A data mining and item response mixture modelling method to retrospectively measure diagnostic and statistical manual-5 attention deficit hyperactivity disorder in the 1970 British Cohort Study

Joanne Cotton and Sara T. Baker

Faculty of Education, University of Cambridge

Corresponding author: Joanne Cotton, jmc236@cam.ac.uk

Acknowledgements

Mapping of DSM-5 ADHD symptoms to BCS70 items was reviewed via survey by 14 ADHD experts and two data analysis experts. Many thanks for taking the time to offer your views: Edmund Sonuga-Barke, Dave Coghill, Rosemary Tannock, Rafaela Marco, Cesar Soutullo, Eric Taylor, Hans-Christoph Steinhausen, Maite Ferrin, Jan R Wiersema, Lisa B Thorell, Mark Stein, Golam Khandaker, Anna Vignoles, Ricardo Sabates, and two anonymous reviewers. We are grateful to Edmund Sonuga-Barke and Tobias Banaschewski from the European Network for Hyperkinetic Disorders (EUNETHYDIS), and Laura Laughlin from the American Professional Society of ADHD and Related Disorders (APSARD), for facilitating distribution of the survey.

Grateful thanks also for assistance from Jung Yeon Park with adapting MPlus code from Wall, Park, & Moustaki, (2015), and Angela Brassett-Harknett with replicating Brassett-Grundy & Butler, (2008).

This study comprises part of Joanne Cotton's PhD, which is funded by the Economic and Social Research Council, award number 1652810.

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Access to data was provided by the UK Data Service.

Abstract

Objective: To facilitate future outcome studies, we aimed to develop a robust and replicable method for estimating a categorical and dimensional measure of DSM-5 ADHD in the 1970 British Cohort Study (BCS70).

Method: Following a data mining framework, we mapped DSM-5 ADHD symptoms to age 10 BCS70 data (*N*=11,426) and derived a 16-item scale (α =0.85). Mapping was validated by an expert panel. A categorical subgroup was derived (*n*=594, 5.2%), and a zero-inflated IRT mixture model fitted to estimate a dimensional measure.

Results: Subgroup composition was comparable to other ADHD samples. Relative Risk Ratios (ADHD/not-ADHD) included: boys = 1.38, unemployed fathers = 2.07, below average reading = 2.58, depressed parent = 3.73. Our estimated measures correlated with two derived reference scales: SDQ hyperactivity (*r*=0.74), and a Rutter/Conners-based scale (*r*=0.81), supporting construct validity. IRT model items (symptoms) had moderate to high discrimination (0.90 – 2.81) and provided maximum information at average to moderate theta levels of ADHD (0.5 – 1.75).

Conclusion: We extended previous work to identify ADHD in BCS70, derived scales from existing data, modeled ADHD items with IRT, and adjusted for a zero-inflated distribution. Psychometric properties were promising and this work will enable future studies of causal mechanisms in ADHD.

Keywords: data-mining, IRT, ADHD, BCS70

1. Introduction

Attention Deficit Hyperactivity Disorder (ADHD) is a disorder of inattention, impulsivity, and hyperactivity that interferes with functioning. It has three presentations: primarily inattentive, primarily hyperactive and impulsive, and combined (American Psychiatric Association, 2013), and affects approximately 6% of children worldwide (Polanczyk, de Lima, Horta, Biederman, & Rohde, 2007; Willcutt, 2012). Lifelong impairment often follows childhood ADHD, but about 50% are not significantly impaired as adults (Caye, Rocha, et al., 2016; Costello & Maughan, 2015). We can gain a better understanding of positive outcomes by studying causal mechanisms in the long term. However, methodological challenges have made it difficult to exploit existing longitudinal datasets to this end. Challenges include insufficient cohort age, sample biases, imprecise measures, and lack of psychosocial data. Here we propose a robust and replicable method to mitigate these challenges and facilitate future causal outcome analyses.

1.1 Methodological challenges

First, longitudinal data sources used in ADHD analyses are limited by cohort age. Most sources report adult ADHD outcomes between ages 18 and 25 (Cadman et al., 2016; Caye, Spadini, et al., 2016; Kuriyan et al., 2013; Lara et al., 2009; Swanson et al., 2017; van Lieshout et al., 2016). However, the brain continues to develop until about age 30 (Sowell et al., 2003), and imaging studies indicate that cortical development in ADHD is slower than average (Shaw et al., 2013). Additionally, there is a trend in Western societies to delay the traditional markers of 'settled' adulthood, such as stability of residence, marriage/partnership, and financial independence from parents (Arnett, 2000, p. 469). Thus, it is our view that long-term outcomes for ADHD should be evaluated after age 30.

Longitudinal data is needed from a cohort born in the mid-1980's or before to support post-age-30 outcomes analysis, but the current ADHD criteria have only been stable since 1987, or the DSM-III-R (American Psychiatric Association, 1987; Barkley, 2015). Yet, ADHD is a latent construct, i.e. not directly observable (Bollen, 2002), and latent constructs lend themselves to data mining, or "…the extraction of implicit, previously unknown, and potentially useful information from data." (Witten, Frank, Hall, & Pal, 2017, p. xxiii). Data mining could be used to retrospectively identify ADHD from data in a long-running, existing study, and mitigate the insufficient cohort age limitation.

Second, samples used for ADHD outcomes studies tend to be small, clinical, or based on retrospective recall (Caye, Spadini, et al., 2016; Cheung et al., 2015; Swanson et al., 2017). Small samples do not provide enough statistical power for complex modelling techniques needed to analyze long term trajectories (Wolf, Harrington, Clark, & Miller, 2013). Clinical samples tend to over-represent boys, cases with severe symptoms, and the combined type presentation of ADHD (Willcutt, 2012). Finally, non-clinical sample studies are often based on retrospective recall of childhood symptoms (Caye, Spadini, et al., 2016; Lara et al., 2009), which is affected by recall ability (Coughlin, 1990) and personality factors (Reuben et al., 2016). Accordingly, Caye et al. (2016) recommended that prospective cohort studies should be implemented. In the meantime, data-mining an existing long-running study could address all three of these biases.

Third, in studies of outcomes, ADHD is typically reported using an imprecise categorical indicator, i.e. 'ADHD' or 'not ADHD'. More sensitive dimensional measures are needed to detect individual differences (American Psychiatric Association, 2013; Gorter, Fox, & Twisk, 2015). A range of ADHD studies support this: in genetics, (Groen-Blokhuis et al., 2014; Thapar, Cooper, Eyre, & Langley, 2013), neural connectivity (Elton, Alcauter, & Gao, 2014), and performance on executive function tasks (Agnew-Blais et al., 2016; Salum et al., 2014). Derivation of a sensitive dimensional measure requires a large, minimally biased dataset.

Finally, identification of ADHD retrospectively in a rich dataset opens the possibility for longitudinal analyses on a variety of outcomes based on psychosocial factors, which are thus far understudied in the ADHD literature (Costello & Maughan, 2015).

In sum, insufficient cohort age, sample biases, imprecise measures, and lack of psychosocial data impede analysis of optimal ADHD outcomes. All could be mitigated by utilizing data from a large, long-term, population-based longitudinal cohort study, rich in psychosocial data. To this end, we short-listed candidate datasets, primarily based on data age, then reviewed in detail the following: Avon Longitudinal Study of Parents and Children (ALSPAC; 1991), 1970 British Cohort Study (BCS70) and Northern Finland Birth Cohort (NFBC) 1986. BCS70 was selected for preferable size, age, representativeness, and richness.

BCS70 is an ongoing population-based study of 17,198 children born from 5-11 April 1970. The study offers a rich array of health, psychological, social, and economic data from nine sweeps between ages 0 and 42 (Centre for Longitudinal Studies, 2015; Elliott & Shepherd, 2006). The third sweep at age 10 includes extensive data on behavior (Butler, Despotidou, & Shepherd, 1997). Age 10 is ideal for assessing ADHD, because it is between 7, the most common age of diagnosis (Centers for Disease Control and Prevention, 2013), and 12, the cutoff for diagnosis of childhood ADHD (American Psychiatric Association, 2013). Also, most of the ADHD-relevant questionnaire items in the BCS70 age 10 sweep were derived from the Rutter (Rutter, 1967) and Conners scales (Conners, 1969) (Butler et al., 1997), which are predecessors to current well-validated ADHD measures (American Academy of Pediatrics, McNeil, & Wolraich, 2002; Conners, 2008). Items were completed by both parents and teachers, providing valuable multiple-setting context (American Psychiatric Association, 2013; Butler et al., 1997). Finally, the age 10 sweep had 14,875 respondents and 11,426 with data on behavior, providing a plenteous sample to support complex statistical models and estimate a robust dimensional ADHD measure.

1.2 Literature review

We found only a handful of studies that derived a scale to measure ADHD, or a similar latent construct, in existing data. Brassett-Grundy & Butler (2008) derived a proxy measure for ADHD and evaluated outcomes at age 30 in BCS70. However, they used a combination of 23 Conners (Conners, 1969) and Rutter items (Rutter, 1967) to measure ADHD, including ten (e.g. “has difficulty using scissors”; Brassett-Grundy & Butler, 2008), which are not part of the current ADHD construct. Therefore, the construct they derived is unlikely to have specifically discerned ADHD as it is currently understood. Also, they calculated a simple sum and applied a clinical cutoff to create a categorical indicator, but did not estimate a dimensional measure.

Other researchers have derived measures of latent constructs like social and emotional skills (Goodman, Joshi, Nasim, & Tyler, 2015), self-control (Daly, Delaney, Egan, & Baumeister, 2015), and hyperactivity (Stuart-Smith, Thapar, Maughan, Thapar, & Collishaw, 2017) in BCS70 or similar datasets. They aggregated items and standardized as a general approach. Garcia-Barrera, Kamphaus, & Bandalos (2011) derived a scale to screen for Executive Function (EF) difficulties using items from the Behavior Assessment System for Children (BASC) in an existing dataset. They mapped BASC items to four EF domains, and estimated dimensional measures using factor analysis. Psychometric properties were evaluated using an expert panel to review the mapping, Cronbach’s alpha, and measurement invariance by age and gender (Garcia-Barrera et al., 2011). A similar factor analysis approach has been used elsewhere to retrospectively measure intelligence, personality, and behavior factors (Gale, Hatch, Batty, & Deary, 2009; Prevoo & ter Weel, 2015; von Stumm, Gale, Batty, & Deary, 2009). These more complex methods address some of the key challenges faced with measuring ADHD in BCS70, including mapping items from an existing scale to an unmeasured construct, estimating with greater precision, and evaluating psychometric properties.

A more complex method is desired here, to provide a robust dimensional measure for use in future work. For our data, Item Response Theory (IRT) is a preferable modelling framework. IRT is a special case of confirmatory factor analysis which builds a model at the item level, leading to better generalizability across samples

than other psychometric methods (Baker, 2001; Embretson & Reise, 2000). IRT fits here because the BCS70 age 10 dataset is large (N>500), the data are categorical (Embretson & Reise, 2000; Van Der Eijk & Rose, 2015), and factor structure evaluation indicates ADHD is most reliably measured as a unidimensional latent trait (Wagner et al., 2016). IRT models have been widely recommended for measuring psychiatric and health-related constructs (Edelen & Reeve, 2007; Gorter et al., 2015; Muthen & Asparouhov, 2006; Sturm, Kuhfeld, Kasari, & Mccracken, 2017). Importantly, other authors have used IRT to evaluate psychometric item properties of DSM ADHD criteria (Arias, Esnaola, & Rodríguez-Medina, 2018; Gomez, 2007, 2008, 2011, 2012; Gomez, Vance, & Gomez, 2011; Li, Reise, Chronis-Tuscano, Mikami, & Lee, 2015), compare model fit in sub-samples (Polanczyk et al., 2010), and provide quantitative verification of diagnosis (Lindhiem, Yu, Grasso, Kolko, & Youngstrom, 2015). These IRT studies reported good indicators of model fit in a variety of clinical and non-clinical samples.

Whilst IRT models are robust to some non-normality, they assume an approximately normal distribution (Reise & Revicki, 2015). We should not assume a normal distribution for ADHD (or any psychiatric disorder) in a population-based sample (Kaat & Farmer, 2017; Reise & Waller, 2009; Wall et al., 2015). A large proportion of respondents are expected to have zero symptoms or very few (Finkelman, Green, Gruber, & Zaslavsky, 2011; Reise & Waller, 2009; Wall et al., 2015). Simulation studies have shown that ignoring non-normality of a latent trait in IRT can lead to significant estimation errors (e.g. inflated discrimination parameters), and adjustments are recommended (Kaat & Farmer, 2017; Sass, Schmitt, & Walker, 2008; Wall et al., 2015; Woods, 2015). There are a few ways to adjust for non-normality in IRT, including the Empirical Histogram, Ramsay Curve, (Woods, 2015), and Zero-Inflated Mixture Model (Wall et al., 2015; ZIMM). The latter method specifically adjusts for the zero-inflation we expect to find with ADHD in BCS70.

1.3 Present study

Our objective was to develop and demonstrate a robust method to derive a categorical and dimensional measure of ADHD in the BCS70 age 10 data, enabling future studies of outcomes. We aimed to incorporate a data-mining framework, apply approximate DSM-5 diagnostic criteria, develop an IRT model adjusted for zero-inflation, and evaluate psychometric properties.

2. Method

2.1 Data

Age 10 BCS70 data were collected in 1980 and 1981 in the United Kingdom. Ten questionnaires were completed by medical professionals, parents, teachers, and participants (Centre for Longitudinal Studies, 2015). Data was accessed through the UK Data Service (University of Manchester, University of Essex, & Jisc, 2015).

In the age 10 sweep, cohort members (*N*=14,875) were 96% 'English, etc.', 51.5% boys, and 63.9% of their parents had jobs in the 'middle' social classes, designated in 1980 as 'III-manual', 'III-non-manual', and 'IV-partly-skilled'. All were born in April 1970. Children with parents born outside Britain, single mothers, teenage mothers, mothers over 40, unemployed fathers, and low parental education level were under-represented due to attrition (Butler et al., 1997, p. 35). The ADHD-relevant behavior questionnaire items were left blank by many respondents (n=3,449); these observations were excluded from our sample (*N*=11,426).

2.2 Ethics

An ethics checklist was approved by the Faculty of Education, University of Cambridge, based on British Educational Research Association (BERA) guidelines (BERA, 2011). Ethical procedures for the original study (BCS70) adhere to BERA and ESRC guidelines (Centre for Longitudinal Studies: IoE/UCL, 2014).

2.3 Tools

Analyses were conducted using Stata 14.2 (StataCorp LLC, 2015), MPlus 8 (Muthen & Muthen, 2017), Microsoft Excel, and Qualtrics (Qualtrics, 2017).

2.4 Measures

2.4.1 DSM-5 ADHD criteria

There are 18 symptoms: nine hyperactive/impulsive, and nine inattentive, plus six additional conditions, totaling 24 items. The diagnostic threshold requires at least six symptoms from either or both lists of nine to be observed 'often', along with all six conditions. Depending on which symptom thresholds are met, presentation types of Primarily Hyperactive and Impulsive (PHI), Primarily Inattentive (PI), or Combined (C) are applicable (American Psychiatric Association, 2013). In the present study we have used abbreviations to refer to the DSM-5 ADHD criteria; for example, 'dh1' refers to the 1st symptom in the DSM-5 list of hyperactive/impulsive symptoms.

2.4.2 BCS70 age 10 behavior items

53 items from the maternal self-completion form and educational questionnaire pertained to child behavior (Butler et al., 1997). The items were completed by a parent and teacher, respectively. Most were based on Rutter (Rutter, 1967) and Conners (Conners, 1969) items, though a handful were written, tested and added by the BCS70 study designers (Butler et al., 1997). An example item was 'Is squirmy or fidgety', and the respondent (parent or teacher) indicated the extent to which the statement applied to the child (see Figure 1).

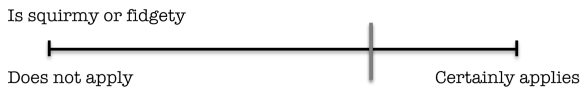

Figure 1. Example of Visual Analog Scale item used in BCS70 age 10 sweep
Respondent indicated the extent of their agreement with the item by marking a vertical line on the horizontal scale

2.4.3 Strengths and Difficulties Questionnaire (SDQ) hyperactivity subscale

The subscale for ages 4-17 consists of five items (abbreviated): restlessness, fidgeting, distractibility, impulsivity, and attention span (Goodman, 1997; youthinmind, 2012). The subscale has been validated for use as a diagnostic screener and in research as a proxy for ADHD diagnosis (Stone, Otten, Engels, Vermulst, & Janssens, 2010; Ullebø, Posserud, Heiervang, Gillberg, & Obel, 2011).

2.5 Approach

Our approach was guided by a data mining framework, and included three phases: 1) data assessment and preparation, 2) modelling, and 3) evaluation (Kurgan & Musilek, 2006, p. 6-7).

2.5.1 Data assessment and preparation

This first phase entailed item mapping, recoding, application of DSM-5 criteria, and model selection.

2.5.1.1 Item mapping and derived scale

Using the 24 DSM-5 ADHD items as a reference point, the 53 BCS70 behavior items were inspected visually for semantically similar content. Next, all the remaining (~2,900) data items from the age 10 sweep were checked for further mapping candidates using keyword searches and visual inspection. We successfully mapped 19 (79%) of the 24 DSM-5 items: five/nine inattentive, nine/nine hyperactive/impulsive, and five/six conditions, to BCS70 items. No mapping could be found for: di1-careless mistakes, di3-doesn't listen, di5-trouble organizing, di7-loses things, or dc6-symptoms > 6 months. Three of the conditions, dc1-symptoms by age 12, dc4-no other psychiatric disorder, and dc5-symptoms not part of another psychiatric disorder, were mapped to the BCS70 data, but had insufficient variation to be useful in a scale, so were excluded from the resultant 16-item scale.

A panel of 16 international experts completed an online survey to review the item mapping. Adjustments were made to reflect their views (Appendix A in the supporting information includes survey instructions, example questions and results, and details of adjustments). The final mapping of DSM-5 to BCS70 items and our derived 16-item scale is reported in Table 1.

No.	DSM-5 criteria	BCS70 questionnaire items †
Inattentive		
Di1	Often fails to give close attention to details or makes careless mistakes in schoolwork, at work, or with other activities.	No mapping found
Di2	Often has trouble holding attention on tasks or play activities.	R-j155 - Pays attention to what is being explained in class m82 - Has difficulty concentrating on any particular task though may return to it frequently j129 - Cannot concentrate on any particular task, even though the child may return to it frequently j077 - How well does this child concentrate on educational tasks, in comparison with the average 10-year-old?
Di3	Often does not seem to listen when spoken to directly.	No mapping found
Di4	Often does not follow through on instructions and fails to finish schoolwork, chores, or duties in the workplace (e.g., loses focus, side-tracked).	m76 - Fails to finish things he/she starts, short attention span R-j174 - Child completes tasks which are started j177 - Fails to finish things he starts
Di5	Often has trouble organizing tasks and activities.	No mapping found
Di6	Often avoids, dislikes, or is reluctant to do tasks that require mental effort over a long period of time (such as schoolwork or homework).	R-j139 - Shows perseverance; persists with difficult or routine work
Di7	Often loses things necessary for tasks and activities (e.g. school materials, pencils, books, tools, wallets, keys, paperwork, eyeglasses, mobile telephones).	No mapping found

Di8	Is often easily distracted	m65 – Inattentive, easily distracted j152 – Is easily distracted
Di9	Is often forgetful in daily activities.	j158 – Is forgetful when given a complex task
Hyperactive		
Dh1	Often fidgets with or taps hands or feet, or squirms in seat.	m44 - Is squirmy or fidgety j151 - Squirmy and fidgety m77 - Given to rhythmic tapping or kicking j165 - Given to rhythmic tapping or rhythmic kicking during class j082 - What percentage of the time is the child fidgeting and indulging other minor distracting activities, when he/she is expected to be working? (paraphrased)
Dh2	Often leaves seat in situations when remaining seated is expected.	j081 - What percentage of the time is the child moving around the classroom, when he/she is expected to be working? (paraphrased)
Dh3	Often runs about or climbs in situations where it is not appropriate (adolescents or adults may be limited to feeling restless).	m43 - Very restless. Often running or jumping up and down. Hardly ever still.
Dh4	Often unable to play or take part in leisure activities quietly.	m57 - Cannot settle to do anything for more than a few moments
Dh5	Is often "on the go" acting as if "driven by a motor".	m72 – Shows restless or overactive behavior j150 - Shows restless or overactive behaviour
Dh6	Often talks excessively.	j080 - What percentage of the time is the child talking to other children, when he/she is expected to be working? *(paraphrased)*
Dh7	Often blurts out an answer before a question has been completed.	m73 – Is impulsive, excitable

Dh8	Often has trouble waiting his/her turn.	m71 - Requests must be met immediately, easily frustrated j175 - Requests must be met immediately - easily frustrated
Dh9	Often interrupts or intrudes on others (e.g., butts into conversations or games)	m74 - Interferes with the activity of other children j142 - Interferes with the activities of other children
Conditions		
Dc1	Several inattentive or hyperactive-impulsive symptoms were present before age 12 years	True for all cases; criteria were evaluated at age 10
Dc2	Several symptoms are present in two or more settings, (such as at home, school or work; with friends or relatives; in other activities)	Both mother and teacher indicated three or more symptoms were present
Dc3	There is clear evidence that the symptoms interfere with, or reduce the quality of, social, school, or work functioning	As a proxy, criterion was considered met if the child was in the 'moderate' or 'severe' behavior problems group based on their (mother) Rutter items score.
Dc4	The symptoms are not better explained by another mental disorder (such as a mood disorder, anxiety disorder, dissociative disorder, or a personality disorder).	Cohort members were excluded if they had been diagnosed with another psychiatric disorder, as per the medical questionnaire, identified by ICD-9 codes. Only two children fulfilled this criterion.
Dc5	The symptoms do not happen only during the course of schizophrenia or another psychotic disorder	Assumed if no diagnosis - See item 4
Dc6	Symptoms should be present for at least six months	No mapping found

Table 1: Mapping of DSM-5 criteria to BCS70 age 10 questionnaire items (paraphrased) (American Psychiatric Association, 2013; Centre for Longitudinal Studies, 2015)
† Note on item codes: 'm' - Maternal Self Completion questionnaire, 'j' - Educational questionnaire, and 'R'- reverse coded

2.5.1.2 Recoding

Most of the mapped BCS70 items were presented to respondents using Visual Analog Scales (VAS; Figure 1). Post-completion, coders assigned values of 1-47 (teacher items), or 0-100 (mother items; Butler et al., 1997). Subsequently, studies have shown that VAS scales function as categorical rather than continuous variables because equal distance cannot be assumed between points; the likely maximum is three to four categories (Svensson, 2001; Wewers & Lowe, 1990). Hence, we recoded VAS items into more plausible categories. Visual inspection of histograms for raw VAS data indicated three roughly-equal-sized response levels. This is consistent with other measures of ADHD (e.g. the SDQ), which use 'not true', 'sometimes true' and 'certainly true' (or similar) as levels. However, the DSM-5 criteria are worded in a dichotomous way: symptoms occur 'often', or 'not often'. Accordingly dichotomous coding has been used in other IRT-based measures of ADHD (Gomez et al., 2011; Lindhiem et al., 2015). Therefore, we divided the scales into thirds and equated the bottom two-thirds to 'not true' and 'sometimes true', recoding both to 'not often' (0). The top third was equated to 'certainly true' and recoded as 'often' (1). Items were reverse coded as appropriate.

Three BCS70 teacher items (j080-talking, j081-moving around, j082-fidgeting) used a different scale ('what percentage of the time does the student spend…'). Precedent could not be found for categorically recoding this type of data. We coded only observations >=3 SDs from the mean as 'often' (1), which was difficult to achieve, but supported conservative inferences.

If more than one BCS70 item from parent *or* teacher mapped to a single DSM-5 criterion, the DSM-5 criterion was considered met if *any* of the mapped BCS70 items were met.

2.5.1.3 Application of DSM-5 ADHD criteria

Next, a categorical ADHD indicator and presentation type were derived by applying (approximated) DSM-5 diagnostic criteria (American Psychiatric Association, 2013) to our 16-item scale (Figure 2).

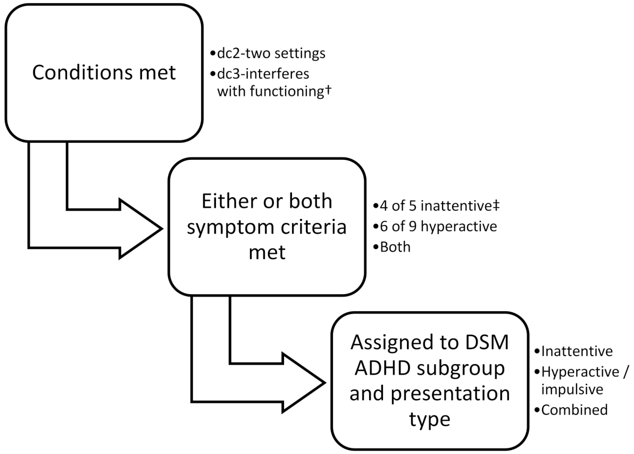

Figure 2. Process used to apply our approximation of DSM-5 ADHD criteria
† Conditions dc4 and dc5 (both based on another psychiatric diagnosis) were omitted from our scale due to insufficient variability. However, two children in our sample were explicitly excluded from the DSM-5-based ADHD subgroup due to another psychiatric diagnosis.
‡ 6/9 is two-thirds, so two-thirds of the of the 5 symptoms was used as a best approximation (3.35, rounded up to 4, to support conservative inferences)

2.5.1.4 Model selection

Descriptive statistics for a simple sum score of the 16 dichotomous items indicated a non-normal, zero-inflated distribution (i.e. a large proportion of the sample had zero symptoms: *n*=2,869, or 25%; see Figure 3). This supported use of a ZIMM model (Wall et al., 2015) for our analyses.

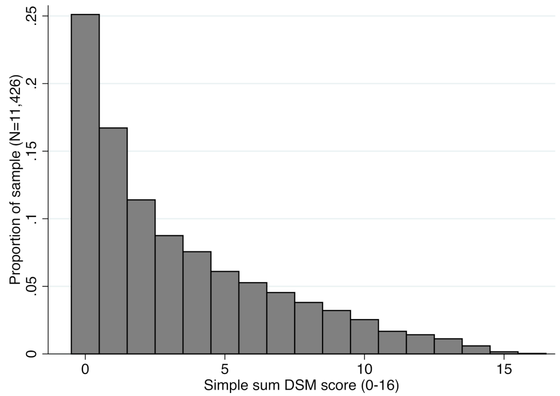

Figure 3. Histogram of mapped DSM-5 ADHD score (simple sum)
Demonstrates zero-inflated distribution

ZIMM is a zero-inflated mixture model, with 'mixture' referring to latent class and factor components. ZIMM uses a degenerate ('non-clinical') class, with an extreme fixed negative mean (μ= -100) and zero variance, to adjust for the influence of the large proportion of observations with zero symptoms (Wall et al., 2015). The second, 'clinical' class is then dominant in the estimation of model parameters, providing a dimensional measure of the latent trait that is less unduly biased by non-clinical cases (Finkelman et al., 2011; Magnus & Thissen, 2017; Wall et al., 2015).

For dichotomous data like ours, IRT models can estimate between one and four parameters: 1PL/2PL/3PL/4PL. The four parameters, building cumulatively, are: difficulty (i.e. location or threshold), discrimination, lower/guessing asymptote, and upper/fatigue asymptote (Embretson & Reise, 2000; Magis, 2013). DSM-5 ADHD items are unequal in their ability to discriminate (see Arias et al., 2018), so slopes will vary and 1PL estimating difficulty only is not adequate. The third and fourth lower and upper asymptote parameters are relevant in educational tests measuring ability, where respondents are motivated to achieve a high score (Embretson & Reise, 2000; Magis, 2013). Accordingly, 3PL and 4PL are not appropriate for psychiatric constructs (Finkelman et al., 2011). Therefore, the two-parameter logistic (2PL) model (Birnbaum, 1968) was used here. The 2PL model is operationalized through an item characteristic curve (ICC) for each item, with the following equation:

$$\Pr(X = 1) = \frac{e^{a(\theta-b)}}{1 + e^{a(\theta-b)}}$$

Where Pr = probability, X = response to the item (either 0 or 1), a = item discrimination, b = item difficulty, and θ = individual scaled factor score. Discrimination is the slope of the ICC at the steepest point, indicating how dramatically the probability of a positive response increases over the range of factor scores (θ). Difficulty is the point on the ICC where the probability of either (0 or 1) response is 50% (Baker, 2001).

2.5.2 Modelling

Within the data mining framework, modelling comprised testing model assumptions, building plausible models, and selecting a model with the best fit to the data.

2.5.2.1 Validation of IRT assumptions

Unidimensionality and local independence were supported by factor analysis on a matrix of tetrachoric correlations for the 16 items, showing clear dominance on a first factor (4.9 times the second factor), and low (<0.30) correlation residuals for each item pair (Embretson & Reise, 2000; Hambleton, Swaminithan, & Rogers, 1991). Tetrachoric correlations were used because they generate less error than Pearson's with categorical data (Embretson & Reise, 2000). Monotonicity was observed using Mokken's rule (Hardouin, Bonnaud-Antignac, & Sebille, 2011). The test indicated that item dh6-talks excessively, fell slightly short (*H*=0.26) of the criteria for being part of a strong scale (Loevinger's *H*>0.30; Hardouin et al., 2011).

The ZIMM models were based on Wall et al. (2015). We compared three variations (Table 2). The log likelihood, AIC, and BIC initially pointed to the ZIMM three class model as the best fit, but Entropy was low (0.45), indicating too many classes (Celeux & Soromenho, 1996). Thus, the ZIMM two-class model was selected, which aligns

with findings from the Wall et al., (2015) study. Mplus code for the ZIMM two-class model is provided in the supporting information (Appendix B).

Model	No. of parameters	logL	BIC	AIC	Entropy
1. 2PL IRT/1 class	32	-71944.82	143953.64	144188.63	NA
2. ZIMM 2 class	33	-71930.19	143926.39	144168.73	0.80
3. ZIMM 3 class	35	-71898.60	143867.21	144124.23	0.45

Table 2: Comparison of three item response models for dimensional measure
logL = log likelihood, BIC = Bayesian Information Criterion, AIC = Akaike's Information Criterion; fit statistics calculated in MPlus

3. Results

Results comprised an evaluation of psychometric properties for the derived 16-item scale, categorical measure based on DSM-5, and dimensional measure based on the ZIMM two-class model. Evaluation is the third phase of our data mining framework.

3.1 Derived 16-item scale

Reliability was good (Cronbach's α=0.85), and face validity was confirmed by an expert panel review (see section 2.5.1.1).

3.2 Categorical measure based on DSM-5 criteria (ADHD subgroup)

The derived ADHD subgroup (*n*=594) was 5.2% of the *N*=11,426 sample.
Since the data were collected in 1980-81 and no validated measures of DSM-5 ADHD were available (Butler et al., 1997), novel approaches were required to assess construct validity. These included comparisons to epidemiology and derived reference scales.

The DSM-5 ADHD subgroup had a similar composition to epidemiology/meta-analyses estimates of overall prevalence, gender, and subtype (Table 3). The subgroup was also comparable to epidemiology reports on ADHD samples, with over-representation of boys, health, social and economic disadvantages, and below average cognitive abilities (Table 4; Costello & Maughan, 2015; Loe & Feldman, 2007; Matza, Paramore, & Prasad, 2005; Willcutt, 2012).

Attribute	ADHD subgroup	Meta-analysis†
Prevalence	5.2%	6.1-7.1%
Ratio of boys to girls	2.3 : 1	2.4 : 1
Combined	35.6%	~32%
Hyperactive	12.4%	~18%
Inattentive	52.0%	~50%

Table 3: DSM-5 categorical subgroup compared to recent meta-analysis estimates
†(Willcutt, 2012, p. 492), data based on estimates from Table 1, only using full DSM-IV criteria data from parents and teachers, as these were most comparable to the method used in the present study. Precise figures were not available for the subtypes, so the '~' symbol indicates an approximation based on the data available.

Attribute	% of ADHD subgroup †	% of non-ADHD subgroup	Relative Risk Ratio (RRR) ‡
Boys	69.90	50.50	1.38
Lives in residential institution	1.90	0.40	4.75
Attends special school	3.20	0.64	5.00
Any medical condition	51.80	24.10	2.15
Family demographics			
Single mother	9.91	5.50	1.80
Unemployed father	6.13	2.96	2.07
Low family income	11.70	6.70	1.74
Cognitive abilities			
Below average reading age (<-1SD)	43.40	16.80	2.58
Below average maths (<-1SD)	44.60	15.20	2.93
Social class			
Professional or Managerial & Technical	16.50	29.80	0.55
Non-manual & manual	52.50	52.40	1.00
Partly skilled or Unskilled	25.30	17.80	1.42
Parent Malaise Inventory			
Severe problems (95+ percentile)	15.30	4.10	3.73

Table 4: Descriptive characteristics of DSM-5 categorically identified ADHD group compared to non-ADHD group
† ADHD Subgroup N = 594, non-ADHD subgroup N = 10,832; denominator in ratio varies as missing data are excluded
‡ $RRR = \frac{Risk\ of\ factor\ in\ ADHD\ group}{Risk\ of\ factor\ in\ non\ ADHD\ group}$

N.B. Relative Risk Ratio (RRR) > 1 indicates disadvantage, and < 1 indicates advantage (e.g. Professional and Managerial Social Class); RRR is also an effect size.

The SDQ hyperactivity subscale items were mapped (youthinmind, 2014b) to items from BCS70 (Table 5) and a sum score was derived for comparison. The simple sum score from our scale was highly correlated with the SDQ subscale score (r = 0.74, p < .001), supporting construct validity.

No.	SDQ	BCS70 questionnaire items
2	Restless, active, cannot stay still for long	M43 - Very restless. Often running about or jumping up and down. Hardly ever still.
10	Constantly fidgeting or squirming	M44 - Is squirmy or fidgety
15	Easily distracted, concentration wanders	M65 - Inattentive, easily distracted
21	Thinks things out before acting	R-M73 - Is impulsive, excitable
25	Sees tasks through to the end, good attention span	R-M76 - Fails to finish things he/she starts, short attention span

Table 5: SDQ hyperactivity subscale mapping to BCS70 items
VAS scores were recoded as follows: 0-32 -> 0 not true, 33-67 -> 1 somewhat true, 68-100 -> 2 certainly true.

Additionally, we replicated part of a study that derived a proxy measure for ADHD in BCS70. Their measure was based on Conners (Conners, 1969) and Rutter (Rutter, 1967) items (Brassett-Grundy & Butler, 2008), including several that are not currently considered part of the DSM-5 ADHD construct (see Literature Review). The replication-based subgroup (*N*=1,102) was much larger than ours (*N*=594) and membership overlapped only 66.5%. However, the simple sum scores from their scale (mother + teacher) and ours were highly correlated (*r*=0.82, *p*<.001), also providing some support for construct validity.

3.3 ZIMM model and estimated dimensional score

The two-class ZIMM was used to estimate a factor score (theta) for our sample; (*N*=11,426, *M* = -0.06; *SD* = 0.91). For cases with zero symptoms (*n*=2,869), *M*=-1.14, and for the remainder (*n*=8,557), *M*=0.30. The overall distribution had a similar shape to the simple sum score, though substantially more nuanced in variation, as predicted (Figure 4; note contrast to Figure 3).

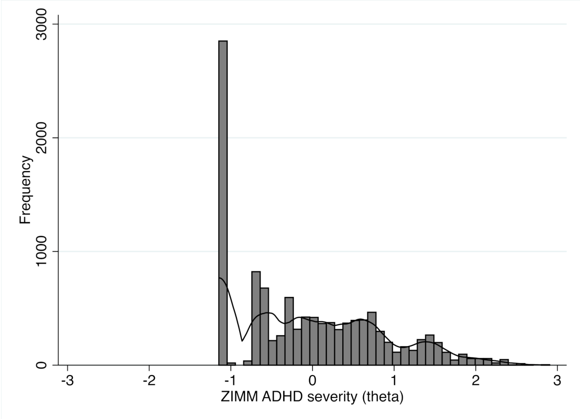

Figure 4. ADHD severity score estimated with ZIMM two-class model (N=11,426)
Showed expected zero inflation but with desired individual variation in ADHD severity

The IRT theta score correlated with the other measures derived, as expected. Logistic regression showed a significant association with the DSM-5-based categorical measure; (*N*=11,426, χ^2=3201.38, *p*<0.001, *df*=1; McFadden's R^2=0.69). Also, there was a large and significant positive correlation between theta and the derived SDQ subscale score (*r*=0.74, *p*<0.001), as well as the derived mother + teacher score (*r*=0.81, *p* < 0.001) calculated by part-replication of Brassett-Grundy & Butler (2008).

All the ZIMM two-class discrimination and difficulty parameters were significant, (*p*<.001; Table 6). Discrimination for symptoms ranged from 0.90 to 2.81, or moderate to very high (Baker, 2001). Difficulty ranged from 0.49 to 3.62, functioning best for individuals just above average to very high on the ADHD trait (Baker, 2001).

Item	Discrimination (α)	Difficulty (b)
Dh1 - fidgets or squirms	1.92	.53
Dh2 - inappropriately leaves seat	1.19	3.62
Dh3 – inappropriately runs about	1.19	1.09
Dh4 – cannot play quietly	1.73	1.57

Dh5 – on the go, 'driven by motor'	1.97	1.09
Dh6 – talks excessively	.90	3.27
Dh7 – blurts answers	1.30	1.38
Dh8 – trouble waiting turn	1.28	1.13
Dh9 – interrupts, intrudes	1.56	1.45
Di2 – trouble holding attention	1.49	.62
Di4 – doesn't follow through	1.74	.49
Di6 – avoids long tasks	1.37	1.05
Di8 – easily distracted	2.81	0.28
Di9 – often forgetful	1.27	1.25
Dc2 – symptoms interfere	1.31	1.40
Dc3 – multiple settings	5.09	1.24

Table 6: ZIMM 2 class 2PL IRT parameters

3.3.1 Information Characteristic Curves (ICC)

All 16 ICC curves visually supported the moderate-to-high ability of the items to discriminate between respondents (Figure 5; Baker, 2001). The most discriminating symptoms were di8-easily distracted (α=2.81) and dh5-'on the go/motor' (α=1.97). The least discriminating was dh6-talks excessively (α=0. 90). Two items had high difficulty: dh2-inappropriately leaves seat (β=3.62) and dh6-talks excessively (β=3.27), only providing information at very high levels of ADHD. Low difficulty items were dh1-fidgets (β=0.53), di2-trouble holding attention (β=0.62) and di4-doesn't follow through (β=0.49).

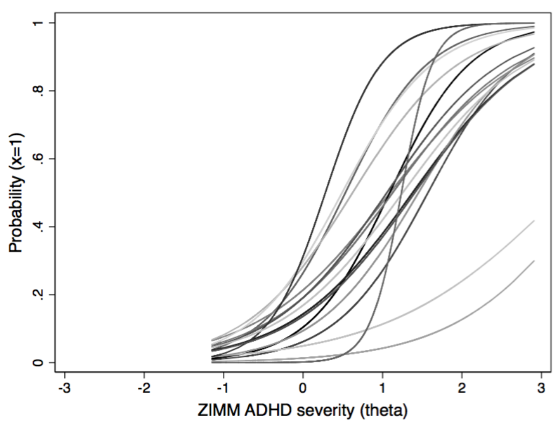

Figure 5. ICC curves of derived 16-item scale based on ZIMM two-class model
ICC = Item Characteristic Curve, ZIMM = Zero-inflated mixture model
Showed that items (other than the two flatter curves) discriminate well between individuals

3.3.2 Test information function

The Test Information Function shows how much information is provided by all items on the 16-item scale or 'test' at varying levels of the latent trait, based on the ZIMM two-class model (Figure 6). The curve shows our model provides the most information between theta values of 0.5 and 1.75, i.e. average to moderate levels of ADHD severity.

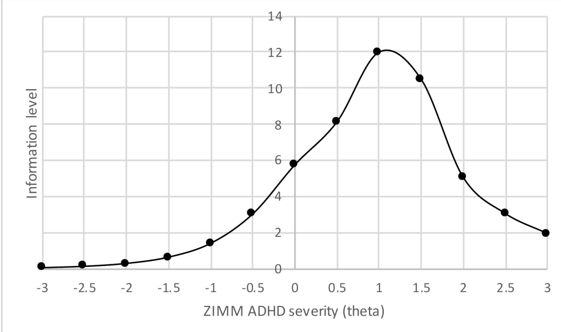

Figure 6. Test Information Function (TIF) for derived 16-item scale based on ZIMM two-class model
Showed the scale provides the most information at moderate levels of ADHD severity

3.3.3 Differential Item Functioning (DIF)

Other child mental health scales evaluate DIF (or measurement invariance) by gender, age and informant (e.g. the SDQ; youthinmind, 2014a). Age and informant were not applicable here because all participants were the same age, and our scale is based on combined responses from parent and teacher informants. Thus, we evaluated DIF by gender. According to the Mantel-Haenszel method, four items had significant DIF (p<0.05): two in favor of males and two in favor of females. However none had a large enough effect size to justify removal based on the Educational Testing Service (ETS) A/B/C classification method (Holland & Thayer, 1986).

3.4 Comparison of categorical and dimensional measures

Our DSM-5-based ADHD subgroup comprised 5.2% (*n*=594) of the sample (*N*=11,426). We compared this group to the top 5.2% (*n*=594) of the sample using the ranked IRT ADHD theta score (Table 7). 425 children (71.5%) were in both groups. Children in the IRT-based subgroup had slightly higher sum and theta scores, and were marginally more likely to be boys, have a medical condition, or a below-average reading age. They were less likely to have an unemployed father, or a parent with severe malaise (depression). 159 of the 169 children in the IRT-based group but *not* in the DSM-V-based group were missing the DSM condition dc3-symptoms interfere with functioning (based on the parent-rated Rutter behavior score). Nine were just under the threshold for both symptom lists (i.e. 3 inattentive symptoms and 5 hyperactive), and one had another psychiatric diagnosis, which was not taken into consideration in the IRT model.

Attribute	IRT ADHD subgroup (*n*=594)	DSM-5 ADHD subgroup (*n*=594)	RRR†
Average sum score	12.1	11.5	
Average IRT score	1.9	1.8	
	%	%	
Boys	73.4	69.9	1.05
Any medical condition	52.8	50.8	1.04
Below average reading age	35.0	31.8	1.10
Unemployed father	.04	.05	.90
Parent with severe malaise	13.1	15.3	.86

Table 7: Comparison of the top 5.2% based on IRT factor scores to the DSM-5-based categorical subgroup

† See notes on RRR (Relative Risk Ratio) with Table 4.

4. Discussion

Our objective was to develop and demonstrate a method to derive a categorical and dimensional measure of ADHD in existing data. We chose the BCS70 to mitigate limitations of insufficient cohort age, sample biases, and imprecise measures typically found in longitudinal studies of ADHD. A data mining framework was used to guide the approach. DSM-5 ADHD criteria were mapped to age 10 data items from BCS70 to derive a 16-item scale, and the mapping was validated by an expert panel. An approximation of the DSM-5 ADHD diagnostic procedure was used to identify a subgroup of children with ADHD symptomatology (*N* = 594; 5.2%). Prevalence is slightly lower than epidemiology estimates of 6%, perhaps because disadvantaged groups were under-represented in our sample, and disadvantaged groups tend to be over-represented in ADHD samples (Russell, Ford, Rosenberg, & Kelly, 2014). A ZIMM two-class model was selected as the optimal model for estimating a dimensional measure of ADHD, based on the non-normal, zero-inflated distribution, and comparison to two other plausible model variations. Psychometric properties tested for the 16-item scale, categorical ADHD measure, and dimensional ADHD measure were promising.

We included five of the six DSM-5 ADHD conditions, which is a strength given that most studies only evaluate symptoms (see Willcutt, 2012). However, four inattentive criteria and one of the conditions could not be mapped (Table 1). Nevertheless, the prevalence of inattentive type presentation in our sample was comparable to meta-analytic findings (Willcutt, 2012). This could be partially explained by findings from Li et al. (2015), who evaluated the full scale and found that two of the items missing from our scale had significant local dependence (di5 and di7; Li et al., 2015). Also Arias et al. (2018) analyzed the full scale and found that the most information was provided by three items (dh5, di2, and di8; Table 1), all of which were in our scale, possibly offsetting the absent items.

Two items, dh2-leaves seat and dh6-talks excessively, were based on BCS70 items from an unusual scale, and to be conservative we only coded an ‘often’ response for values 3SDs above the mean. Both items were accordingly high on difficulty parameters, and dh6 appeared as a weaker item per Mokken's rule and Loevinger's H. We accepted the high difficulty because it provides information at higher levels of the trait, which is desirable for our purposes. Regarding the relative weakness of dh6, we did not consider this an aberration, because others studies using typical levels of scale measurement also found dh6 to be a weaker item in terms of information provided (Arias et al., 2018; Gomez, 2011; Li et al., 2015).

The two approaches used to identify an ADHD subgroup (DSM diagnostic rules vs. top 5.2% based on IRT theta score) overlapped substantially in membership. Some difference was expected because the DSM-5 diagnostic rules assume all items are weighted equally, whilst the IRT model weights items according to their relative prevalence. Interestingly the IRT subgroup had a lower proportion of cases with an unemployed father or depressed parent. Non-overlapping cases were mostly (94%) explained by the parent rating of moderate to severe behavior problems (condition dc2-symptoms interfere). Children with an unemployed father or depressed parent may have been more likely to receive this rating, thus meeting the condition. This bias may indicate our mapped item dc2 is not an ideal indicator of the DSM condition. Moreover, endorsement for the mapping of this item, whilst acceptable, was somewhat mixed amongst expert panel members. These findings illuminate an interesting area for future work.

Our method extends previous work that aimed to identify ADHD in BCS70 (Brassett-Grundy & Butler, 2008) by adhering more closely to the current definition of ADHD, and estimating a more precise dimensional measure. We also built upon the work of Garcia-Barrera et al. (2011) by incorporating a data mining framework, more nuanced modelling technique, and validation through comparisons to mapped reference scales (e.g. SDQ), and epidemiology. Furthermore, we have replicated part of Wall et al. (2015) by re-using the ZIMM model, strengthening their findings, and applying the model to a different psychiatric construct (ADHD).

The present study adds to the literature on IRT models of ADHD, which has primarily focused on evaluating psychometric properties of items (e.g. Arias et al., 2018; Gomez et al., 2011; Li et al., 2015; G Polanczyk et al., 2010). Our approach aimed additionally to minimize error and estimate a theta score as precisely as possible, through use of a large non-clinical sample and adjustment for the zero-inflated distribution of symptomatology. Also, building a model within the longitudinal context of the BCS70 provides a previously untapped opportunity for future exploration of a wide range of antecedents to long-term outcomes.

Finally, our method is clearly documented and uses mainstream software, making it easy to replicate or adapt (see Appendix C in the supporting information regarding sharing of data). Thus, in addition to supporting our future work on causal mechanisms in long-term outcomes for ADHD, similar knowledge gains could be pursued by other authors applying our method in existing large datasets with numerous unmeasured psychiatric constructs.

The authors have no conflicts of interest to declare.

References

Agnew-Blais, J. C., Polanczyk, G. V., Danese, A., Wertz, J., Moffitt, T. E., & Arseneault, L. (2016). Evaluation of the Persistence, Remission, and Emergence of Attention-Deficit/Hyperactivity Disorder in Young Adulthood. *JAMA Psychiatry*, *73*(7), 1–8. http://doi.org/10.1001/jamapsychiatry.2016.0465

American Academy of Pediatrics, McNeil, & Wolraich, M. (2002). NICHQ Vanderbilt Assessment Scales. NICHQ. Retrieved from http://www.nichq.org/sites/default/files/resource-file/NICHQ_Vanderbilt_Assessment_Scales_0.pdf

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders (DSM-III-R)* (3rd, rev ed.). Washington DC: American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)*. Washington DC: American Psychiatric Association.

Arias, V. B., Esnaola, I., & Rodríguez-Medina, J. (2018). Identifying potentially marker symptoms of attention-deficit/hyperactivity disorder. *PeerJ*, *6*, e4820. http://doi.org/10.7717/peerj.4820

Arnett, J. J. (2000). Emerging adulthood: A theory of development from the late teens through the twenties. *American Psychologist*, *55*(5), 469–480. http://doi.org/10.1037/0003-066X.55.5.469

Baker, F. B. (2001). *The Basics of Item Response Theory* (2nd ed.). ERIC Clearing House on Assessment and Evaluation.

Barkley, R. A. (2015). *Attention-Deficit Hyperactivity Disorder, Fourth Edition: A Handbook for Diagnosis and Treatment*. (R. A. Barkley, Ed.) (Fourth). New York: The Guildford Press.

BERA. (2011). Ethical Guidelines for Educational Research. London: BERA.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee’s ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397–472). Reading, MA: Addison-Wesley.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, *53*, 605–634.

Brassett-Grundy, A., & Butler, N. (2008). *Prevalence and adult outcomes of Attention-Deficit / Hyperactivity Disorder* (BG occasional paper No. 2). *Bedford Group for Lifecourse & Statistical Studies*. Retrieved from http://cls.ioe.ac.uk/library-media%5Cdocuments%5CBG_OccPaper2_ADHD_outcomes_2008.pdf

Butler, N., Despotidou, S., & Shepherd, P. (1997). *1970 British Cohort Study Ten-year Follow-up: Guide to data available at the ESRC archive*. London.

Cadman, T., Findon, J., Eklund, H., Hayward, H., Howley, D., Cheung, C., … Asherson, P. (2016). Six-year follow-up study of combined type ADHD from childhood to young adulthood: Predictors of functional impairment and comorbid symptoms. *European Psychiatry*, *35*, 47–54. http://doi.org/10.1016/j.eurpsy.2015.08.007

Caye, A., Rocha, T. B.-M., Anselmi, L., Murray, J., Menezes, A. M. B., Barros, F. C., … Rohde, L. A. (2016). Attention-Deficit/Hyperactivity Disorder Trajectories From Childhood to Young Adulthood: Evidence From a Birth Cohort Supporting a Late-Onset Syndrome. *JAMA Psychiatry*, *73*(7). http://doi.org/10.1001/jamapsychiatry.2016.0383

Caye, A., Spadini, A. V., Karam, R. G., Grevet, E. H., Rovaris, D. L., Bau, C. H. D., … Kieling, C. (2016). Predictors of persistence of ADHD into adulthood: a systematic review of the literature and meta-analysis. *European Child and Adolescent Psychiatry*, *25*(11), 1–9. http://doi.org/10.1007/s00787-016-0831-8

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the numbers of clusters in a mixture model. *Journal of Classification*, *13*(2), 195–212. http://doi.org/10.1016/S0167-8655(98)00144-5

Centers for Disease Control and Prevention. (2013). ADHD, Symptoms and Diagnosis. Retrieved February 16, 2014, from http://www.cdc.gov/ncbddd/adhd/diagnosis.html

Centre for Longitudinal Studies: IoE/UCL. (2014). *CLS Cohort Studies Code of Practice*. London.

Centre for Longitudinal Studies. (2015). 1970 British Cohort Study. *Home of the 1958, 1970, Millennium and Next Steps Cohort Studies*. London. Retrieved from http://www.cls.ioe.ac.uk/page.aspx?&sitesectionid=795&sitesectiontitle=Welc

ome+to+the+1970+British+Cohort+Study

Cheung, C. H. M., Rijdijk, F., McLoughlin, G., Faraone, S. V., Asherson, P., & Kuntsi, J. (2015). Childhood predictors of adolescent and young adult outcome in ADHD. *Journal of Psychiatric Research*, *62*, 92–100. http://doi.org/10.1016/j.jpsychires.2015.01.011

Conners, C. K. (1969). A teacher rating scale for use in drug studies with children. *The American Journal of Psychiatry*, *126*(6), 884–888.

Conners, C. K. (2008). *Conners Comprehansive Behavior Rating Scale (CBRS)*. Toronto, Canada: Multi-Health Systems.

Costello, E. J., & Maughan, B. (2015). Annual Research Review: Optimal outcomes of child and adolescent mental illness. *Journal of Child Psychology and Psychiatry*, *56*(3), 324–341. http://doi.org/10.1111/jcpp.12371

Coughlin, S. S. (1990). Recall bias in epidemiologic studies. *Journal of Clinical Epidemiology*, *43*(I), 87–91.

Daly, M., Delaney, L., Egan, M., & Baumeister, R. F. (2015). Childhood Self-Control and Unemployment Throughout the Life Span. *Psychological Science*, *26*(6), 709–723. http://doi.org/10.1177/0956797615569001

Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development , evaluation , and refinement. *Quality of Life Research*, *16*, 5–18. http://doi.org/10.1007/s11136-007-9198-0

Elliott, J., & Shepherd, P. (2006). Cohort profile: 1970 British Birth Cohort (BCS70). *International Journal of Epidemiology*, *35*(4), 836–843. http://doi.org/10.1093/ije/dyl174

Elton, A., Alcauter, S., & Gao, W. (2014). Network Connectivity Abnormality Profile Supports a Categorical-Dimensional Hybrid Model of ADHD. *Human Brain Mapping*, *35*(9), 4531–4543. http://doi.org/10.1016/j.neuron.2009.10.017.A

Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. *Multivariate Applications Series*. Mahwah, NJ USA: Lawrence Erlbaum Associates Inc.

Finkelman, M. D., Green, J. G., Gruber, M. J., & Zaslavsky, A. M. (2011). A zero- and K-inflated mixture model for health questionnaire data. *Statistics in Medicine*,

30(9), 1028–1043. http://doi.org/10.1002/sim.4217

Gale, C. R., Hatch, S. L., Batty, G. D., & Deary, I. J. (2009). Intelligence in childhood and risk of psychological distress in adulthood: The 1958 National Child Development Survey and the 1970 British Cohort Study. *Intelligence*, *37*(6), 592–599. http://doi.org/10.1016/j.intell.2008.09.002

Garcia-Barrera, M. A., Kamphaus, R. W., & Bandalos, D. (2011). Theoretical and Statistical Derivation of a Screener for the Behavioral Assessment of Executive Functions in Children. *Psychological Assessment*, *23*(1), 64–79. http://doi.org/10.1037/a0021097

Gomez, R. (2007). Australian Parent and Teacher Ratings of the DSM-IV ADHD Symptoms. *Journal of Attention Disorders*, *11*(1), 17–27. http://doi.org/10.1177/1087054706295665

Gomez, R. (2008). Parent ratings of the ADHD items of the disruptive behavior rating scale: Analyses of their IRT properties based on the generalized partial credit model. *Personality and Individual Differences*, *45*(2), 181–186. http://doi.org/10.1016/j.paid.2008.04.001

Gomez, R. (2011). Item response theory analyses of adult self-ratings of the ADHD symptoms in the current symptoms scale. *Assessment*, *18*(4), 476–486. http://doi.org/10.1177/1073191110386341

Gomez, R. (2012). Item response theory analyses of adolescent self-ratings of the ADHD symptoms in the Disruptive Behavior Rating Scale. *Personality and Individual Differences*, *53*(8), 963–968. http://doi.org/10.1016/j.paid.2012.07.012

Gomez, R., Vance, A., & Gomez, A. (2011). Item response theory analyses of parent and teacher ratings of the ADHD symptoms for recoded dichotomous scores. *Journal of Attention Disorders*, *15*(4), 269–285. http://doi.org/10.1177/1087054709356404

Goodman, A., Joshi, H., Nasim, B., & Tyler, C. (2015). *Social and emotional skills in childhood and their long-term effects on adult life*. London.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *38*(5), 581–6.

Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9255702

Gorter, R., Fox, J. P., & Twisk, J. W. (2015). Why item response theory should be used for longitudinal questionnaire data analysis in medical research. *BMC Medical Research Methodology*, *15*, 55–015–0050–x. http://doi.org/10.1186/s12874-015-0050-x [doi]

Groen-Blokhuis, M. M., Middeldorp, C. M., Kan, K. J., Abdellaoui, A., Van Beijsterveldt, C. E. M., Ehli, E. A., … Boomsma, D. I. (2014). Attention-deficit/hyperactivity disorder polygenic risk scores predict attention problems in a population-based sample of children. *Journal of the American Academy of Child and Adolescent Psychiatry*, *53*(10), 1123–1129. http://doi.org/10.1016/j.jaac.2014.06.014

Hambleton, R. K., Swaminithan, H., & Rogers, H. (1991). *Fundamentals of Item Response Theory*. Newbury Park: SAGE Publications. Retrieved from https://books.google.co.uk/books?id=gW05DQAAQBAJ&printsec=frontcover#v=onepage&q&f=false

Hardouin, J.-B., Bonnaud-Antignac, A., & Sebille, V. (2011). Nonparametric item repsonse theory using Stata. *The Stata Journal*, *1*, 30–51. Retrieved from http://www.stata-journal.com/sjpdf.html?articlenum=st0216

Holland, P. W., & Thayer, D. T. (1986). *Differential Item Functioning and the Mantel-Haenszel Procedure*. *Program Statistics Research*. Princeton, NJ.

Kaat, A. J., & Farmer, C. (2017). Commentary : Lingering questions about the Social Responsiveness Scale short form. A commentary on Sturm et al. (2017). *The Journal of Child Psychology and Psychiatry*, *58*(9), 1062–1064. http://doi.org/10.1111/jcpp.12801

Kurgan, L. A., & Musilek, P. (2006). A survey of Knowledge Discovery and Data Mining process models. *The Knowledge Engineering Review*, *21*(01), 1. http://doi.org/10.1017/S0269888906000737

Kuriyan, A. B., Pelham, W. E., Molina, B. S. G., Waschbusch, D. a, Gnagy, E. M., Sibley, M. H., … Kent, K. M. (2013). Young adult educational and vocational outcomes of children diagnosed with ADHD. *Journal of Abnormal Child Psychology*, *41*(1), 27–41. http://doi.org/10.1007/s10802-012-9658-z

Lara, C., Fayyad, J., de Graaf, R., Kessler, R. C., Aguilar-Gaxiola, S., Angermeyer, M., … Sampson, N. (2009). Childhood Predictors of Adult Attention-Deficit/Hyperactivity Disorder: Results from the World Health Organization World Mental Health Survey Initiative. *Biological Psychiatry*, *65*(1), 46–54. http://doi.org/10.1016/j.biopsych.2008.10.005

Li, J. J., Reise, S. P., Chronis-Tuscano, A., Mikami, A. Y., & Lee, S. S. (2015). Item Response Theory Analysis of ADHD Symptoms in Children With and Without ADHD. *Assessment*, *23*(6). http://doi.org/10.1177/1073191115591595

Lindhiem, O., Yu, L., Grasso, D. J., Kolko, D. J., & Youngstrom, E. (2015). Adapting the Posterior Probability of Diagnosis (PPOD) Index to Enhance Evidence-Based Screening: An Application to ADHD in Primary Care. *Assessment*, *22*(2), 198–207. http://doi.org/10.1097/MPG.0b013e3181a15ae8.Screening

Loe, I. M., & Feldman, H. M. (2007). Academic and educational outcomes of children with ADHD. *Journal of Pediatric Psychology*, *32*(6), 643–54. http://doi.org/10.1093/jpepsy/jsl054

Magis, D. (2013). A Note on the Item Information Function of the Four-Parameter Logistic Model. *Applied Psychological Measurement*, *37*(4), 304–315. http://doi.org/10.1177/0146621613475471

Magnus, B. E., & Thissen, D. (2017). Item Response Modeling of Multivariate Count Data with Zero Inflation, Maximum Inflation, and Heaping. *Journal of Educational and Behavioral Statistics*, *42*(5), 531–558. http://doi.org/10.3102/1076998617694878

Matza, L. S., Paramore, C., & Prasad, M. (2005). A review of the economic burden of ADHD. *Cost Effectiveness and Resource Allocation : C/E*, *3*, 5. http://doi.org/10.1186/1478-7547-3-5

Muthen & Muthen. (2017). Mplus.

Muthen, B., & Asparouhov, T. (2006). Item response mixture modeling: Application to tobacco dependence criteria. *Addictive Behaviors*, *31*(6), 1050–1066. http://doi.org/10.1016/j.addbeh.2006.03.026

Polanczyk, G., de Lima, M. S., Horta, B. L., Biederman, J., & Rohde, L. A. (2007). The worldwide prevalence of ADHD: a systematic review and metaregression analysis. *The American Journal of Psychiatry*, *164*(6), 942–8. http://doi.org/10.1176/appi.ajp.164.6.942

Polanczyk, G., Laranjeira, R., Zaleski, M., Pinksy, I., Caetano, R., & Rohde, L. A. (2010). ADHD in a representative sample of the Brazilian population: estimated prevalence and comparative adequacy of criteria between adolescents and adults according to the item response theory. *International Journal of Methods in Psychiatric Research*, *19*(3), 177–184. http://doi.org/10.1002/mpr.319

Prevoo, T., & ter Weel, B. (2015). The importance of early conscientiousness for socio-economic outcomes: Evidence from the british cohort study. *Oxford Economic Papers*, *67*(4), 918–948. http://doi.org/10.1093/oep/gpv022

Qualtrics. (2017). Qualtrics. Provo, Utah, USA. Retrieved from http://www.qualtrics.com

Reise, S. P., & Revicki, D. A. (Eds.). (2015). *Handbook of Item Response Theory Modelling*. New York and London: Routledge.

Reise, S. P., & Waller, N. G. (2009). Item Response Theory and Clinical Measurement. *Annual Review of Clinical Psychology*, *5*(1), 27–48. http://doi.org/10.1146/annurev.clinpsy.032408.153553

Reuben, A., Moffitt, T. E., Caspi, A., Belsky, D. W., Harrington, H., Schroeder, F., … Danese, A. (2016). Lest we forget: comparing retrospective and prospective assessments of adverse childhood experiences in the prediction of adult health. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *57*(10), 1103–1112. http://doi.org/10.1111/jcpp.12621

Russell, G., Ford, T., Rosenberg, R., & Kelly, S. (2014). The association of attention deficit hyperactivity disorder with socioeconomic disadvantage: Alternative explanations and evidence. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *55*(5), 436–445. http://doi.org/10.1111/jcpp.12170

Rutter, M. (1967). A children's behaviour questionnaire for completion by teachers: Preliminary findings. *Journal of Child Psychology and Psychiatry*, *8*, 1–11.

Salum, G. A., Sonuga-Barke, E., Sergeant, J., Vandekerckhove, J., Gadelha, A., Moriyama, T. S., … Rohde, L. A. P. (2014). Mechanisms underpinning inattention and hyperactivity: Neurocognitive support for ADHD dimensionality.

Psychological Medicine, *44*(15), 3189–3201. http://doi.org/10.1017/S0033291714000919

Sass, D. A., Schmitt, T. A., & Walker, C. M. (2008). Estimating non-normal latent trait distributions within Item Response Theory using true and estimated item parameters. *Applied Measurement in Education*, *21*(1), 67–89. http://doi.org/10.1080/08957340701796415

Shaw, P., Malek, M., Watson, B., Greenstein, D., De Rossi, P., & Sharp, W. (2013). Trajectories of cerebral cortical development in childhood and adolescence and adult attention-deficit/hyperactivity disorder. *Biological Psychiatry*, *74*(8), 599–606. http://doi.org/10.1016/j.biopsych.2013.04.007

Sowell, E. R., Peterson, B. S., Thompson, P. M., Welcome, S. E., Henkenius, A. L., & Toga, A. W. (2003). Mapping cortical change across the human life span. *Nature Neuroscience*, *6*(3), 309–315. http://doi.org/10.1038/nn1008

StataCorp LLC. (2015). Stata Statistics/Data Analysis SE. College Station, Texas, USA.

Stone, L. L., Otten, R., Engels, R. C. M. E., Vermulst, A. a, & Janssens, J. M. a M. (2010). Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4- to 12-year-olds: a review. *Clinical Child and Family Psychology Review*, *13*(3), 254–74. http://doi.org/10.1007/s10567-010-0071-2

Stuart-Smith, J., Thapar, A., Maughan, B., Thapar, A., & Collishaw, S. (2017). Childhood hyperactivity and mood problems at mid-life: evidence from a prospective birth cohort. *Social Psychiatry and Psychiatric Epidemiology*, *52*(1), 87–94. http://doi.org/10.1007/s00127-016-1285-5

Sturm, A., Kuhfeld, M., Kasari, C., & Mccracken, J. T. (2017). Development and validation of an item response theory-based Social Responsiveness Scale short form. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *9*, 1053–1061. http://doi.org/10.1111/jcpp.12731

Svensson, E. (2001). Guidelines to Statistical Evaluation of Data From Scales and Questionnaires. *Journal of Rehabilitation Medicine*, *33*, 47–48. http://doi.org/10.1080/165019701300006542

Swanson, J. M., Arnold, L. E., Molina, B. S. G., Sibley, M. H., Hechtman, L. T., Hinshaw,

S. P., … Stern, K. (2017). Young adult outcomes in the follow-up of the multimodal treatment study of attention-deficit/hyperactivity disorder: symptom persistence, source discrepancy, and height suppression. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *58*(6), 663–678. http://doi.org/10.1111/jcpp.12684

Thapar, A., Cooper, M., Eyre, O., & Langley, K. (2013). Practitioner review: What have we learnt about the causes of ADHD? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *54*(1), 3–16. http://doi.org/10.1111/j.1469-7610.2012.02611.x

Ullebø, A. K., Posserud, M. B., Heiervang, E., Gillberg, C., & Obel, C. (2011). Screening for the attention deficit hyperactivity disorder phenotype using the strength and difficulties questionnaire. *European Child and Adolescent Psychiatry*, *20*(123), 451–458. http://doi.org/10.1007/s00787-011-0198-9

University of Manchester, University of Essex, & Jisc. (2015). UK Data Service. Retrieved October 16, 2015, from https://ukdataservice.ac.uk

Van Der Eijk, C., & Rose, J. (2015). Risky business: Factor analysis of survey data - Assessing the probability of incorrect dimensionalisation. *PLoS ONE*, *10*(3), 1–31. http://doi.org/10.1371/journal.pone.0118900

van Lieshout, M., Luman, M., Twisk, J. W. R., van Ewijk, H., Groenman, A. P., Thissen, A. J. A. M., … Oosterlaan, J. (2016). A 6-year follow-up of a large European cohort of children with attention-deficit/hyperactivity disorder-combined subtype: outcomes in late adolescence and young adulthood. *European Child and Adolescent Psychiatry*, *25*(9). http://doi.org/10.1007/s00787-016-0820-y

von Stumm, S., Gale, C. R., Batty, G. D., & Deary, I. J. (2009). Childhood intelligence, locus of control and behaviour disturbance as determinants of intergenerational social mobility: British Cohort Study 1970. *Intelligence*, *37*(4), 329–340. http://doi.org/10.1016/j.intell.2009.04.002

Wagner, F., Martel, M. M., Cogo, H., Renato, C., Maia, M., Pan, P. M., … Salum, G. A. (2016). Attention - deficit / hyperactivity disorder dimensionality : the reliable ‘ g ’ and the elusive ‘ s ’ dimensions. *European Child & Adolescent Psychiatry*, *25*, 83–90. http://doi.org/10.1007/s00787-015-0709-1

Wall, M. M., Park, J. Y., & Moustaki, I. (2015). IRT modeling in the presence of zero-inflation with application to psychiatric disorder severity. *Applied Psychological Measurement*, *39*(8), 583–597. http://doi.org/10.1177/0146621615588184

Wewers, M. E., & Lowe, N. K. (1990). A critical review of visual analogue scales in the measurement of clinical phenomena. *Research in Nursing & Health*, *13*(4), 227–236. http://doi.org/10.1002/nur.4770130405

Willcutt, E. G. (2012). The Prevalence of DSM-IV Attention-Deficit/Hyperactivity Disorder: A Meta-Analytic Review. *Neurotherapeutics*, *9*(3), 490–499. http://doi.org/10.1007/s13311-012-0135-8

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2017). *Data Mining: Practical Machine Learning Tools and Techniques* (Fourth). Cambridge, Massachusetts USA: Elsevier. Retrieved from https://books.google.co.uk/books?hl=en&lr=&id=1SylCgAAQBAJ&oi=fnd&pg=PP1&dq=data+mining+approach&ots=8IALqiiAua&sig=ZBdxa-_QuQ-3_bz7oMdeno9EJwo#v=onepage&q=data mining approach&f=false

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. *Educational and Psychological Measurement*, *73*(6), 913–934. http://doi.org/10.1177/0013164413495237

Woods, C. M. (2015). Estimating the Latent Density in Unidimensional IRT to Permit Non-Normality. In S. P. Reise & A. Revicki, Dennis (Eds.), *Handbook of Item Response Theory Modelling* (pp. 60–84). New York and London: Routledge.

youthinmind. (2012). SDQ: Information for Researchers and Professionals about the Strengths & Difficulties Questionnaire. Retrieved December 28, 2013, from http://www.sdqinfo.com

youthinmind. (2014a). SDQ Norms (Britain). Retrieved from http://www.sdqinfo.org/UKNorms.html

youthinmind. (2014b). SDQ Scoring Site. Retrieved from http://www.sdqscore.org/Amber

Supporting Information

Appendix A - Expert panel survey details
Survey instructions and first question:

Qualtrics Survey Software 04/02/2018, 18:50

UNIVERSITY OF CAMBRIDGE

Background

Background
I am developing a new method to identify Attention Deficit Hyperactivity Disorder (ADHD) symptomatology in existing longitudinal birth cohort data, to enable further study of long-term outcomes. The method relies on a mapping of DSM-5 (American Psychiatric Association, 2013) ADHD criteria to questionnaire items. In order to evaluate validity of this mapping, I am asking contacts with expertise in ADHD or related subjects to provide an independent assessment.

Please treat the content as confidential. If you would like to forward the survey link to someone else who would be well-suited to answering the questions, that would be fantastic, but please inform me you are doing so via email.

The survey should take between 5 and 10 minutes to complete.

Instructions, and DSM-V to questionnaire item mapping

Instructions
The questions below link DSM-5 ADHD criteria (wording slightly modified, see notes at end of survey) to questionnaire items designed for parent or teacher response regarding a 10-year old child (University College London, 2017). The items are mostly derived from Rutter (1967) and Conners (1969). There are five inattentive questions, nine hyperactive-impulsive, two

https://cambridge.eu.qualtrics.com/ControlPanel/Ajax.php?action=GetSurveyPrintPreview Page 1 of 11

Qualtrics Survey Software 04/02/2018, 18:50

conditional, three comments, and five questions about you and your role; 24 items in total. Please complete your response **by 30 September, 2017.** Responses are required for all statements; comments are optional.

Use the rating scale to indicate ***how well*** you think each statement ***captures the same meaning*** as the (slightly modified) DSM-5 criterion. Where there is more than one, evaluate each statement separately.

Inattentive items:

Has trouble holding attention to tasks or play activities (DSM-5 I2)
***(How well** does each statement below, evaluated separately, **capture the same meaning**...)*

	Extremely well	Very well	Moderately well	Slightly well	Not well at all
Does not pay attention to what is being explained in class	○	○	○	○	○
Becomes bored during class	○	○	○	○	○
Cannot concentrate on any particular task, even though the child may return to it frequently	○	○	○	○	○
Child concentrates poorly on educational tasks, in comparison with the average 10 year old	○	○	○	○	○

https://cambridge.eu.qualtrics.com/ControlPanel/Ajax.php?action=GetSurveyPrintPreview Page 2 of 11

Example of survey results:

Fidgets with or taps hands or feet, or squirms in seat (DSM-5 H1)

#	Question	Extremely well		Very well		Moderately well		Slightly well		Not well at all		Total
1	Is squirmy or fidgety	42.86%	6	50.00%	7	7.14%	1	0.00%	0	0.00%	0	14
2	Given to rhythmic tapping or kicking	0.00%	0	14.29%	2	57.14%	8	14.29%	2	14.29%	2	14
3	Fidgeting and indulging in minor distracting activities	14.29%	2	42.86%	6	42.86%	6	0.00%	0	0.00%	0	14

Four of the initially mapped BCS70 items were removed as a result of the review: di2/j138-bored during class, di4/j087-persevere with difficult tasks, di6/m241-sits still and concentrates more than 5 minutes, and di6/ j143-confused/hesitant with complex task (University of Bristol & National, 1980; University of Bristol & National Birthday Trust, 1980).

Those four items were clearly indicated by a majority of the experts as mapping 'not well at all'. Coincidentally, all four were somewhat redundant, as there were other BCS70 items that did map well to the relevant DSM-5 criteria. Two further items: dh7/m73-impulsive excitable, and dc3/moderate or severe behavior problems on the Rutter scale, did not have a clear majority of opinion from the experts, but mixed views. These two were the only candidate items from BCS70 that could map to the two relevant DSM-5 criteria, so we decided to keep them and adhere as closely to the DSM-5 as possible.

Appendix B - Mplus code used for 2-class zero-inflated mixture model

Derived from Wall et al., (2015), adapted with advice from Jung Yeon Park

```
TITLE:
ZI Mixture IRT (2 latent classes) based on derived scale of 16 DSM-5 criteria mapped to BCS70 Age 10 behaviour data;

DATA:
FILE IS <<filename.dat>>

VARIABLE:
NAMES = rowid dh1 dh2 dh3 dh4 dh5 dh6 dh7 dh8 dh9 di2 di4 di6 di8 di9 dc1 dc2;
IDVARIABLE IS rowid;
USEVARIABLES = dh1 dh2 dh3 dh4 dh5 dh6 dh7 dh8 dh9 di2 di4 di6 di8 di9 dc1 dc2;
CATEGORICAL = dh1 dh2 dh3 dh4 dh5 dh6 dh7 dh8 dh9 di2 di4 di6 di8 di9 dc1 dc2;
MISSING = ALL(999);
CLASSES = c (2);

ANALYSIS:
ESTIMATOR = MLR;
TYPE = MIXTURE;
ALGORITHM=INTEGRATION ODLL;
! the algorithm = odll is needed because of the model constraint command
STARTS 400 50;
STSEED 170056;
PROCESS = 6 (STARTS);
```

```
MODEL:
%OVERALL%
f BY dh1* dh2 dh3 dh4 dh5 dh6 dh7 dh8 dh9 di2 di4 di6 (lam1-lam12)
 di8 di9 dc1 dc2 (lam13-lam16);

[dh1$1* dh2$1 dh3$1 dh4$1 dh5$1 dh6$1 dh7$1 dh8$1 dh9$1] (tau1-tau9)
[di2$1 di4$1 di6$1 di8$1 di9$1 dc1$1 dc2$1] (tau10-tau16);

[c#1] (logitp1);

%c#1%
f*  (phi1);
[f*] (m1);

%c#2%
f*  (phi2);
[f*] (m2);

MODEL CONSTRAINT:
new(b1,b2,b3,b4,b5,b6,b7,b8,b9,b10,b11,b12,b13,b14,b15,b16);

m1= -100;
phi1= 0.0001;

m2= 0;
phi2= 1;

b1 = tau1/lam1;
```



```
b2 = tau2/lam2;
b3 = tau3/lam3;
b4 = tau4/lam4;
b5 = tau5/lam5;
b6 = tau6/lam6;
b7 = tau7/lam7;
b8 = tau8/lam8;
b9 = tau9/lam9;
b10 = tau10/lam10;
b11 = tau11/lam11;
b12 = tau12/lam12;
b13 = tau13/lam13;
b14 = tau14/lam14;
b15 = tau15/lam15;
b16 = tau16/lam16;

OUTPUT: TECH1 TECH8;
! plots for ICC curves
PLOT: TYPE = PLOT3;

SAVEDATA: FILE IS <<filename.dat>>;
          SAVE = FSCORES;
          SAVE = CPROB;
```

Appendix C – Note on sharing of derived categorical and dimensional measures data

Other researchers may wish to use the categorical and dimensional ADHD indicators we derived without replicating the entire analysis. The indicators would not be useful without the related identifier (BCSID) to allow linking with other variables in the BCS70 datasets. The BCSID is owned by the Centre for Longitudinal Studies (CLS). They encourage sharing of derived data, and we will share our variables with them, which they may share more widely at their discretion. In the meantime, requests for access to our data will be coordinated between us and the CLS on a case-by-case basis.