Item response drift in the Family Affluence Scale: A study on three consecutive surveys of the Health Behaviour in School-aged Children (HBSC) survey

Schnohr, Christina; Makransky, Guido; Kreiner, Svend; Torsheim, Torbjørn; Hofmann, Felix; De Clercq, Bart; Elgar, Frank J.; Currie, Candace

Published in:
Measurement

DOI:
10.1016/j.measurement.2013.06.016

Publication date:
2013

Document Version
Early version, also known as pre-print

Citation for published version (APA):
Schnohr, C. W., Makransky, G., Kreiner, S., Torsheim, T., Hofmann, F., De Clercq, B., ... Currie, C. (2013). Item response drift in the Family Affluence Scale: A study on three consecutive surveys of the Health Behaviour in School-aged Children (HBSC) survey. Measurement, 46, 3119-26.
https://doi.org/10.1016/j.measurement.2013.06.016

Measurement 46 (2013) 3119–3126

Contents lists available at SciVerse ScienceDirect

Measurement

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

Item response drift in the Family Affluence Scale: A study on three consecutive surveys of the Health Behaviour in School-aged Children (HBSC) survey

C.W. Schnohr [a,*], G. Makransky [b], S. Kreiner [a], T. Torsheim [c], F. Hofmann [d], B. De Clerq [e] F.J. Elgar [f], C. Currie [g]

[a] Institute of Public Health, University of Copenhagen, Copenhagen, Denmark
[b] University of Southern Denmark, Odense, Denmark
[c] University of Ghent, Ghent, Belgium
[d] Ludwig Boltzmann Institute Health Promotion Research, Vienna, Austria
[e] University of Bergen, Bergen, Norway
[f] McGill University, Montreal, Canada
[g] University of St Andrews, Scotland, United Kingdom

ARTICLE INFO

Article history:
Received 19 November 2012
Received in revised form 14 May 2013
Accepted 3 June 2013
Available online 27 June 2013

Keywords:
Family affluence scale
Differential item functioning
Health Behaviour in School-aged Children (HBSC)
Comparability
Validity

ABSTRACT

Comparable data on socio-economic position (SEP) is essential to international studies on health inequalities. The Health Behaviour in School-aged Children (HBSC) has used the Family Affluence Scale (FAS) on material assets. The present study used data collected from adolescents in eight countries in 2002, 2006 and 2010, and examined the construct validity of the FAS, by focusing on changes in item responses over time. The analyses reported the changes in means in item responses, and fitted models which estimated differential item functioning (DIF), and local dependency (LD) between items. DIF and LD were analysed by Graphical Log-Linear Rasch Models (GLLRMs), and changes in the measurement properties of the FAS over time and between countries were assessed. The results showed that the FAS items have changed their measurement properties between 2002 and 2010, and caution is warranted in studies comparing the FAS between different time points.

1. Introduction

International variations and trends in social inequalities in health are a staple of social epidemiological research [1]. Most of the research done in this area has involved adult populations, despite a broad consensus that reducing inequalities in child and adolescents health are essential to the health of future generations. A challenge in researching socio-economic differences in child and adolescent health lies in providing valid and comparable measurements on socio-economic status. Questionnaires to adolescents on their parents' education, occupation or income usually result in low completion rates and high misclassification rate [2–4] so various alternatives have been examined. Since 1998, the Health Behaviour in School-aged Children (HBSC) study has measured family affluence in more than 40 countries by collecting information on specific material assets in the Family Affluence Scale (FAS).

Abbreviations: HBSC, Health Behaviour in School-aged Children; SEP, socio economic position; FAS, Family Affluence Scale; HASC, home affluence scale; DIF, differential item functioning; LD, local dependence; GLLRM, Graphical Log-Linear Rasch Models; CLR, Conditional Likelihood Ratio test; IRT, item response theory.

* Corresponding author. Address: Øster Farimagsgade 5, P.O. Box 2099, DK – Copenhagen 1014 K, Denmark. Tel.: +45 29 35 70 94 (work)/28 23 37 75 (private).

E-mail address: cwsc@sund.ku.dk (C.W. Schnohr).

0263-2241/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.measurement.2013.06.016

Different theoretical perspectives can be taken on the FAS [2,5–7]. One implies that the items that make up the FAS may be considered as indicators of the material assets that produce affluence, which is called formative. Another is that the FAS items reflect a latent construct, which is called reflexive. Either way, the procedure for using the FAS in analysis often involved adding the scores on each item to produce a sum score. A formative index considers the items to be separate independent (albeit correlated) exposures of the family's material wealth, whereas a reflexive assumes correlation between items and to an underlying (latent) construct such as material wealth or affluence. This approach assumes a cumulative effect of the different items, and that items can be combined to a reliable scale measuring an underlying construct. Scales constructed by several items potentially suffer from differential item functioning and this is exacerbated when used in cross-national and trend analyses [8,9]. Differential item functioning (DIF) occurs when item responses vary across different subgroups and local dependency (LD) occurs when items correlate beyond the underlying construct expressed by the sum score. Earlier studies have shown large country variations in the FAS in terms of DIF, and a cautious approach has been warranted on using the FAS for international comparisons, for validity reasons [6,10].

In our experience, the FAS is no different to many scales. DIF and LD appears to be the rule rather than the exception in health related scales; recently developed software tools can be useful for identifying validity problems such as DIF. Scales are validated to make sure that measurements are not confounded by DIF and/or items measuring something other than intended, but in most cases, the task of validating a scale is overlooked perhaps because it is considered subordinate to addressing empirical research questions. It could be argued that this is true in the case of the FAS which was developed as a measure of deprivation in order to study health inequalities among adolescents [2,11,12].

1.1. The Health Behaviour in School-aged Children and the Family Affluence Scale

HBSC is an international study with 41 participating countries in the European Region and North America. The study aims at providing comparable data on young people's health and lifestyle from countries with different economic conditions and cultural, societal and political systems.

The HBSC study includes 11-, 13- and 15-year-old school children in representative samples of schools in the participating countries. The students answer the standardised questionnaire during a school lesson after instruction from the teacher or survey administrator. HBSC has been collecting data on adolescents every fourth year since 1982, the most recent cross-national survey was conducted in 2009/10.

In the HBSC-study socio-economic position (SEP) is measured by the FAS. Assessing adolescent's absolute socio-economic status based on material markers provides an alternative to the more traditional social class [2,4] and is conceptually related to common consumption indices of material deprivation [13] and home affluence [4,14]. FAS items ask students about things they are likely to know about in their family (car, bedrooms, vacations, and computers), thus limiting the number of non-responses in the study. When the scale was introduced in 1998 it was used in a national context only and contained three items (family car, bedroom and telephone) [2]. In 2001/02 it was used cross-nationally and comprised family car, bedroom, holiday and computer. The items, their response categories, and their rationale are the following:

- *Does your family own a car, van or truck*? (No = 0, Yes, one = 1, Yes, two or more = 2). This item is a component of the Scottish deprivation index developed by Carstairs and Morris [13], which is used widely in health inequalities research.
- *Do you have your own bedroom for yourself*? (No = 0, Yes = 1). This item is a simple proxy for overcrowding, classified by Townsend [14] as housing deprivation, and is also a component of the Scottish deprivation index.
- *During the past 12 months, how many times did you travel away on holiday with your family*? (Not at all = 0, Once = 1, Twice = 2, More than twice = 3). This item is a measure of 'deprivation of home facilities' [14].
- *How many computers does your family own*? (None = 0, One = 1, Two = 2, More than two = 3). This item has been introduced to differentiate SEP in affluent countries.

In previous research the FAS has been included differently depending on the research question, the data being analysed and the statistical technique in use. The FAS has both been included at an individual, and as an aggregated school and country level variable [12], and is widely used in educational research [10].

1.2. Validity of FAS

Based on the thorough development of the survey items and the theoretical rationale of the scale, it is assumed that the scale is content valid [15]. From its early development, the validity of FAS has been discussed in several papers, and validated at both national and international levels. Studies found that the FAS has good criterion validity, on the basis of showing graded associations between SEP and various health outcomes [12,16,17]. Adolescents and parents report similarly to the FAS items [18] and the FAS is less affected by non-response bias than SEP measures that rely on child reports of household income or parental occupation [2,7,11]. Additionally, data from the FAS are far less burdensome to collect and manage than other sources on SEP, e.g. data on parental occupation. Within educational research, the FAS is considered valuable as a tool when assessing children's eligibility for free school meals, but so far it has been concluded that the FAS does not have good overall reliability [10]. Boyce et al. has discussed limitations in the use of FAS in relation to cross-national studies. The authors compared the FAS measurements internationally and recommended the inclusion of the FAS in aggregate analyses in studies on health inequalities [16].

Scale measurement properties are crucial when relying on a scale that is derived from items presumed to be equally sensitive to an attribute (in this case family affluence) [19]. Collecting data according to a standardised protocol as done in HBSC allow researchers to compare data from different countries. A critical evaluation of the comparability of questionnaires across cultures and languages is only partly achieved by back translation, review by focus groups and co-ordinated multinational questionnaire development [9]. However, statistical methods for translation evaluation can also be used, one of which is the analysis of DIF also known as item bias. DIF methods require that items should function in the same way, whatever subgroups are investigated. Individuals at the same level of affluence ideally respond the same on the FAS items. DIF occurs when groups at the same level of affluence respond differently to an item, e.g. number of computers in boys and girls. The method chosen to investigate for DIF in the present paper can also be used to correct for DIF, if analyses are successful in fitting a model satisfactorily [9].

Batista-Foguet and colleagues found country specific variations in the relative contribution of the FAS-items, and their work provided a new technique for weighting scale items in a revised FAS. They concluded that cross-cultural studies were possible when constructing a country specific index [8]. This method, however, has not been widely used so far. Schnohr et al. also found country specific variations in the measurement properties of the FAS, and showed that modelling its DIF was possible [6]. By use of the statistical procedures described for the present study, equated scores were generated, and analyses on health inequalities were conducted for six countries [20]. However, neither of these two methods for handling DIF, has been adopted by other HBSC researchers, presumably due to its perceived complexity and to a lack of consistency with the more widely used approach to working with the FAS. Even though the recommendations from these two studies has not been applied, they were critical to the position presented by Boyce et al. [16] from a theoretical point of view. It is therefore still relevant to question the comparability of the FAS between HBSC countries because of DIF [8].

The overall aim of the work reported here was to further investigate the construct validity of the FAS in order to support its use cross-national time trends analyses with HBSC data. Specifically, the authors set out to assess how the properties of the FAS changed over time within each country. This was done by examining the drift in item responses as well as examining DIF and LD in the FAS in five HBSC countries and across three survey cycles spanning 12 years. Since the results of the model used can be used to correct for DIF and LD, the study examine the possibilities of constructing adjusted DIF and LD FAS-scores for future use in health inequality research.

1.3. Validation of FAS

The construct validity and objectivity of the FAS was investigated by Rasch́s model for item analysis and its applicability to Graphical Log Linear Rasch Models (GLLRM), analysing five countries in three consecutive surveys to determine the patterns and magnitude of DIF and LD.

A number of technical demands, need to be met when validating the measurement properties of a scale. The Rasch model is an item response model [21], which can be regarded as a formalisation of perfect measurement, by meeting the requirements of criterion related *construct validity* [22], statistical *sufficiency* [23] and *objectivity* [24]. Rosenbaum's definition of construct validity is but one of many attempts to define construct validity. We refer to Cronbach and Meehl [32], Borsboom [25], Kane [26] and Zumbo [27] for discussions of validity.

In this analysis the FAS is considered a latent construct. Analysis addressing the dependence of the *latent* variable, or the way the latent variable influences or is associated to other variables, is referred to as *latent regression analyses* or *latent structure analyses*. GLLRM are useful for such purposes, and preferable to analyses where the score – DIF equated or not – is used as a proxy for the latent variable. Latent regression in GLLRM has been described by Christensen and colleagues [28]. GLLRM are also useful because conditional inference and Mantel–Haenszel techniques apply for these models in the same way as for ordinary Rasch models. Two of the requirements of construct validity according to Rosenbaum are that items are mutually locally independent (no local dependence, LD) and conditionally independent of exogenous variables given the latent variable being measured (no differential item functioning, DIF). The GLLRM relaxes these two requirements, replacing them with the requirements that LD and DIF are uniform in the strength of association between locally dependent items, and the effect of exogenous variables on items does not depend on the trait being measured. It is therefore relevant to test the degree of DIF and LD and examine the patterns in these before concluding on the validity of the FAS. For a more detailed description of the method, DIF equating, reliability and validity claims, see Schnohr et al. [6]. The analyses were supplemented by a description of the changes over time and between countries in the range of item responses expressed by means and standard deviations.

2. Method

The HBSC study has been described in the introduction. The basis for the present study was data from surveys in 2001/02, 2005/06 and 2009/10 in Austria, Belgium (Flemish region), Canada, Norway and Scotland including a total of 24,426 adolescents from the survey in 2001/02, 25,715 from the survey in 2005/06 and 35,969 from the survey in 2009/10. The sample characteristics on the basis of age-group and gender are presented in Table 1.

2.1. The fitting of Graphical Log-Linear Rasch Models (GLLRM's)

The statistical procedures chosen to analyse for DIF and LD in the present paper were the GLLRM. The model fit of the FAS to GLLRM was initially performed for each of the five countries. Estimates of GLLRM parameters and *partial*

Table 1
Data characteristics.

Year of survey	Gender		Age			Total
	Boys	Girls	11-year olds	13-year olds	15-year olds	
Austria						
2001/02	2164	2020	1527	1562	1277	4366
2005/06	2340	2435	1694	1587	1494	4775
2009/10	2456	2547	1457	1726	1820	5003
Belgium						
2001/02	2996	3293	2153	2106	2030	6289
2005/06	2198	2113	1291	1404	1616	4311
2009/10	2086	2094	1501	1453	1226	4180
Canada						
2001/02	1996	2365	1641	1513	1207	4361
2005/06	2732	3055	1466	2032	2289	5787
2009/10	7711	7999	4490	5779	5441	15,710
Norway						
2001/02	2550	2465	1657	1736	1622	5015
2005/06	2428	2269	1578	1585	1534	4697
2009/10	2171	2167	1679	1320	1339	4338
Scotland						
2001/02	2240	2155	1740	1506	1149	4395
2005/06	3032	3113	1691	2256	2198	6145
2009/10	3319	3419	2055	2116	2567	6738
Total	42,419	43,691	27,620	29,681	28,809	86,110

Table 2
Results of Graphical Log-Linear Rasch Models disclosing DIF and LD stratified by country and survey years.*

"–" = n.s.	Differential item functioning								Local dependence					
	Gender				Age				Room			Holiday		Computer
	Room	Car	Holiday	Computer	Room	Car	Holiday	Computer	Car	Holiday	Computer	Car	Computer	Car
2001/02														
Austria	–	–	–	−0.22	0.19	–	−0.20	–	0.19	−0.05	−0.06	−0.09	−0.01	0.07
Belgium	–	–	–	–	0.24	–	−0.12	–	–	0.06	−0.11	−0.09	−0.08	0.19
Canada	0.18	–	–	−0.10	0.27	–	−0.17	–	0.12	−0.06	−0.09	−0.01	−0.13	0.21
Norway	–	–	–	−0.22	0.44	–	−0.19	–	0.11	−0.09	0.02	−0.09	0.01	0.06
Scotland	–	–	–	−0.22	0.21	0.16	−0.12	−0.10	0.13	−0.08	−0.10	0.04	−0.05	0.05
2005/06														
Austria	–	–	–	−0.09	–	0.09	−0.21	0.08	0.29	−0.07	−0.04	−0.09	−0.09	0.16
Belgium	–	–	–	–	0.20	–	−0.23	–	0.05	−0.04	−0.03	0.04	−0.12	0.20
Canada	–	–	–	–	–	–	−0.25	–	0.17	−0.03	−0.09	0.04	−0.10	0.09
Norway	0.24	–	0.26	–	0.42	–	–	0.10	0.21	–	−0.06	−0.09	0.05	0.10
Scotland	–	–	–	−0.22	0.18	0.14	−0.22	–	0.07	0.06	−0.12	0.05	−0.17	0.17
2009/10														
Austria	–	–	–	–	0.17	–	–	0.18	0.21	−0.11	0.01	−0.14	−0.06	0.19
Belgium	–	–	–	–	0.24	–	−0.26	0.22	0.16	0.05	−0.13	–	−0.17	0.19
Canada	–	–	–	–	0.11	−0.07	−0.14	0.14	0.18	−0.02	−0.08	0.05	0.14	0.11
Norway	–	–	–	−0.13	0.20	–	−0.28	0.32	0.05	0.10	−0.02	−0.16	−0.11	0.25
Scotland	–	–	–	–	0.20	–	−0.26	0.17	0.08	0.06	−0.14	0.07	−0.20	0.16

* Partial gamma coefficients are denoted γ^p measuring the strength of the conditional association between pairs of items, and between items and exogenous variables, Interpretations in the present paper are γ^p of 0.00–0.15 = weak, 0.16–0.30 = moderate and $\gamma^p > 0.30$ = strong correlations.

gamma coefficients (denoted γ^p) measuring the strength of the conditional association between pairs of items, and between items and exogenous variables were used to analyse the patterns of DIF and LD. In this paper, γ^p correlations of 0.00–0.15 were regarded as weak correlations, 0.16–0.30 were regarded as moderate correlations and $\gamma^p > 0.30$ were regarded as strong correlations. Table 2 provides with an overview of the γ^p calculated on the basis of analyses for each of the three survey cycles in all five countries, a total of 15 analyses (Table 2).

Unsuccessful attempts were made to fit a model including all the countries. Results revealed that neither for all surveys (CLR = 495.09, df = 51, p-value = 0.000) nor when stratifying each survey year, was it possible to fit a model across all countries (data not shown). Hence, the following statistical procedures focused on the fitting of models for

Table 3
Fit statistics (Conditional Likelihood Ratio) of country-year-specific models.

	CLR	df (Parameters)	*p*-Value
Austria			
2001/02	24.24	16	0.0843
2005/06	14.77	26	0.9614
2009/10	23.63	16	0.0978
Belgium			
2001/02	30.47	18	0.0331
2005/06	14.02	16	0.5971
2009/10	17.30	26	0.8998
Canada			
2001/02	17.76	19	0.5338
2005/06	16.07	20	0.7121
2009/10	21.16	20	0.3878
Norway			
2001/02	14.51	18	0.6954
2005/06	4.86	17	0.9981
2009/10*	–	–	–
Scotland			
2001/02	33.14	28	0.2306
2005/06**	67.62	28	0.0000
2009/10	19.94	26	0.7945

* Norway 2009/10 fitted a model with 43 parameters, but it was considered too complex to be included as meaningful.
** Scotland 2005/06 had very poor fit, but was included in the table to show fit statistics.

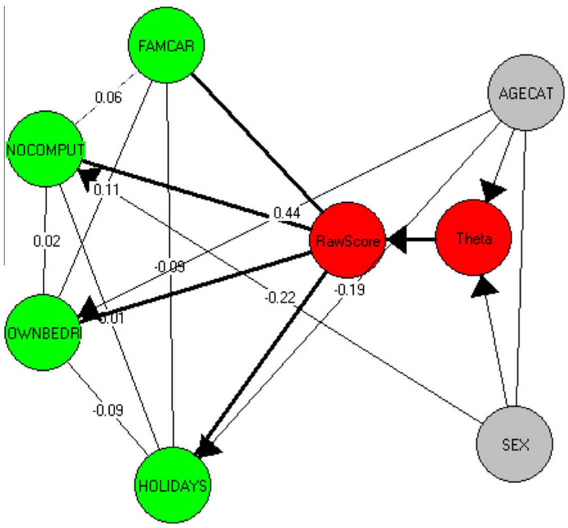

Fig. 1. Norway 2001/2002.

each country and year, assessing the item response drift for the FAS items, and the patterns in DIF and LD.

3. Results

Table 3 shows the fit statistics of all the country specific models created. In most cases, successful attempts were made in fitting a satisfactory model, but as shown, in two cases the magnitude of identified DIF and LD were too large to consider the model meaningful. The degrees of freedom reveal the number of parameters, and show the complexity of the models created. When almost all items were correlated in a model it was not considered sensible to accept it with an assumption of a latent trait. This was the case for Norway in 2009/10 and Scotland 2005/06. The γ^p correlations were included in Table 2 even though the model showed very poor fit.

Table 3 summarises the findings across countries and survey years.

To illustrate the constructed models and the findings of correlation coefficients within item (LD) and between items and exogenous variables (DIF), the models of Norway in 2001/02 (Fig. 1) and Scotland in 2009/10 (Fig. 2) are shown. Due to space constraints for this paper, only two models out of the 15 sets of analyses performed are presented.

3.1. Survey year 2001/02

Fig. 1 shows the GLLRM-model for Norway in the first survey. DIF was identified between gender and computer (γ^p between −0.10 and −0.22) in four of the five countries, indicating the families with boys were more likely to have computers in the home. No other DIF-associations were found with regards to gender, except a γ^p of 0.18 indicating that girls in Canada were more likely to have their own bedroom than boys. Findings on DIF between age and the FAS-items, results were consistent and strong in showing DIF between age and bedroom (γ^p between 0.19 to 0.44), indicating that age positively related to having your own bedroom. Likewise, results were consistent and strong in showing DIF between age and going on holiday (γ^p between −0.12 and −0.20), indicating that age negatively related to the likelihood of going on holiday with your parents. In the fitted models, LD was identified on all between-item associations, but with few exceptions, these were weak.

3.2. Survey year 2005/06

As seen in 2001/02, DIF between age and bedroom were strong in three of the five countries (γ^p between 0.18 and 0.42). The same but a bit stronger associations were found between age and going on holiday (γ^p between −0.21 and −0.25), indicating that age was positively related to having your own bedroom, and negatively related to going on holiday with your parents. The association between gender and computer had diminished to be weak (γ^p of −0.09 and −0.21), and only identified in two countries. In the fitted models, LD was identified on all between-item associations, but weak as seen in the first survey.

3.3. Survey year 2009/10

Fig. 2 shows the GLLRM-model for Scotland in the third survey. In comparison to the two first surveys, no DIF was found with regards to gender. Findings on DIF with regards to age were consistent, and showing DIF between age and bedroom (γ^p between 0.11 and 0.24), and between age and going on holiday (γ^p between −0.14 and −0.28). DIF between age and computer had emerged, being moderate to strong (γ^p between 0.14 and 0.32). In the fitted models,

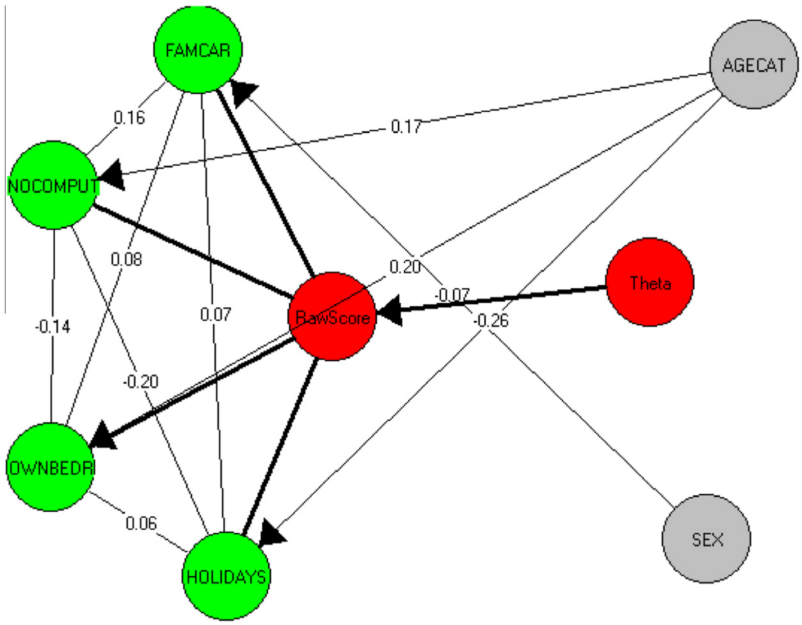

Fig. 2. Scotland 2009/2010.

Table 4
Range of item responses: means and standard deviations across survey years and country.

Item	Year of survey	Austria	Belgium	Canada	Norway	Scotland
Car (range 0–2)	2001/02	1.44 (.58)	1.44 (.59)	1.69 (.53)	1.50 (.57)	1.38 (.66)
	2005/06	1.50 (.57)	1.48 (.59)	1.70 (.53)	1.56 (.55)	1.43 (.65)
	2009/10	1.51 (.58)	1.54 (.58)	1.66 (.56)	1.64 (.53)	1.49 (.64)
Holiday (range 0–3)	2001/02	1.53 (1.03)	1.42 (1.01)	1.68 (1.09)	2.16 (.95)	1.43 (.98)
	2005/06	1.45 (1.04)	1.52 (1.03)	1.55 (1.09)	2.12 (.97)	1.47 (1.00)
	2009/10	1.57 (1.05)	1.65 (1.02)	1.57 (1.08)	1.99 (1.03)	1.55 (1.03)
Bedroom (range 0–1)	2001/02	0.82 (.38)	0.84 (.37)	0.88 (.32)	0.93 (.27)	0.78 (.41)
	2005/06	0.85 (.36)	0.86 (.35)	0.89 (.31)	0.94 (.23)	0.80 (.40)
	2009/10	0.85 (.36)	0.85 (.36)	0.86 (.35)	0.94 (.24)	0.83 (.38)
Computers (range 0–3)	2001/02	1.42 (.85)	1.45 (.87)	1.48 (.81)	1.66 (.84)	1.51 (.90)
	2005/06	1.74 (.85)	1.79 (.88)	1.74 (.83)	2.12 (.83)	1.79 (.85)
	2009/10	2.10 (.84)	2.17 (.82)	2.09 (.85)	2.64 (.63)	2.30 (.78)

weak or moderate associations on LD were identified on all between-item associations with only one exception.

3.4. Fitting of models across survey years

In the statistical process, a model was fitted for each country, where survey-year was included as a potential (third) exogenous variable. This maneuver revealed models with massive DIF and LD between the surveys, where a model was fitted with acceptable test-results for Austria (CLR[1] = 56.44, $p = 0.0274$) and good fit for Canada (CLR = 50.09, $p = 0.1317$) and Norway (CLR = 30.78, $p = 0.9185$). No overall model with acceptable fit was accomplished for Belgium and Scotland (data not shown).

[1] Conditional Likelihood Ratio-test, *p*-value > 0.05 indicating good model fit.

3.5. Item response drift in the FAS

Another crude but intuitive measure of item response is shown in Table 4, where the range in item responses are calculated for each country and year of survey.

Most inconsistent finding which is interpreted as the highest item response drift is seen for the computer item, which increases from 1.42 in Austria and 1.66 in Norway to 2.10 and 2.64 respectively (Table 4).

4. Discussion

Earlier studies have discussed how scales such as the FAS perform differently in different countries, and these findings have now been examined in HBSC. This study found, that the computer item has changed from being different between genders, to varying across age. The assumption is that owning a computer in earlier 2000s related to a game-culture led on by boys. It seems that during the past 8 years, this item measures differently, now being related to the educational span adolescents go through during their schooling; there is a need for older students to have computer for educational purposes instead of a computer being used for leisure primarily for boys. Computers have also become a tool for social interaction between 2002 and 2010, which does not have a traceable gender difference. The consistency of this finding across all five countries results in an emphasis to these results as reliable.

Even though housing space can be seen as a sociological issue concerning family size, there is a clear tendency that children have their own room as they grow older, irrespective of the family's level of affluence.

The failure to fit GLLRM's across survey years warrants a cautious approach when comparing data from the FAS between surveys. The indication of changing LD over time emphasises this conclusion. LD is also dependent on development in items over time, which is part of the explanation to the poor model fit between surveys. However, it does not necessarily mean that within country measurement is invalid. The comprehensive finding of LD indicate that further analyses are warranted on the loading of the items, and which are most optimal to include in the scale. Further development of the FAS in this respect would require one or two items, which function in the same way between surveys, and this (these) items will benefit from functioning as "anchor items" when adjusting for DIF and LD, as previously described [6]. A development of additional FAS items, which functions in the same way between countries and survey years, is consequently of high importance if part of the validation process around the FAS is to support the claims that the FAS can be used for valid comparisons between countries or surveys.

It is important to note that the problem of DIF would not be solved by adjusting for age and gender as a confounder, since this modification does not adjust on an item level, but on the sum score. Older children more often had their own bedroom, independent of family affluence, and they were also less likely to go on holiday with their parents.

The nature of relationships between constructs and measures has been discussed in the literature on construct validity [29]. There are basically two theoretical approaches with the question whether constructs are specified as causes or effects of their measures. If constructs are viewed as causes of measures, a variation in a construct will lead to a variation in its measures [30]. This type of measure is labelled reflexive, because they represent reflections of a construct. The theoretical approach of this study is that the FAS indicators are effect indicators, an approach providing the opportunity for testing construct validity by means of Rasch models, as done in the present paper. It is also supported by the changing of items across survey years, which is not advisable for causal indicators, that are not interchangeable and usually essential [5]. Additionally, the rationale is also based on the fitting of a model. Although the fit of a latent variable model to the data may not prove the existence of causally operating latent variables, the model does formulate this as a hypothesis, and the fitting of the models can be adduced as evidence supporting this hypothesis [31]. Discussions on the directions and relationship between constructs and measures are important, and must be addressed before models can be empirically tested [29]. The view on the FAS presented in this study may be challenged by researchers presupposing that the FAS is a formative scale with causal indicators, and future validation work taking on this view is welcomed to extend the methodological discussion even further.

The concluding remarks of the present paper based on the analyses is in support of previous studies [6,8,10], which urge caution when using the FAS to compare across different countries and different times. Future development in FAS items could benefit from using this knowledge, and further methodological developments should be undertaken before comparisons can be done with confidence to that conclusions drawn are reliable.

Acknowledgements

The HBSC study is an international study carried out in collaboration with WHO Europe. The international coordinator is Professor Candace Currie from the University of Edinburgh, and the international databank manager is Dr. Oddrun Samdal from Bergen University. The HBSC survey is conducted by personal responsibility of principal investigators (PI's) in 41 countries.

References

[1] J.P. Mackenbach, M.J. Bakker, Reducing Inequalities in Health: A European Perspective, Routledge, London, 2002.
[2] C. Currie, R.A. Elton, J. Todd, S. Platt, Indicators of socioeconomic status for adolescents: the WHO Health Behaviour in School-aged Children survey, Health Educ. Res. 12 (3) (1997) 385–397.
[3] N. Lien, C. Friestad, K.I. Klepp, Adolescents' proxy reports of parents' socioeconomic status: How valid are they?, J Epidemiol. Community Health 55 (10) (2001) 731–737.
[4] J. Wardle, K. Robb, F. Johnson, Assessing socioeconomic status in adolescents: the validity of a home affluence scale, J. Epidemiol. Community Health 56 (8) (2002) 595–599.
[5] K.A. Bollen, S. Bauldry, Three Cs in measurement models: causal indicators, composite indicators, and covariates, Psychol. Methods 16 (3) (2011) 265–284.
[6] C.W. Schnohr, S. Kreiner, E. Due, C. Currie, W. Boyce, F. Diderichsen, Differential item functioning of a family affluence scale: validation study on data from HBSC 2001/02, Soc. Indic. Res. 89 (1) (2008) 79–95.
[7] T. Torsheim, C. Currie, W. Boyce, I. Kalnins, M. Overpeck, S. Haugland, Material deprivation and self-rated health: a multilevel study of adolescents from 22 European and North American countries, Soc. Sci. Med. 59 (1) (2004) 1–12.
[8] J.M. Batista-Foguet, J. Fortiana, C. Currie, J.R. Villalbi, Socio-economic indexes in surveys for comparisons between countries, Soc. Indic. Res. 67 (3) (2004) 315–332.
[9] J.B. Bjorner, S. Kreiner, J.E. Ware, M.T. Damsgaard, P. Bech, Differential item functioning in the Danish translation of the SF-36, J. Clin. Epidemiol. 51 (11) (1998) 1189–1202.
[10] S. Kehoe, L. O'Hare, The reliability and validity of the family affluence scale, Eff. Educ. 2 (2) (2010) 155–164.
[11] C. Currie, M. Molcho, W. Boyce, B. Holstein, T. Torsheim, M. Richter, Researching health inequalities in adolescents: the development of the Health Behaviour in School-Aged Children (HBSC) family affluence scale, Soc. Sci. Med. 66 (6) (2008) 1429–1436.
[12] T. Torsheim, U. Ravens-Sieberer, J. Hetland, R. Valimaa, M. Danielson, M. Overpeck, Cross-national variation of gender differences in adolescent subjective health in Europe and North America, Soc. Sci. Med. 62 (4) (2006) 815–827.

[13] V. Carstairs, R. Morris, Deprivation: explaining differences in mortality between Scotland and England and Wales, Brit. Med. J. 299 (1989) 886–889.
[14] P. Townsend, Deprivation, J. Soc. Policy 16 (02) (1987) 125–146.
[15] C. Currie, S.N. Gabhainn, E. Godeau, W. Picket, M. Richter, C. Roberts, A. Morgan, V. Barnekow, R. Smith, Inequalities in young people's health. Health Behaviour in School-aged Children international report from the 2005/2006 survey, WHO Regional Office for Europe, 5, CAHRU, 2008.
[16] W. Boyce, T. Torsheim, C. Currie, A. Zambon, The family affluence scale as a measure of national wealth: validation of an adolescent self-report measure, Soc. Indic. Res. 78 (3) (2006) 473–487.
[17] A. Zambon, W. Boyce, E. Cois, C. Currie, P. Lemma, P. Dalmasso, A. Borraccino, F. Cavallo, Do welfare regimes mediate the effect of socioeconomic position on health in adolescence? A Cross-national comparison in Europe, North America, and Israel, Int. J. Health Serv. 36 (2) (2006) 309–329.
[18] A. Andersen, R. Krolner, C. Currie, L. Dallago, P. Due, M. Richter, A. Orkenyi, B.E. Holstein, High agreement on family affluence between children's and parents' reports: international study of 11-year-old children, J. Epidemiol. Community Health 62 (12) (2008) 1092–1094.
[19] J.B. Bjorner, M.T. Damsgaard, T. Watt, M. Groenvold, Tests of data quality, scaling assumptions, and reliability of the Danish SF-36, J. Clin. Epidemiol. 51 (11) (1998) 1001–1011.
[20] C.W. Schnohr, S. Kreiner, M. Rasmussen, P. Due, F. Diderichsen, School-related mediators in social inequalities in smoking: a comparative cross-sectional study of 20,399 adolescents, Int. J. Equity. Health 8 (2009) 17.
[21] G. Rasch, Probablistic Models for Some Intelligence and Attainment Tests, Danish Institute for Educational Research, Copenhagen, 1960.
[22] P.R. Rosenbaum, Criterion-related construct validity, Psychometrika 54 (1989) 625–633.
[23] E.B. Andersen, A goodness of fit test for the Rasch model, Psychometrika 38 (1973) 123–140.
[24] G. Rasch, An item analysis which takes individual differences into account, Br. J. Math. Stat. Psychol. 19 (1) (1966) 49–57.
[25] D. Borsboom, Measuring the Mind, Cambridge University Press, United Kingdom, 2005.
[26] M.T. Kane, Validation, in: R.L. Brennan (Ed.), Educational Measurement, Praeger Publishers, Westport, 2006, pp. 17–65.
[27] B.D. Zumbo, Validity: foundational issues and statistical methodology, in: C.R. Rao, S. Sinharay (Eds.), Handbook of Statistics, Psychometrics, vol. 26, Elsevier, Amsterdam, 2007.
[28] K.B. Christensen, J.B. Bjorner, S. Kreiner, J.H. Petersen, Latent Regression in Log Linear Rasch Models, vol. 33, Department of Biostatistics, University of Copenhagen, 2004.
[29] J.R. Edwards, R.P. Bagozzi, On the nature and direction of relationships between constructs and measures, Psychol. Methods 5 (2) (2000) 155–174.
[30] K.A. Bollen, Structural Equations with Latent Variables, Wiley, New York, 1989.
[31] D. Borsboom, G.J. Mellenbergh, H.J. van, The theoretical status of latent variables, Psychol. Rev. 110 (2) (2003) 203–219.
[32] L.J. Cronbach, P.E. Meehl, Construct validity in psychological tests, Psychological Bulletin, 52 (1955/2006) 281–302. Reprinted, in: D.J. Bartholomew (Ed.), Measurement, vol. 1, Sage Publications, London, 2006, pp. 277–306.