and small Kappa1

values. Difficulties in making changes to routines also had fair agreement, but the Kappa1 value was moderate.

4. Discussion

To the authors' knowledge, this is the first time the CFM has been tested without proxy in a special educational setting. The use of the CFM for self-reported disabilities is, overall, an acceptable measure. This is an important finding because previous work on these items have been based on proxy reporting [28,29] and there is a need to include self-reported disabilities in national health surveys [5]. More specifically, in this study the items on the self-reported version of the child functioning module were completed by over 85% of the pupils with special support needs. There were specific items, most notably the item on 'self-care', 'concentration', and 'changes in routines' and may need to undergo further development to ensure acceptable reliability, especially in a Finnish special education setting. These new findings are discussed in this paper.

4.1 Reliability as a scale

The self-reported version of the CFM was designed to have the same response options as the proxy report version [28]. In eight of the 11 items, the four response categories were answered with substantial or moderate agreement. However, the level of agreement on items on difficulties with self-care, concentrating on things the child enjoys, and having changes to the routine were only fair. Similar problems with the item on difficulties with self-care were reported in an inter-rater reliability study between parents and teachers found poor agreement [29]. This would suggest these items may have different meanings at different times of survey completion, and may need to be interpreted with caution [16].

Upon inspection of the item concerning 'self-care', one of the problems may be the examples of self-care presented in the item itself. The examples consisted of two different types of behaviours namely, eating or dressing up. The functions in relation to eating are vast. These can include fine motor coordination, such as the ability to use cutlery, means of swallowing, as well as desire to eat food. According to the ICF-child and youth version, there are five different codes related to just eating [10]. The other example of self-care; dressing up, may consist of differing functions. These may include as gross motor coordination, such as putting arms through cloths, fine motor coordination to do buttons or pull up the zip, as well as other functions such as selecting clothes. Again, when mapped against the ICF, various different body functions as well as contextual factors are involved with this task of 'self-care' [30]. Therefore, it may not be surprising this item had low levels of reliability. It may be worthwhile to use only one concrete example of self-care that exemplifies child behaviour. For the purpose of international comparability, modifications to the scale need to be explicitly stated when reporting the prevalence of children with self-care difficulties [22,28].

Another item with fair agreement levels was the item, concentrating on things the individual enjoys. Children's enjoyment of activities may change from one moment to another [31]. The instrument could easily be misinterpreted when there is a lack of consistency of behaviours being reported [24]. Naturally, the item was designed for reporting by the parents, and it is assumed the parents would know what the child enjoys doing [12]. However, this notion has been challenged as reported by Mactaggart and colleagues [13], who reported adults over reported the functional difficulties from the child perception of difficulties. This could be because social interactions increase with peers and decline with family during adolescence [32]. More critical considerations are needed for this item when using both self-report and use by proxy among adolescents.

The item on concentrating was created as an extension of the Washington Group short set of six items, whereby one of them was related to 'difficulties in remembering and concentration' [33]. It was not featured in an draft reliability study (collected in 2015) of the CFM in a special education setting [29], indicating the possibility of low level of evidence of the item. One observation of the items, as a whole, is there are more child related domains in the CFM and it is assumed they are uniform across the ages of 5-18 years old [13]. Although the CFM is divided into early childhood – between 2-4 years, there are no different question sets between pre- and post-puberty, or pre- and

early adolescence. This may be regarded as a weakness of the CFM, and differences in either adult or child perceptions during adolescence may need to be encouraged for another separate package. An example of how survey items change is the Harter's self-perception scale [34]. Harter's scale was originally tested for 7-9 year olds, and was later adapted for adolescents [35]. Perhaps such a convention maybe required for the self-report version of the CFM.

4.2 Reliability as two types of cut-offs

In reporting groups of children with disabilities, there are variable cut-off options. In our study, we carried out test-retest on two cut-off values. The items on seeing, hearing, remembering things, controlling own behaviour, and making friends could be interpreted to have had large agreement when the first level cut-off (at least some difficulties) was administered. These results partly contrast the earlier evidence from the draft reliability tests of the CFM, whereby the levels agreements between parents and teachers was poor when reporting children who have at least some difficulty in making friends [29]. Further evidence is needed whereby the aim of triangulating the data from three main sources, the pupil, their parents, and teacher is examined before conclusions about these differing items can be made.

The first level cut-off value of at least some difficulties gives an indication on the number of adolescents who perceive any type of difficulties in performing the function. Aggregating this information may yield high prevalence of disability and may serve a purpose for providing indicators of trends over time. The second cut off, has been used as an indicator of disability prevalence in national based studies [36]. In the case of hearing difficulties, there were not enough study participants to give a reliability statistic. There was large agreement for reporting behaviour difficulties, and moderate agreement for seeing, walking, and speaking difficulties. The cut-off points for disability prevalence in the functions of seeing, walking, speaking, and controlling own behaviours may be used among young adolescents in the special school environment.

This research offers new insight into the way children may self-report their own functional difficulties as an indicator for disabilities. The development of the work has been a long process, from the point of view of population statistics [11], to transfer the context for children [28], before it was converted to self-report for adolescents [37]. Through these steps, it would be possible to create data pooling for future big data sets. This could be a cost-effective answer to the problem where typically, group sizes are insufficiently large enough to make statistical comparisons and other analyses. For example, in the Finnish national monitoring study from over 6000 children and adolescents on physical activity behaviours, there were not enough cases to report difficulties in walking after stratifying by age and gender [38]. Children with walking difficulties are in an important group as they have reportedly been considered to have the lowest levels of physical activity [39]. Although the results from our study may suggest caution is required when interpreting some of the items, researchers and policy makers who use these items may need to consider which variables can actually be used to describe the prevalence of disabilities [40,41].

4.3 Limitations

The sample was limited to children only in Finnish-speaking special schools in a region of Finland who were able to complete an online questionnaire. Different concepts of functional difficulties may exist in other environments. The study was on the intra-rater stability of the items, and reflects the perception of the child's functional abilities, rather than corroborate with other data from other sources to validate the actual abilities. It may be necessary to examine the construct and face validity of the items when interpreting the findings in future studies that adopt the child functioning module.

5. Conclusions

Self-reporting of functional difficulties is a subtle way of measuring childhood disabilities. A common approach to reporting disabilities has been through dichotomous variables comprising of

adolescents with or without disabilities. There were large and moderate agreements when a cut off at least a lot of difficulties in four of the variables; seeing, walking, speaking, and controlling own behaviour difficulties were used. Most of the self-report version of the Child Functioning Module can be used for adolescents, even if they have special or intensified support needs. The stability of two items (self-care and concentration) were poor and these items should be carefully considered in other studies. Further validation is required to support these findings.

Author Contributions: Conceptualization, KN.; methodology, KN, PA, NL, PR; formal analysis, KN.; writing—original draft preparation, KN, PR; writing—review and editing, KN, PA, NL, PR; funding acquisition, PR. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. 1. UNESCO. The Salamanca Statement and Framework for Action on Special Needs Education. **1994, June**.
2. 2. Finland. Basic Education Act. **2010**, *628/1998. Amendments up to 1136/2010*.
3. 3. Official Statistics of Finland. Special Education - Concepts and Definitions [E-Publication]. **2018, January**.
4. 4. Pesonen, H.; Itkonen, T.; Jahnukainen, M.; Kontu, E.; Kokko, T.; Ojala, T.; Pirttimaa, R. The Implementation of New Special Education Legislation in Finland. Educational Policy **2015**, *29*, 162-178.
5. 5. Tardi, R.; Njelesani, J. Disability and the Post-2015 Development Agenda. Disabil. Rehabil. **2015**, *37*, 1496-1500.
6. 6. Rathmann, K.; Vockert, T.; Bilz, L.; Gebhardt, M.; Hurrelmann, K. Self-Rated Health and Wellbeing among School-Aged Children with and without Special Educational Needs: Differences between Mainstream and Special Schools. Res Dev Disabil **2018**, *81*, 134-142.
7. 7. Cappa, C.; Petrowski, N.; Njelesani, J. Navigating the Landscape of Child Disability Measurement: A Review of Available Data Collection Instruments. ALTER **2015**, *9*, 317-330.
8. 8. World Bank. Disability - Statistics and Data Collection. **2009, November**, *2010*.
9. 9. WHO. *International Classification of Functioning, Disability and Health (ICF)*.; World Health Organization: Geneva, Switzerland, 2001.
10. 10. WHO. *International Classification of Functioning, Disability and Health: Children and Youth Version: ICF-CY.*; World Health Organization: Geneva, 2007; pp. 322.
11. 11. Mont, D. Measuring Health and Disability. The Lancet **2007**, *369*, 1658-1663.
12. 12. Schutz, M.E.; Durkin, M.S.; Gottlieb, C.; Maenner, M.; Cappa, C.; Loaiza, E. Monitoring Child Disability in Developing Countries; Results from the Multiple Indicator Cluster Surveys. **2008**.
13. 13. Mactaggart, I.; Cappa, C.; Kuper, H.; Loeb, M.; Polack, S. Field Testing a Draft Version of the UNICEF/Washington Group Module on Child Functioning and Disability. Background, Methodology and Preliminary Findings from Cameroon and India. ALTER - European Journal of Disability Research / Revue Européenne de Recherche sur le Handicap **2016**, *10*, 345-360.
14. 14. Marker, A.M.; Steele, R.G.; Noser, A.E. Physical Activity and Health-Related Quality of Life in Children and Adolescents: A Systematic Review and Meta-Analysis. Health Psychology **2018**, *37*, 893-903.
15. 15. Simeoni, M.; Schmidt, S.; Muehlan, H.; Debensason, D.; Bullinger, M. Field Testing of a European Quality of Life Instrument for Children and Adolescents with Chronic Conditions: The 37-Item DISABKIDS Chronic Generic Module. Quality of Life Research **2007**, *16*, 881-93.
16. 16. Cohen, J.; Cohen, P. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*.; Lawrence Erlbaum, Oxford: Oxford, England, 1975; pp. 490-xxi, 490.
17. 17. Lenzner, T.; Kaczmirek, L.; Lenzner, A. Cognitive Burden of Survey Questions and Response Times: A Psycholinguistic Experiment. Applied Cognitive Psychology **2010**, *24*, 1003-1020.
18. 18. Sallis, J.F.; Saelens, B.E. Assessment of Physical Activity by Self-Report: Status, Limitations, and Future Directions. Res. Q. Exerc. Sport **2000**, *71*, 1-14.
19. 19. Donner, A.; Eliasziw, M. Sample Size Requirements for Reliability Studies. Statist. Med. **1987**, *6*, 441-448.

20.   20. Ravens-Sieberer, U.; Torsheim, T.; Hetland, J.; Vollebergh, W.; Cavallo, F.; Jericek, H.; Alikasifoglu, M.; Välimaa, R.S.; Ottova, V.; Erhart, M. Subjective Health, Symptom Load and Quality of Life of Children and Adolescents in Europe. International Journal of Public Health **2009**, *54*, 151-159.

21.   21. Kämppi, K.; Välimaa, R.S.; Tynjälä, J.A.; Haapasalo, I.; Villberg, J.; Kannas, L.K. *Peruskoulun 5., 7. Ja 9. Luokan Oppilaiden Koulukokemukset Ja Koettu Terveys; WHO-Koululaistutkimuksen Trendejä Vuosina 1994-2006* [Basic School 5th, 7th, and 9th Grade Pupils School Experience and Perceived Health: HBSC Study Trends 1994-2006]. **2008**, *Terveyden edistämisen tutkimuskeskus ja tekijät*.

22.   22. Loeb, M.; Mont, D.; Cappa, C.; De Palma, E.; Madans, J.; Crialesi, R. The Development and Testing of a Module on Child Functioning for Identifying Children with Disabilities on Surveys. I: Background. Disabil Health J **2018**, *11*, 495-501.

23.   23. Bergendorff, S. Contextual Translation – Concepts and Practices among the Mekeo of Papua New Guinea. Oceania **2009**, *79*, 225-237.

24.   24. Lee, S. Self-Rated Health in Health Surveys. In *Health Survey Methods*.; Johnson, T.P., Ed.; John Wiley & Sons: Hoboken, New Jersey, 2014, pp. 193-216.

25.   25. Boyce, W.F.; Davies, D.; Raman, S.R.; Tynjälä, J.A.; Välimaa, R.S.; King, M.; Gallupe, O.; Kannas, L.K. Emotional Health of Canadian and Finnish Students with Disabilities Or Chronic Conditions. International Journal of Rehabilitation Research **2009**, *32*, 154-161.

26.   26. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edition ed.; Lawrence Earlbaum Associates: Hillsdale, NJ, 1988.

27.   27. Landis, J.R.; Koch, G.G. The Measurement of Observer Agreement for Categorical Data. Biometrics **1977**, *33*, 159-174.

28.   28. Cappa, C.; Mont, D.; Loeb, M.; Misunas, C.; Madans, J.; Comic, T.; de Castro, F. The Development and Testing of a Module on Child Functioning for Identifying Children with Disabilities on Surveys. III: Field Testing. Disabil Health J **2018**, *11*, 510-518.

29.   29. Sprunt, B.; McPake, B.; Marella, M. The UNICEF/Washington Group Child Functioning Module-Accuracy, Inter-Rater Reliability and Cut-Off Level for Disability Disaggregation of Fiji's Education Management Information System. Int. J. Environ. Res. Public. Health. **2019**, *16*, 10.3390/ijerph16050806.

30.   30. Kostanjsek, N. Use of the International Classification of Functioning, Disability and Health (ICF) as a Conceptual Framework and Common Language for Disability Statistics and Health Information Systems. BMC Public Health **2011**, *11*, S3.

31.   31. Rosenbaum, P.; Gorter, J.W. The 'F-Words' in Childhood Disability: I Swear this is how we should Think! Child Care Health Dev **2012**, *38*, 457-463.

32.   32. Viner, R.M.; Ozer, E.M.; Denny, S.; Marmot, M.; Resnick, M.; Fatusi, A.O.; Currie, C. Adolescence and the Social Determinants of Health. Lancet **2012**, *379*, 1641-1652.

33.   33. Madans, J.H.; Loeb, M.E.; Altman, B.M. Measuring Disability and Monitoring the UN Convention on the Rights of Persons with Disabilities: The Work of the Washington Group on Disability Statistics. BMC Public Health **2011**, *11(Suppl 4)*, 8.

34.   34. Harter, S. The Perceived Competence Scale for Children. Child Development **1982**, *53*, 87-97.

35.   35. Hagborg, W.J. The Rosenberg Self-Esteem Scale and Harter's Self-Perception Profile for Adolescents: A Concurrent Validity Study. Psychology in the Schools **1993**, *30*, 132-136.

36.   36. Kanste, O.; Sainio, P.; Halme, N.; Nurmi-Koikkalainen, P. Toimintarajoitteisten Nuorten Hyvinvointi Ja Avun Saaminen – Toteutuuko Yhdenvertaisuus? Kouluterveyskyselyn Tuloksia [Well-being of Youth with Disabilities and Access to Help - is Equality Going to Happen? the Results of the School Health Promotion Survey]. **2017**, *Research Reports 24*, 1-8.

37.   37. Leppä, N.; Asunta, P.; Rintala, P. Tukea Tarvitsevat Lapset Vahvemmin Mukaan Lasten Ja Nuorten Liikuntakäyttäytymistutkimuksiin. [Children with Supportive Needs have to be Included Stronger in Physical Activity Behaviour Research]. Liikunta & Tiede **2018**, *55*, 13-15.

38.   38. Ng, K.W.; Rintala, P.; Asunta, P. Toimintarajoitteiden yhteydet liikuntakäyttäytymiseen [Associations between physical activity behaviours and functional difficulties]. In *Lasten Ja Nuorten Liikuntakäyttäytyminen Suomessa: LIITU-Tutkimuksen Tuloksia 2018 [Finnish Children and Adolescents Physical Activity Behaviours: Results of the F-SPA Study 2018]*.; Kokko, S. and Martin, L., Eds.; Valtion Liikuntaneuvosto [Finnish Sports Council] 2019:1: Helsinki, 2019, pp. 107-116.

39.   39. Ng, K.W.; Sainio, P.; Sit, C. Physical Activity of Adolescents with and without Disabilities from a Complete Enumeration Study (N = 128,803): School Health Promotion Study 2017. IJERPH **2019**, *16*, 3156.

40.   40. Madans, J.H.; Mont, D.; Loeb, M. Comments on Sabariego Et Al. Measuring Disability: Comparing the Impact of Two Data Collection Approaches on Disability Rates. Int. J. Environ. Res. Public Health, 2015, 12, 10329–10351. Int J Environ Res Public Health **2016**, *13*.

41.   41. Sabariego, C.; Oberhauser, C.; Posarac, A.; Bickenbach, J.; Kostanjsek, N.; Chatterji, S.; Officer, A.; Coenen, M.; Chhan, L.; Cieza, A. Measuring Disability: Comparing the Impact of Two Data Collection Approaches on Disability Rates. Int J Environ Res Public Health **2015**, *12*.