Measuring Severity of Mental Disorders with the Young Minds Matter Parent/Carer-Reported Impact Items

Technical report

Stephen R Zubrick

David Lawrence

Sarah Johnson

Jennifer Hafekost

Centre for Child Health Research, The University of Western Australia, Perth, Australia

23 March 2015

Corresponding Author: Stephen Zubrick, Telethon Kids Institute, The University of Western Australia, PO Box 855, West Perth WA 6872, Australia. Tel: +61 8 9489 7714, Fax: +61 8 9489 7700, Email: Stephen.Zubrick@telethonkids.org.au

Australian Government Department of Health

THE UNIVERSITY OF WESTERN AUSTRALIA

TELETHON KIDS INSTITUTE Discover. Prevent. Cure.

Roy Morgan Research

Acknowledgments

The survey was funded by the Australian Government Department of Health. The authors wish to express their gratitude to the 6,310 families who participated in the survey.

The authors would like to particularly thank Bill Buckingham and Suzy Saw for their advice and support throughout the entire project. We would also like to thank the following staff of the Telethon Kids Institute who were vital to the success of the project: Allyson Browne, Daniel Christensen, Katherine Hafekost, Monique Robinson, Francis Mitrou, Wavne Rikkers, Kirsten Hancock, Judy Straton, Leanne Scott, Karina Allen, Beth Veitch and Donna Cross.

The authors wish to thank the following people who helped with the cognitive testing of the questionnaire: Caroline Goosens, Margaret Jones, Simon Davies, Adrienne Wills, Jennifer Griffiths, Julie Lambert, Haylee Clark, Vanessa Watson, Shannon McNeair, Naomi D'Souza and Kym Thomas.

Professors Lawrence and Zubrick are supported by a grant from the Australian Research Council (CE140100027).

Suggested citation: Zubrick S.R., Lawrence D., Johnson S. and Hafekost J. (2015) *Measuring Severity of Mental Disorders with the Young Minds Matter Parent/Carer-Reported Impact Items: Technical Report*. Centre for Child Health Research, University of Western Australia: Perth, Australia, ISBN 978-1-74052-332-5.

Table of contents

Introduction	**1**
What are the impact items?	**1**
How were the items scaled?	**2**
Who were the respondents that completed the impact assessments?	**2**
Which modules required the use of the impact items?	**3**
How were the impact items created?	**3**
How were the impact items used to create composite scores?	**4**
How suitable are the items for estimating composite Domain Impact Scores and Diagnosis Impact Scores?	**5**
How well do the items measure Impact?	**6**
How were impact scores calculated?	**11**
How were the Diagnostic Impact Scores created?	**12**
How was Level of Severity determined?	**12**
Which method was finally chosen to classify severity?	**23**
How were the severity classifications applied to the Impact Domain Scores?	**26**
Conclusions	**30**
APPENDIX A: Confirmatory Factor Analyses, Scree Plots, Path Diagrams and Graded Response Models for the Impact Assessment Items: By Diagnostic group by Domain	**32**
References	**75**

Measuring Severity of Mental Disorders with the Young Minds Matter Parent/Carer-Reported Impact Items

Introduction

This document describes the calibration of the parent/carer reported impact items developed for use in the Second Australian Child and Adolescent Survey of Mental Health and Wellbeing ("Young Minds Matter " – YMM). These items are used to set the cut-points for defining levels of severity where children reach the diagnostic threshold on the parent-reported Diagnostic and Statistical Manual of Mental Disorders (Fourth Edition)[1] modules as operationalised by the Diagnostic Interview Schedule for Children (DISC)[2,3].

What are the impact items?

There are 17 items used to assess impact.

Table 1. The Impact Item Set

Variable		abbreviation
	Impact on schooling/work	
V1	In the last 12 months, when these problems were at their worst, how often did [child] not want to go to school?	nogo
V2	As a result of these problems, how many days has [child] been absent from [school/work] in the last 12 months?	days
V3	In the last 12 months, when these problems were at their worst, did they affect [child's] grades or ability to do [his/her] [schoolwork/work]?	grades
V4	When these problems were at their worst, did these difficulties limit child in participating in voluntary school activities such as leading a group, volunteering for an activity or contributing to class discussions?	volun
V5	When these problems were at their worst, did these difficulties limit child in participating in extracurricular activities, such as sports, music, arts or drama activities?	extra
V6	When these problems were at their worst, how often [was [child] in trouble at school/did [child's] boss get annoyed or upset with [him/her]]?	trouble
V7	When these problems were at their worst, how often did [child] have difficulties completing school work or home work on time?	schwork
	Impact on friends	
V8	In the last 12 months, when these problems were at their worst, did these difficulties cause problems with [child's] ability to make or maintain friendships?	breakup
V9	How often have these difficulties stopped [child] from doing things or going places with other children [his/her] age?	places
V10	When these problems were at their worst how much difficulty did [child] have dealing with people [he/she] didn't know well?	people
	Impact on family	
V11	In the last 12 months how often have [child's] difficulties prevented you from taking [him/her] places or going out in public?	taking
V12	How often have [child's] difficulties interrupted everyday family activities such as eating meals or watching TV?	famact
V13	How much distress do [child's] difficulties cause you and other members of the family?	famdis
V14	How much do [child's] difficulties impact on your other family and household responsibilities, such as time to spend with other children or family members?	respon

Table 1. The Impact Item Set

	Impact on child (self)	
V15	In the last 12 months, when these problems were at their worst, did these difficulties distress [child] or make [him/her] feel bad or upset? How distressed?	childd
V16	When these problems were at their worst, how much did these difficulties prevent [child] from concentrating on things [he/she] was supposed to be doing?	childcon
V17	When these problems were at their worst, how much did these difficulties impact on [child's] sleeping?	sleep

How were these items scaled?

A 5-point ordinal Likert scale was used to measure each item.

Table 2: Impact item scale response categories

Items 1, 5, 6, 7, 9, 11, 12	**Items 2**	**Items 13, 14, 17**	**Items 3, 4, 8, 15, 16**	**Item 10**
1 Never	1 None	1 None	1 Not at all	1 No difficulty
2 Hardly ever	2 1-2 days	2 A little	2 A little	2 Mild difficulty
3 Sometimes	3 3-5 days	3 Some	3 Somewhat	3 Moderate difficulty
4 Most of the time	4 6-15 days	4 A lot	4 A lot	4 Severe difficulty
5 All of the time	5 15+ days	5 Extreme	5 Extremely	5 They completely avoided people they didn't know

Who were the respondents that completed the impact assessments?

Not all survey respondents needed to respond to the impact items. The impact items were part of the diagnostic assessment for mental disorder and were administered only when children reached at least the diagnostic sub-threshold for the disorder. This is typically set at half the number of symptoms required to meet the diagnostic criteria. This means that any survey child or young person whose interview reached *at least* the sub-threshold for diagnosis on one or more of the DISC modules was subsequently administered the impact questions. This of course includes any child who went on to score at or above the diagnostic threshold to be classified with the given disorder.

The impact items were administered principally to the parent/carer of the child or young person. However, because of the known under-reporting of depression by parents of young people, the Major Depressive Disorder Module was administered to young people also. If they reached at least the sub-threshold for a diagnosis on the Major Depressive disorder Module, the young person was given the impact items to complete.[1]

[1] A total of 16 of the 17 items was administered. Item 14 (see Table 1) was not included for the young person.

This document presents analyses based upon the parent/carer-reported impact assessments only. The youth-reported impact assessments are not included in this report and are the subject of separate analyses.

Which modules required the use of the impact items?

There were eight diagnostic modules used in the DISC. These comprised Anxiety Disorders (Generalised Anxiety Disorder, Social Phobia, Separation Anxiety and Obsessive Compulsive Disorder), Major Depressive Disorder, Attention Deficit Disorder, Conduct Disorder and Oppositional Defiant Disorder.[2]

For some respondents where the child had substantial co-morbidity, the impact items could potentially be administered several times. To prevent over-burdening the respondent the impact items were administered for each of the following diagnostic groupings:

1. Attention Deficit Disorder

2. Conduct Disorder and/or Oppositional Problem Behaviours[2]

3. Any Anxiety disorder and/or Major Depressive Disorder

This means that a parent/carer whose child was comorbid across these three groupings, would receive the set of 17 impact items 3 times while a respondent whose child was co-morbid with Social Phobia (e.g. Anxiety) and Depression would receive the impact items once.

How were the impact items created?

The DSM-IV diagnostic classification is largely silent on the processes for determining severity of disorder for children. In commissioning the YMM the Department of Health required the survey team to address this.

In the First Child and Adolescent Mental Health Survey in 1998, the DISC was used to assess diagnostic status for three mental disorders - depression, ADHD and conduct disorder[4]. The DISC algorithms available at the time did not operationalise the clinical significant impairment criterion within the DSM criteria for each disorder. The original DISC included 6 questions about impairment that used the presence of specific symptoms to assess the level of impairment associated with them. These 6 questions probed:

1. How often <the given symptom(s)> annoyed or upset the parent.
2. How often <the given symptom(s)> prevented the child from going places or doing things with the family.
3. How often the <given symptom(s)> prevented the child from going places or doing things with the friends.
4. How difficult was it for the child to do their school work when the <given symptoms(s)> were at their worst,
5. How often the <given symptom(s)> annoyed or upset the teacher?

[2] Owing to the need for clinician administration, Oppositional Defiant Disorder could not be operationalized in the YMM. Instead, the YMM gathered information about Oppositional Problem Behaviours.

6. How bad did the symptoms made the child feel when <the given symptoms(s)> were at their worst.

Four algorithms (i.e. A thru D) were produced by the DISC team at Columbia University that could be used to assess impairment. The recommended algorithm for use in assessing impairment with the DISC is Algorithm D. Algorithm D requires either at least one of the six impairment questions (above) to be endorsed at the severe level; or, at least two of these impairment questions to be endorsed at the moderate level.

It is important for readers to appreciate that these 6 questions were used to assist the process of diagnostic classification. In other words, these six questions are used to determine if the minimum level of impairment required as part of the process for *assigning* diagnostic status is present or absent. The original DISC questions were not designed to assess levels of severity (Mild, Moderate and Severe) within those who met diagnostic criteria nor were there enough items to probe the range of domains (School, Friends, Family and Self) that could possibly be impaired when children had a mental disorder diagnosis.

In needing to assess level of severity, *as independent from diagnosis*, the YMM team needed to develop new items to expand the range of domains of possible impairment. This was undertaken by reviewing existing mental health assessment methods and developing new items to reflect the range of practice used to assess severity. Therefore, many of the items used (see Table 1) to measure impact in the YMM are being used for the first time.

As a result, the YMM team was required to undertake investigations into the nature of these items, their response profiles, and their potential utility for creating severity classifications.

This is a technical document that describes the process for determining whether the developed items can be used to classify severity.

How were the impact items used to create composite scores?

Several levels of analyse are possible given the nature of the item set and the method of its administration.

The 17 impact items probed respondents for information about four **domains of impairment**; that is, whether the child or young person was impaired 1) at school, or 2) with their friends, or 3) with their family and 4) how distressed with the symptoms they were themselves. The items were administered in response to the three specific **diagnostic grouping** examined by the DISC; 1) anxiety/depression, 2) ADHD and 3) conduct disorder. Because of the experimental nature of the item pool used to assess impact on function, we prepared the data for analyses as follows:

1. All item distributions were examined for skewness and kurtosis, missing data and outliers.
2. Item pools were then assembled into **Impact Domains** as follows:
    a) Impact at school (7 items)
    b) Impact with friends (3 items)
    c) Impact with family (4 items)
    d) Impact with self (3 items)

3. We then assessed the item and scale characteristics in each of these four domains for each of the following **Diagnostic Groupings**: Anxiety/Depression, Attention Deficit Disorder and Conduct Disorder. To do this we created composites representing **12 Domain Impact Scores** (4 Impact Domains by 3 Diagnostic Groupings).
4. We then created **3 Diagnosis Impact Scores** representing composites of the 17 items:
   a) Impact Score for Anxiety/Depression (17 items)
   b) Impact Score for Attention Deficit Disorder (17 items)
   c) Impact Score for Conduct Disorder (17 items)
5. We undertook an analysis that pooled the 3 Diagnosis Impact Scores to classify levels of **Severity: Mild, Moderate and Severe**.

To create the composites for each of the 12 Domain Impact Scores and 3 Diagnosis Impact Scores we:

1. Conducted principal components as a first screen to broadly assess whether the respective item pools were unidimensional;
2. Fitted ordinal confirmatory factor models using Robust Weight Least Squares methods to confirm unidimensionality of the Impact Domain Scales and the Total Impact Scores;
3. Fitted Graded Response Models using Item Response Theory (IRT) to the Impact Domain Scales and the Total Impact Scores;
4. Generated IRT scores for each of the Domain Impact Scores and Diagnosis Impact Scores;
5. For each of the 12 Domain Impact Scores and 3 Diagnosis Impact Scores, we calibrated **Severity Thresholds** for Mild, Moderate and Severe impairment.

Appendix A contains the detailed statistical summary of the analyses for the initial examination of the Domain Impact Scores.

How suitable are the items for estimating composite Domain Impact Scores and Diagnosis Impact Scores?

Do the item sets in each Domain actually measure the underlying concept of Impact?

In order for items to be aggregated to create a meaningful composite, the items need to validly measure the same underlying concept or dimension (i.e. “Impact”). Unidimensionality is a prerequisite for creating a composite score from a set of items and can be assessed in several ways. We used two methods to assess unidimensionality: Principal components factor analysis and confirmatory ordinal factor analysis.

Four domains of impact were assessed using principal components analysis: Impact at school, with friends, in the family and with respect to distress to the child. Each of these domains was assessed for each of the Diagnostic Groupings and we provide here the 12 scree tests for each. This is a descriptive technique and if the items are unidimensional a scree test should reveal one leading principal component (eg. “factor”) that explains the underlying relationships among the item set.

Principal components analysis revealed that all 12 of the Domain Impact Scores could be considered unidimensional using the broad criteria of the first Eigenvalue >4 relative to the second Eigenvalue and/or only one Eigenvalue >=1.0 (see Figure 1).

We then carried out the same procedure for each of the three Diagnostic Impact Scores. This procedure used all 17 of the items to examine their unidimensionality within each of the three diagnostic groups. With 17 items there is more variation in the possible underlying relationships. This is shown in the scree plots (see Figure 2). Notwithstanding this, there is good evidence that the 17 items broadly measure the same underlying dimension or factor of impact with each demonstrating a very large first factor with a substantially smaller second factor.

How well do the items measure Impact?

A stronger statistical test of unidimensionality was undertaken using ordinal confirmatory factor analysis. Ordinal confirmatory factor models were fitted for each of the 12 domain impact item sets using diagonally adjusted robust weighted least squares estimation. In this method, the composite score may be derived using item factor score regression weights, which are estimated as part of the confirmatory factor analysis (CFA) process. Instead of assuming that each item contributes equally to the factor, this approach adjusts the weighting of each item on the factor.

The selection of appropriate fit indices for SEM has been extensively reviewed notably by Hu and Bentler[5-7] and Yuan and Bentler[8]. A 'combinational' rule, in which two or possibly three fit indices are used to judge model fit, is recommended. The selection of the recommended fit indices is reliant upon sample size, distributional characteristics of the data, and model complexity. Hu and Bentler suggest the use of the Standardized Root Mean Residual (SRMR) supplemented with one of either the NNFI, (Non-Normal Fit Index; also called the TLI, Tucker-Lewis Index) or the Comparative Fit Index (CFI)[6] . These have been used here.

In addition to diagonally adjusted robust weighted least squares estimation the H-index of scale reliability is also calculated[9]. This is a measure of the *proportion of variance* accounted for in the underlying factor and is selected for reporting here rather than the traditional Cronbach's alpha. The H-index is the preferred indicator of scale reliability for ordinal measures (see Hancock and Mueller, 2006). It represents the squared correlation (i.e. variance) between the underlying latent construct (i.e. factor) and the optimum linear composite formed by its indicators (i.e. items). If the items perfectly measured the concept of Impact, then the H index would equal 1.0 (i.e. 100% of the variance in Impact is measured by the items). Broadly speaking, magnitudes of H >= 0.80 are considered desirable with respect to scale performance.

There are limitations to the ability to statistically fit some of the models (Table 3). This is because single factor models require at least 4 items in order to estimate degree of fit. Measures of Impact on Friends and Impact on Self only have 3 items and these models fit the data perfectly having no degrees of freedom (i.e. the models are saturated), so it's not possible to estimate fit indices.

With this in mind, results broadly indicated acceptable fit where these could be estimated. Scale reliabilities (H indices) were more variable with weaker scale reliabilities consistently produced across all Diagnostic Groups for Impact on Self. The most robust impact scale, in terms of both fit and scale reliability, was the Impact on School scale. This undoubtedly reflects the size of the item pool (7 items).

In summary, the results of the principal component scree tests and the confirmatory factor analyses suggest that unidimensionality for each of the 12 Domain Impact Scores was reasonable and on this basis we took the next step to estimate Domain Impact Scores using IRT methods.

Table 3: Ordinal Confirmatory Factor Analysis: Robust Weighted Least Squares, Fit Indices

Anxiety/Depression							
Domain	**N**	**df**	**SRMR**	**NNFI**	**CFI**	**H**	**Fit**
School	1651	14	0.06	0.96	0.97	0.94	Acceptable
Friends	1862	0	-	-	-	0.92	Saturated model
Family	1903	2	0.04	0.97	0.99	0.91	Acceptable
Self	1849	0	-	-	-	0.82	Saturated model
Attention-Deficit/Hyperactivity Disorder							
School	1762	14	0.07	0.95	0.97	0.93	Acceptable
Friends	1975	0	-	-	-	0.95	Saturated model
Family	2025	2	0.03	0.99	0.99	0.91	Acceptable
Self	1963	0	-	-	-	0.82	Saturated model
Conduct Disorder							
School	889	14	0.08	0.95	0.97	0.93	Acceptable
Friends	993	0	-	-	-	0.86	Saturated model
Family	1015	2	0.04	0.98	0.98	0.90	Acceptable
Self	978	0	-	-	-	0.76	Saturated model

Fit criteria: Acceptable fit Standardized Room Mean Residual (SRMR) <= 0.09; Non-Normal Fit (NNFI) >= 0.95; Comparative Fit Index (CFI) >=0.95; H-index of reliability.

We then undertook to fit the Diagnostic Impact Scores using robust diagonally weighted least squares factor analysis. Table 4 contains a summary of the fit indices for the 3 Total Domain Impact Scores.[3]

Overall, the Diagnostic Impact Scores fit the underlying measurement model for assessing Total Impact.

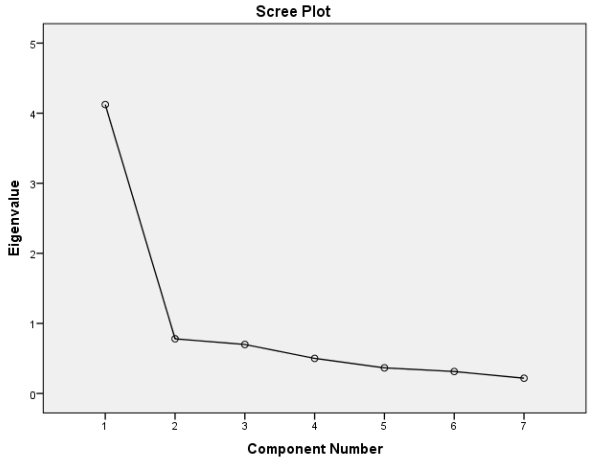
In summary the composites for the 12 Domain Impact Scores and the 3 Diagnostic Impact Scores demonstrate a reasonable level of unidimensionality. We then undertook to use IRT methods to calculate these composites.

[3] Factor loadings and regression coefficients for the Total Domain Scores are not reproduced here to conserve space. They are available on request.

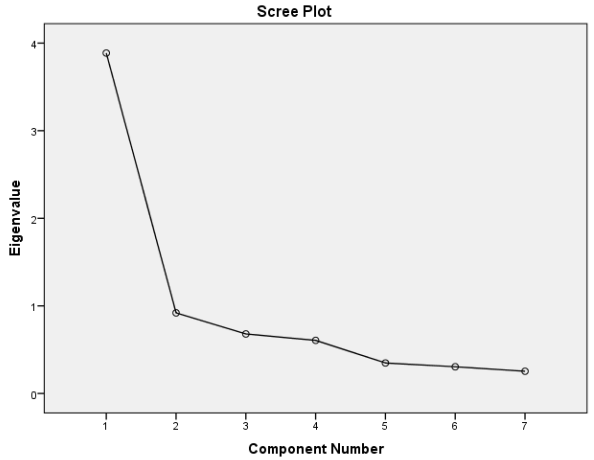
Figure 1: Unidimensionality of Domain Impact Scores: Principal Components Analysis

Impact at School (7 items)

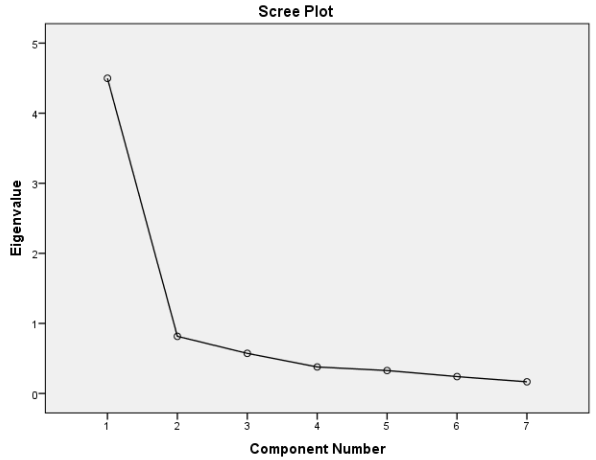
Anxiety/Depression



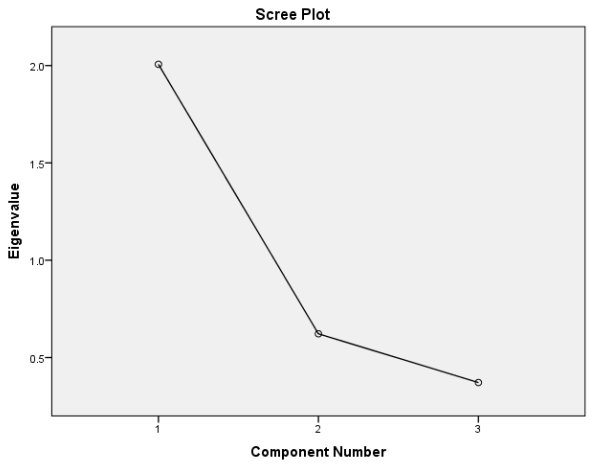
Attention-Deficit/Hyperactivity Disorder

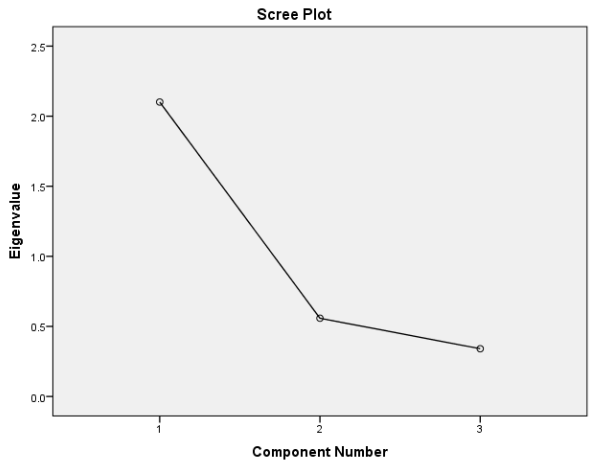


Conduct Disorder



Impact with Friends (3 items)





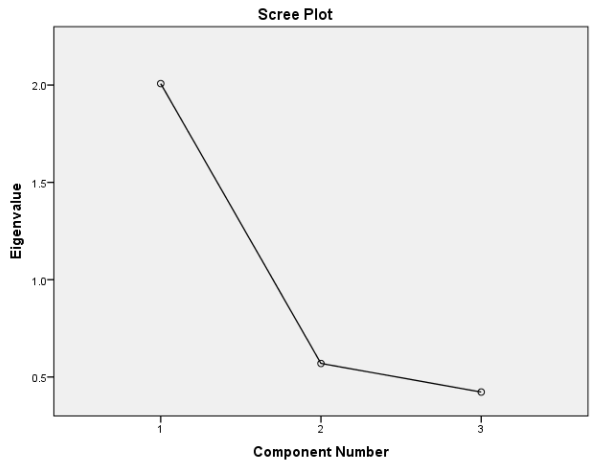
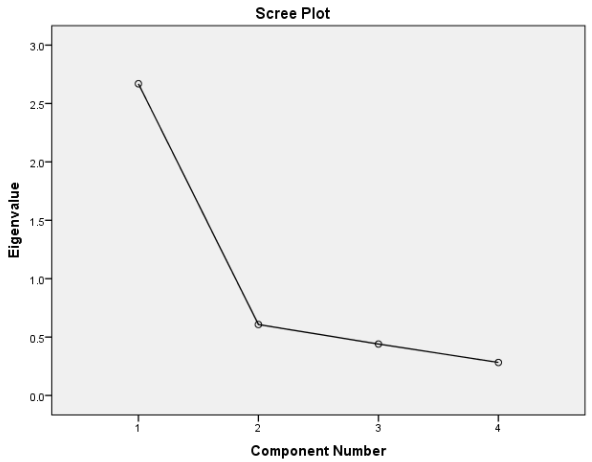


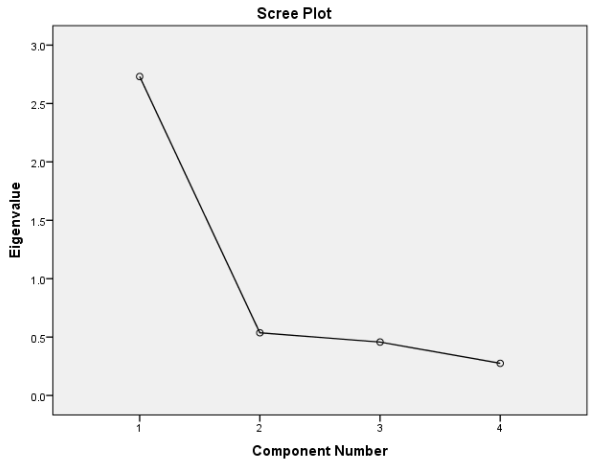
Figure 1, con't Unidimensionality of Domain Impact Scores: Principal Components Analysis

Impact on Family

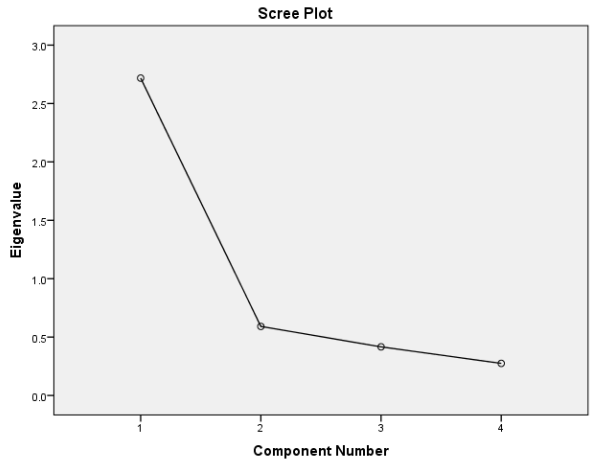
Anxiety/Depression



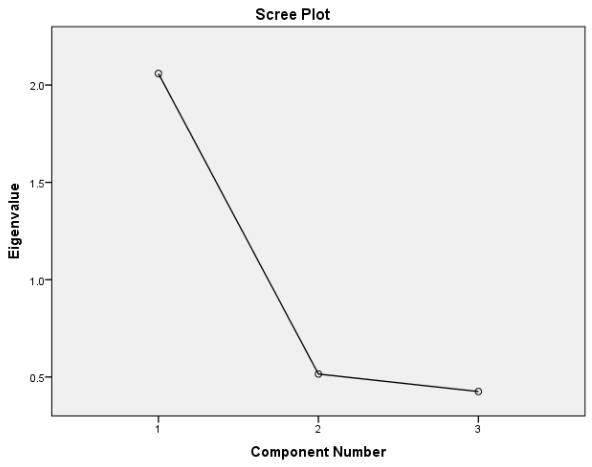
Attention-Deficit/Hyperactivity Disorder

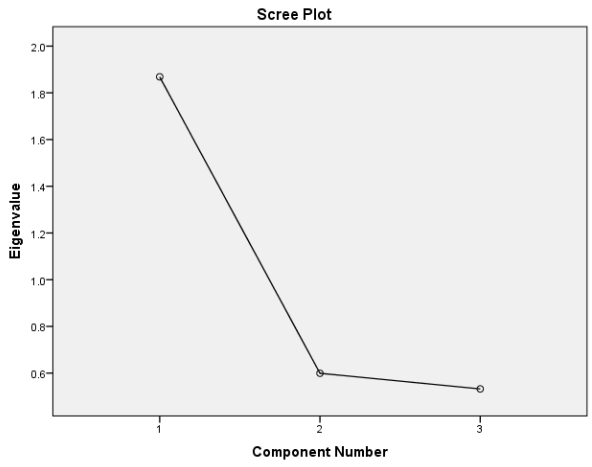


Conduct Disorder



Impact on Self





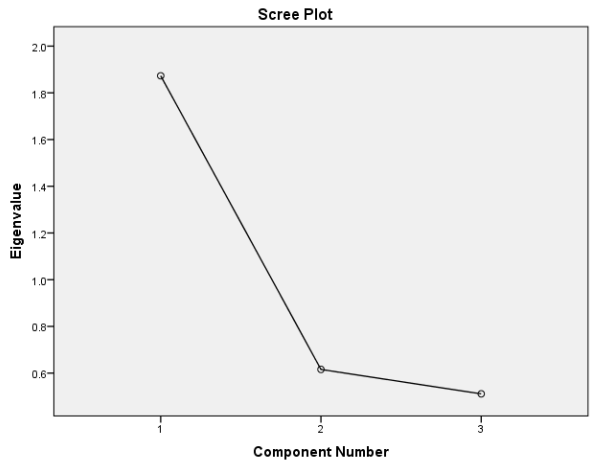
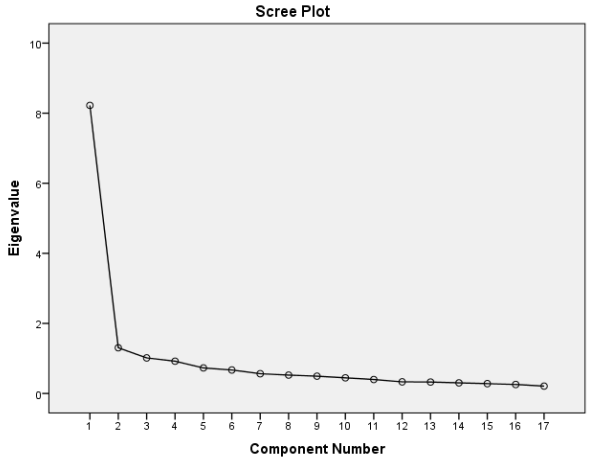


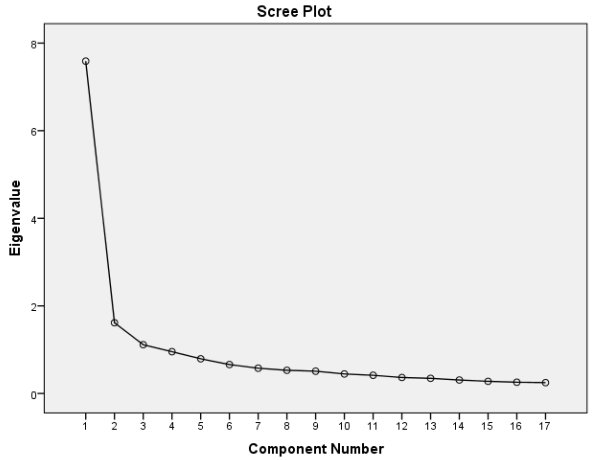
Figure 2, Unidimensionality of Diagnostic Impact Scores: Principal Components Analysis

Total Domain Impact Score

Anxiety/Depression



Attention-Deficit/Hyperactivity Disorder



Conduct Disorder

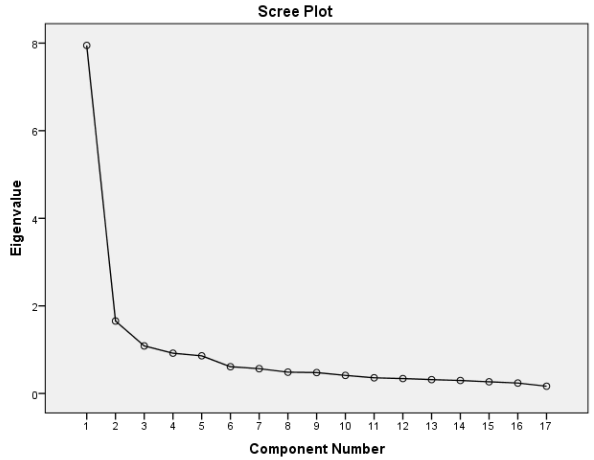


Table 4: Ordinal Confirmatory Factor Analysis: Robust Weighted Least Squares, Fit Indices Diagnostic Impact Scores

Anxiety/Depression						
N	**Df**	**SRMR**	**NNFI**	**CFI**	**H**	**Fit**
1578	119	0.07	0.97	0.97	0.96	Acceptable
Attention-Deficit/Hyperactivity Disorder						
1696	119	0.09	0.96	0.97	0.96	Acceptable
Conduct Disorder						
848	119	0.09	0.96	0.97	0.96	Acceptable

Fit criteria: Acceptable fit Standardized Room Mean Residual (SRMR) <= 0.09; Non-Normal Fit (NNFI) >= 0.95; Comparative Fit Index (CFI) >=0.95; H-index of reliability.

How were impact scores calculated?

To generate Domain Impact Scores and Diagnostic Impact Scores we fitted Graded Response Models (GRM) using Item Response Theory (IRT)[10-12].

Several views of the items were derived from fitting GRMs to them.

First, the items varied considerably in their power to discriminate underlying impact and supply information about the impact of the given diagnostic classification. Not all items performed well.

Table 5: Impact Domain by Diagnostic Grouping – Most discriminating impact items*

Impact Domain	**Anxiety/Depression**	**ADHD**	**Conduct**
School	Grades Completing school work Being absent	Volunteering Grades Being absent	Grades Volunteering Completing school work
Friends	Going places with friends	Going places with friends	Going places with friends Maintaining friendships
Family	Family household responsibilities Difficulties for other family members Taking the child places	Family household responsibilities Difficulties for other family members Taking the child places	Family household responsibilities Difficulties for other family members Interrupting family activities
Self	Distress child Prevent doing things	-	Prevent doing things

**Items were selected on the basis of the IRT GRM discrimination coefficient (a) and are rank ordered in this table. Coefficients >= 2.0 are selected here and are considered to have "very high" discrimination[13,14].*

Table 5 shows the most discriminating items selected from the GRMs of each of the Impact Domains for each of the Diagnostic Groupings. A discriminating item is one in which each level of the response (i.e from 1 to 5) carries a significantly unique proportion of the information about impact. So, for example, there were seven items in the Impact on School Domain, but only three of the items yielded a high amount of information about impact. Broadly speaking, only when the given disorder affected the child's grades, attendance at school, being able to volunteer for activities and completing school work did the parent judge the disorder to have impact on school function.

Impact on Friends broadly reduced to being able to go places with friends. Impact on Self was a particularly weak measure.

Second, in addition to discrimination, the IRT analysis also provides item response curves which permit assessing how the scales for each for the items perform as a measure of increasing impact. Broadly speaking, as impact increases, the probability that a parent will indorse a higher response category should also increase. This should occur monotonically. Along with the quantitative tables (Appendix A) the item response curves showed that the item impact scales were not monotonically sensitive to increasing impact. While respondents were asked to rate each of the items on a scale of 1 to 5, their actual responses revealed that many categories were not equally sensitive to impact and that categories could be collapsed without loss of discriminating power. Table 6 provides a visual schematic to show where the optimal "bin breaks" or scale boundaries are located according to the IRT GRMs.

Finally, the IRT analysis permitted examination of the overall fit of the GRM model to each of the 12 scales. In other words, despite the variation in item discrimination and the lack of monotonicity, how well did the Domain Impact Scales perform?

This analysis needs to be interpreted with some caution.

Global fit indices can be problematic for models with a large number of parameters. A large number of items and a large sample size can conceal poor item fit in the presence of an overall global fit. However, the models here are small in terms of items – in some cases probably too small, comprising only three items. We examined the global fit statistics for the GRM models (G^2 using the likelihood ration chi square to compare observed and expected frequencies where full item classification was possible and M_2 statistic based on the one- and two-way marginal tables as a proxy for G^2 where models were too large). Each statistic was accompanied by an estimated Root Mean Squared Error of Approximation (criterion <= 0.06). Appendix A shows that the global fit indices for the 12 models were acceptable.

How were the Diagnostic Impact Scores created?

We then turned to estimating the three Diagnostic Impact Scores. How well do the 17 items measure the total impact for each Diagnostic Grouping? In other words, can the 17 items be aggregated to create an impact score for each of the three Diagnostic Groupings?

We undertook a full assessment of this using the same procedures described above.

Scree tests and distributions for the Diagnostic Impact Scores are provided in Figure 3. The general modelling from both the principal components and from the Robust Weighted Least Squares (not shown) indicated that the each of the 17 item scales were adequate measures of the underlying concept of impact. On this basis we proceeded to estimate impact scores by pooling the assessments and fitting a graded response model using IRT.

How was Level of Severity determined?

The pooled data permitted a single IRT analysis of the 17 items which were administered 4950 times to the parents of 3153 survey children. Fifty-seven percent (N=1805) had only one administration of

the impact items, 28.5% (N=899) had 2 administrations, and 14.2% (N=449) had 3 administrations of the impact items. Cut-points were set by applying the National Mental Health Service Planning Framework standard ratio of severity for mental disorders to the pooled data for those who reached the diagnostic threshold.

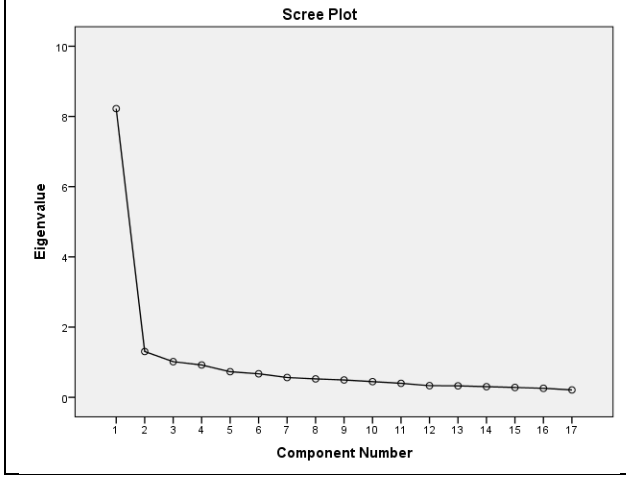
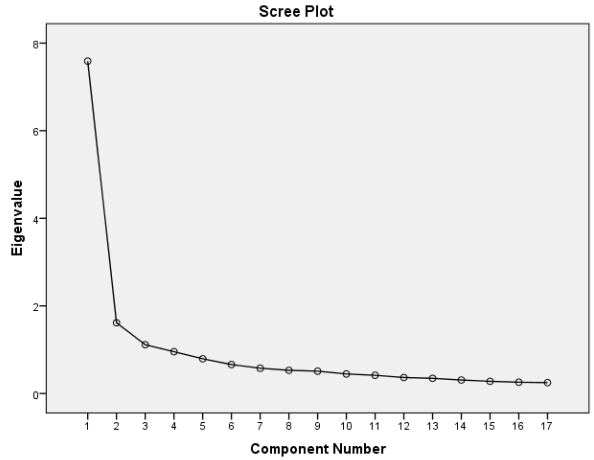
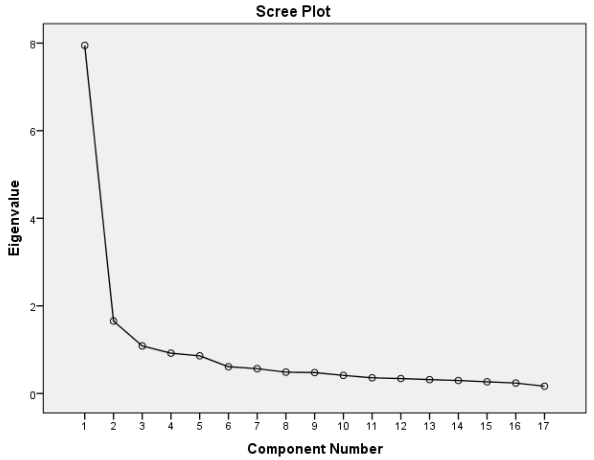
The 4950 responses were first analysed using principal components factor analysis. The 17 variables demonstrated excellent factorability (KMO = 0.93). Communalities ranged from 0.43 to 0.78.

Table 6: Item Response Curves: Optimal Category Bins (|)

Impact of School			
	ANX/DEP	ADHD	CONDUCT
asv001 In the last 12 months, when these problems were at their worst, how often did child not want to go to school	1 2 | 3 | 4 5	1 2 | 3 | 4 5	1 2 | 3 | 4 5
ASV002r As a result of these problems, how many days has [child] been absent from school in the last 12 months?	1 2 3 | 4 5	1 2 3 | 4 5	1 2 3 | 4 5
asv003r In the last 12 months, when these problems were at their worst, did they affect child's grades or ability to do work?	1 | 2 | 3 | 4 | 5	1 | 2 | 3 | 4 | 5	1 | 2 | 3 | 4 | 5
asv004 When these problems were at their worst, did these difficulties limit child in participating in volunteer activities?	1 2 | 3 4 | 5	1 2 | 3 4 | 5	1 2 | 3 4 | 5
asv005r When these problems were at their worst, did these difficulties limit child in participating in extracurricular activities?	1 2 | 3 | 4 5	1 2 | 3 | 4 5	1 2 | 3 | 4 5
asv006r When these problems were at their worst, did these difficulties cause trouble at school (or at work)?	1 2 | 3 | 4 5	1 2 | 3 | 4 5	1 2 | 3 | 4 | 5
asv007 When these problems were at their worst, how often did child have difficulties completing school work?	1 2 | 3 | 4 | 5	1 2 | 3 | 4 | 5	1 2 | 3 | 4 | 5
Impact on Friends			
	ANX/DEP	ADHD	CONDUCT
ASV012r In the last 12 months, when these problems were at their worst, did these difficulties cause problems with [child's] ability to make or maintain friendships?	1 2 3 4 | 5	1 2 3 4 | 5	1 2 3 4 | 5
ASV014 How often have these difficulties stopped [child] from doing things or going places with other children [his/her] age?	1 | 2| 3 | 4 | 5	1 2 | 3 | 4 | 5	1 2 | 3 | 4 | 5
ASV015 When these problems were at their worst how much difficulty did [child] have dealing with people [he/she] didn't know well?	1 | 2 | 3 | 4 5	1 2 | 3 | 4 5	1 2 | 3 | 4 5

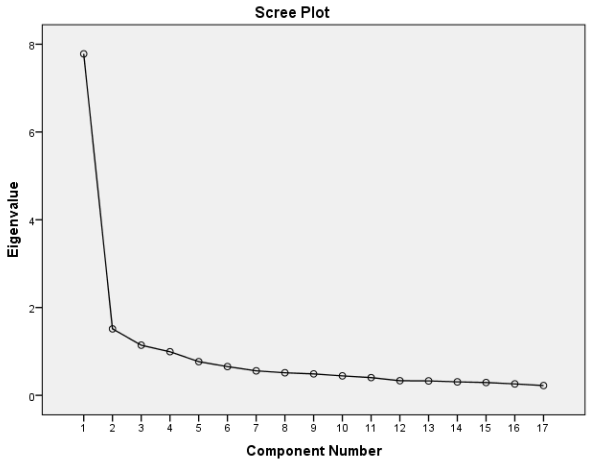
Table 6: Item Response Curves: Optimal Category Bins (|)

	Impact on Family		
	ANX/DEP	ADHD	CONDUCT
ASV016 In the last 12 months how often have [child's] difficulties prevented you from taking [him/her] places or going out in public?	1 2 | 3 | 4 5	1 2 | 3 | 4 5	1 2 | 3 | 4 | 5
ASV017 How often have [child's] difficulties interrupted everyday family activities such as eating meals or watching TV?	1 2 | 3 | 4 | 5	1 2 | 3 | 4 | 5	1 2 | 3 | 4 | 5
ASV018 How much distress do [child's] difficulties cause you and other members of the family?	1 | 2 | 3 | 4 | 5	1 | 2 | 3 | 4 | 5	1 | 2 | 3 | 4 | 5
ASV019 How much do [child's] difficulties impact on your other family and household responsibilities, such as time to spend with other children or family members?	1 | 2 | 3 | 4 | 5	1 | 2 | 3 | 4 | 5	1 | 2 | 3 | 4 | 5
	Impact on Self		
	ANX/DEP	ADHD	CONDUCT
ASV020r In the last 12 months, when these problems were at their worst, did these difficulties distress [child] or make [him/her] feel bad or upset? How distressed?	1 2 | 3 | 4 | 5	1 2 | 3 | 4 | 5	1 2 | 3 | 4 | 5
ASV021 When these problems were at their worst, how much did these difficulties prevent [child] from concentrating on things [he/she] was supposed to be doing?	1 | 2 | 3 | 4 | 5	1 | 2 | 3 | 4 | 5	1 2 | 3 | 4 | 5
ASV022 When these problems were at their worst, how much did these difficulties impact on [child's] sleeping?	1 2 | 3 4 | 5	1 2 | 3 4 | 5	1 2 | 3 4 | 5

Figure 3: Diagnostic Impact Scores – Scree plots		
Anxiety/Depression	**ADHD**	**Conduct Disorder**
		

Three factors were extracted representing 61% of the common factor variance. While there is a case to be made for extracting more than one factor, the extremely large Eigenvalue of Factor 1 suggested a dominant factor underlies the item set. Examination of the "simple structure" (varimax rotation) revealed multiple item loadings across three potentially possible factors (NB: only three variables demonstrated a single loading on one factor).

Figure 4. Scree plot of the Diagnostic Impact Items – pooled sample (N = 4950)



A subsequent confirmatory factor analysis using robust diagonally weighted least squares resulted in a single factor model with a good fit (df = 119; SRMR = 0.09; NNFI=0.96; CFI = 0.96). Scale reliability was very good (H = 0.96). This suggested that the pooled data offered a broadly unidimensional item set suitable for inputting to IRT analysis.

Table 7 contains the results from the Graded Response Model. Item slopes (a_i) and item thresholds (b_{k-1}) along with their standard errors are shown. The item level diagnostic statistics are shown in Table 9.

Item slopes (a) ranged from 1.23 to 2.78 and provide a general indication of how informative each item is of severity. Volunteering for activities and participating at school, not being able to do things or go places with other children and an impact on grades or school work were the three items with the highest slopes indicating that they were highly informative of severity. In contrast, the item measuring the level of distress the symptoms caused the child was much less informative of severity.

Item locations (b_i) represent the given severity level at the threshold *between* the response-option categories. The 17 items each have 5 possible response categories, so there are 4 thresholds separating these categories. Each value of b_i represents the severity level necessary to respond above the category threshold with a 0.50 probability. So for example, the severity score for the "places" variable is b_4 = 2.56. This means that once the severity score reaches a level of 2.56 there is a 50% probability that the respondent will have rated the "How often have these difficulties stopped the child from going places with other children" at the level of "All of the time" (i.e. a value of 5, which is the above the category threshold of 4. Examination of the b_i values shows that some impact

items that are rated by respondents at a lower rating level actually carry a greater level of severity than some items rated at a higher level.

Table 7: Graded Response Model Item Parameter Estimates for the Diagnostic Impact Items – pooled sample

Item	Label	*a*	*s.e.*	b_1	*s.e.*	b_2	*s.e.*	b_3	*s.e.*	b_4	*s.e.*
1	nogo	1.58	0.05	0.06	0.02	0.62	0.03	1.49	0.04	2.27	0.06
2	days	1.44	0.05	0.68	0.03	1.00	0.03	1.46	0.04	2.02	0.06
3	grades	2.28	0.06	-0.09	0.02	0.47	0.02	1.02	0.03	1.97	0.04
4	volun	2.73	0.08	0.32	0.02	0.73	0.02	1.23	0.03	2.07	0.04
5	extra	2.30	0.07	0.54	0.02	0.87	0.02	1.56	0.03	2.25	0.05
6	trouble	1.33	0.04	0.26	0.03	0.91	0.03	2.11	0.06	3.22	0.10
7	schwork	1.95	0.05	-0.25	0.02	0.11	0.02	1.03	0.03	1.89	0.04
8	breakup	1.79	0.06	0.54	0.02	0.94	0.03	1.24	0.03	1.40	0.04
9	places	2.32	0.07	0.55	0.02	0.93	0.02	1.78	0.04	2.56	0.06
10	people	1.50	0.04	0.23	0.03	1.20	0.04	2.16	0.06	2.76	0.08
11	taking	1.66	0.05	0.72	0.03	1.13	0.03	2.20	0.06	3.27	0.10
12	famact	1.23	0.04	0.05	0.03	0.56	0.03	2.14	0.06	3.52	0.11
13	famdis	1.93	0.05	-0.94	0.03	0.12	0.02	0.91	0.03	2.29	0.05
14	respon	1.85	0.05	-0.16	0.02	0.66	0.02	1.48	0.04	2.76	0.07
15	childd	1.31	0.04	-0.37	0.03	0.26	0.03	1.26	0.04	2.60	0.07
16	childcon	2.15	0.05	-0.85	0.03	0.18	0.02	0.93	0.03	2.23	0.05
17	sleep	1.50	0.04	0.04	0.02	0.76	0.03	1.42	0.04	2.78	0.08

Table 8 contains item level diagnostic information. This information provides some indication of how well each item fits the Graded Response Model. Items that fit well have smaller, non-significant Chi Square values. Six of the items have p=values < 0.01 suggesting lack of fit, however, the ratio of X^2/df < 3.0 suggesting for these items suggesting that item fit is probably reasonable. The global test of fit, based on the M_2 statistic produced an RMSEA = 0.04, which was deemed acceptable.

Appendix B contains the item characteristic curves, the item information curves, and the total information curve for the 17 item impact assessment estimated from the pooled sample of assessment forms (N = 4950).

Table 8: Item level diagnostic statistics for the Diagnostic Impact Items – pooled sample

Item	Label	X^2	*d.f.*	Probability
1	nogo	224.63	207	0.1905
2	days	267.53	209	0.0039
3	grades	215.47	184	0.0559
4	volun	220.78	173	0.0082
5	extra	207.82	189	0.1654
6	trouble	400.22	220	0.0001
7	schwork	190.86	186	0.3878
8	breakup	276.07	194	0.0001
9	places	185.21	184	0.4615
10	people	297.18	213	0.0001
11	taking	185.29	197	0.7154
12	famact	250.50	214	0.0441
13	famdis	182.91	182	0.4676
14	respon	208.92	194	0.2198
15	childd	404.65	208	0.0001
16	childcon	192.84	176	0.1824
17	sleep	207.34	211	0.5587

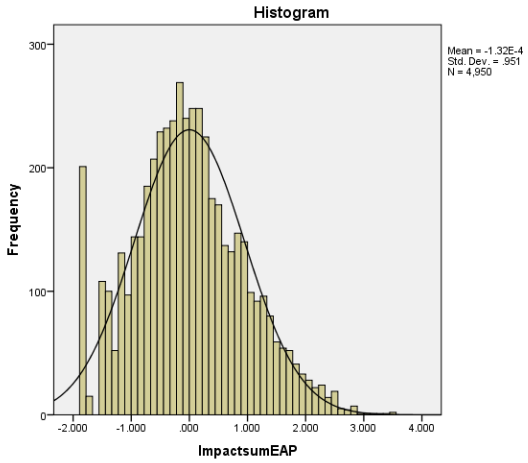
Figure 5 shows the distribution of pooled Impact Score. Estimated scores have a mean of approximately zero and a standard deviation of 1.0. Lower scores represent low impact/severity and higher scores represent high levels of impact/severity. The distribution is bimodal in the sense that it contains children with no level of impact (far left of distribution).

Of more relevance are the sub-distributions contained within the pooled distribution for the three Diagnostic Impact Scores. Figure 6 contains each of the three Diagnostic Impact Scores – for Anxiety/Depression, ADHD and Conduct Disorder. Because these total distributions contain children who only reached sub-threshold status on the DISC-IV, distributions are also provided for those children who had reached a diagnostic threshold and required a severity classification (rightmost column, Figure 6).

Distributions for both the Sub-threshold and Diagnostic Positive children were very good and broadly normally distributed. This latter feature is an important finding: There are no natural "breaks" in the distribution of impact: Impact is continuously and relatively normally distributed. Relative to the individual Domain Impact Scores the three Diagnostic Impact Scores provide a much more robust and well-distributed estimate of impact.

Figure 5. IRT Score Distribution for the pooled impact assessments

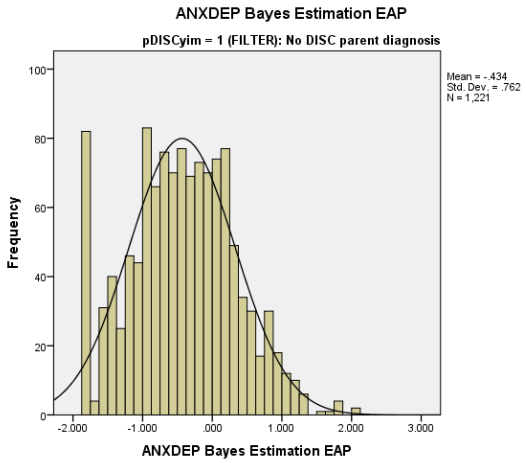
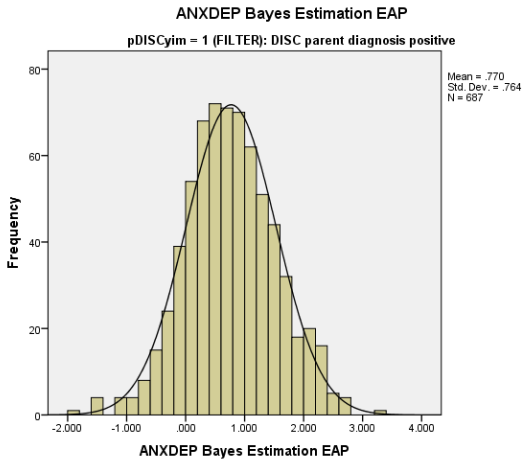
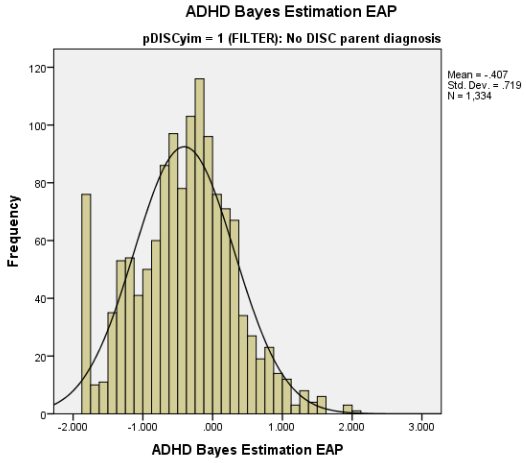
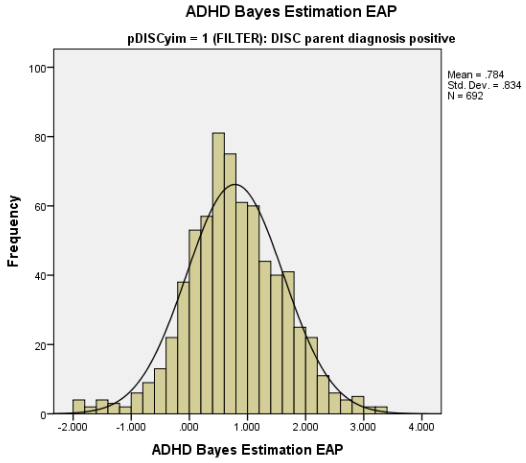
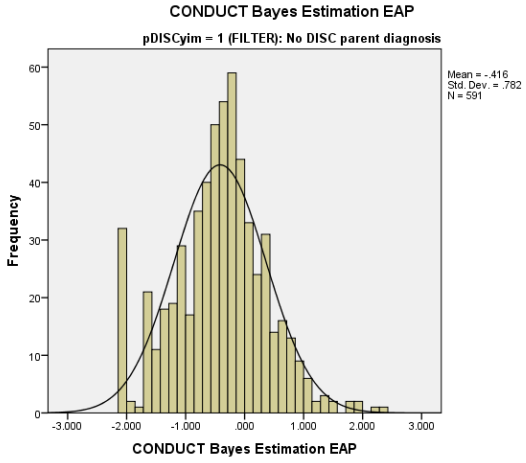
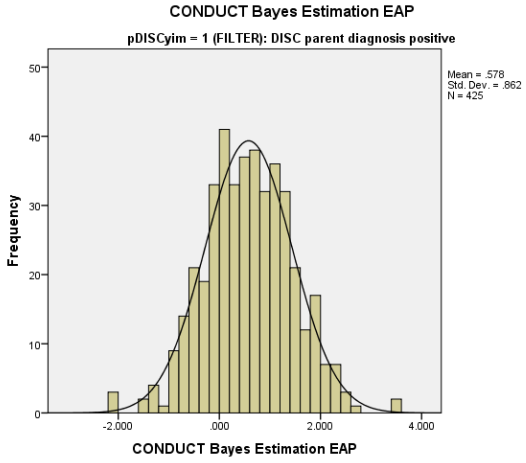
(N=4950)



With no "natural" break in the continuous distributions of the three Diagnostic Impact Scores, several options were canvassed for categorising these scores into Severity groups. One of the many advantages of the IRT method is the calculation of a continuous weighted composite measure of severity that takes into account the relative contribution of the 17 impact items in measuring the underlying concept of impact on functioning. The IRT process also allows for the differential role of some items dependent on the 3 groupings of disorder. Not all items are equally sensitive to changes in severity (see Table 5) nor are the scales equally differentiated (see Table 6).

The next step required imposing categorical breaks (or "cut-points") on the continuous IRT-derived scores to classify these distributions into three broad categories representing mild, moderate and severe impact on functioning. The challenge in undertaking this is to work out where to sub-divide the continuous measure to determine severity of disorder without any external benchmarks (eg. an external "clinical" validation sample or a gold standard severity measure) to guide this process. Clinical cut-points demarking a threshold of clinically meaningful impairment of a given disorder are simply lacking, there is no "gold standard" tool that could have been used in the survey process, and the survey process did not permit a separate clinical calibration study to assist with setting independent thresholds for severity.

Figure 6: Diagnostic Impact Scores

Diagnostic impact score- Sub-threshold cases	Diagnostic impact score – Diagnosis positive
	
	
	

One approach considered entailed establishing cut points based on standard deviation units (0.8; 1.4; 1.8) which would deliver approximately 3% severe, 6% moderate and 12% mild disorders. This recommendation is based on a precedent first established in the work of Thomas Achenbach, and further supported in results from the Epidemiologic Catchment Area surveys in the United States. However, it was not possible to directly apply these distribution cut-points to the YMM scores created by the IRT process. This was because the YMM scores were created *only* for those that had reached at least a sub-threshold level on one of the 8 disorders assessed. The 17 impact items were only asked of parents where at least a minimum level of symptoms in the child or young person was reported. As such, a different proportion of the survey sample was asked the impact on functioning items in each of the three disorder groups. So although the IRT scores are scaled to have a common mean and standard deviation, these scores represent the distribution of different proportions of the total population in each disorder group.

In addition to this, another problem with applying the same cuts-points across the three total impact scales is that the derived IRT scales are all essentially normally distributed. Applying the same cut points would result in very similar proportions of severity across different disorders. For instance, if the same cut-points were applied in each of the three disorder groups, a different proportion of the total sample (as opposed to the sub-threshold and above sample) would be expected to above the cut-off, because of the different proportions of the sample being filtered into each of the sets of questions.

No clinical calibration study has been conducted and as a result there is no external measure of severity to use as a benchmark for validating any of the measures of severity of functional impact derived from the survey. In the absence of a clinical calibration study, the survey team explored a few approaches that were largely data-driven.

An early approach to imposing severity categories on the impact item set entailed subjectively fitting each item and its respective rating-level to the Child Global Assessment of Severity Scale (C-GAS) (Table 9)[15]. No composite score was produced. Instead, a decision algorithm based loosely on the wording in the C-GAS was applied to each of the items in their respective domains to judge each domain with respect to mild, moderate and severe impairment. This resulted in a scaling that coincidentally produced approximately a ratio of 1:2:4 across the categories of severe, moderate and mild impairment. This result paralleled the expected planning ratios for levels of severity.

Two other approaches were developed that were quantitative and that broadly imposed the expected planning ratio (1:2:4) on the IRT distributions. In the first approach the planning ratios were applied in turn to each of the 3 Diagnostic Scores. In the second, the planning ratios were applied to the IRT distribution for the pooled impact data. In both instances, provision was made to include suicide attempt as an indicator that resulted in the assignment to the “severe” category.

Table 9. Child-Global Assessment of Severity (C-GAS) Scale

100-91	Superior functioning in all areas (at home, at school and with peers); involved in a wide range of activities and has many interests (eg., has hobbies or participates in extracurricular activities or belongs to an group such as Scouts, etc); likeable, confident; 'everyday' worries never get out of hand; doing well in school; no symptoms.
90-81	Good functioning in all areas; secure in family, school, and with peers; there may be transient difficulties and 'everyday' worries that occasionally get out of hand (eg., mild anxiety associated with an important exam, occasional 'blowups' with siblings, parents or peers).
80-71	No more than slight impairments in functioning at home, at school, or with peers; some disturbance of behaviour or emotional distress may be present in response to life stresses (eg., parental separations, deaths, birth of a sib), but these are brief and interference with functioning is transient; such children are only minimally disturbing to others and are not considered deviant by those who know them.
70-61	Some difficulty in a single area but generally functioning pretty well (eg., sporadic or isolated antisocial acts, such as occasionally playing hooky or petty theft; consistent minor difficulties with school work; mood changes of brief duration; fears and anxieties which do not lead to gross avoidance behaviour; self-doubts); has some meaningful interpersonal relationships; most people who do not know the child well would not consider him/her deviant but those who do know him/her well might express concern.
60-51	Variable functioning with sporadic difficulties or symptoms in several but not all social areas; disturbance would be apparent to those who encounter the child in a dysfunctional setting or time but not to those who see the child in other settings.
50-41	Moderate degree of interference in functioning in most social areas or severe impairment of functioning in one area, such as might result from, for example, suicidal preoccupations and ruminations, school refusal and other forms of anxiety, obsessive rituals, major conversion symptoms, frequent anxiety attacks, poor to inappropriate social skills, frequent episodes of aggressive or other antisocial behaviour with some preservation of meaningful social relationships.
40-31	Major impairment of functioning in several areas and unable to function in one of these areas (ie., disturbed at home, at school, with peers, or in society at large, eg., persistent aggression without clear instigation; markedly withdrawn and isolated behaviour due to either mood or thought disturbance, suicidal attempts with clear lethal intent; such children are likely to require special schooling and/or hospitalization or withdrawal from school (but this is not a sufficient criterion for inclusion in this category).
30-21	Unable to function in almost all areas eg., stays at home, in ward, or in bed all day without taking part in social activities or severe impairment in reality testing or serious impairment in communication (eg., sometimes incoherent or inappropriate).
20-11	Needs considerable supervision to prevent hurting others or self (eg., frequently violent, repeated suicide attempts) or to maintain personal hygiene or gross impairment in all forms of communication, eg., severe abnormalities in verbal and gestural communication, marked social aloofness, stupor, etc.
10-1	Needs constant supervision (24-hour care) due to severely aggressive or self-destructive behaviour or gross impairment in reality testing, communication, cognition, affect or personal hygiene.

Which method was finally chosen to classify severity?

All three methods of deriving impact on functioning provided very similar classifications into mild, moderate and severe groups. The three methods deliver highly correlated impact assessments. This level of convergence is reassuring. All three methods also showed a similar level of discrimination in comparison with the SDQ impact scale, which was also administered to the respondents.

We examined the classification results of these methods by comparing them with the measures of perceived demand for services collected in the YMM data. While similar classifications were achieved with each of the three methods, the two IRT-based approaches provided a higher degree of separation between level of perceived demand for services between mild, moderate and severe disorder when compared with the CGAS-based approach. So, if a higher degree of separation between levels of perceived demand for services with rising levels of severity is seen as a valid criteria for choosing methods, then, of the three methods, the CGAS-based approach performed the least efficiently.

In contrast, both IRT-based assessments provide broadly similar results. There are some distinctive features between these methods. The second method using a single pooled IRT model, assigns more ADHD cases to mild impact compared with the other approaches. This results in a greater differential in level of perceived demand for services between mild and severe ADHD cases.

We concluded that the IRT-based methods of deriving impact on functioning showed superior discrimination in level of demand for services between mild, moderate and severe cases suggesting that these are better methods for deriving impact on functioning. There was very little difference in the results produced by the two different approaches to the IRT analysis. **However, the single pooled IRT analysis suggested that more ADHD cases would be classified as mild compared with the other approaches, resulting in a higher level of discrimination in level of perceived service demand for ADHD. As the single pooled IRT model is conceptually simpler and provides an improved result for ADHD this methodology was adopted for the Young Minds Matter publication.**

With this in mind, children and adolescents were classified into three levels of severity by applying the National Mental Health Service Planning Framework standard ratio of severity for mental disorders to the IRT score (1:2:4 for severe, moderate and mild cases). In addition suicide “plans” or “attempts in the past 12 months” were also included in the severity classification. The three levels are:

- Severe: A positive diagnosis plus an impact score greater than or equal to 1.75 and/or a history of suicide attempt in the 12 months prior to interview;
- Moderate: A positive diagnosis plus an impact score greater than or equal to 0.95 or a history of suicide plans in the 12 months prior to interview; and
- Mild: All other cases with a positive diagnosis.

Table 10 shows the results of applying this severity classification to the distribution of the total impact score and Table 11 shows these distributions stratified by Diagnostic Group.

Table 10. Final severity classification for the pooled impact assessments

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 No impact	2518	50.9	50.9	50.9
	1 Mild	1578	31.9	31.9	82.7
	2 Moderate	549	11.1	11.1	93.8
	3 Severe	305	6.2	6.2	100.0
	Total	4950	100.0	100.0	

Table 11. Final severity classification by Diagnostic Grouping

Module DISC Module administered			Frequency	Percent	Valid Percent	Cumulative Percent
ANXIETY/ DEPRESSION	Valid	1 Mild	603	31.6	61.0	61.0
		2 Moderate	246	12.9	24.9	85.9
		3 Severe	139	7.3	14.1	100.0
		Total	988	51.8	100.0	
		0 No diagnosis	920	48.2		
	Total		1908	100.0		
ADHD	Valid	1 Mild	585	28.9	67.5	67.5
		2 Moderate	188	9.3	21.7	89.2
		3 Severe	94	4.6	10.8	100.0
		Total	867	42.8	100.0	
		0 No diagnosis	1159	57.2		
	Total		2026	100.0		
CONDUCT DISORDER	Valid	1 Mild	390	38.4	67.6	67.6
		2 Moderate	115	11.3	19.9	87.5
		3 Severe	72	7.1	12.5	100.0
		Total	577	56.8	100.0	
		0 No diagnosis	439	43.2		
	Total		1016	100.0		

Table 12 provides the mean impact scores for each of the diagnostic groups stratified by level of severity while Table 13 provides the weighted prevalence of each DISC-IV disorder by final level of severity classification.

Table 12: Mean Total Impact Scores by Diagnostic Group by final severity classification

Module DISC Module administered	Severity category	Mean	N	Std. Deviation	Minimum	Maximum
1 ANXDEP	1 Mild	0.408	603	0.322	-1.066	0.946
	2 Moderate	1.248	246	0.251	0.179	1.747
	3 Severe	1.928	139	0.571	-0.077	3.475
	Total	0.831	988	0.668	-1.066	3.475
2 ADHD	1 Mild	0.372	585	0.328	-1.047	0.945
	2 Moderate	1.267	188	0.288	-0.269	1.747
	3 Severe	1.747	94	0.798	-1.182	3.236
	Total	0.715	867	0.648	-1.182	3.236
3 CONDUCT	1 Mild	0.201	390	0.437	-1.271	0.949
	2 Moderate	1.239	115	0.265	-0.129	1.733
	3 Severe	1.696	72	0.782	-0.714	3.475
	Total	.594	577	0.747	-1.271	3.475

Table 13. Prevalence of DISC disorder by final severity classifications

Disorder	Mild		Moderate		Severe		Total	
	Per cent	95% CI	Per cent	95% CI	Per cent	95% CI	Per cent	95% CI
Depressive disorder	0.6	(0.4 - 0.8)	1.0	(0.7 - 1.3)	1.2	(0.9 - 1.5)	2.8	(2.4 - 3.2)
Conduct disorder	1.2	(0.9 - 1.6)	0.5	(0.3 - 0.7)	0.4	(0.2 - 0.6)	2.1	(1.6 - 2.5)
ADHD	4.9	(4.3 - 5.5)	1.8	(1.4 - 2.2)	0.8	(0.5 - 1.0)	7.4	(6.6 - 8.2)
Any anxiety disorder	3.7	(3.2 - 4.2)	1.9	(1.5 - 2.3)	1.3	(1.0 - 1.6)	6.9	(6.2 - 7.6)
Social phobia	0.9	(0.7 - 1.2)	0.7	(0.5 - 0.9)	0.7	(0.5 - 1.0)	2.3	(1.9 - 2.7)
Separation anxiety disorder	2.3	(1.9 - 2.7)	1.1	(0.8 - 1.4)	0.8	(0.5 - 1.1)	4.3	(3.7 - 4.8)
Generalised anxiety disorder	0.8	(0.6 - 1.1)	0.7	(0.5 - 0.9)	0.7	(0.5 - 0.9)	2.2	(1.8 - 2.6)
Obsessive compulsive disorder	0.3	(0.1 - 0.4)	0.3	(0.1 - 0.4)	0.3	(0.1 - 0.4)	0.8	(0.6 - 1.1)
Any mental disorder	8.3	(7.6 - 9.1)	3.5	(3.0 - 4.1)	2.1	(1.6 - 2.5)	13.9	(12.9 -15.0)

How were the severity classifications applied to the Impact Domain Scores?

Although fitting a single IRT model between the three disorder groups simplifies the comparison between severity scores across disorder groups, the challenge still remains in how to compare severity of impact on functioning in individual domains of functioning. A single IRT model cannot be used for this approach, as different variables have been used to assess impact on functioning in each domain.

The final step was to create cut points in each of the scores for the 4 domains of functioning (school or work/family/friends/self). The IRT process scales the individual domain scores to be on the same metric. If the same cut points for severity in each of the domains were applied as for the total impact scales the result would be that each disorder would have roughly equal impact in each domain of functioning. There is no *a priori* reason to believe that each mental disorder would have equal impact in all domains. In order to determine and build in the different weighting of domains across disorders, an index item with the same 5-point response options (not at all, a little, somewhat, a lot, extremely) from each domain was chosen (Q3 from school; Q12 from friends; Q18 from family and Q20A from self) and responses assessed within each domain. That is, an index item was identified in each domain that has been rated by parents on the same scale. A regression model was fitted to model the domain specific IRT scores by the responses to the index item in order to determine a relativity between the IRT scores in each domain. Through the regression modelling process it was possible to create coefficients that represented the relative value of an item score across domains and these coefficients were used to adjust the cut points in each of the domains.

The existing C-GAS domain scores and the two methods of IRT-derived domain severity scores for anxiety, depression, ADHD, and conduct disorder are presented in the Tables 14-17.

Table 14. Severity Classification Methods for Anxiety by Domain of Impact

A) Level of impact on functioning - CGAS approach	School/work	Friends	Family	Self	Overall severity
N/A	4.7				
None	20.9	20.4	22.4	6.6	
Mild	25.0	33.1	45.8	31.1	52.9
Moderate	29.6	30.0	20.4	42.1	27.6
Severe	19.8	16.5	11.4	20.2	19.5

B) Level of impact on functioning - IRT method (i)	School/work	Friends	Family	Self	Overall severity
N/A	4.7				
None	25.0	28.1	18.7	18.1	
Mild	24.7	26.0	32.8	34.2	55.1
Moderate	18.7	28.0	23.1	25.4	24.9
Severe	26.9	17.9	25.4	22.3	20.0

C) Level of impact on functioning - IRT method (ii) – final pooled result	School/work	Friends	Family	Self	Overall severity
N/A	4.7				
None	20.0	23.5	17.4	9.9	
Mild	31.7	28.2	36.2	35.0	53.8
Moderate	23.9	34.6	27.2	36.7	27.5
Severe	19.7	13.7	19.2	18.4	18.7

Table 15. Severity Classification Methods for Depression by Domain of Impact

A) Level of impact on functioning - CGAS approach	School/work	Friends	Family	Self	Overall severity
Does not go to school or work	4.8				
None	11.0	10.7	10.8	1.2	
Mild	17.7	23.8	43.6	23.7	29.8
Moderate	34.8	39.5	30.9	38.9	26.6
Severe	31.8	26.1	14.8	36.2	43.6

B) Level of impact on functioning - IRT method (i)	School/work	Friends	Family	Self	Overall severity
Does not go to school or work	4.8				
None	10.7	16.3	8.4	5.8	
Mild	14.9	18.9	28.1	23.3	24.6
Moderate	23.0	34.2	27.2	27.3	30.4
Severe	46.6	30.6	36.3	43.6	45.0

C) Level of impact on functioning - IRT method (ii) – final pooled result	School/work	Friends	Family	Self	Overall severity
None	9.6	14.4	8.4	0.2	
Mild	17.3	19.1	30.6	21.9	21.4
Moderate	34.1	43.1	33.5	37.5	35.8
Severe	34.3	23.4	27.4	40.3	42.8
Does not go to school or work	4.8				

Table 16. Severity Classification Methods for ADHD by Domain of Impact

A) Level of impact on functioning - CGAS approach	School/work	Friends	Family	Self	Overall severity
Does not go to school or work	3.1				
None	10.4	41.6	23.8	23.4	
Mild	26.6	26.7	39.5	45.4	60.7
Moderate	40.5	19.9	25.9	26.2	28.2
Severe	19.4	11.8	10.8	5.0	11.1

B) Level of impact on functioning - IRT method (i)	School/work	Friends	Family	Self	Overall severity
Does not go to school or work	2.8				
None	16.4	31.1	19.5	26.6	
Mild	37.8	28.9	36.0	46.3	59.6
Moderate	27.4	29.3	30.8	23.1	29.1
Severe	15.6	10.6	13.8	3.9	11.3

C) Level of impact on functioning - IRT method (ii) – final pooled result	School/work	Friends	Family	Self	Overall severity
None	13.3	40.9	18.1	29.3	
Mild	40.0	24.9	35.8	45.7	65.7
Moderate	31.1	23.6	28.8	21.3	23.8
Severe	12.8	10.6	17.3	3.7	10.5
Does not go to school or work	2.8				

Table 17. Severity Classification Methods for Conduct Disorder by Domain of Impact

A) Level of impact on functioning - CGAS approach	School/work	Friends	Family	Self	Overall severity
Does not go to school or work	5.4				
None	30.6	32.7	13.3	21.1	
Mild	19.1	29.3	38.8	46.3	50.1
Moderate	29.3	21.5	29.3	18.5	30.1
Severe	15.6	16.4	18.7	14.1	19.7

B) Level of impact on functioning - IRT method (i)	School/work	Friends	Family	Self	Overall severity
Does not go to school or work	5.4				
None	39.5	29.2	20.9	36.2	
Mild	15.4	29.3	28.7	29.1	53.7
Moderate	22.3	27.0	34.2	29.7	25.4
Severe	22.8	14.5	16.2	5.1	20.9

C) Level of impact on functioning - IRT method (ii) – final pooled result	School/work	Friends	Family	Self	Overall severity
None	35.6	32.5	4.3	28.3	
Mild	18.1	26.2	30.7	43.6	58.7
Moderate	25.2	31.4	35.5	24.6	22.4
Severe	15.8	10.0	29.5	3.6	18.9
Does not go to school or work	5.4				

Limitations

At the outset there are limitations with the design and method used to produce these findings.

It was not possible to undertake the needed scientific development to produce an approach to measuring mental health severity in an epidemiological survey. Such scientific studies still remain elusive and in high need. Methods for measuring the severity of a mental disorder have remained a long-outstanding need in both the clinical and applied (epidemiological) settings. The findings here reveal just how complex this need is and demonstrate that more and better research is still required.

The 17-items used here have expanded the original six items used in the DISC-IV implementation of the DSM-IV taxonomy. Face validity and clinical practice have largely guided the selection of the items – neither are robust defences for implementing better science to develop valid, reliable and calibrated measures. There have been no opportunities to undertake qualitative development of a wider pool of measures, using clinicians, parents and patients. While cognitive testing was conducted with clinical and non-clinical cases, this was not conducted within a laboratory setting. Measuring mental health severity is a critical goal worthy of this type of undertaking.

With the in mind, the current analysis suggests that the 17 items are broadly unidimensional – although more could be done to produce greater unidimensionality in the item set. Certainly the CFA analyses using an appropriate weight matrix to adjust for the non-normal, ordinal nature of the item set, produced confirmations of unidimensional fit. The violation of local independence is certainly evident to some degree. Subject's responses to items showed tendencies to depend upon their responses to other items in the item set – a certain feature of the phenomenology of the severity of a mental disorder. The best defence against violations of local independence is, chiefly, preliminary item development trials that would have allowed the minimization of local dependence. These were not within the reach of the YMM survey process. One method of addressing local dependence when it does occur is to combine items into "testlets" post-hoc[16]. Unfortunately, this is best undertaken with binary items. The Likert-scaling of the existing 17 items did not facilitate this.

Individual item model fit was deemed to be reasonable as was global fit to the GRM. Inspection of the item characteristic curves, item information curves, and total test curve provides a high degree of detail on how survey participants are responding to the items (Appendix A). Many five-category items are actually effectively binary in their capture of the "no/mild severity" and "severe" end of the impact score distribution but otherwise carry no information in the mid-range of the scale: Days out of school (V2) and difficulties that cause problems with making or maintaining friendships (V8) were clear examples. Other items showed clear and progressive increments in impact across the range of severity: How much distress is caused to the parent/carer or other family members (V13) and how much the problems prevented the child from concentrating (V16).

The calculation of the continuous impact scores appears to have produced a distribution of respondents along the continuum of severity. However, neither an independently collected clinical sample nor a gold standard reference tool was available to use for benchmarking and setting categories of severity. This necessitated adopting a pragmatic approach to impose category boundaries. Our selection of the National Mental Health Service Planning Framework standard ratio of severity for mental disorders is the method we chose. Clearly though, benchmarking severity categories against *existing* service-delivery expectations inverts the causal logic of first determining

the true proportions of the population that are functionally impaired and *then* fitting the service delivery configuration to the phenomenology, although the CGAS-based approach produced very similar proportions of mild, moderate and severe cases. Until the primary scientific work is undertaken, this is the best that time and funding permitted.

Conclusions

This is the first time severity of child and adolescent disorders has been assessed in a population-wide survey. The strong empirical association between levels of severity and use of services suggests that the severity scoring process produces a useful relative metric even if there is some uncertainty about exact cut-points in the absence of a clinical validation. This is encouraging both in the sense of the interpretation of the YMM results and in advancing the feasibility of the development of better metrics. The final distributions of the continuous total impact scores, and diagnostic scores, along with their severity classifications suggest that there are true differences in the functional impact that is produced by mental disorders in very young and older children as well as young people. The results here are encouraging enough to motivate better scientific study of the nature of severity and certainly informative, if not fully representative, of the nature of the distress and burden that these disorders impose on families, schools, and, most particularly the children and young people themselves.

APPENDIX A

For each Impact Domain the following is provided:

1. A summary of the confirmatory factor analysis along with the item path diagram;
2. A scree plot from the principal components analysis for each Domain Impact Score
3. The results of the Graded Response Model from the IRT analysis
4. At the end of each of the three diagnostic groupings, a frequency diagram of the Total Impact Score (this summarises all 17 items) for each of the Diagnostic Groupings.

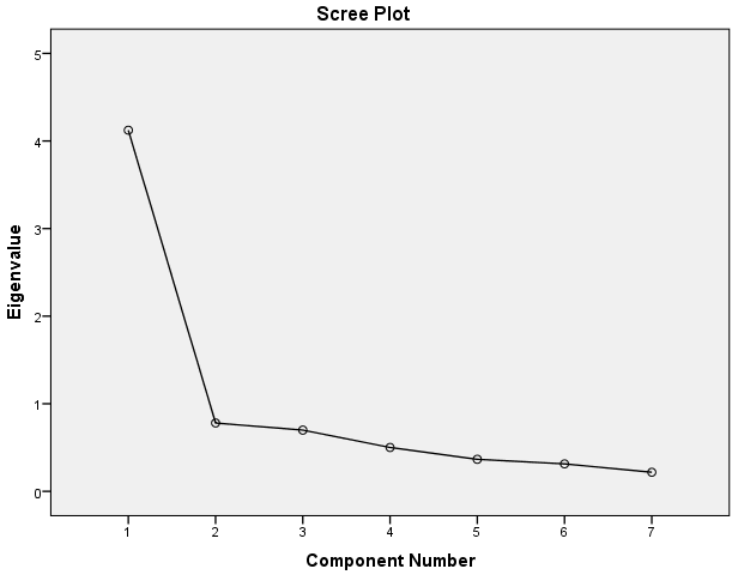
The selection of appropriate fit indices for SEM has been extensively reviewed notably by Hu and Bentler (1995; 1998; 1999) and Yuan and Bentler (1997). A 'combinational' rule, in which two or possibly three fit indices are used to judge model fit, is recommended. The selection of the recommended fit indices is reliant upon sample size, distributional characteristics of the data, and model complexity. Hu and Bentler suggest the use of the Standardized Root Mean Residual (SRMR) supplemented with one of either the NNFI, (Non-Normal Fit Index; also called the TLI, Tucker-Lewis Index) or the Comparative Fit Index (CFI) (Hu & Bentler, 1998). These have been used here.

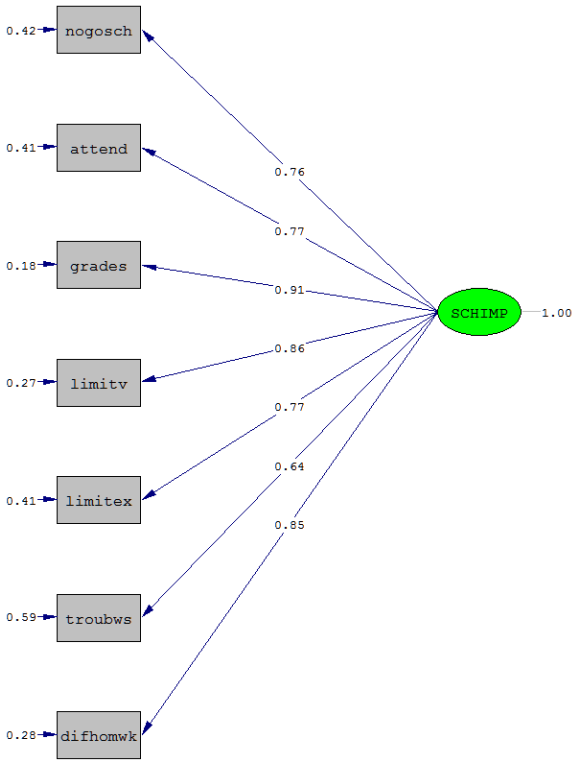
Confirmatory Factor Analyses, Scree Plots, Path Diagrams and Graded Response Models for the Impact Assessment Items: By Diagnostic group by Domain

Impact on Schooling **ANXIETY/DEPRESSION**	**Item loadings[a]** **λ_x**	**Regression weights[b]**	**Model characteristics[c]**
asv001 In the last 12 months, when these problems were at their worst, how often did child not want to go to school	0.760	0.113 *0.100*	N = 1651 df = 14 χ^2 = 382.2[d] **SRMR = 0.06** **NNFI = 0.96** **CFI = 0.97** H = 0.94 Acceptable
ASV002r As a result of these problems, how many days has [child] been absent from school in the last 12 months?	0.768	0.118 *0.104*	
asv003r In the last 12 months, when these problems were at their worst, did they affect child's grades or ability to do work?	0.907	0.322 *0.285*	
asv004 When these problems were at their worst, did these difficulties limit child in participating in volunteer activities?	0.856	0.202 *0.179*	
asv005r When these problems were at their worst, did these difficulties limit child in participating in extracurricular activities?	0.769	0.118 *0.104*	
asv006r When these problems were at their worst, did these difficulties cause trouble at school (or at work)?	0.639	0.068 *0.060*	
asv007 When these problems were at their worst, how often did child have difficulties completing school work?	0.848	0.190 *0.168*	

a. Partial regression coefficients of the item on the underlying construct. b. Upper figures are raw factor score indices and lower figures are proportionally adjusted factor score regression indices. c. Models were fitted via Robust Weighted Least Squares using polychoric correlations and their asymptotic covariance matrix via LISREL 9.1 (SSI Inc., 2007). d. Satorrra-Bentler Adjusted Chi-Square

Impact on Schooling
ANXIETY/DEPRESSION





Impact on Schooling
ANXIETY/DEPRESSION

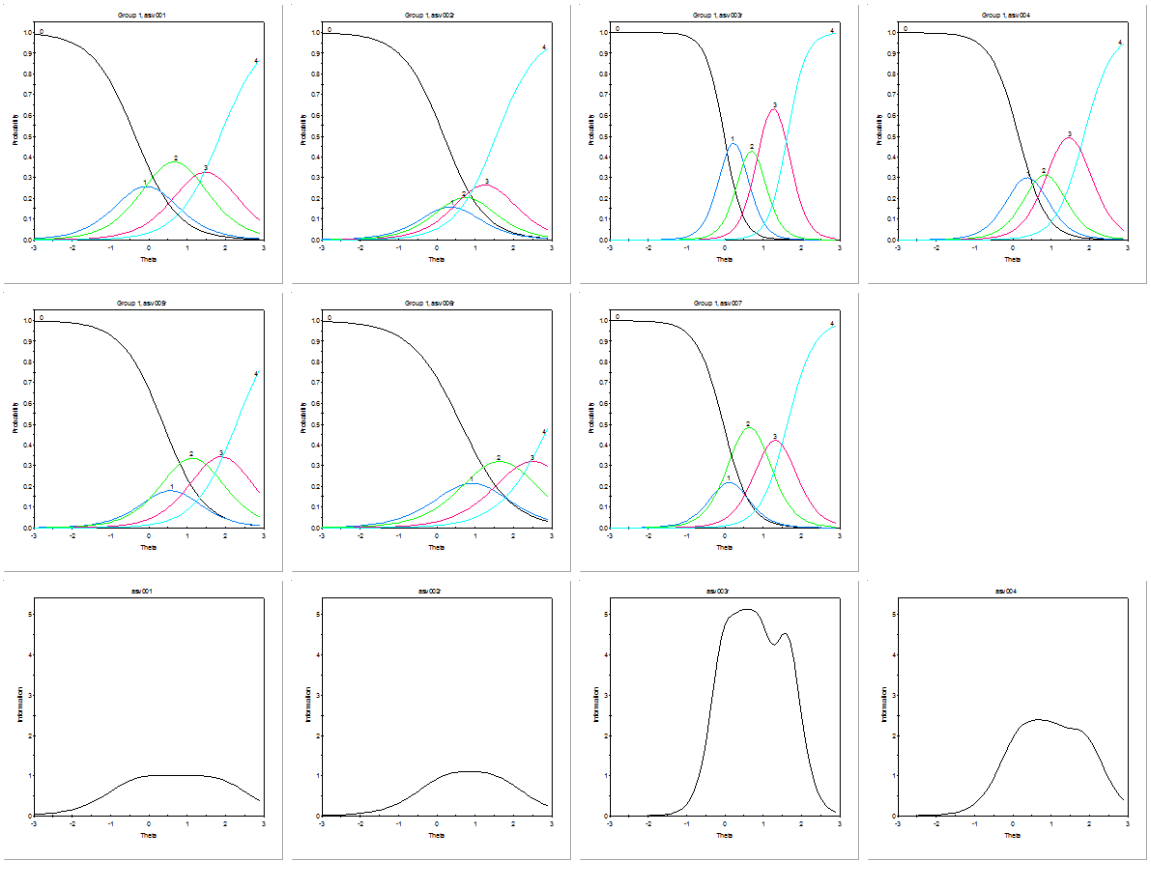
Graded Model Item Parameter Estimates for Group 1, logit: $a(\theta - b)$ (Back to TOC)

Item	Label		a	s.e.	b_1	s.e.	b_2	s.e.	b_3	s.e.	b_4	s.e.
1	asv001	5	1.80	0.08	-0.36	0.04	0.23	0.04	1.11	0.05	1.86	0.08
2	asv002r	10	1.86	0.09	0.18	0.04	0.52	0.04	0.97	0.05	1.55	0.07
3	asv003r	15	4.16	0.23	-0.02	0.03	0.47	0.03	0.91	0.04	1.63	0.05
4	asv004	20	2.76	0.13	0.15	0.03	0.60	0.04	1.07	0.04	1.85	0.06
5	asv005r	25	1.87	0.09	0.37	0.04	0.76	0.04	1.51	0.06	2.28	0.09
6	asv006r	30	1.52	0.08	0.63	0.05	1.20	0.06	2.07	0.09	2.95	0.14
7	asv007	35	2.88	0.14	-0.04	0.03	0.27	0.03	1.00	0.04	1.63	0.06

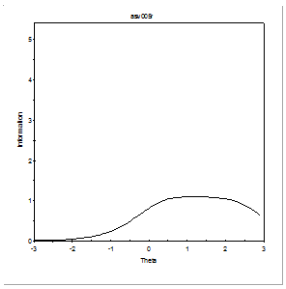
M^2 = 1824.06; df = 329; p = 0.0001; RMSEA = 0.03

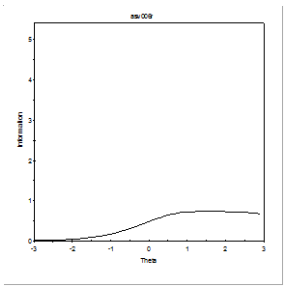
S-X^2 Item Level Diagnostic Statistics

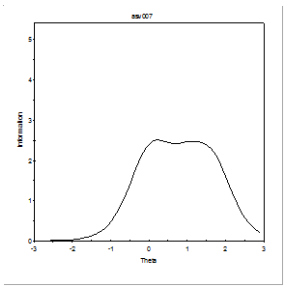
Item	Label	X^2	*d.f.*	Probability
1	asv001	112.40	83	0.0175
2	asv002r	107.48	83	0.0367
3	asv003r	102.12	66	0.0029
4	asv004	83.96	77	0.2743
5	asv005r	86.50	87	0.4958
6	asv006r	172.49	91	0.0001
7	asv007	74.21	72	0.4055

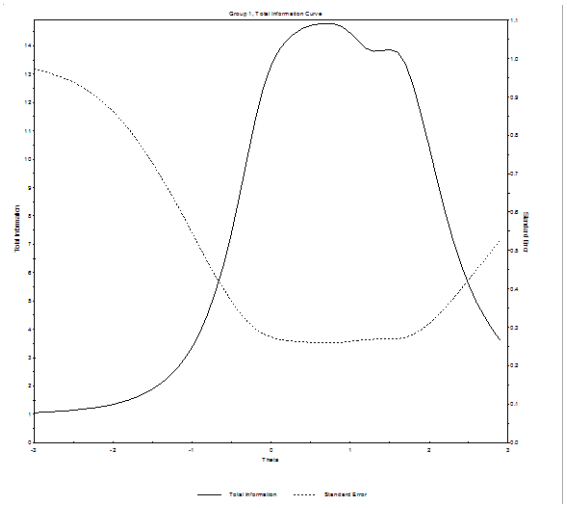


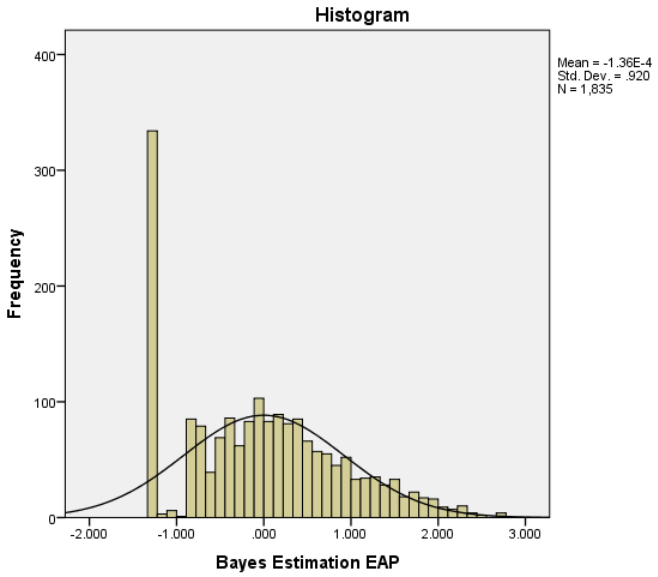
Impact on Schooling
ANXIETY/DEPRESSION





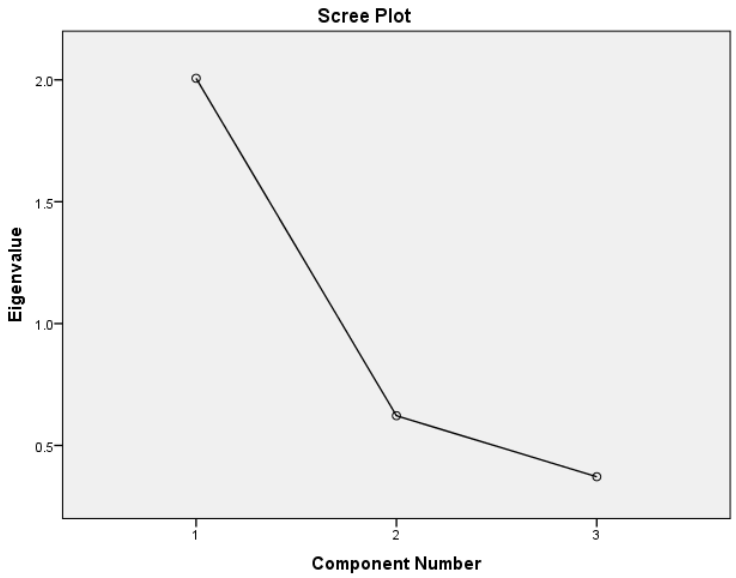






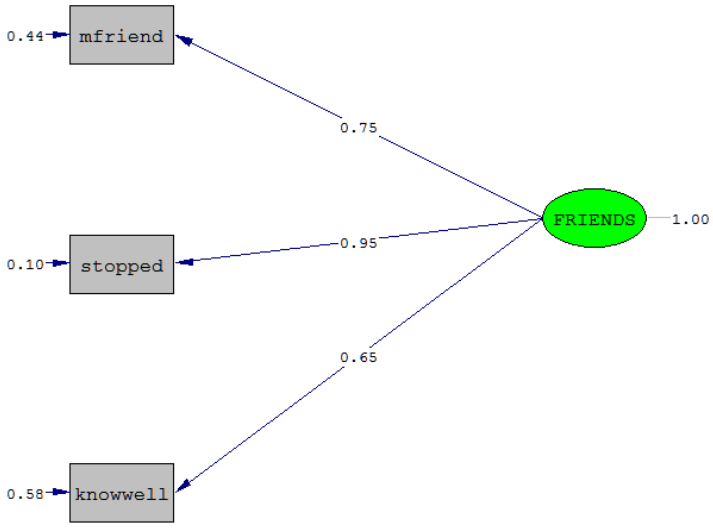
Impact on Friends ANXIETY/DEPRESSION	**Item loadings[a] λ_x**	**Regression weights[b]**	**Model characteristics[c]**
ASV012r In the last 12 months, when these problems were at their worst, did these difficulties cause problems with [child's] ability to make or maintain friendships?	0.747	0.135 *0.132*	N = 1862 df = 0 χ^2 = NA
ASV014 How often have these difficulties stopped [child] from doing things or going places with other children [his/her] age?	0.951	0.799 *0.780*	**SRMR = NA** **NNFI = NA** **CFI = NA**
ASV015 When these problems were at their worst how much difficulty did [child] have dealing with people [he/she] didn't know well?	0.650	0.090 *0.088*	H = 0.92 Saturated model

a. Partial regression coefficients of the item on the underlying construct. b. Upper figures are raw factor score indices and lower figures are proportionally adjusted factor score regression indices. c. Models were fitted via Robust Weighted Least Squares using polychoric correlations and their asymptotic covariance matrix via LISREL 9.1 (SSI Inc., 2007). d. Satorrra-Bentler Adjusted Chi-Square



Impact on Friends

ANXIETY/DEPRESSION



Graded Model Item Parameter Estimates for Group 1, logit: $a(\theta - b)$ (Back to TOC)

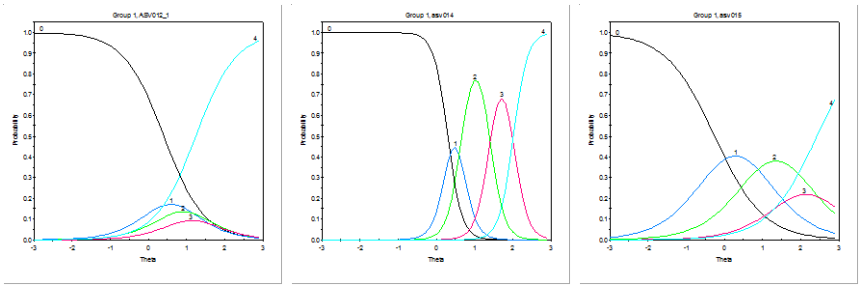
Item	Label		*a*	*s.e.*	b_1	*s.e.*	b_2	*s.e.*	b_3	*s.e.*	b_4	*s.e.*
1	ASV012_1	5	1.91	0.11	0.41	0.04	0.77	0.05	1.05	0.05	1.24	0.06
2	asv014	10	5.44	0.37	0.30	0.03	0.65	0.03	1.41	0.04	2.02	0.07
3	asv015	15	1.56	0.08	-0.26	0.04	0.84	0.05	1.86	0.09	2.43	0.11

G^2 = 187.71 df = 109 p = 0.0001 RMSEA =0.01

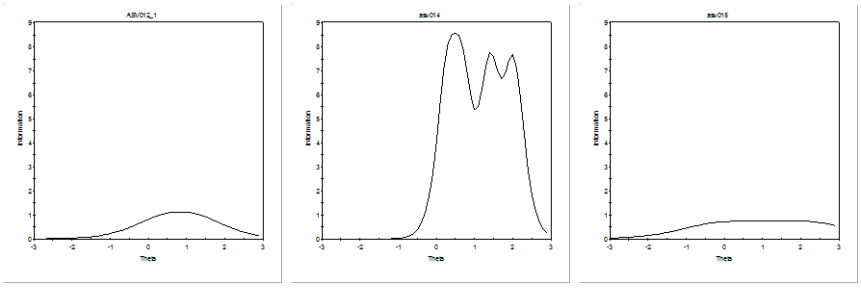
Summed-Score Based Item Diagnostic Tables and X^2s for Group 1 (Back to TOC)

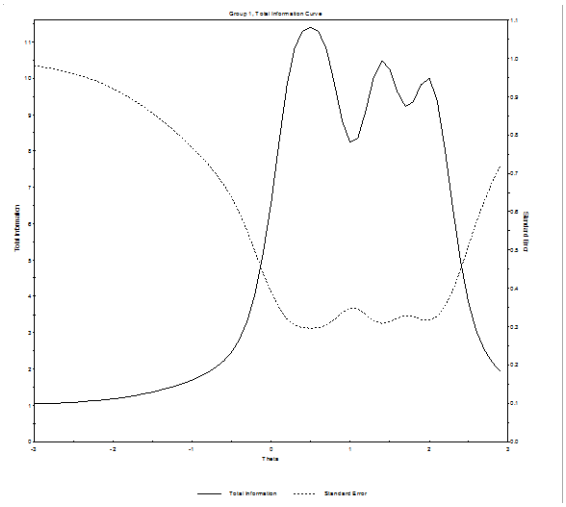
S-X^2 Item Level Diagnostic Statistics

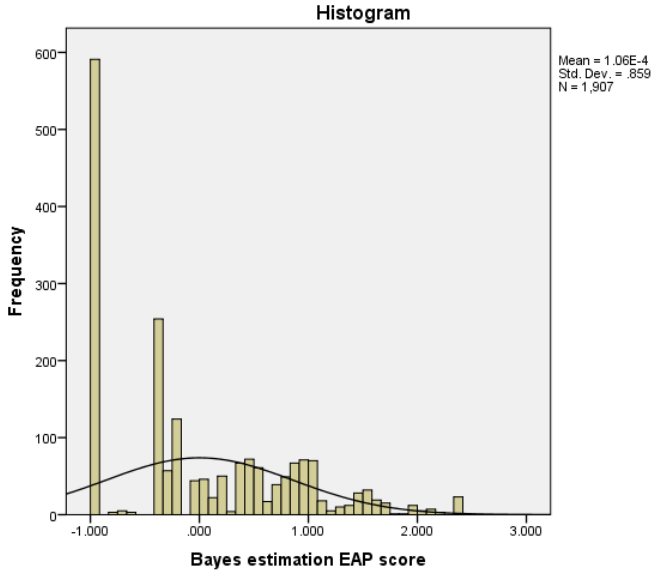
Item	Label	X^2	*d.f.*	Probability
1	ASV012_1	57.81	29	0.0011
2	asv014	49.43	29	0.0104
3	asv015	68.19	30	0.0001



Impact on Friends
ANXIETY/DEPRESSION

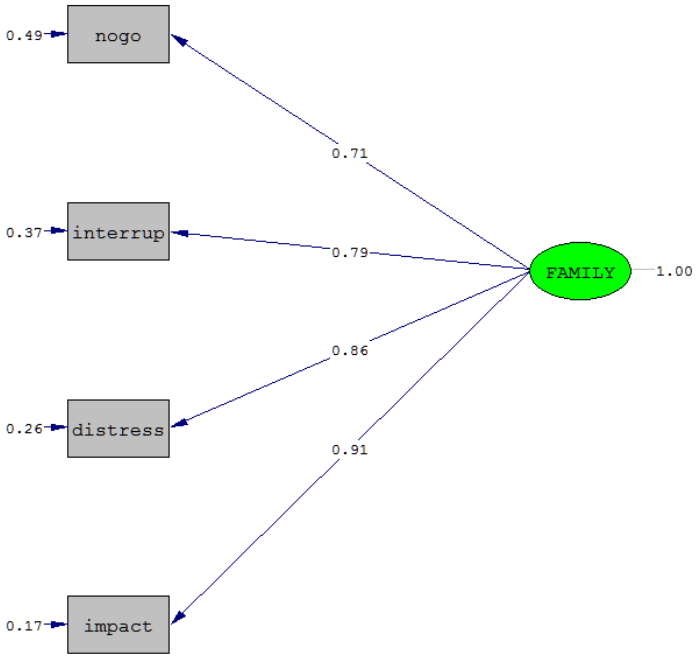






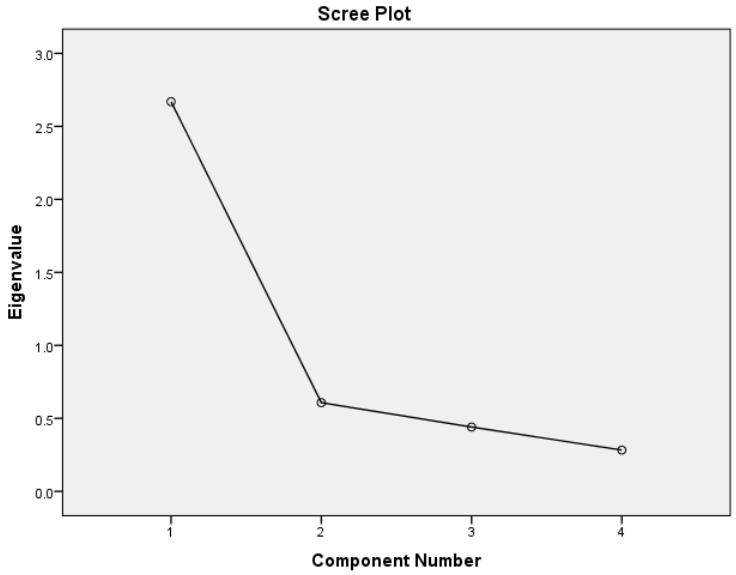
Impact on Family ANXIETY/DEPRESSION	Item loadings[a] λ_x	Regression weights[b]	Model characteristics[c]
ASV016 In the last 12 months how often have [child's] difficulties prevented you from taking [him/her] places or going out in public?	0.715	0.128 *0.120*	N = 1903 df = 2 $\chi^2 = 55.48$[d] **SRMR = 0.04** **NNFI = 0.97** **CFI = 0.99** H = 0.91 Acceptable
ASV017 How often have [child's] difficulties interrupted everyday family activities such as eating meals or watching TV?	0.792	0.186 *0.174*	
ASV018 How much distress do [child's] difficulties cause you and other members of the family?	0.861	0.291 *0.272*	
ASV019 How much do [child's] difficulties impact on your other family and household responsibilities, such as time to spend with other children or family members?	0.910	0.464 *0.434*	

a. Partial regression coefficients of the item on the underlying construct. b. Upper figures are raw factor score indices and lower figures are proportionally adjusted factor score regression indices. c. Models were fitted via Robust Weighted Least Squares using polychoric correlations and their asymptotic covariance matrix via LISREL 9.1 (SSI Inc., 2007). d. Satorrra-Bentler Adjusted Chi-Square



Impact on Family

ANXIETY/DEPRESSION



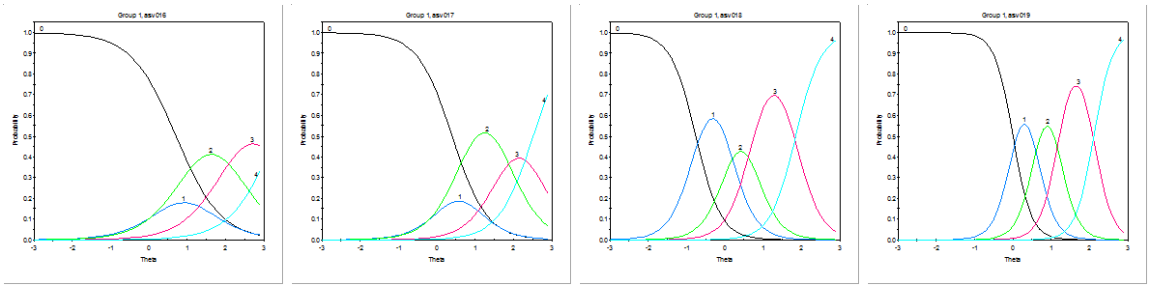
Graded Model Item Parameter Estimates for Group 1, logit: *a*(*θ* - *b*)

Item	Label		*a*	*s.e.*	b_1	*s.e.*	b_2	*s.e.*	b_3	*s.e.*	b_4	*s.e.*
1	asv016	5	1.72	0.09	0.71	0.04	1.13	0.05	2.14	0.09	3.31	0.17
2	asv017	10	2.22	0.11	0.40	0.04	0.74	0.04	1.77	0.07	2.52	0.10
3	asv018	15	3.05	0.15	-0.76	0.04	0.12	0.03	0.72	0.04	1.85	0.06
4	asv019	20	4.19	0.29	0.00	0.03	0.60	0.03	1.19	0.04	2.11	0.07

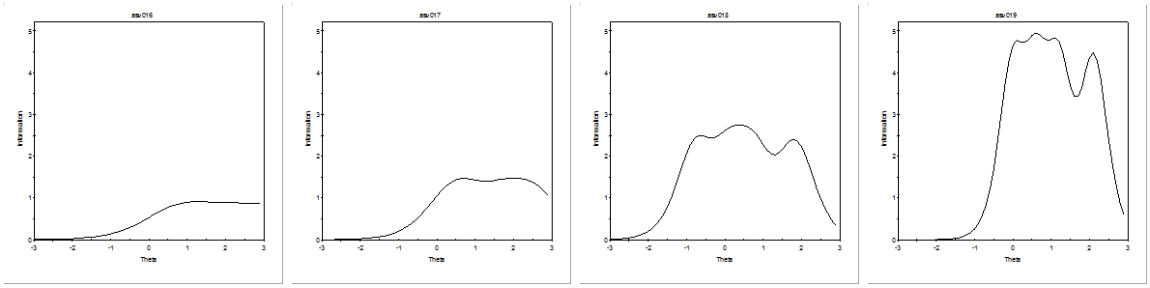
M_2 = 1197.04; df = 92; Probability = 0.0001; RMSEA = 0.04

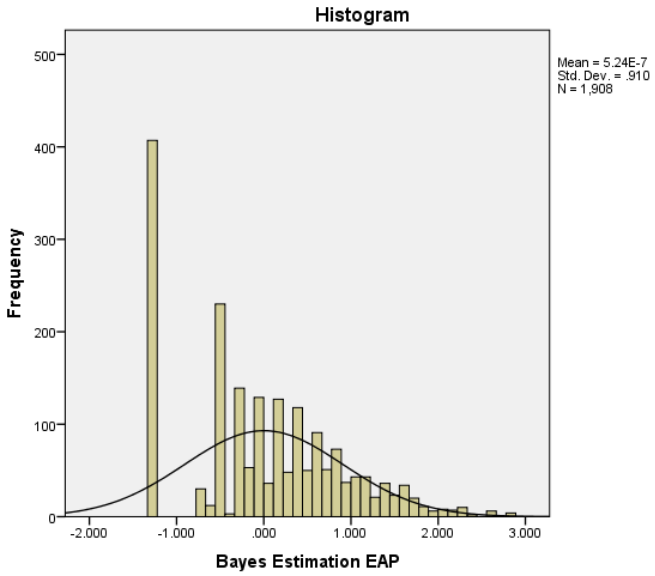
S-X^2 Item Level Diagnostic Statistics

Item	Label	X^2	*d.f.*	Probability
1	asv016	86.42	42	0.0001
2	asv017	61.83	39	0.0114
3	asv018	80.31	32	0.0001
4	asv019	81.51	32	0.0001



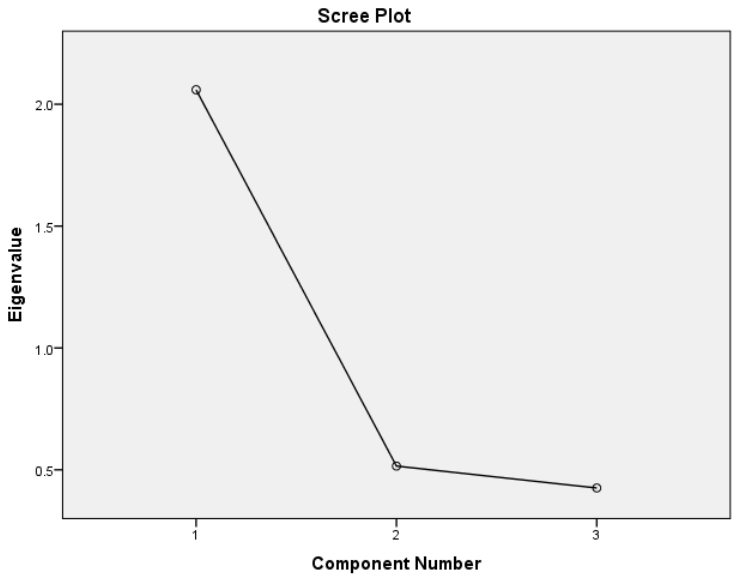
Impact on Family
ANXIETY/DEPRESSION





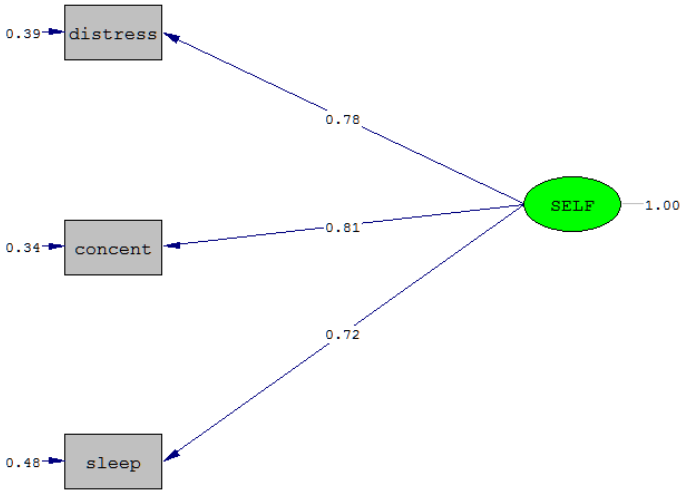
Impact on Self ANXIETY/DEPRESSION	Item loadings[a] λ_x	Regression weights[b]	Model characteristics[c]
ASV020r In the last 12 months, when these problems were at their worst, did these difficulties distress [child] or make [him/her] feel bad or upset? How distressed?	0.779	0.357 *0.338*	N = 1849 df = 0 χ^2 = na
ASV021 When these problems were at their worst, how much did these difficulties prevent [child] from concentrating on things [he/she] was supposed to be doing?	0.811	0.426 *0.404*	**SRMR = na** **NNFI = na** **CFI = na**
ASV022 When these problems were at their worst, how much did these difficulties impact on [child's] sleeping?	0.723	0.272 *0.258*	H = 0.82 Saturated model

a. Partial regression coefficients of the item on the underlying construct. b. Upper figures are raw factor score indices and lower figures are proportionally adjusted factor score regression indices. c. Models were fitted via Robust Weighted Least Squares using polychoric correlations and their asymptotic covariance matrix via LISREL 9.1 (SSI Inc., 2007). d. Satorrra-Bentler Adjusted Chi-Square



Impact on Self

ANXIETY/DEPRESSION



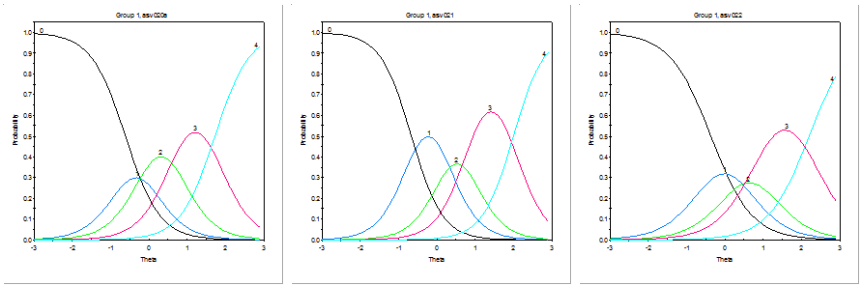
Graded Model Item Parameter Estimates for Group 1, logit: *a*(*θ* - *b*) (Back to TOC)

Item	Label		*a*	*s.e.*	b_1	*s.e.*	b_2	*s.e.*	b_3	*s.e.*	b_4	*s.e.*
1	asv020a	5	2.24	0.13	-0.61	0.04	-0.06	0.03	0.70	0.04	1.72	0.07
2	asv021	10	2.48	0.15	-0.67	0.04	0.21	0.03	0.83	0.04	2.00	0.08
3	asv022	15	1.87	0.10	-0.37	0.04	0.34	0.04	0.94	0.05	2.20	0.09

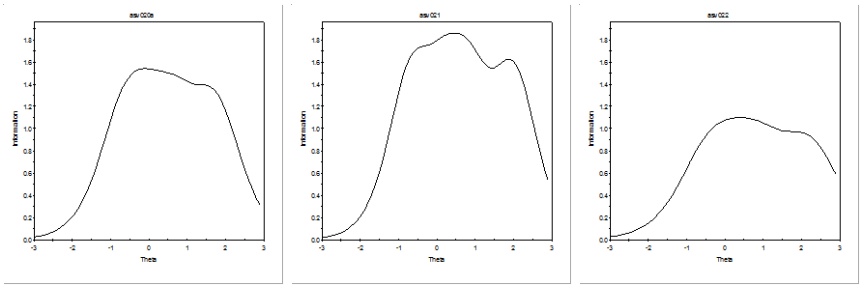
G^2 = 227.68; df = 109; p = 0.0001; RMSEA = 0.01

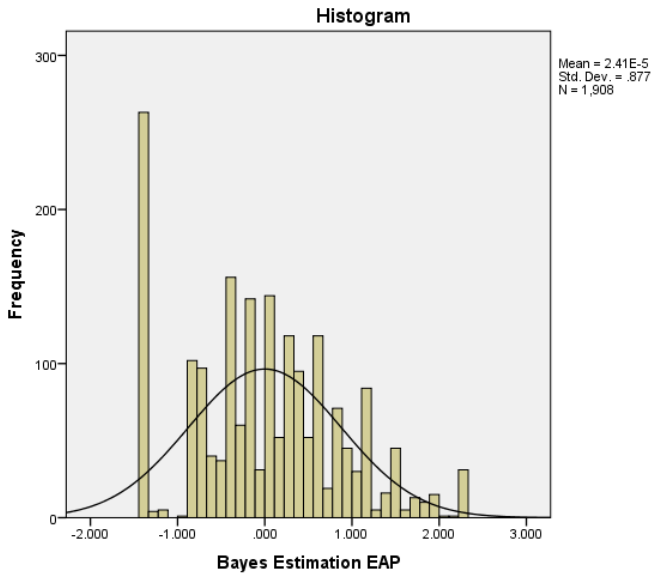
S-X^2 Item Level Diagnostic Statistics

Item	Label	X^2	*d.f.*	Probability
1	asv020a	93.46	29	0.0001
2	asv021	58.12	29	0.0011
3	asv022	47.59	30	0.0217



Impact on Self
ANXIETY/DEPRESSION

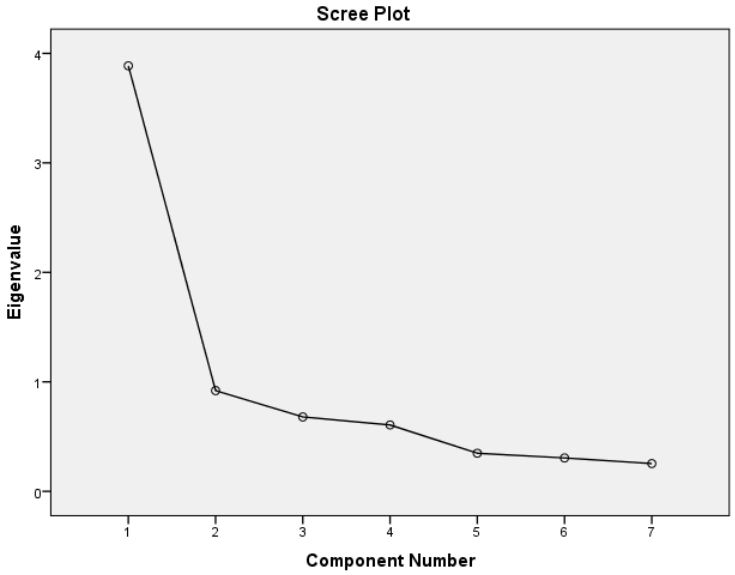


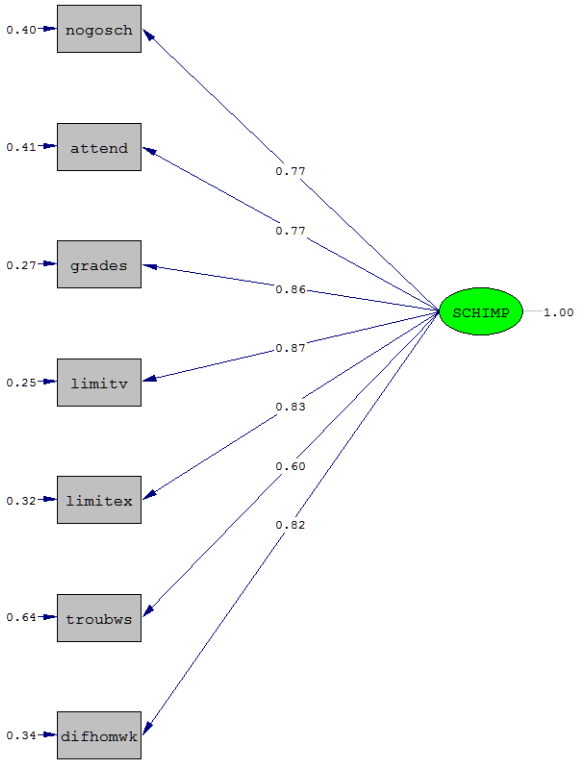


Impact on Schooling ADHD	Item loadings[a] λ_x	Regression weights[b]	Model characteristics[c]
hsv001 In the last 12 months, when these problems were at their worst, how often did child not want to go to school	0.772	0.133 *0.116*	
hsv002 As a result of these problems, how many days has [child] been absent from school in the last 12 months?	0.769	0.131 *0.114*	
hsv003 In the last 12 months, when these problems were at their worst, did they affect child's grades or ability to do work?	0.856	0.222 *0.194*	N = 1762 df = 14 χ^2 = 464.90[d]
hsv004 When these problems were at their worst, did these difficulties limit child in participating in volunteer activities?	0.868	0.244 *0.213*	**SRMR = 0.07** **NNFI = 0.95** **CFI = 0.97**
hsv005 When these problems were at their worst, did these difficulties limit child in participating in extracurricular activities?	0.825	0.180 *0.157*	H = 0.93 Acceptable
asv006 When these problems were at their worst, did these difficulties cause trouble at school (or at work)?	0.602	0.066 *0.058*	
hsv007 When these problems were at their worst, how often did child have difficulties completing school work?	0.815	0.169 *0.148*	

a. Partial regression coefficients of the item on the underlying construct. b. Upper figures are raw factor score indices and lower figures are proportionally adjusted factor score regression indices. c. Models were fitted via Robust Weighted Least Squares using polychoric correlations and their asymptotic covariance matrix via LISREL 9.1 (SSI Inc., 2007). d. Satorrra-Bentler Adjusted Chi-Square

Impact on Schooling
ADHD





Impact on Schooling

ADHD

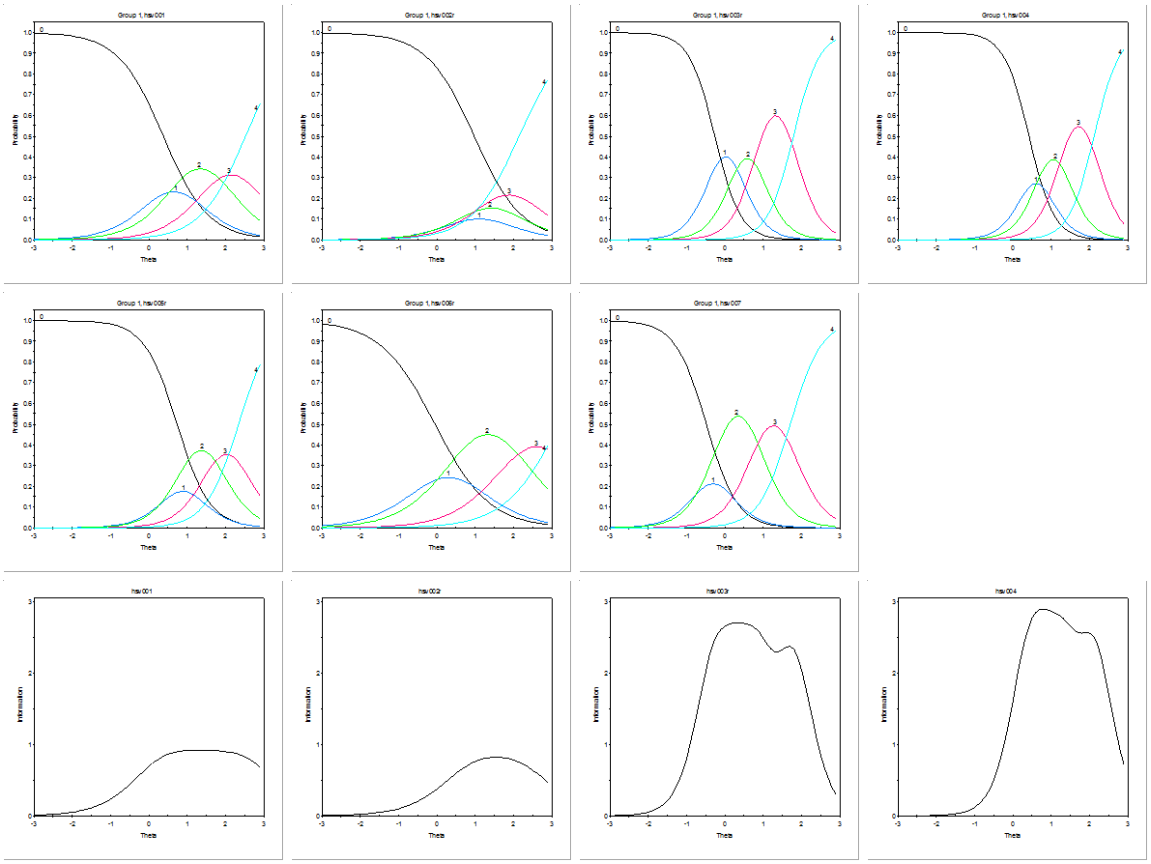
Graded Model Item Parameter Estimates for Group 1, logit: $a(\theta - b)$ (Back to TOC)

Item	Label		*a*	*s.e.*	b_1	*s.e.*	b_2	*s.e.*	b_3	*s.e.*	b_4	*s.e.*
1	hsv001	5	1.71	0.09	0.37	0.04	0.93	0.05	1.76	0.07	2.52	0.11
2	hsv002r	10	1.62	0.09	0.97	0.05	1.22	0.06	1.60	0.07	2.15	0.10
3	hsv003r	15	2.98	0.14	-0.26	0.03	0.31	0.03	0.86	0.04	1.79	0.06
4	hsv004	20	3.06	0.17	0.42	0.03	0.78	0.04	1.32	0.05	2.12	0.07
5	hsv005r	25	2.34	0.13	0.73	0.04	1.04	0.04	1.71	0.06	2.34	0.09
6	hsv006r	30	1.41	0.07	-0.05	0.04	0.65	0.05	2.02	0.09	3.20	0.16
7	hsv007	35	2.47	0.11	-0.49	0.04	-0.14	0.03	0.83	0.04	1.71	0.06

M2 = 2232.21; df = 329; p = 0.0001; RMSEA = 0.03

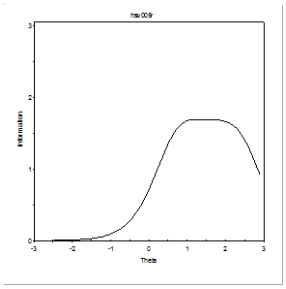
S-X^2 Item Level Diagnostic Statistics

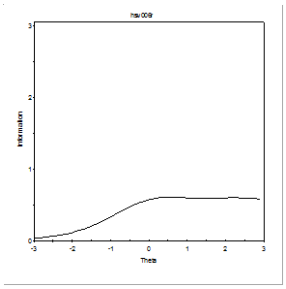
Item	Label	X^2	*d.f.*	Probability
1	hsv001	102.32	85	0.0971
2	hsv002r	85.12	84	0.4463
3	hsv003r	109.67	66	0.0006
4	hsv004	117.06	69	0.0003
5	hsv005r	104.32	78	0.0249
6	hsv006r	201.80	86	0.0001
7	hsv007	95.44	67	0.0128

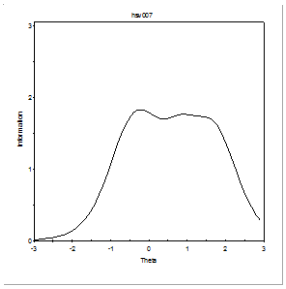


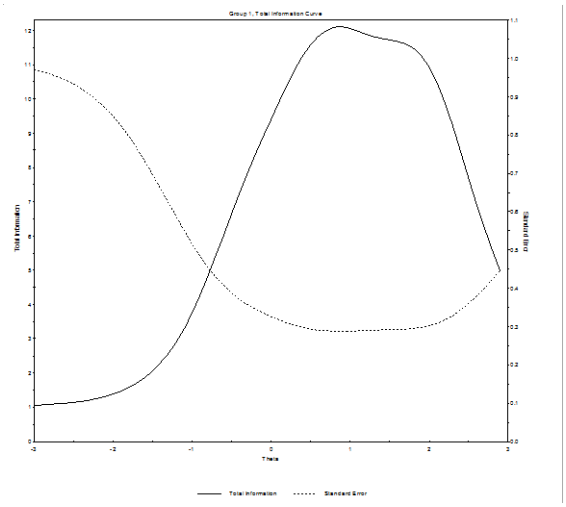
Impact on Schooling

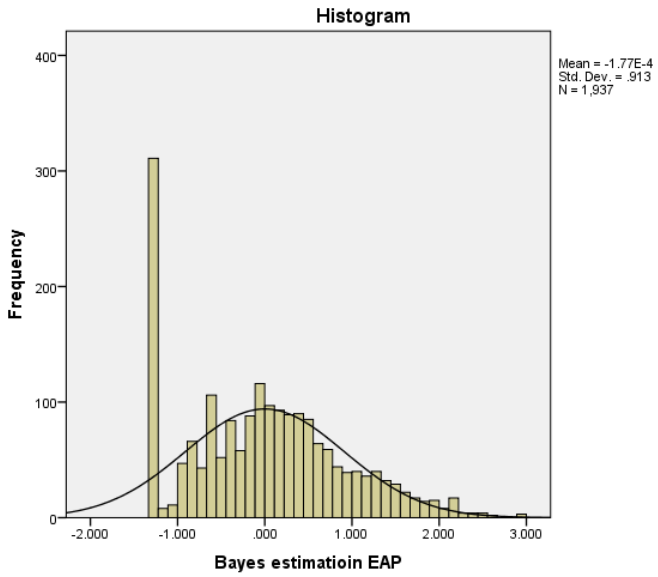
ADHD





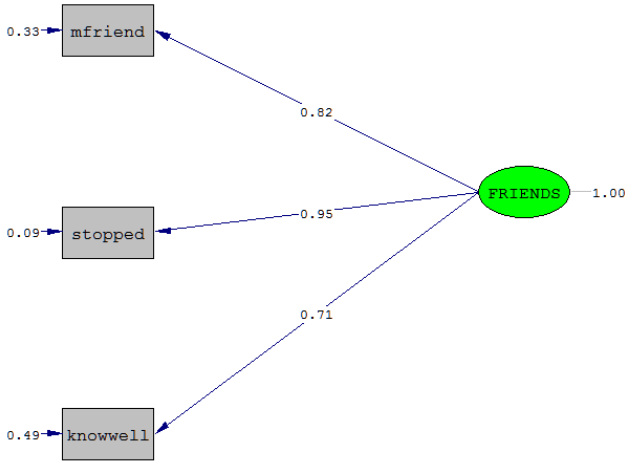






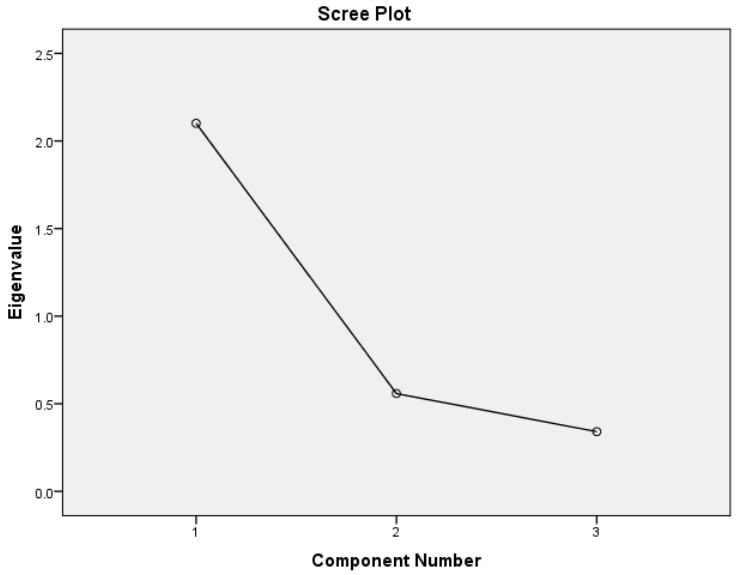
Impact on Friends ADHD	Item loadings[a] λ_x	Regression weights[b]	Model characteristics[c]
HSV012 In the last 12 months, when these problems were at their worst, did these difficulties cause problems with [child's] ability to make or maintain friendships?	0.822	0.175 *0.171*	N = 1975 df = 0 χ^2 = na
HSV014 How often have these difficulties stopped [child] from doing things or going places with other children [his/her] age?	0.955	0.749 *0.731*	**SRMR = na** **NNFI = na** **CFI = na**
HSV015 When these problems were at their worst how much difficulty did [child] have dealing with people [he/she] didn't know well?	0.714	0.101 *0.099*	H = 0.94 Saturated model

a. Partial regression coefficients of the item on the underlying construct. b. Upper figures are raw factor score indices and lower figures are proportionally adjusted factor score regression indices. c. Models were fitted via Robust Weighted Least Squares using polychoric correlations and their asymptotic covariance matrix via LISREL 9.1 (SSI Inc., 2007). d. Satorrra-Bentler Adjusted Chi-Square



Impact on Friends

ADHD



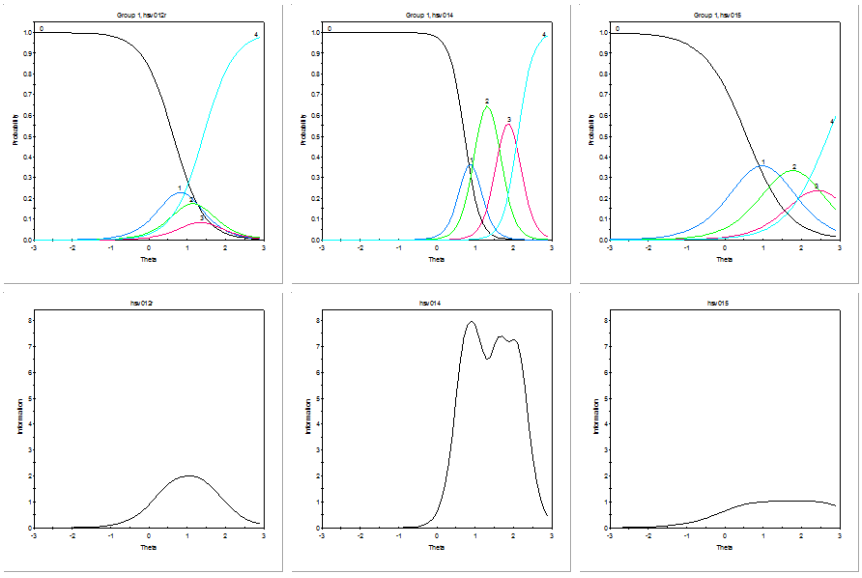
Graded Model Item Parameter Estimates for Group 1, logit: *a*(*θ* - *b*) (Back to TOC)

Item	Label		*a*	*s.e.*	b_1	*s.e.*	b_2	*s.e.*	b_3	*s.e.*	b_4	*s.e.*
1	hsv012r	5	2.52	0.17	0.64	0.04	1.01	0.05	1.29	0.05	1.42	0.06
2	hsv014	10	5.18	0.62	0.72	0.03	1.02	0.04	1.62	0.05	2.11	0.07
3	hsv015	15	1.81	0.10	0.57	0.04	1.39	0.06	2.15	0.09	2.69	0.13

G^2 = 216.71; df = 216.7; p= .0001; RSMEA = 0.01

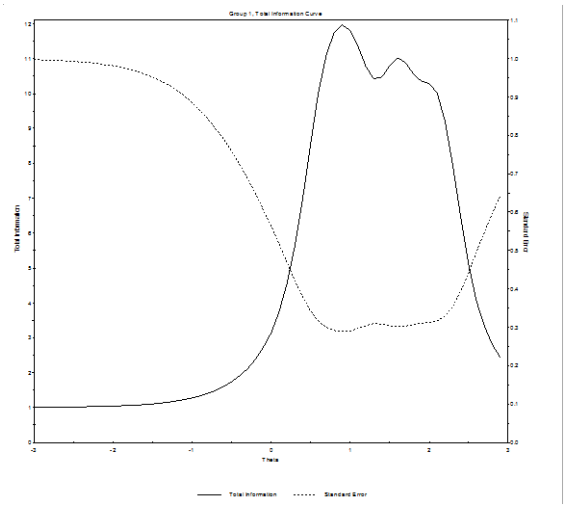
S-X^2 Item Level Diagnostic Statistics

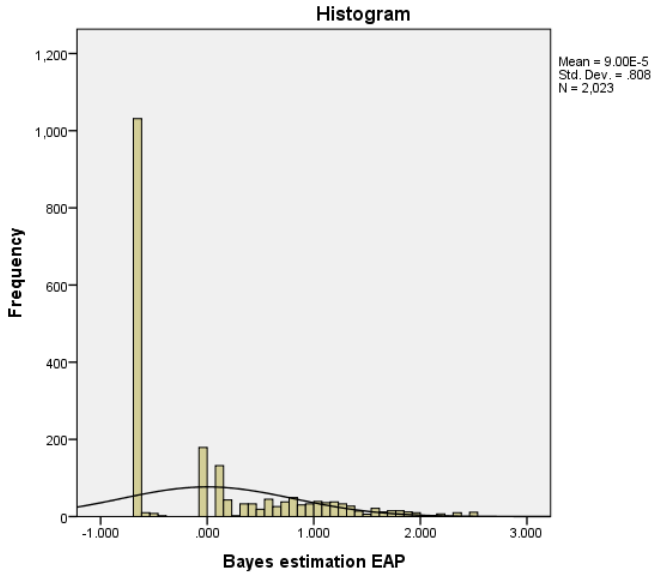
Item	Label	X^2	*d.f.*	Probability
1	hsv012r	80.37	27	0.0001
2	hsv014	67.62	27	0.0001
3	hsv015	73.47	31	0.0001



Impact on Friends

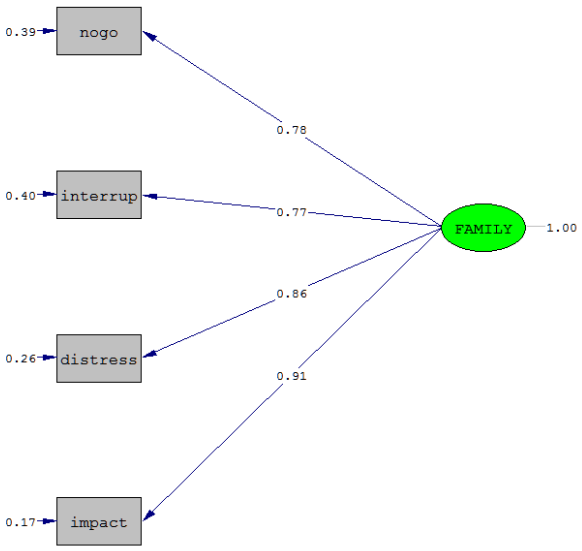
ADHD





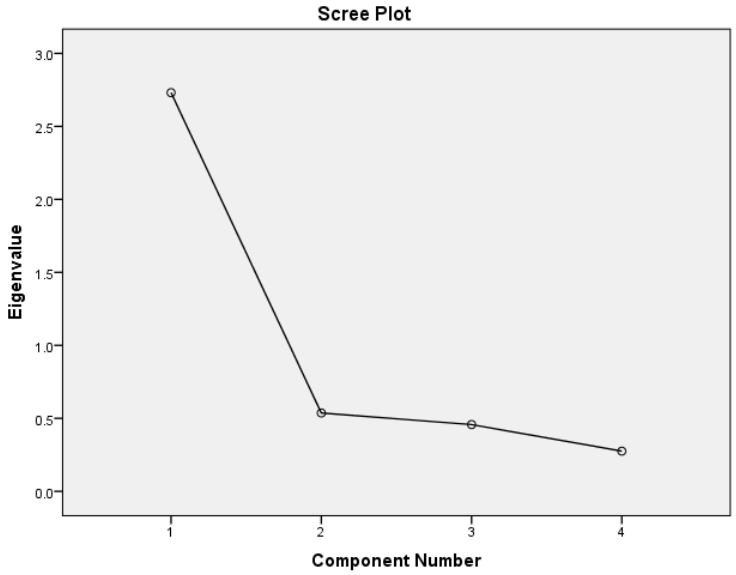
Impact on Family ADHD	Item loadings[a] λ_x	Regression weights[b]	Model characteristics[c]
hSV016 In the last 12 months how often have [child's] difficulties prevented you from taking [him/her] places or going out in public?	0.782	0.172 *0.508*	N = 2025 df = 2 χ^2 = 44.5[d] **SRMR = 0.03** **NNFI = 0.99** **CFI = 0.99** H = 0.91 Acceptable
hSV017 How often have [child's] difficulties interrupted everyday family activities such as eating meals or watching TV?	0.775	0.166 *0.211*	
hSV018 How much distress do [child's] difficulties cause you and other members of the family?	0.857	0.278 *0.199*	
hSV019 How much do [child's] difficulties impact on your other family and household responsibilities, such as time to spend with other children or family members?	0.910	0.454 *0.082*	

a. Partial regression coefficients of the item on the underlying construct. b. Upper figures are raw factor score indices and lower figures are proportionally adjusted factor score regression indices. c. Models were fitted via Robust Weighted Least Squares using polychoric correlations and their asymptotic covariance matrix via LISREL 9.1 (SSI Inc., 2007). d. Satorrra-Bentler Adjusted Chi-Square



Impact on Family

ADHD



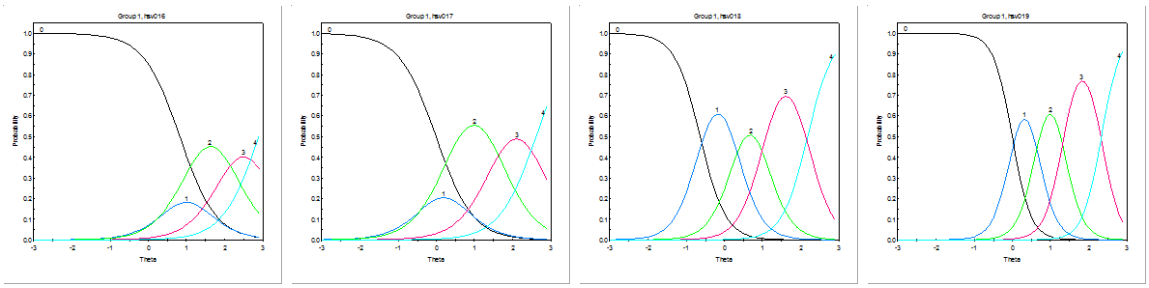
Graded Model Item Parameter Estimates for Group 1, logit: *a*(*θ* - *b*) (Back to TOC)

Item	Label		*a*	*s.e.*	b_1	*s.e.*	b_2	*s.e.*	b_3	*s.e.*	b_4	*s.e.*
1	hsv016	5	2.12	0.11	0.82	0.04	1.17	0.05	2.09	0.08	2.89	0.13
2	hsv017	10	2.09	0.10	-0.01	0.04	0.39	0.04	1.59	0.06	2.61	0.10
3	hsv018	15	3.03	0.14	-0.63	0.04	0.30	0.03	1.04	0.04	2.18	0.07
4	hsv019	20	4.12	0.26	-0.01	0.03	0.65	0.03	1.34	0.04	2.33	0.08

G2 = 662.12; df = 604; p = 0.0504; RMSEA = 0.01

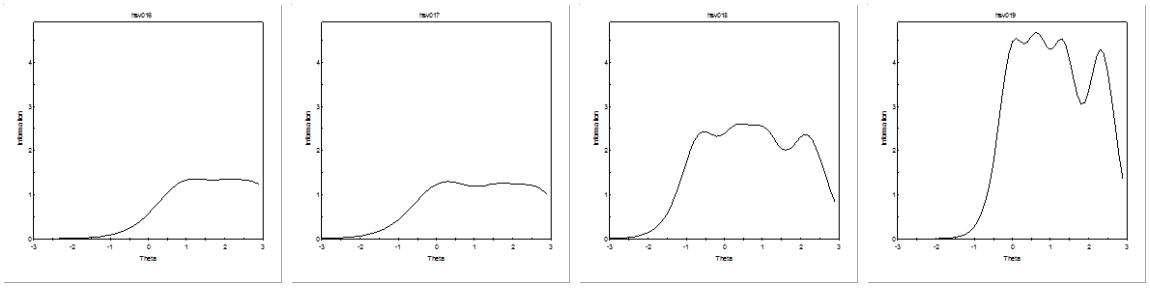
S-X^2 Item Level Diagnostic Statistics

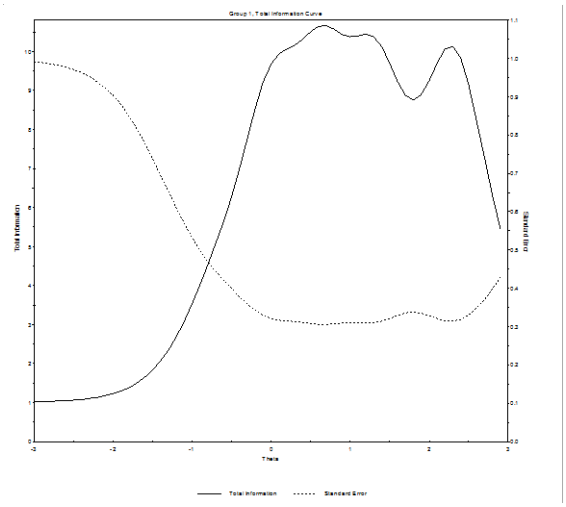
Item	Label	X^2	*d.f.*	Probability
1	hsv016	51.18	41	0.1321
2	hsv017	41.80	39	0.3495
3	hsv018	80.84	34	0.0001
4	hsv019	72.23	31	0.0001

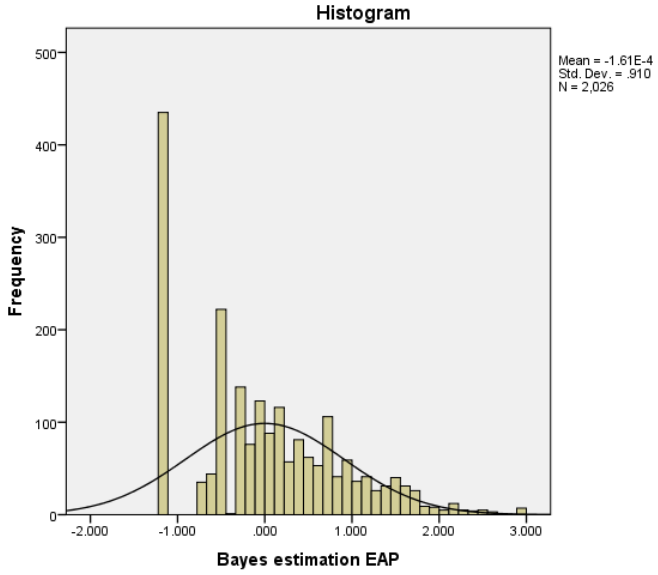


Impact on Family

ADHD

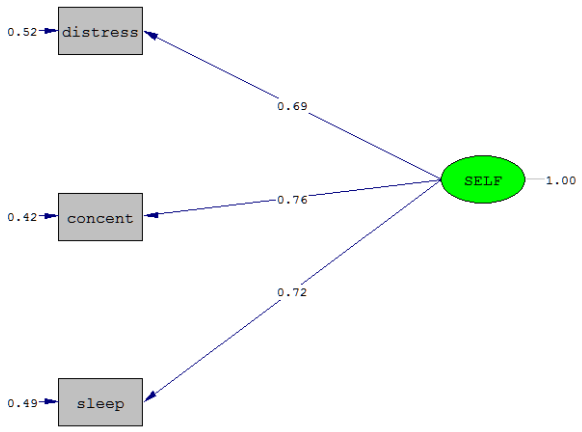






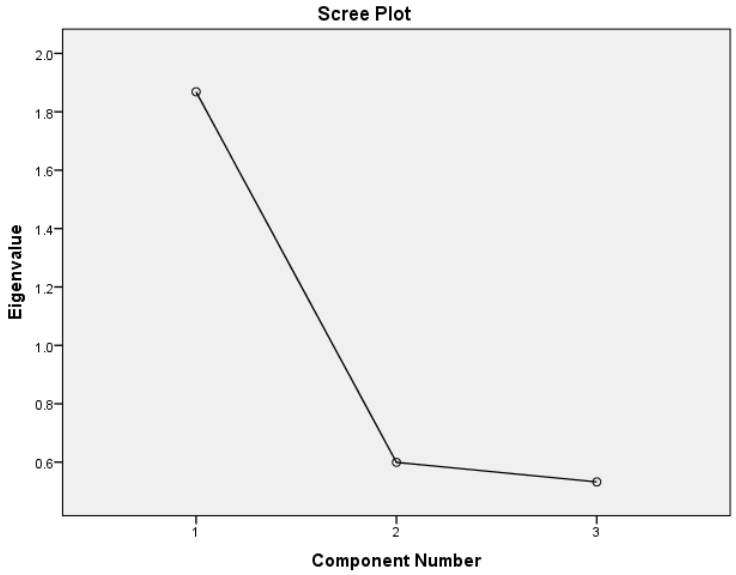
Impact on Self ADHD	Item loadings[a] λ_x	Regression weights[b]	Model characteristics[c]
hSV020/a In the last 12 months, when these problems were at their worst, did these difficulties distress [child] or make [him/her] feel bad or upset? How distressed?	0.694	0.308 *0.291*	N = 1963 df = 0 χ^2 = na
hSV021 When these problems were at their worst, how much did these difficulties prevent [child] from concentrating on things [he/she] was supposed to be doing?	0.759	0.413 *0.390*	**SRMR = na** **NNFI = na** **CFI = na**
hSV022 When these problems were at their worst, how much did these difficulties impact on [child's] sleeping?	0.716	0.338 *0.319*	H = 0.82 Saturated model

a. Partial regression coefficients of the item on the underlying construct. b. Upper figures are raw factor score indices and lower figures are proportionally adjusted factor score regression indices. c. Models were fitted via Robust Weighted Least Squares using polychoric correlations and their asymptotic covariance matrix via LISREL 9.1 (SSI Inc., 2007). d. Satorrra-Bentler Adjusted Chi-Square



Impact on Self

ADHD



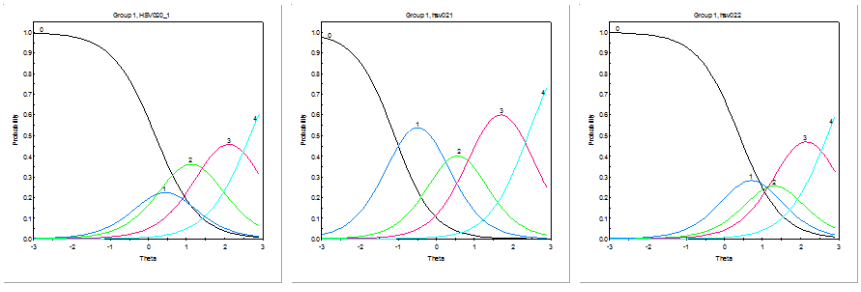
Graded Model Item Parameter Estimates for Group 1, logit: *a*(*θ* - *b*) (Back to TOC)

Item	Label		*a*	*s.e.*	b_1	*s.e.*	b_2	*s.e.*	b_3	*s.e.*	b_4	*s.e.*
1	HSV020_1	5	1.78	0.12	0.18	0.04	0.70	0.05	1.55	0.07	2.66	0.13
2	hsv021	10	1.97	0.12	-1.10	0.05	0.12	0.04	0.99	0.05	2.39	0.10
3	hsv022	15	1.84	0.13	0.39	0.04	1.02	0.05	1.59	0.07	2.70	0.13

G2 = 315.68; df = 109 0; p = .0001; RSMEA = 0.02

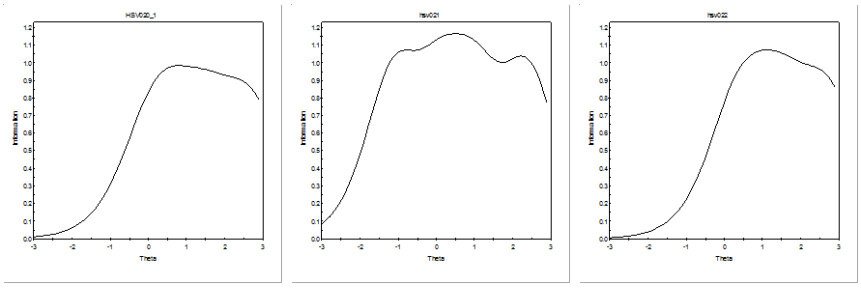
S-X^2 Item Level Diagnostic Statistics

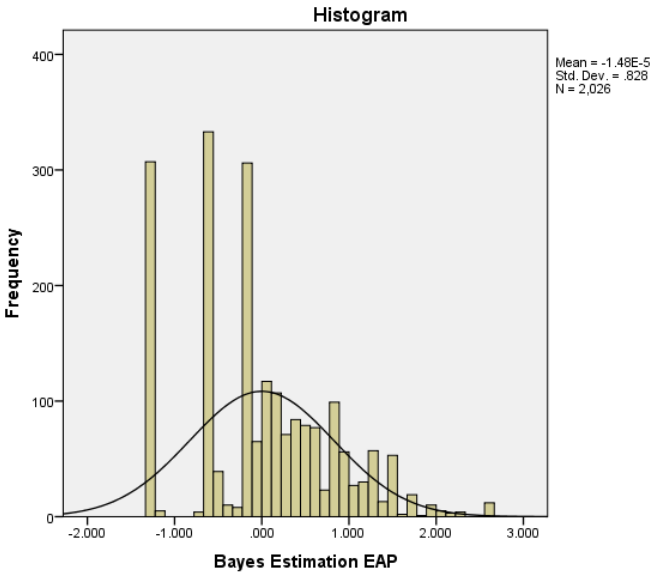
Item	Label	X^2	*d.f.*	Probability
1	HSV020_1	194.66	30	0.0001
2	hsv021	104.04	28	0.0001
3	hsv022	59.17	30	0.0012



Impact on Self

ADHD

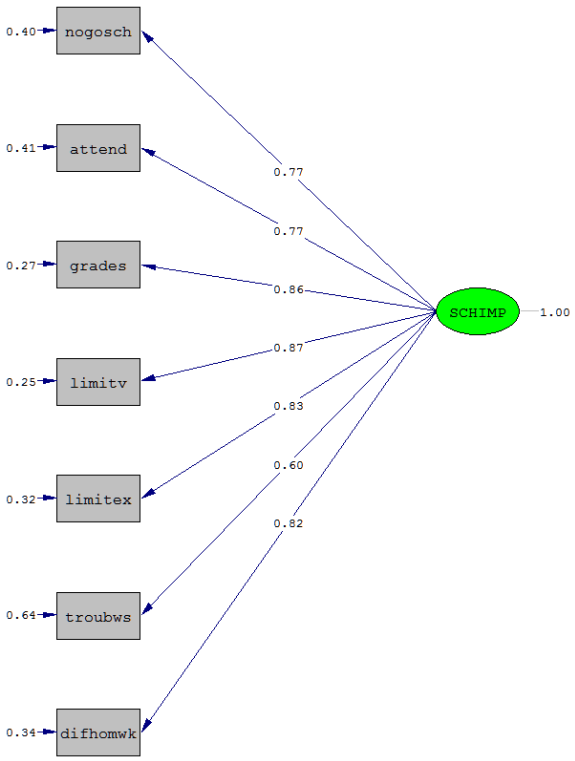


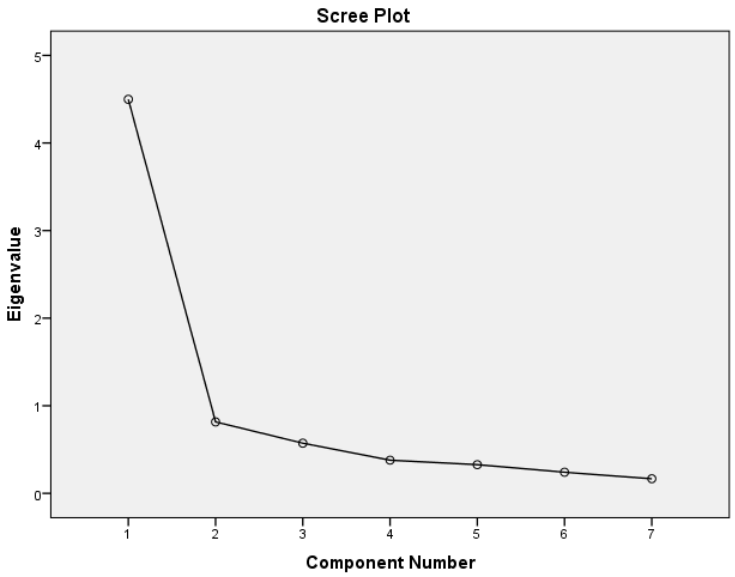


Impact on Schooling CONDUCT DISORDER	Item loadings[a] λ_x	Regression weights[b]	Model characteristics[c]
csv001 In the last 12 months, when these problems were at their worst, how often did child not want to go to school	0.772	0.133 *0.116*	
csv002 As a result of these problems, how many days has [child] been absent from school in the last 12 months?	0.769	0.131 *0.114*	
csv003 In the last 12 months, when these problems were at their worst, did they affect child's grades or ability to do work?	0.856	0.222 *0.194*	N = 889 df = 14 χ^2 = 175.0[d]
csv004 When these problems were at their worst, did these difficulties limit child in participating in volunteer activities?	0.868	0.244 *0.213*	SRMR = 0.08 **NNFI = 0.95** **CFI = 0.97**
csv005 When these problems were at their worst, did these difficulties limit child in participating in extracurricular activities?	0.825	0.180 *0.157*	H = 0.93 Acceptable
csv006 When these problems were at their worst, did these difficulties cause trouble at school (or at work)?	0.602	0.066 *0.058*	
csv007 When these problems were at their worst, how often did child have difficulties completing school work?	0.815	0.169 *0.148*	

a. Partial regression coefficients of the item on the underlying construct. b. Upper figures are raw factor score indices and lower figures are proportionally adjusted factor score regression indices. c. Models were fitted via Robust Weighted Least Squares using polychoric correlations and their asymptotic covariance matrix via LISREL 9.1 (SSI Inc., 2007). d. Satorrra-Bentler Adjusted Chi-Square

Impact on Schooling
CONDUCT DISORDER





Impact on Schooling

CONDUCT DISORDER

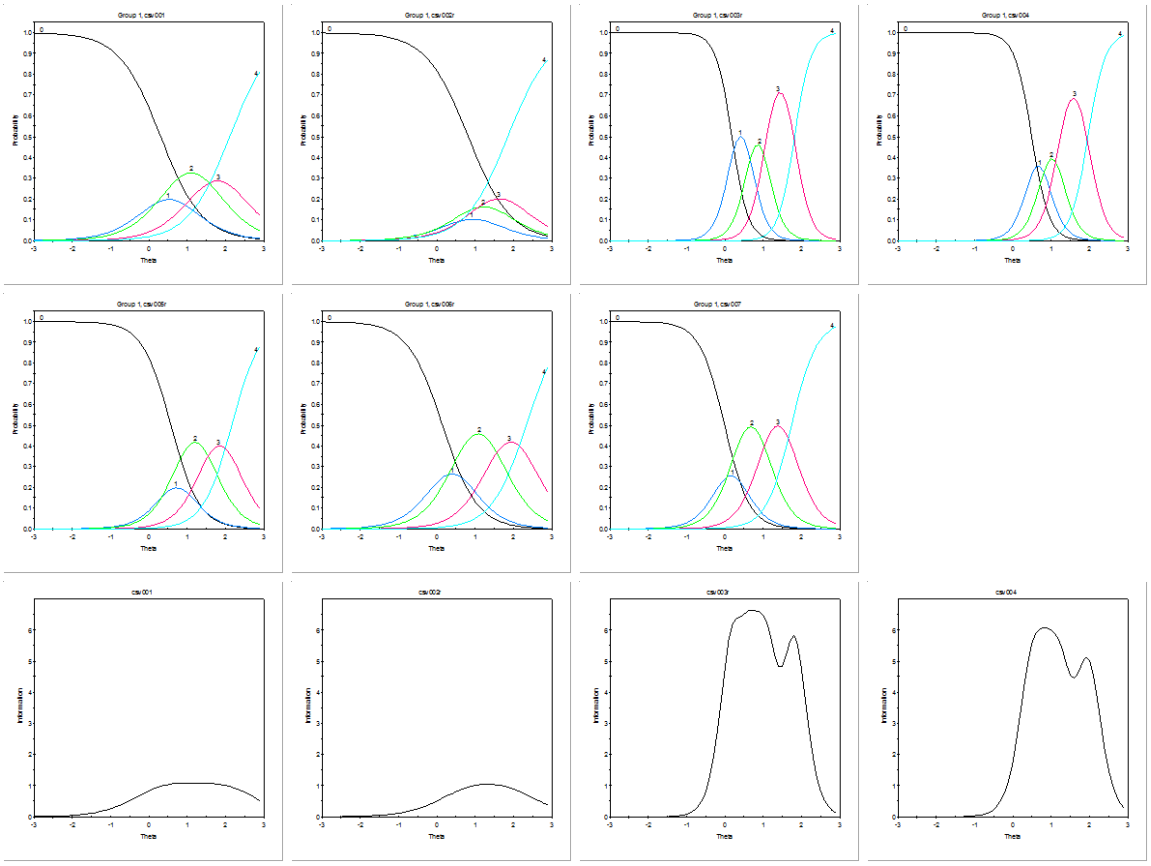
Graded Model Item Parameter Estimates for Group 1, logit: $a(\theta - b)$ (Back to TOC)

Item	Label		a	s.e.	b_1	s.e.	b_2	s.e.	b_3	s.e.	b_4	s.e.
1	csv001	5	1.85	0.12	0.31	0.05	0.74	0.06	1.47	0.09	2.11	0.12
2	csv002r	10	1.82	0.14	0.82	0.06	1.04	0.07	1.40	0.09	1.85	0.11
3	csv003r	15	4.76	0.37	0.18	0.04	0.65	0.04	1.07	0.05	1.82	0.08
4	csv004	20	4.45	0.36	0.49	0.04	0.83	0.05	1.20	0.06	1.96	0.09
5	csv005r	25	2.69	0.19	0.58	0.05	0.87	0.05	1.53	0.08	2.16	0.11
6	csv006r	30	2.21	0.14	0.15	0.05	0.63	0.05	1.53	0.08	2.33	0.13
7	csv007	35	3.08	0.20	-0.01	0.05	0.33	0.05	1.03	0.06	1.74	0.08

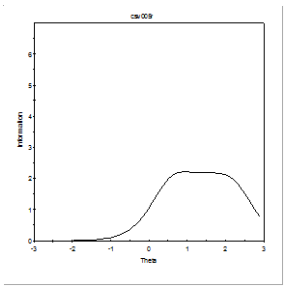
M2 = 957.54; df = 329; p < 0.0001; RMSEA = 0.02

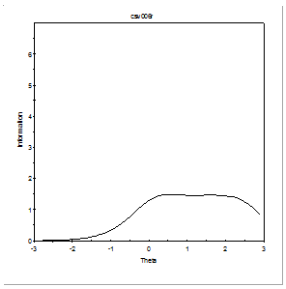
S-X^2 Item Level Diagnostic Statistics

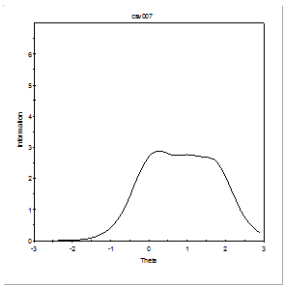
Item	Label	X^2	*d.f.*	Probability
1	csv001	83.44	78	0.3154
2	csv002r	69.92	79	0.7579
3	csv003r	48.20	52	0.6247
4	csv004	74.28	53	0.0284
5	csv005r	86.26	69	0.0779
6	csv006r	93.92	74	0.0588
7	csv007	74.40	63	0.1538

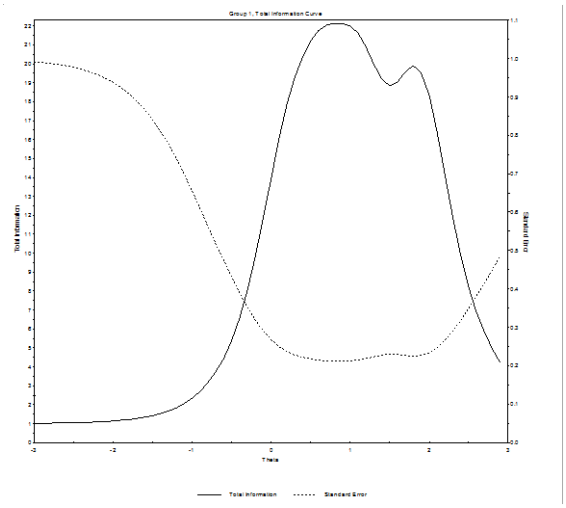


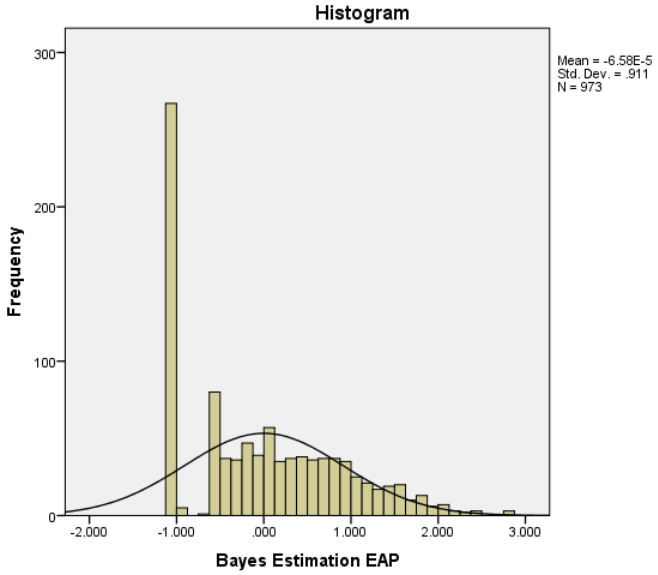
Impact on Schooling
CONDUCT DISORDER





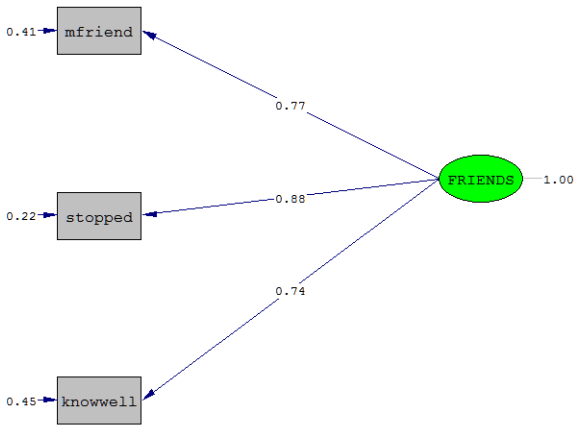




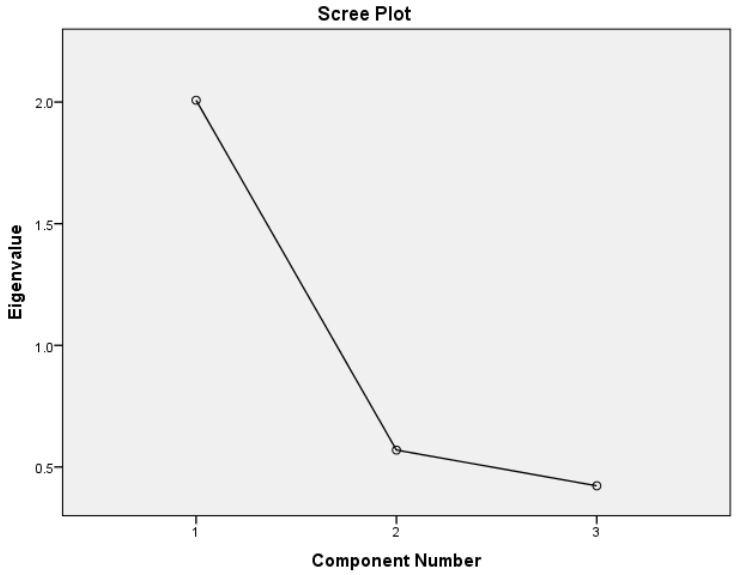


Impact on Friends CONDUCT DISORDER	Item loadings[a] λ_x	Regression weights[b]	Model characteristics[c]
CSV012 In the last 12 months, when these problems were at their worst, did these difficulties cause problems with [child's] ability to make or maintain friendships?	0.770	0.263 *0.251*	N = 993 df = 0 χ^2 = na
CSV014 How often have these difficulties stopped [child] from doing things or going places with other children [his/her] age?	0.883	0.557 *0.532*	**SRMR = na** **NNFI = na** **CFI = na**
CSV015 When these problems were at their worst how much difficulty did [child] have dealing with people [he/she] didn't know well?	0.739	0.226 *0.216*	H = 0.86 Saturated model

a. Partial regression coefficients of the item on the underlying construct. b. Upper figures are raw factor score indices and lower figures are proportionally adjusted factor score regression indices. c. Models were fitted via Robust Weighted Least Squares using polychoric correlations and their asymptotic covariance matrix via LISREL 9.1 (SSI Inc., 2007). d. Satorrra-Bentler Adjusted Chi-Square



Impact on Friends
CONDUCT DISORDER



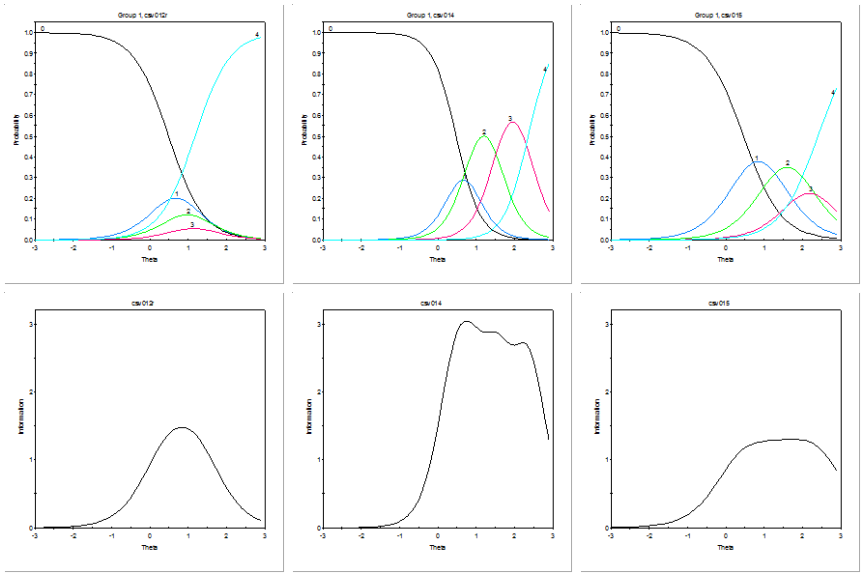
Graded Model Item Parameter Estimates for Group 1, logit: *a*(*θ* - *b*)

Item	Label		*a*	*s.e.*	b_1	*s.e.*	b_2	*s.e.*	b_3	*s.e.*	b_4	*s.e.*
1	csv012r	5	2.19	0.20	0.49	0.05	0.86	0.06	1.08	0.07	1.18	0.07
2	csv014	10	3.17	0.41	0.48	0.05	0.85	0.05	1.54	0.08	2.36	0.14
3	csv015	15	2.04	0.18	0.46	0.05	1.24	0.08	1.96	0.11	2.40	0.15

G^2 = 194.57; df = 109 0; p <.000; RSMEA = 0.01

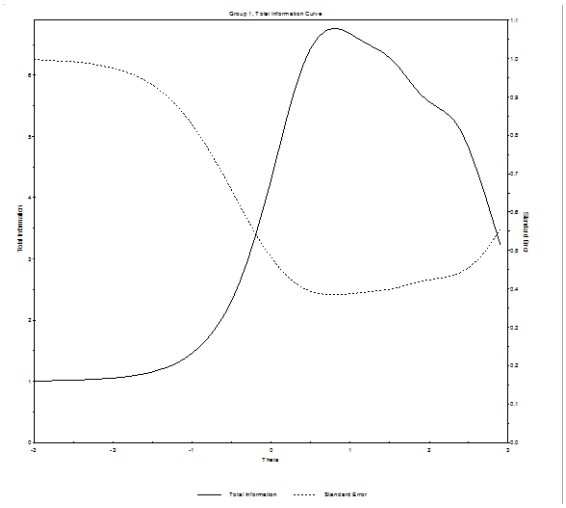
S-X^2 Item Level Diagnostic Statistics

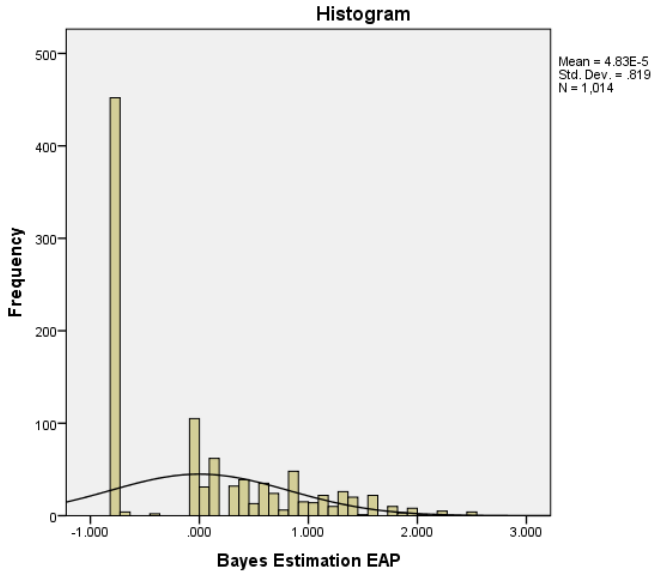
Item	Label	X^2	*d.f.*	Probability
1	csv012r	55.16	26	0.0007
2	csv014	41.68	26	0.0264
3	csv015	66.44	30	0.0001



Impact on Friends

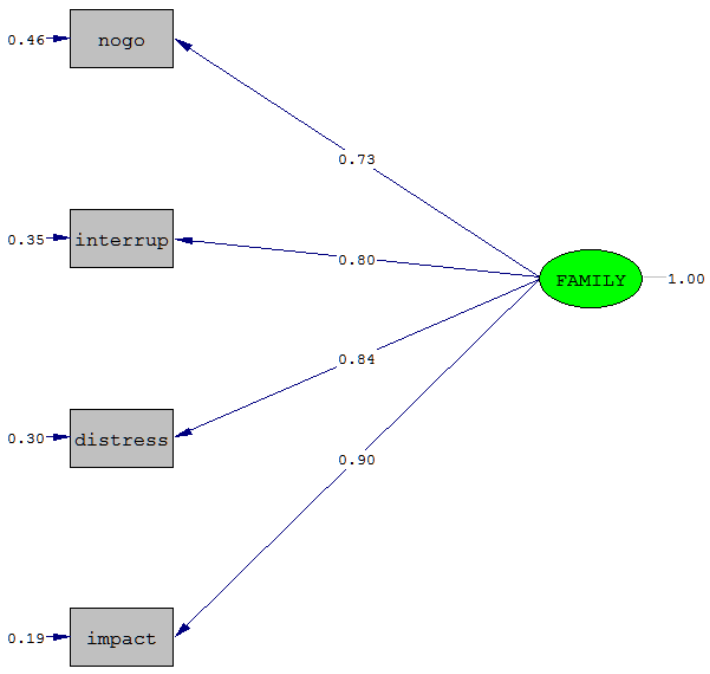
CONDUCT DISORDER



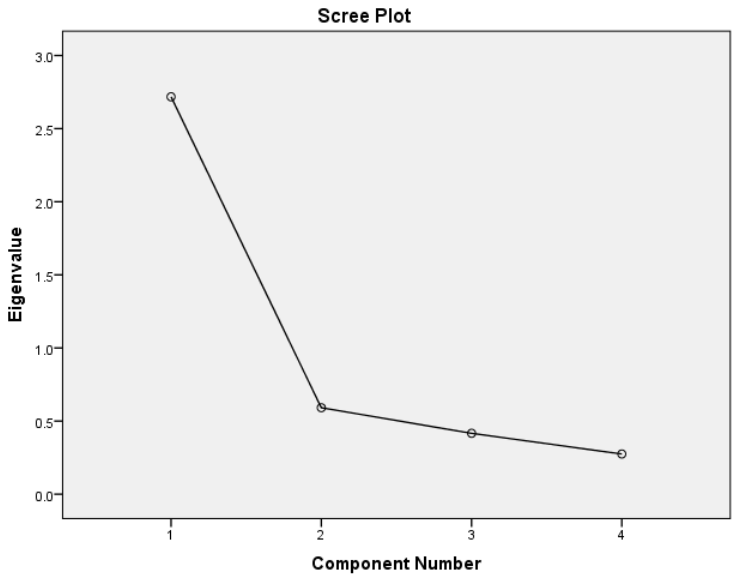


Impact on Family CONDUCT DISORDER	Item loadings[a] λ_x	Regression weights[b]	Model characteristics[c]
CSV016 In the last 12 months how often have [child's] difficulties prevented you from taking [him/her] places or going out in public?	0.732	0.151 *0.508*	N = 1015 df = 2 χ^2 = 41.2[d] **SRMR = 0.04** **NNFI = 0.98** **CFI = 0.98** H = 0.90 Acceptable
CSV017 How often have [child's] difficulties interrupted everyday family activities such as eating meals or watching TV?	0.804	0.218 *0.211*	
CSV018 How much distress do [child's] difficulties cause you and other members of the family?	0.836	0.265 *0.199*	
CSV019 How much do [child's] difficulties impact on your other family and household responsibilities, such as time to spend with other children or family members?	0.897	0.442 *0.082*	

a. Partial regression coefficients of the item on the underlying construct. b. Upper figures are raw factor score indices and lower figures are proportionally adjusted factor score regression indices. c. Models were fitted via Robust Weighted Least Squares using polychoric correlations and their asymptotic covariance matrix via LISREL 9.1 (SSI Inc., 2007). d. Satorrra-Bentler Adjusted Chi-Square



Impact on Family
CONDUCT DISORDER



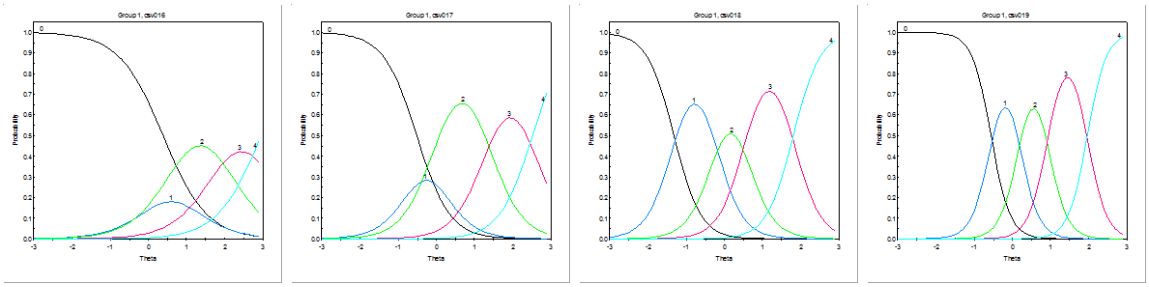
Graded Model Item Parameter Estimates for Group 1, logit: *a*(*θ* - *b*) (Back to TOC)

Item	Label		*a*	*s.e.*	b_1	*s.e.*	b_2	*s.e.*	b_3	*s.e.*	b_4	*s.e.*
1	csv016	5	1.74	0.12	0.39	0.06	0.80	0.06	1.92	0.11	2.96	0.19
2	csv017	10	2.31	0.15	-0.51	0.05	-0.01	0.05	1.35	0.07	2.52	0.13
3	csv018	15	2.87	0.18	-1.32	0.07	-0.23	0.05	0.55	0.05	1.81	0.09
4	csv019	20	4.04	0.37	-0.56	0.04	0.19	0.04	0.93	0.06	1.96	

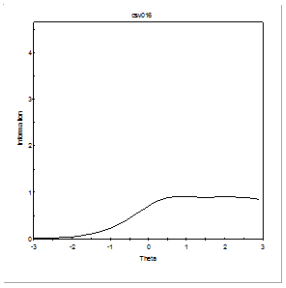
G2= 654.14; df = 604; p < 0.0773; RMSEA= 0.01

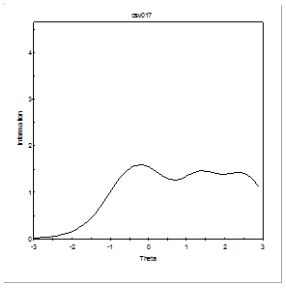
S-X^2 Item Level Diagnostic Statistics

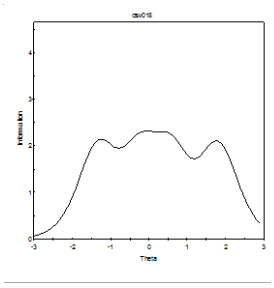
Item	Label	X^2	*d.f.*	Probability
1	csv016	62.72	38	0.0070
2	csv017	56.48	34	0.0091
3	csv018	89.36	30	0.0001
4	csv019	64.86	29	0.0001

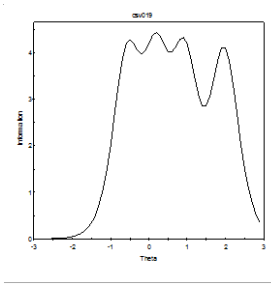


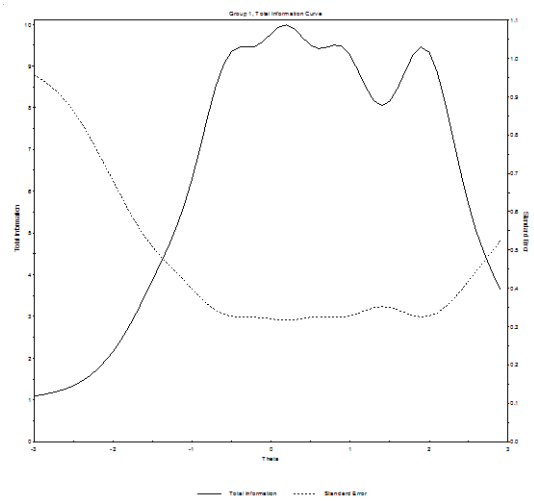
Impact on Family
CONDUCT DISORDER

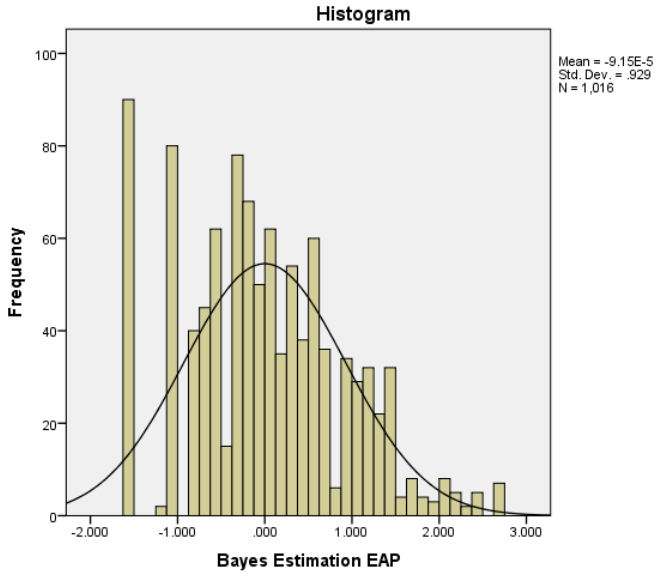






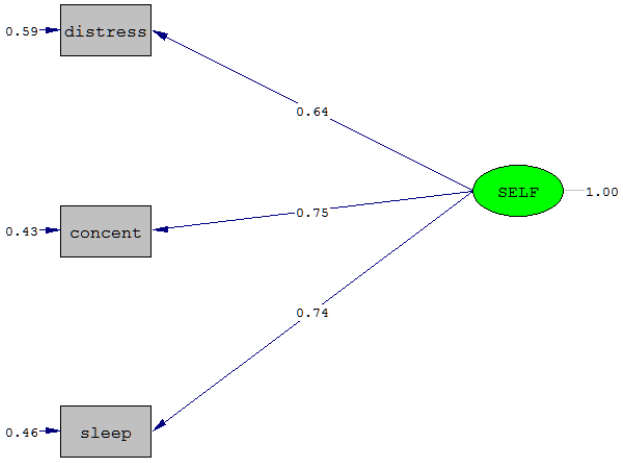






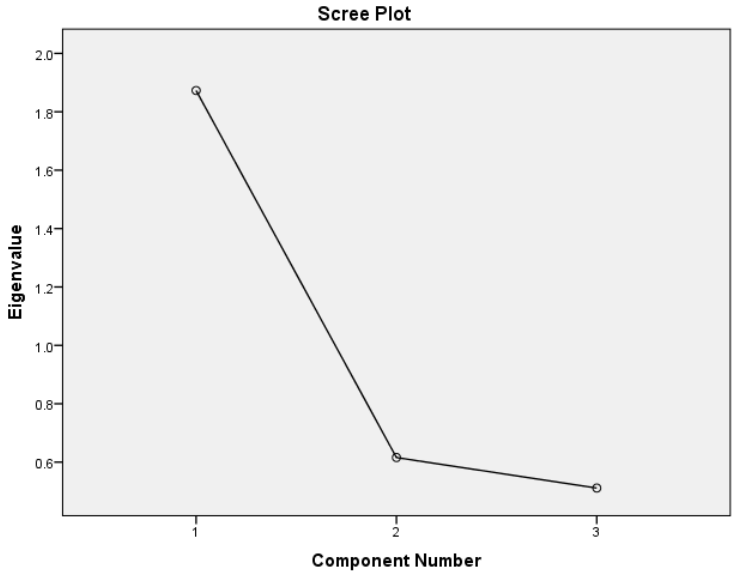
Impact on Self CONDUCT DISORDER	Item loadings[a] λ_x	Regression weights[b]	Model characteristics[c]
CSV020/a In the last 12 months, when these problems were at their worst, did these difficulties distress [child] or make [him/her] feel bad or upset? How distressed?	0.643	0.260 *0.246*	N = 978 df = 0 χ^2 = na
CSV021 When these problems were at their worst, how much did these difficulties prevent [child] from concentrating on things [he/she] was supposed to be doing?	0.755	0.415 *0.392*	**SRMR = na** **NNFI = na** **CFI = na**
CSV022 When these problems were at their worst, how much did these difficulties impact on [child's] sleeping?	0.738	0.384 *0.363*	H = 0.76 Saturated model

a. Partial regression coefficients of the item on the underlying construct. b. Upper figures are raw factor score indices and lower figures are proportionally adjusted factor score regression indices. c. Models were fitted via Robust Weighted Least Squares using polychoric correlations and their asymptotic covariance matrix via LISREL 9.1 (SSI Inc., 2007). d. Satorrra-Bentler Adjusted Chi-Square



Impact on Self

CONDUCT DISORDER



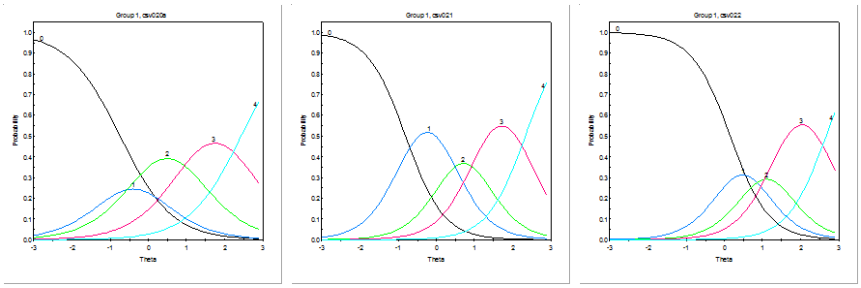
Graded Model Item Parameter Estimates for Group 1, logit: *a*(*θ* - *b*) (Back to TOC)

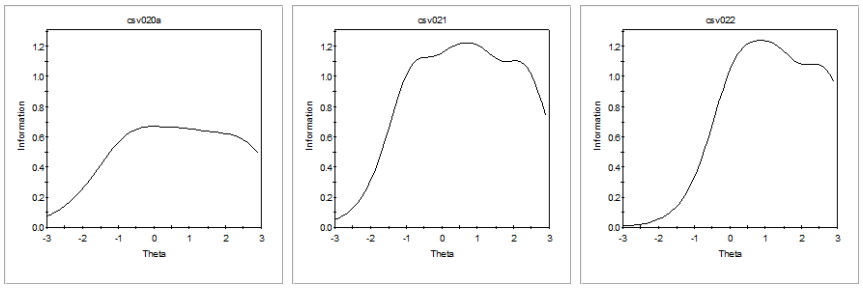
Item	Label		*a*	*s.e.*	b_1	*s.e.*	b_2	*s.e.*	b_3	*s.e.*	b_4	*s.e.*
1	csv020a	5	1.47	0.12	-0.74	0.07	-0.07	0.06	1.06	0.08	2.43	0.16
2	csv021	10	2.01	0.20	-0.81	0.07	0.33	0.05	1.10	0.08	2.33	0.15
3	csv022	15	1.99	0.20	0.15	0.05	0.80	0.06	1.41	0.09	2.66	0.18

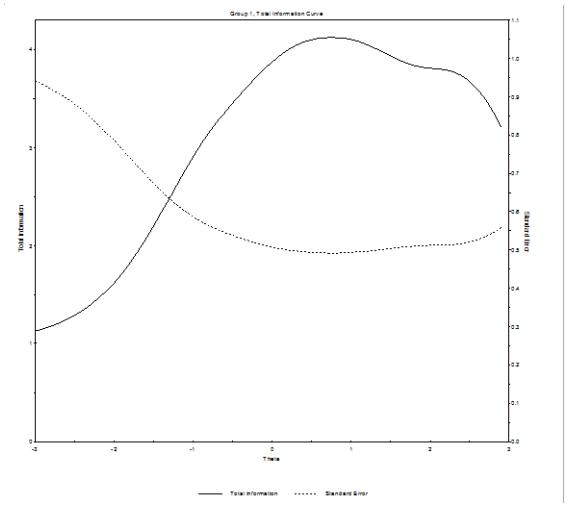
G2 = 162.75; df = 109; p < 0.0007; RMSEA = 0.01

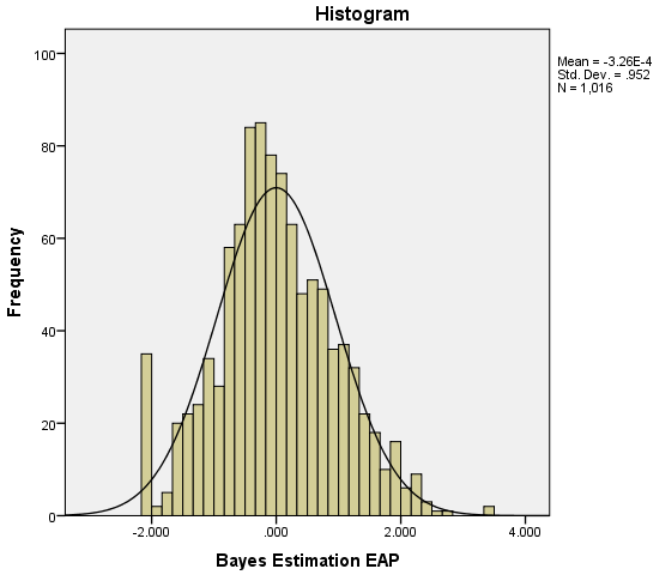
S-X^2 Item Level Diagnostic Statistics

Item	Label	X^2	*d.f.*	Probability
1	csv020a	36.29	28	0.1351
2	csv021	27.97	27	0.4142
3	csv022	44.72	28	0.0235




csv020a
csv021
csv022
Information
Theta


Total Information
Standard Error
Theta

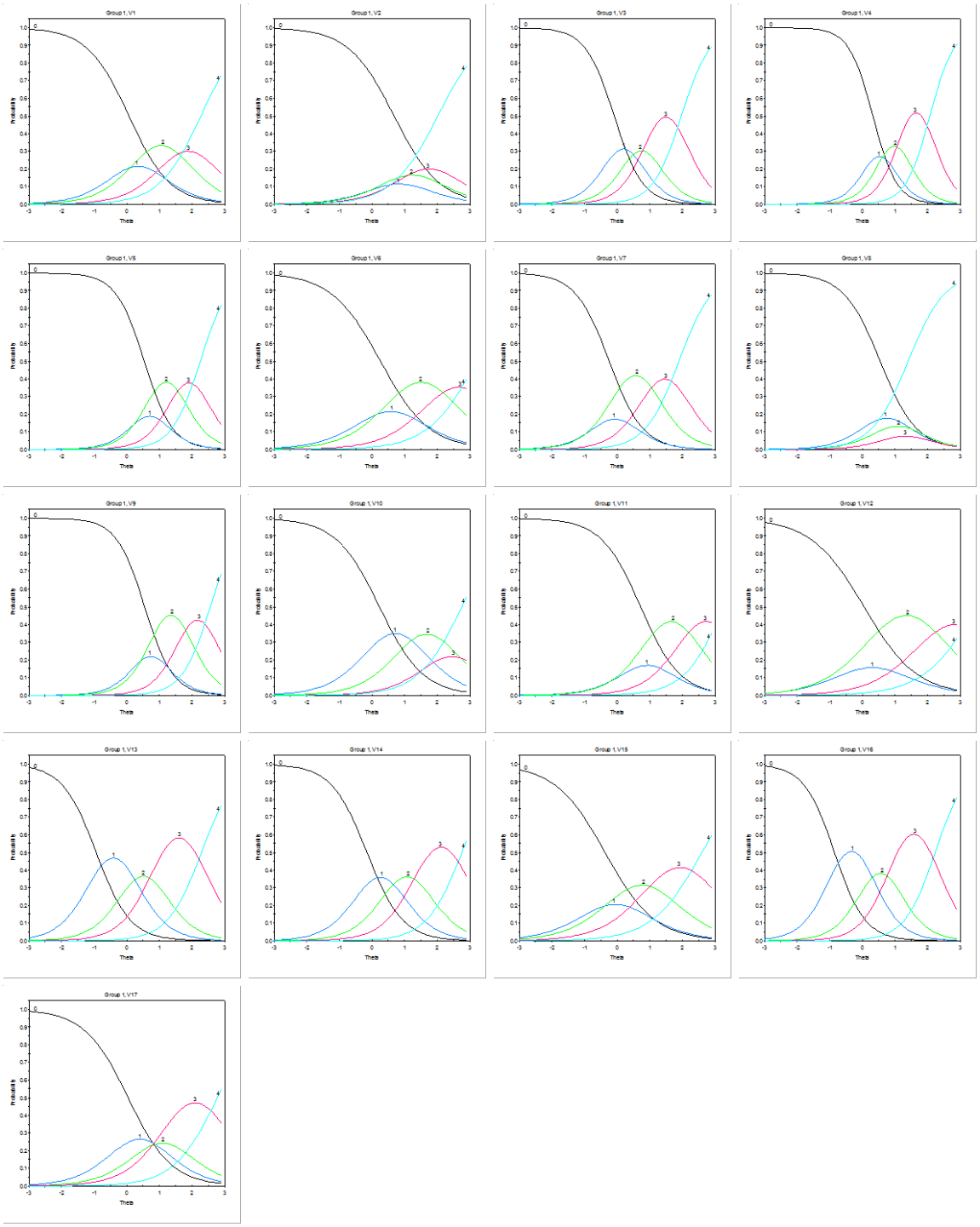

Histogram
Mean = -3.26E-4
Std. Dev. = .952
N = 1,016
Frequency
Bayes Estimation EAP

APPENDIX B

Item Characteristic Curves, Item Information Curves, and Total Information Curves for the Diagnostic group

Pooled results (N = 4950)

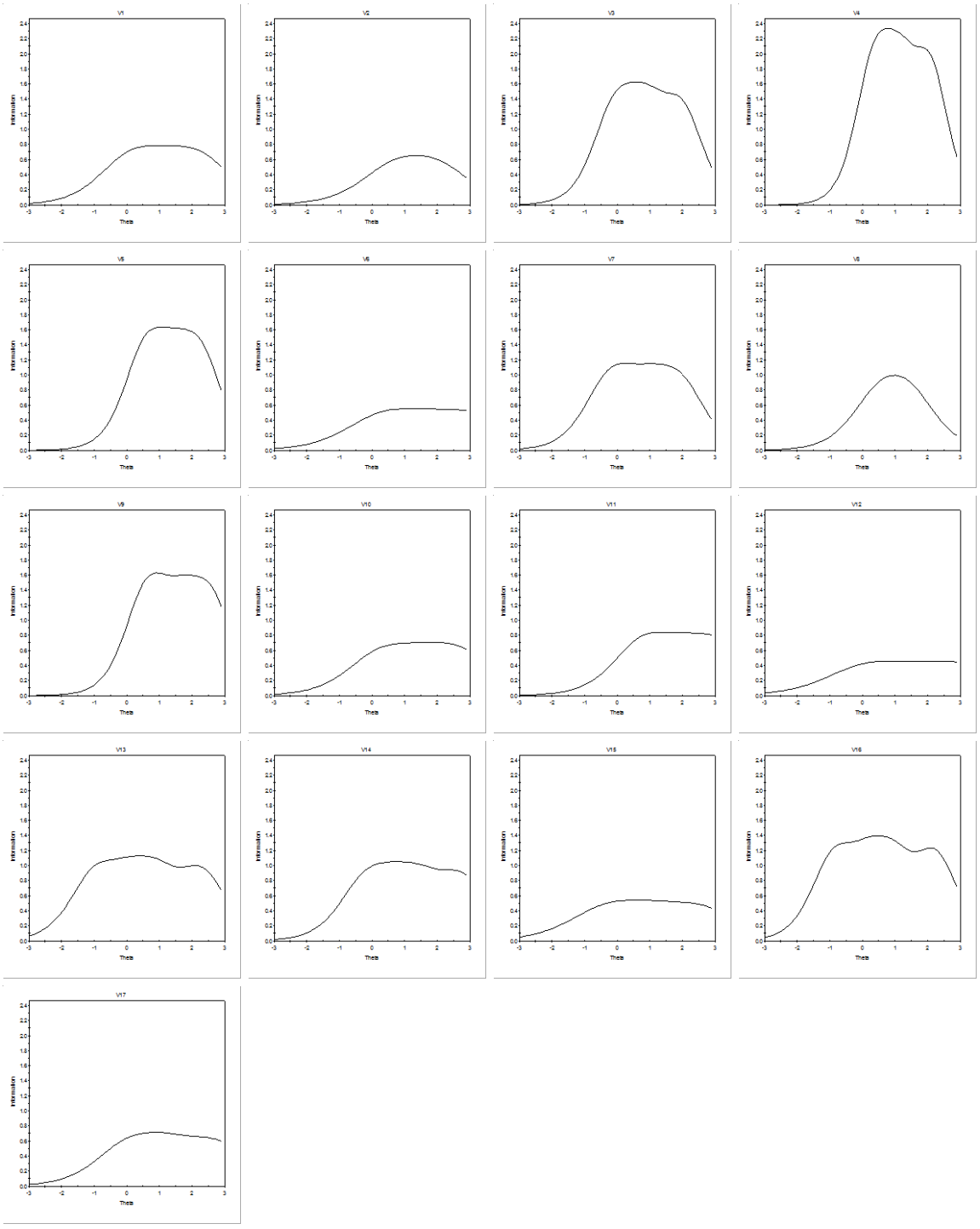
Item characteristic curves[4]



[4] Please see Tables 1 and 2 for a guide to the variable descriptions.

Pooled results (N = 4950)

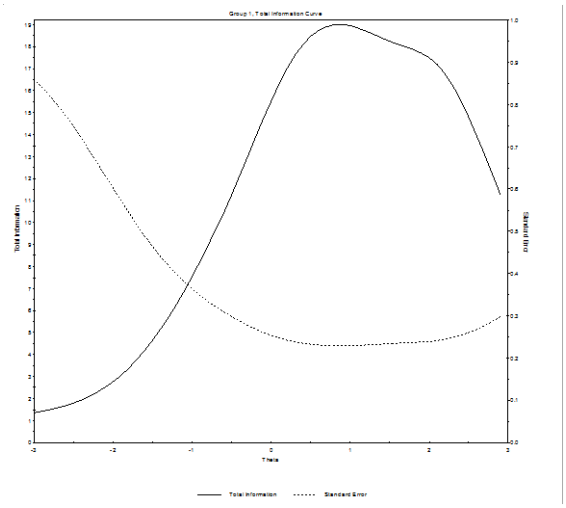
Item information curves[5]



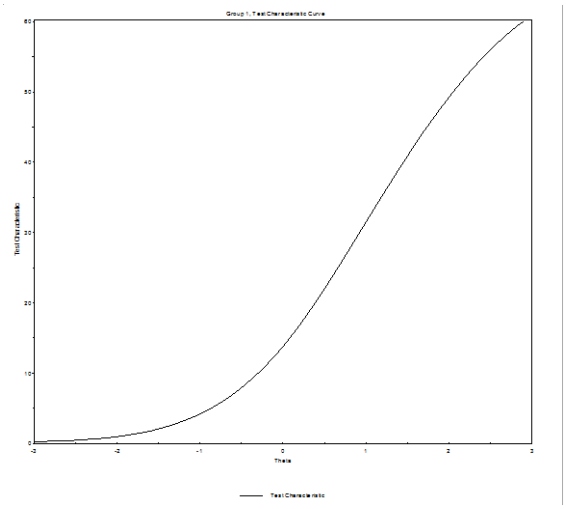
[5] Please see Tables 1 and 2 for a guide to the variable descriptions.

Pooled results (N = 4950)

Total information curve



Test characteristic curve



References

1. American Psychiatric Association (2000) Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision. Washington, DC, American Psychiatric Association.

2. Shaffer D, Fisher P, Lucas C, Dulcan MK, Schwab-Stone ME (2000) NIMH Diagnostic Interview Schedule for Children, Version IV (NIMH DISC-IV): description, differences from previous versions and reliability of some common diagnoses. Journal of the American Academy of Child and Adolescent Psychiatry. 39: 28-38.

3. Fisher PW, Shaffer D, Paicentini JC, Lapkin J, Kafantaris V, Leonard H, Herzog DB (1993) Sensitivity of the Diagnostic Interview Schedule for Children, 2nd edition (DISC-2.1) for specific diagnoses of children and adolescents. Journal of the American Academy of Child and Adolescent Psychiatry. 32: 666-673.

4. Sawyer MG, Arney PA, Baghurst PA, Clark JJ, Graetz BW, Kosky RJ, Nurcombe B, Patton GC, Raphael B, Rey JM, Whaites LC, Zubrick SR (2001) The mental health of young people in Australia: key findings from the Child and Adolescent Component of the National Survey of Mental Health and Well-being. Australian and New Zealand Journal of Psychiatry. 35: 806-814.

5. Hu, L., & Bentler, P. M. (1995). Evaluating model fit. Structural equation modeling: concepts, issues, and applications. In R. H. Hoyle (Ed.), *Structural equation modeling: concepts, issues, and applications* (pp. 76–99). Thousand Oaks, CA: Sage Publications.

6. Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods 3*(4), 424–53.

7. Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A multidisciplinary journal, 6*(1), 1–55.

8. Yuan, K.-H., & Bentler, P. M. (1997). Mean and covariance structure analysis: Theoretical and practical improvements. *Journal of the American Statistical Association, 92*(438), 767–74.

9. Hancock, G. R., & Mueller, R. O. (Eds.). (2006). *Structural equation modeling: A second course*. Greenwich, CT: Information Age Publishing.

10. du Toit, Mathilda. (2003). IRT from SSI. Scientific Software International: Lincolnwood,IL.

11. Scientific Software International. (2011). IRTPRO (Version 2.2.1). SSI: Lincolnwood,IL.

12. Embretson, S.E. and Reise, S.P. (2000). Item response theory for psychologists. Lawrence Erlbaum Associates: Mahwah, NJ.

13. Baker, F.B. & Kim, S.-H. (2004). Item response theory : Parameter estimation techniques (2nd ed.). New York, Marcel Dekker.

14. Fraley, R. C., Waller, N. G., & Brennan, K. A. (2000). An item-response theory analysis of self-report measures of adult attachment. *Journal of Personality and Social Psychology*, *78*, 350-365.

15. Schaffer D, Gould MS, Brasic J, et al. (1983) A children's global assessment scale (CGAS). Archives of General Psychiatry, 40, 1228-1231.

16. Steinberg L, Thissen D. (1996) Uses of item response theory and the testlet concept in the measurement of psychopathology. Psychological Methods, 1, 81-97.