Research Article

iMedPub Journals
http://www.imedpub.com

DOI: 10.4172/2472-1786.100006

Journal of Childhood & Developmental Disorders
ISSN 2472-1786

2015
Vol. 1 No. 1:6

Item Response Theory Properties of the Internalizing Disorders in Adolescents

Rapson Gomez[1] and Alasdair Vance[2]

1 School of Health Sciences and Psychology, Federation University Australia
2 Royal Children's Hospital, The University of Melbourne, Australia

Corresponding author:
Rapson Gomez

rapson.gomez@federation.edu.au

Professor, School of Health Sciences and Psychology, Federation University Australia, University Drive, Mt Helen, PO Box 663, Ballarat, Victoria, Australia.

Tel: (03) 5327 6087

Citation: Gomez R, Vance A. Item Response Theory Properties of the Internalizing Disorders in Adolescents. J Child Dev Disord. 2015, 1:1.

Abstract

Context: Although there is evidence for a single internalizing dimension for the major anxiety and depressive disorders, there are little data on psychometric properties of this dimension in adolescents.

Objective: The study examined this using item response theory.

Method: The 2-parameter logistic model was used to examine the properties for the common internalizing disorders (depressive and anxiety) in a group of 625 clinic-referred adolescents.

Results: All disorders were strong discriminators of the internalizing dimension, and generally more representative of and measured the internalizing dimension with more precision in the upper half of the trait continuum. There was also support for the concurrent validity of the internalizing dimension, in that it had large to medium effect size correlations with the internalizing scores of other measures.

Conclusion: The findings explain the high comorbidity of anxiety and depressive disorders, and support the organization of these disorders in a single overarching category.

Keywords: Depressive and Anxiety Disorders; Adolescents; Item Response Theory; Comorbidity

Received: September 18, 2015; **Accepted:** November 18, 2015; **Published:** November 25, 2015

Introduction

The 4th edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) [1] and its text revision edition (DSM-IV TR) [2] have separation anxiety disorder (SAD), social phobia (SOP), specific phobia (SPP), panic disorder (PD), agoraphobia (AG), generalized anxiety disorder (GAD), obsessive compulsive disorder (OCD), post-traumatic stress disorder (PTSD), dysthymia (DYS), and major depressive disorder (MDD) as the major anxiety and depression disorders. The anxiety and depressive disorders were organized into two different diagnostic groups. For these anxiety disorders, the recently published DSM-5 [3] has them in three different groups: anxiety disorders (comprising SAD, SOP, SPP, PD, AG, and GAD); obsessive-compulsive and related disorders (comprising OCD, body dysmorphic disorder, trichotillomania, hoarding disorder, and excoriation disorder); and trauma- and stressor-related disorders (comprising PTSD, acute stress disorder, adjustment disorders, and reactive attachment disorder). For depressive disorders, DSM-5 has MDD and DYS combined together under the category of persistent depressive disorder. Thus DSM-5 suggests four groups for the DSM-IV anxiety and depressive disorders.

In both DSM-IV/DSM-IV TR and DSM-5, the different disorders are considered to be distinct categories. However there are emerging empirical data to indicate that the major internalizing disorders (as listed in DSM-IV) could be viewed as indicators of a single underlying dimension that has generally been referred to as internalizing disorders [4–6]. Corresponding to this there are data indicating support for a one-factor model for some of the internalizing disorders in adults [7–9], and for all the ten internalizing disorders mentioned earlier in adolescents [10].

Based on the support for the one-factor model for the internalizing disorders, at least two studies have examined the psychometric properties of the internalizing factor or dimension using item response theory (IRT) analysis [7,8]. Before we discuss the findings in these studies we will discuss IRT briefly and explain how IRT can be used to evaluate the properties of the internalizing dimension. IRT is model-based, and although there are many IRT models, a general feature of all IRT models is that they all show the relationship between the response to an item and the latent trait the item is measuring [11]. The studies by Krueger and Finger [7], and McGlinchey and Zimmerman [8] used the 2-parameter logistic model (2-PLM). In this model, a graph called item characteristic curve (ICC) is generated for each item showing the probability of a positive response to the item as a function of the underlying trait. For each item, the model estimates the item difficulty (*β*) and discrimination (α) parameters. The difficulty parameter (β) indicates the point on the scale of the latent trait where a person has a 0.5 probability of endorsing or responding positively to the item. The item discrimination parameter (α) is the ability of an item to discriminate people with different levels of the underlying trait [12]. Higher values would mean better ability to discriminate different levels of the trait in question, and consequently stronger associations with the latent construct. According to de Ayala [13], α values of .80 or more can be considered reasonably high.

The two graphs in the top panel of Figure 1 show the ICCs for two hypothetical items. For these ICC graphs and all the other ICCs graphs, the *x*-axis is the trait (θ) scale (mean = 0, *SD* = 1). For each of the two items there is one ICC, which shows the probability of endorsing that item (*y*-axis), given the person's θ score. The slope of each ICC for an item is determined by its α value. The β value of each item determines the point at which the ICC curves intersect on the θ-scale. As an illustration for interpreting these ICC graphs, compared to the right graph, the left graph shows a higher difficult parameter (shown as *b*), thereby indicating that higher trait values are needed for a positive response to the item shown on the left side. The ICC of the graph on the left also shows a higher item discrimination parameter (shown as *a*), suggesting that this item will be better able to discriminate the underlying trait, compared to the item shown on the right side.

IRT models can also generate item information function (IIF), test information function (TIF) and the standard error of measurement (*SEM*) of the TIF. The IIF indicates the effectiveness or precision of an item to measure the latent trait at different levels of the trait continuum, while the TIF provides the effectiveness or precision of the test (i.e., all items together) to measure the latent trait at different levels of the trait continuum. The peak of the TIF is the point at which all the items combined provides the most precision in estimating the latent trait. The *SEM* of the TIF provides a measure of the imprecision of the TIF along the trait continuum. Thus, IRT can show the continuous reliability and precision of the individual items in a measure across the entire trait spectrum.

The bottom panel of **Figure 1** shows examples of an IIF (left graph) and a TIF (right graph). For these graphs, the *x*-axis is the trait- (θ) scale (mean = 0, *SD* = 1), provided across the trait continuum, from -3.00 to 3.00. The TIF (continuous line) is the combined value of all the IIF values. The dotted line in the TIF graph shows the corresponding curve for the *SEM*. As an illustration for interpreting the IIF graph, this graph shows that the IIF values are relatively higher for trait values from -1 *SD* to +1 *SD*. This implies more information function for the item between these trait levels, compared to other trait levels. For the TIF graph, the TIF values are relatively higher for trait values from -0.5 *SD* to +2 *SD*. This implies more information function for the test as a whole between these trait levels, compared to other trait levels. This can also be seen by the relatively low *SEM* values for trait values from -0.5 *SD* to +2 *SD.*

When the presence of the internalizing disorders is viewed as items, IRT can model the association (in probabilistic terms) between the internalizing disorders and the underlying latent trait, in this case the internalizing dimension. In such a case, the β value of a disorder indicates the point on the scale of the latent trait where that disorder has a 0.5 probability of being diagnosed as present. Thus, compared to an internalizing disorder with a low β value, an internalizing disorder with a high β value would have to have more of the internalizing trait for a diagnosis. A disorder with a high α value would mean that the disorder has strong association with the internalizing dimension, and has good ability to discriminate different levels of the internalizing dimension. Also, disorders with roughly the same α values would have more associations (or more comorbidity) with each other. The IIF of a disorder indicates its precision to measure the internalizing dimension at different levels of the trait continuum, while the TIF indicates the precision of all the disorders together to measure the internalizing dimension at different levels of the trait continuum. The *SEM* of the TIF will show the imprecision of all the disorders together to measure the internalizing dimension at different levels of the trait continuum. IRT models can also compute the latent trait scores for participants, based on their specific pattern of endorsements for the set of items in the model. Thus when presence/absence of the internalizing disorders are the "items" in the analysis, it will be possible to compute the internalizing dimension scores for the participants, based on their presence/absence of the different internal disorders in the model.

As noted above, there have been two studies that have examined the psychometric properties of the internalizing dimension [7,8]. The findings in these studies showed that all the disorders were strong discriminators of the underlying internalizing dimension (high α values, or all values in both studies greater than .75), and were more representative of this dimension in the upper half of the internalizing trait spectrum than the lower half (β values above the mean level of the latent trait). In both studies the β values for MDD were the closest to the mean of the internalizing trait spectrum, and the other disorders had increasing values above the mean. Also, the TIF showed more precision in the upper half of the internalizing trait continuum than the lower half (TIF values higher above the mean level of the latent trait), peaking at around 1 *SD* from the mean. Also, in both studies, the internalizing dimension scores correlated almost perfectly with the number of internalizing disorders diagnosed, and were associated with several measures of social burden. These findings were interpreted as providing external validity for the internalizing dimension.

Besides these general findings, there were also some notable differences across the Krueger and Finger [7] and McGlinchey and

Zimmerman [8] studies. The study by Krueger and Finger included MDD, GAD, SOP, simple phobia (SIP – a diagnosis in DSM-III-R that is comparable to SPP in DSM-IV), PD, AG and DYS, whereas they were MDD, SOP, PD/AG, SPP and GAD in the study by McGlinchey and Zimmerman. There were differences across these studies for the βs, especially for GAD. While GAD was one of the highest in the study by McGlinchey and Zimmerman, it had the lowest value in the study by Krueger and Finger, and the ordinal relationship for GAD, SOP and SIP/SPP was reversed across the studies. According to McGlinchey and Zimmerman, although such differences may be due to differences in the frequencies of comorbidity in the samples examined, they could also be related to the different sets of disorders examined in these studies.

Overall, although we have some data on the IRT properties of the internalizing dimension, the data are limited. Both the previous IRT studies [7,8] included only a limited range of the internalizing disorders. As pointed out by McGlinchey and Zimmerman [8], IRT properties of the internalizing dimension would vary as a function of the disorders included in the analysis. This suggests that for a better understanding of the properties of the internalizing factor it would be necessary to include all, if not most of major internalizing disorders in the analysis. In this respect, SAD and OCD were not included in the two past studies. Second, there is no IRT data on the internalizing dimension in children and adolescents. The IRT study by McGlinchey and Zimmerman [8] involved adults. Although the participants (*n* = 251) age ranged from 15 and 54 years in the Krueger and Finger [7] study, only 20.5% were between 15 and 24 years. Thus the sample was predominantly an adult sample. Overall, therefore, it is can be argued that there are important gaps in our understanding of the IRT psychometric properties of the internalizing dimension, especially among children and adolescents.

Given the limitations, the first aim of the current study was to use the 2-PLM IRT procedure to examine the item response theory properties for a wider range of the common DSM-IV/TR internalizing disorders (depressive and anxiety) for a large group of clinic-referred adolescents. Diagnoses were derived from interviews of adolescents. The disorders included were SAD, SOB, SPP, PD, AG, GAD, OCD, PTSD, DYS and MDD. As will be evident, this study included more internalizing disorders than the two previous IRT studies [7,8], including OCD and SAD, that were omitted in the past studies. The second aim of the study was to examine the external validity of the internalizing dimension. This was done by examining the concurrent and discriminant validities of the internalizing dimension by correlating participants' internalizing traits scores (obtained through the 2-PLM analysis) with the internalizing and externalizing scores derived from other measures.

Method

Participant

The data for all participants were collected archivally from the Academic Child Psychiatry Unit (ACPU) of the Royal Children's Hospital, Melbourne, Australia. The ACPU is an out-patient psychiatric unit that provides services for children and adolescents with behavioral, emotional and learning problems. Referrals are generally from other medical services, schools, and social and welfare organizations. For the current study we used the records of adolescents, aged between 12 and 18 years, referred between 2004 and 2010, who had been interviewed for clinical diagnosis.

In all, there were 625 adolescents[1], comprising 416 males (66.6%) and 209 females (33.4%). The overall mean age of participants was 13.90 years (*SD* = 1.52). The percentages of father's employment status were as follows: employed = 78.2%, home duties = 2.1%, pensioner/retired = 5.7%, unemployed = 8.8%, others/unknown = 5.2%. The percentages of father's highest education level were as follows: tertiary = 15.5%, high school/some years in secondary school or equivalent = 63.0%, technical certificate or equivalent = 18.3%, primary school = 2.5%, and no schooling = 0.7%. Thus, most fathers of participants were employed, and more than two-third of participants had fathers who had attended at least secondary school. In terms of parental relationship, about 50% of parents were living together and 43% were separated or divorced.

Table 1 shows the percentages of different disorders for the participants. As shown, Attention Deficit/Hyperactivity Disorder (ADHD) and Oppositional Defiant Disorder/Conduct Disorder (ODD/CD) were more prevalent than the other disorders, with around 69% and 68%, respectively. Around 75% of participants had at least one depressive or anxiety disorder, and about 79% of these participants had either ADHD or CD/ODD or both. GAD and DYS were more prevalent than the other depressive and anxiety disorders. Around 48% and 50% of the participants had GAD and DYS, respectively. For those with an anxiety disorder, 55% had a depressive disorder, and for those with a depressive disorder, 90% had an anxiety disorder. SOP, SPP and MDD were also relatively high, and AG and PD were relatively rare.

Ethics

The study was approved by the RCH ethics committee as part of our group's comprehensive examination of children and adolescent referred for psychological problems. Each legal guardian and participant provided informed written consent for any data provided by them to be used in future research studies. This is a standard part of the ACPU assessment procedure. The RCH ethics committee adheres to the ethical guidelines set by the Australian National Medical Research Council that in turn confirms to the World Medical Association Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects.

Measures

Anxiety Disorders Interview Schedule for *Children* (ADISC-IV) [14]. The ADISC-IV is a semi-structured interview, based on the DSM-IV/DSM-IV TR diagnostic systems. Although ADISC-IV has been designed primarily to facilitate the diagnosis of the major anxiety disorders, it can also be used for diagnosing the other major childhood disorders, including the depressive disorders,

[1] These participants were the same participants involved in an earlier study that use CFA to examine different factor models of the internalizing disorders, including a one-factor model [10]. For the one-factor model in that paper, only the factor loadings that are analogous to the discrimination parameters were provided. It did not provide any other IRT relevant properties that are reported in the current paper.

Table 1 Frequency and Percentage of Different Disorders of Participants

	Frequency	Percentage
Separation Anxiety Disorder	129	20.6
Social Phobia	180	28.8
Specific Phobia	182	29.1
Panic Disorder	68	10.9
Agoraphobia	31	5.0
Generalized Anxiety Disorder	300	47.9
Obsessive Compulsive Disorder	130	20.8
Post Traumatic Stress Disorder	105	16.8
Dysthymia	314	50.2
Major Depressive Disorder	165	26.4
AttentionDeficit/Hyperactivity Disorder	429	68.5
OppositionalDefiant/Conduct Disorder	424	67.7
No diagnosis	21	3.4

ADHD, ODD and CD (in only the parent version), and also a range of other behavior problems. The ADISC-IV guideline for diagnosis is that the child be given a diagnosis of all disorders meeting the diagnostic criteria. Clinical diagnosis can be based either on parent (using the ADISC-IV/P) or child/adolescent (using the ADISC-IV/C) interview or on both interviews together. The scores of ADISC-IV/P and ADISC-IV/C have sound psychometric properties [15]. Test-retest reliability for the ADISC-IV scores over a 7 to 14-day interval has shown good to excellent reliability. Kappa values for interviews with parents and children (7 and 16 years) range from 0.65 to 1.00, and 0.61 to 0.80, respectively [15]. Since the ADISC-IV/P, but not the ADISC-IV/C, allows for diagnosis of additional disorders, the disorders reported earlier **(Table 1)** were derived using ADISC-IV/P. The diagnoses derived from the interviews of adolescents using the ADISC-IV/C were however used for the IRT analysis. Like the two previous IRT studies [7,8], the hierarchical exclusionary rules in DSM-IV were not taken into account for diagnoses.

Achenbach System of Empirically Based Assessment (ASEBA) [16]. The Child Behavior Checklist/6-18 (CBCL), the Teacher Report Form (TRF) and the Youth Self-Report (YSR), which are part of the ASEBA, were used to obtain internalizing and externalizing scores for testing the concurrent and discriminant validities of the internalizing dimension. The CBCL, completed by parents, has 113 items, while the TRF has 120 items for teacher completion. Both are used to rate children between 4 and 18 years of age. The YSR, completed by individuals between 11 and 18 years, has 112 items, worded in the first person. For all three versions, respondents indicate the frequency of each behavior described in the item on a scale of 0 (not true), 1 (somewhat or sometimes true) or 2 (very true or often true). The standard rating period is 6 months for the CBCL and YSR, and 2 months for the TRF. All three scales have excellent psychometric properties [16]. Among other scores, these scales provide scores for internalizing behavior problems, and externalizing behavior problems. These scores were used for the concurrent and discriminant validity analyses.

Procedure

All adolescents and their parents participated in separate interviews and testing sessions with breaks as needed over two consecutive days. Information was also obtained from teachers using various checklists and questionnaires. In all cases, parental consent forms were completed prior to the assessment. The data collected covered a comprehensive demographic, medical (primarily neurological and endocrinological), educational, psychological, familial and social assessment of the children and their families. All psychological data were collected by research assistants, who were advanced students in clinical psychology or in child psychiatry, and under the supervision of registered clinical psychologists/child and adolescent psychiatrists. The research assistants were provided with extensive supervised training and practice by the psychologists/child and adolescent psychiatrists prior to them collecting data. This training for the ADISC-IV/P and ADISC-IV/C included observations of them being administered by the psychologists/child and adolescent psychiatrists. The research assistants commenced administering the ADISC-IV/P and ADISC-IV/C only after they attained competence in their administration, as assessed by the registered psychologists/child and adolescent psychiatrists. There was adequate inter-rater reliability for the diagnoses made between the research assistants and their supervisors, and between research assistants (kappa values generally more than .88). Standard procedures were used for the administration of all measures. Where necessary, researchers read the items to participants who then completed their responses. Approximately 95% of the ADISC-IV/P interviews involved mothers only, and the rest involved fathers or both fathers and mothers together. Clinical diagnosis was confirmed by two consultant child and adolescent psychiatrists who independently reviewed these data. The inter-rater reliability for diagnoses of the two psychiatrists was high for both the parent and child versions (kappa = .90).

Statistical Procedures

This study used Multilog 7.0.3 [17] to perform the 2-PLM analyses. For each disorder, the following psychometric properties were examined: ICC; (graphically),α, β, and IIF (graphically). In addition, for the overall internalizing dimension, the TIF was also examined (graphically). As the 2-PLM is model-based, it is necessary to test if there is model-data fit. This was ascertained by examining the residuals (differences between the observed proportion and the model-based expected proportion of the responses in each category for each item) provided by Multilog. Low residual values suggest good model-data fit. As a further confirmation of model-data fit, fit plots derived from Modfit [18], using the 2-PLM item parameters estimated from Multilog, were also examined. When there is good model-data fit, the response curve for the observed data will correspond closely to the response curve predicted by the 2-PLM. The 2-PLM assumes unidimensionality and local independence. Local independence implies that associations between items are only caused by the underlying latent trait. Unidimensionality and local independence were examined using a 1-factor CFA model, comprising the ten disorders in the IRT analysis. Support for unidimensionality is inferred when there is good model fit, and support for local independence is inferred when no residual correlation is more than .20.

The latent trait scores for participants, based on their specific pattern of endorsements for the set of disorders were computed here using expected a posteriori (EAP) [19], obtained as part of the IRT analysis. To examine concurrent and discriminant validities, these scores were correlated with the number of diagnoses endorsed, and the CBCL, TRF and YSR internalizing and internalizing problem behavior scores. The effect sizes of these correlations were interpreted using the guidelines suggested by Hemphill [20]: correlations < .2 = small, correlations of .2 to .3 = medium and correlations >.30 = large.

Results

Unidimensionality, Model-Data Fit and Local Independence

Table 2 presents the tetrachoric correlation matrix between the disorders in the IRT model. As shown, these were all significantly and positively correlated ($p < .001$).

Unidimensionality for the ten disorders in the IRT model was examined via a 1-factor CFA model, comprising a single latent factors on which the ten disorders loaded. The fit values for this model, computed in M*plus* (Version 7) [21], using mean and variance-adjusted weighted least squares or WLSMV were (df = 35) = 58.98, $p < .01$; root mean square error of approximation (RMSEA) = .033; Comparative Fit Index (CFI) = .981. Both the RMSEA and CFI values showed good fit, based on guidelines suggested by Hu and Bentler [22] that RMSEA values close to 0.06 or below, and CFI values close to .95 or above indicate good fit. Thus there was support for unidimensionality.

MULTILOG indicated that the residuals (differences between the observed proportion and the expected proportion of the responses in each category) for the disorders ranged from .00 to .001. Fit plots derived from MODFIT [18] indicated that all the observed category curves for the ten disorders showed good fit. The residuals and fit-plots suggest good model-data fit for the 2-PLM in this study for the disorders. As already mentioned, there was good fit for the 1-factor model. Also for this model, the highest residual correlation was .19, and the remaining residual correlations ranged from .00 to .14. Taken together, these findings can be taken as indicating acceptable support for model-data fit and for the local independence of the disorders in the IRT model examined.

2-PLM Analysis

The α and β parameters for the disorders are provided in **Table 3**. **Figure 2** shows the ICCs curves for these disorders. **Table 3** shows that although there was wide variability, the α values for all disorders were high **(Figure 2)**, thereby suggesting that each disorder was good at discriminating the underlying internalizing factor. The order in terms of increasing discrimination values were SPP, SAD, PTSD, OCD, SOP, MDD, DYS, PD, GAD and AG. Using the equal option available in Multilog we examined the equality across these discrimination values. Although details are not shown, the discrimination values were equal across SPP, SAD, PTSD, OCD and SOP (between .85 and 1.35); MDD, DYS and PD (between 1.81, 1.82 and 1.90, respectively); and GAD and AG (2.38 and 2.65, respectively).

Table 3 and **Figure 2** show that although there was variability for the β values, all disorders, except DYS and GAD were located close to or above the mean trait level. The order in terms of increasing difficulty values were DYS, GAD, MDD, SPP, SOP, PD, OCD, SAD, AG and PTSD. Using the equal option available in Multilog we examined the equality across these difficulty values. Although details are not shown, the difficulty values were equal across DYS and GAD (.48 and .53, respectively); MDD, SPP and, SOP (between .97 and 1.24); PD and OCD (1.49 and 1.56, respectively); and SAD, AG and PTSD (between 1.87 and 2.18).

The TIF graph in **Figure 3** shows that for the internalizing

Table 2 Tetrachoric Correlations for Ten Unipolar Mood and Anxiety Diagnoses (N =625)

	1	2	3	4	5	6	7	8	9
SAD (1)									
SOP (2)	.36								
SPP (3)	.31	.29							
PD (4)	.28	.41	.42						
AG (5)	.46	.53	.39	.65					
GAD (6)	.43	.45	.31	.55	.62				
OCD (7)	.34	.45	.31	.40	.61	.48			
PTST (8)	.23	.21	.28	.42	.47	.35	.30		
DYS (9)	.29	.47	.27	.41	.44	.60	.35	.38	
MDD (10)	.23	.39	.24	.57	.40	.61	.32	.38	.64

Note. SAD = separation anxiety disorder; SOP = social phobia; SPP = specific phobia; PD = panic disorder; AG = agoraphobia; GAD = generalized anxiety disorder; OCD = obsessive compulsive disorder; PTSD = post traumatic stress disorder; DYS = dysthymia; MDD = major depressive disorder.
$p < .001$ for all correlations.

Table 3 Two-Parameter Logistic Item Response Model Parameter Estimates (N = 625).

	SAD	SOP	SPP	PD	AG	GAD	OCD	PTSD	DYS	MDD
α: Estimate	1.03	1.35	.85	1.90	2.65	2.38	1.32	1.13	1.82	1.81
SE	(.19)	(.19)	(.15)	(.28)	(.50)	(.27)	(.21)	(.22)	(.20)	(.22)
β: Estimate	1.87	1.24	1.18	1.49	1.94	.53	1.56	2.18	.48	.97
SE	(.28)	(.15)	(.22)	(.14)	(.15)	(.07)	(.20)	(.33)	(.08)	(.10)

Note. SAD = separation anxiety disorder; SOP = social phobia; SPP = specific phobia; PD = panic disorder; AG = agoraphobia; GAD = generalized anxiety disorder; OCD = obsessive compulsive disorder; PTSD = post traumatic stress disorder; DYS = dysthymia; MDD = major depressive disorder.

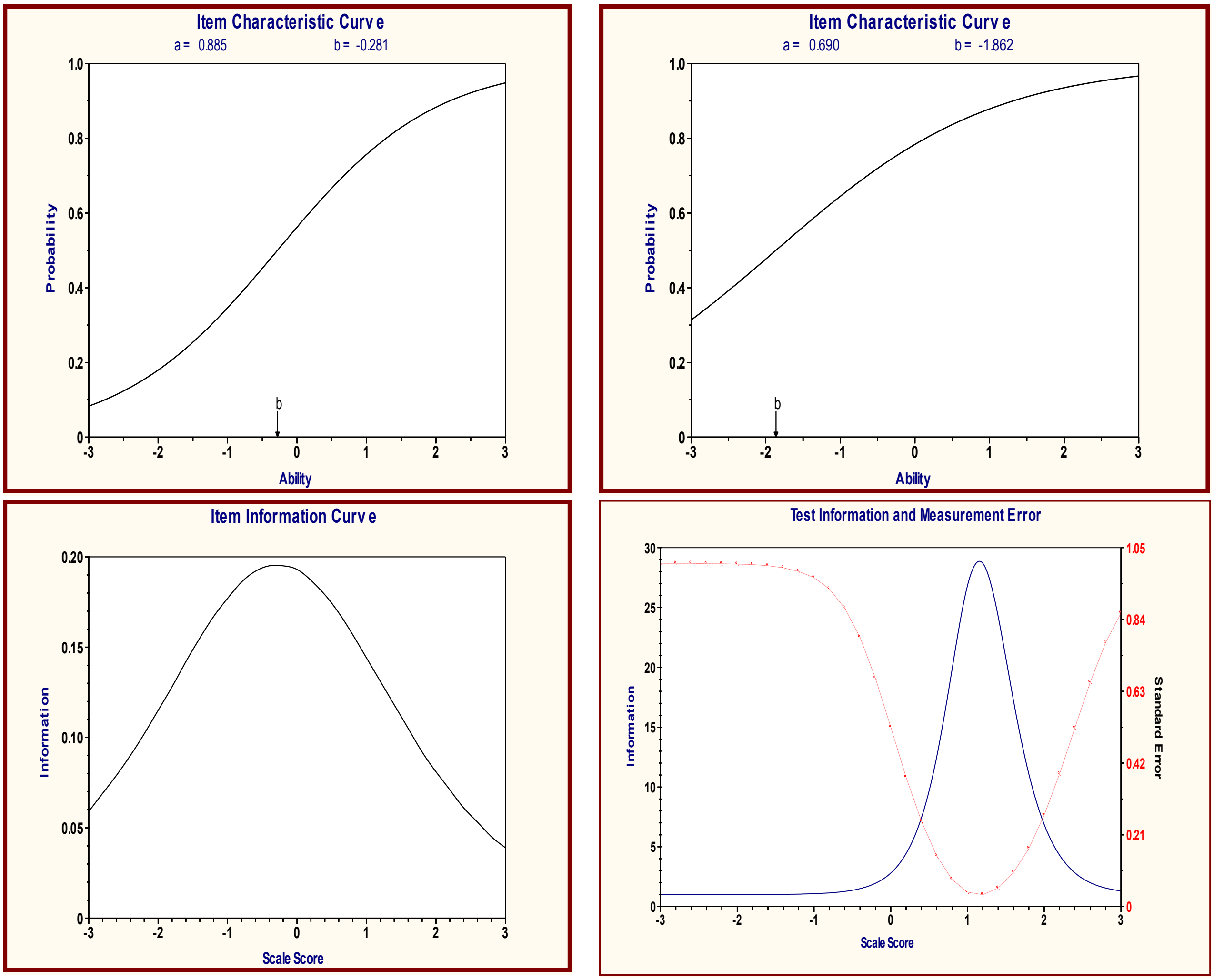

Figure 1 Graphs used to explain item response theory in the text.

dimension as a whole, the TIF values were relatively low up to around the mean trait level. They were relatively high from the mean level onwards. Thus, taken together, the ten internalizing disorders measured the higher half of the internalizing continuum better than the lower half. The TIF was highest at 1.4 SD from the mean, and the SEM at this level was .40. However as shown in the IIF graphs in Figure 4, MDD, DYS, PD, GAD and AG contributed relatively more information to the internalizing dimension than SPP, SAD, PTSD, OCD and SOP. Also, the IIF values for all other disorders, except GAD and DYS, were very low up to around the mean trait level. GAD and DYS had relative high IIF values from around -1 SD to around 2 SD from the mean, with the values being higher for most of this region for GAD than DYS.

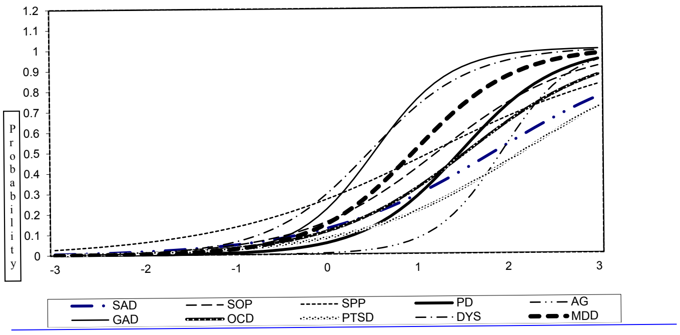

Figure 2 Item characteristic curve (ICC) for the internalizing disorders.

Concurrent and Discriminant Validities of the Internalizing Factor

Correlation analyses indicated that the EAP scores were highly correlated with the number of diagnoses (r = .98, p < .001). They were also significantly correlated with the internalizing scores of the CBCL (r = .35, p < .001, N = 617), TRF (r = .21, p < .001, N = 293), and YSR (r = .21, p < .001, N = 578). Based on guidelines suggested by Hemphill, all these correlations were of large or medium effect sizes. These findings are supportive of the concurrent validity of the internalizing latent dimension. There were also significant, but negative correlations with the externalizing scale scores of the CBCL (r = -.08, p < .05; N = 617) and the TRF (r = -.14, p < .05, N = 293). The correlation with YSR was significant and positive (r = .09, p < .05; N = 578). However the effect size was negligible. Together these findings provide support for the discriminant validity of the internalizing latent dimension.

Discussion

The first aim of the study was to examine the IRT properties of the internalizing factor, comprising SAD, SOB, SPP, PD, AG, GAD, OCD, PTSD, DYS and MDD as indicators. For the 2-PLM analysis, the general findings were that all the disorders had high discrimination values, thereby indicating that they were all strong discriminators of the internalizing dimension. These general findings were as hypothesized, and are also consistent with the two previous studies involving adults [7,8]. The discrimination values can be inferred as indications of the strength of the associations of the indicators with the underlying latent factor. Although the findings that all the disorders had high discrimination values mean that they all have strong associations with the internalizing factor, a closer examination of these values suggest differences that are worth noting. SPP, SAD, PTSD, OCD and SOP (which all had equal associations with the latent factor) were not as strongly associated as the other disorders. Of the others, MDD, DYS and PD (which all had equal associations with the latent factor) were not as strongly associated as GAD and AG (which had equal associations with the latent factor). The findings here for the discrimination values correspondence partially with the previous IRT studies. Like this study, Krueger and Finger [7] found relatively higher values for PD and GAD, and McGlinchey and Zimmerman [8] found relatively higher value for GAD. Like this study, McGlinchey and Zimmerman found the lowest value for SPP. Taken together with the findings in the current study, it would appear that GAD and AG, and to a lesser degree, PD have better ability to identify individuals with different levels of the internalizing dimension than the other disorders. In contrast, SPP has relatively lower ability.

For the difficulty values, all disorders were located close to or above the mean trait level. This indicates that most of the disorders were more representative of the internalizing dimension in the upper half of the internalizing trait continuum. These general findings were as hypothesized, and are also consistent with the two previous studies involving adults [7,8]. Also, like this study, the previous studies found relatively low difficulty value for MDD. Despite this, there was little similarity for the other disorders. While both this and the McGlinchey and Zimmerman studies found low difficulty values for GAD, this value was high in the study by Krueger and Finger. DYS had the lowest value in the current study, while it had the highest value in the study by Krueger and Finger. As already noted there was also much variability in the findings across the Krueger and Finger study and McGlinchey and Zimmerman study. McGlinchey and Zimmerman explained these differences in terms of the frequencies of comorbidity in the samples examined, and the different sets of disorders examined. These explanations are also likely to be applicable for the differences in this and the previous studies. In addition it is also likely that developmental differences may account for these differences since this study examined adolescents, while the previous studies examined predominantly adults. Clearly this is an area worthy of future research.

Another general finding was that the TIF values were also much higher in upper half of the internalizing trait continuum, thereby suggesting that the disorders, as a whole, provided more measurement precision in the upper half of the internalizing trait continuum, but not the lower half. Also it peaked at around 1.4 *SD* from the mean. All these findings are also consistent with that reported in the two previous studies involving adults [7,8]. The findings here also showed that although the disorders, as a whole, provided more measurement precision in the upper half of the trait continuum, but not the lower half, this was not the case for GAD and DYS. These disorders had relatively higher precision from around -1 *SD* to around 2 *SD* from the mean, with

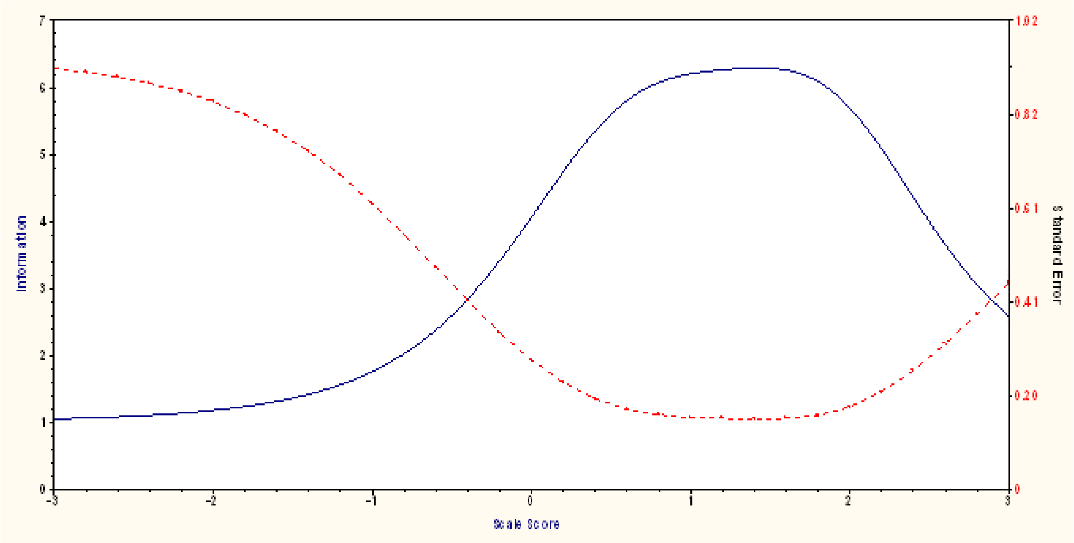

Figure 3 Test information function (TIF) and standard error (dotted line) curves for the internalizing factor.

the values being higher for most of this region for GAD than DYS. Taken together these findings suggest that relative to the other disorders, GAD and DYS are more reliable for measuring the internalizing dimensions across a broader trait spectrum, with GAD being more reliable than DYS.

The findings here showed that participants' EAP scores correlated close to unity with their number of diagnoses and with either large or medium effect sizes with the internalizing scores of the CBCL, TRF and YSR. In contrast, the correlations with the externalizing scores of the CBCL, TRF and YRF were either negative or had a negligible effect size. These findings can be interpreted as supporting the concurrent and discriminant validities, respectively, of the internalizing dimension. The support for the concurrent validity of this dimension is consistent with the findings of past studies in this area [7,8]. The previous studies have also shown that this dimension is correlated close to unity with the number of diagnoses, and is positively associated with social burden.

The findings here also have implications for broad groupings and understanding of the comorbidity of the internalizing disorders. Support for the 1-factor model suggests that from a psychometric viewpoint all the disorders examined in this study could be grouped together in one broad overarching category of emotional disorders [23]. The high α values (all above.85) for all the disorders in the 2-PLM analysis can be inferred as indications of the strength of the associations of the disorders with the underlying latent factor, and by extension the comorbidity of the disorders. Consistent with past studies (for a meta-analysis studies, see [24,25]), these findings suggest high comorbidity among the internalizing disorders. However as SPP, SAD, PTSD, OCD and SOP had comparable discrimination values, it can be assumed that there will be relatively higher comorbidity between these disorders. Similarly as MDD, DYS and PD had equal discrimination values, these disorders could have relatively higher comorbidity with each other. For the same reason GAD and AG could have relatively higher comorbidity with each other. The difficulty values found for the internalizing disorders provide further insights into the nature of the comorbidity. In general, the difficulty value of a disorder indicates the point on the scale of the latent trait where that disorder has a 0.5 probability of being diagnosed as present. Thus, compared to an internalizing disorder with a low difficulty value, an internalizing disorder with a high difficulty value would have to have more of the internalizing trait for a diagnosis. For this study, the difficulty values for DYS and GAD were equal and close to 0.5 *SD* from the mean; MDD, SPP and SOP were equal and close to around the mean; PD and OCD were equal and close to around 1.5 *SD* from the mean; and SAD, AG and PTSD were equal and around 2 *SD* from the mean. Seen in relation to the comorbidities suggested by the discrimination values, these findings would imply that (1) PTSD will be more comorbid with SPP, SOP, OCD, SAD than SPP, SOP, OCD, SAD with PTSD; (2) SAD will be more comorbid with SPP, SOP and OCD than SPP, SOP and OCD with SAD; (3) OCD will be more comorbid with SPP and SOP than SPP and SOP with SPP; (4) SOP will be more comorbid with SPP than SPP with SOP; (5) PD will be more comorbid with DYS and MDD than DYS and MDD with PD; (6) MDD will be more comorbid with DYS than DYS with MDD; and (7) AG will be more comorbid with GAD than GAD with AG.

The findings also have implications for clinical practice Firstly, the close associations between the anxiety and depressive disorders found in the study highlight the need for a comprehensive evaluation of all the internalizing disorders for a better understanding of an adolescent's psychopathology. As this study found very close associations between DYS and GAD; MDD, SPP and SOP; PD and OCD; and AG and PTSD, differential diagnosis of these sets of disorders would be challenging. Secondly, the findings imply that treatment of anxiety and depressive disorders may have to focus on general distress with special focus on the range of associated abnormal mood, anxiety and fear responses rather than the individual disorders. In this respect, recently developed transdiagnostic treatment approaches for anxiety and depression disorders in children and adolescents would be valuable [26]. In brief, transdiagnostic approaches focus on common factors that produce symptoms in related classes of disorders, such as anxiety and depression, thereby addressing multiple concerns or disorders within an individual [27].

There are several strengths to the current study. First, unlike previous studies that omitted one or more of the anxiety disorders, this study included all the DSM-IV/DSM-IV TR anxiety disorders, and thus the finding here are more comprehensive. Second, this study examined the IRT properties for a clinic-referred sample of adolescents, and therefore the findings can be seen as being more useful from a clinical viewpoint. Despite these strengths, there are limitations that need to be considered when interpreting the findings here. First, since the nature, incidence, age of onset, course, stability and comorbidity of the depressive and anxiety disorders are different across children, adolescents, and adults [28,29], the findings may not be applicable to either children or adults. Second, about 70% of the participants in the current study had either ADHD or CD/ODD or both. As they were not controlled it is uncertain if they exerted any influence on the findings. Third, all the participants in this study were from the same clinic. Thus it is possible that this may constitute an additional bias thereby, limiting the findings and conclusions made in this study. Fourth, as this study examined a clinic-referred adolescent sample, the findings here may not be applicable to comorbidity of the depressive and anxiety disorders in adolescents from the general community. Fifth, is the appropriateness of the application of 2-PLM in the study. This model assumes that traits are bipolar, that is, both ends of the trait continuum scale represent meaningful variation of the trait. Thus the mean score of the latent trait is defined as zero, with low scores reflecting levels below the average levels. According to Reise and Waller [30], many clinical constructs could be unipolar where one end of the trait continuum represents severity and the other end represents its absence. Lucke [31] has suggested that for such traits, the person with certain amount of the trait (or disorder) has to be reference to the level of no trait (or disorder), and not the mean. This implies that low scores represent the absence of the trait and not scores below the average, and thus zero is the lowest possible latent trait score. He developed new IRT models (called unipolar item response models), and illustrated their applications with reference to a gambling addiction scale. Although such models may seem as viable alternative to the 2-PLM for application in the current study, Lucke has pointed out that the assumption in unipolar item response models that the probability of item endorsement is zero for those persons with a trait level at zero does not necessarily apply to other unipolar traits. Given this, it does not make sense to diminish the relevance of the 2-PLM for the current study. Additionally, the application of the 2-PLM in the current study allowed us to compare the findings in the current study with the findings in the previous studies [7, 8]. In concluding, given the limitations highlighted here, there is a need for cross-validation of the findings before they can be generalized. It will be useful for future studies to examine samples from several clinics and from the general community, keeping in mind the limitations mentioned here.

References

1 American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders. (4th ed.) Arlington, VA, US: American Psychiatric Publishing, Inc.

2 American Psychiatric Association (2000) Diagnostic and statistical manual of mental disorders. (4th ed.) Text Revision. Washington, DC: American Psychiatric Publishing, Inc.

3 American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders. (5th ed.) Washington, DC: American Psychiatric Publishing, Inc.

4 Clark LA, Watson D (1991) Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. J Abnorm Psychol 100: 316-336.

5 Krueger RF (1999) The structure of common mental disorders. Arch Gen Psychiatry 56: 921-926.

6 Mineka S, Watson D, Clark LA (1998) Comorbidity of anxiety and unipolar mood disorders. Annu Rev Psychol 49: 377-412.

7 Krueger RF, Finger MS (2001) Using item response theory to understand comorbidity among anxiety and unipolar mood disorders. Psychol Assess 13: 140-151.

8 McGlinchey JB, Zimmerman M (2007) Examining a dimensional representation of depression and anxiety disorders' comorbidity in psychiatric outpatients with item response modeling. J Abnorm Psychol 116: 464-474.

9 Seeley JR, Kosty DB, Farmer RF, Lewinsohn PM (2011) The modeling of internalizing disorders on the basis of patterns of lifetime comorbidity: associations with psychosocial functioning and psychiatric disorders among first-degree relatives. J Abnorm Psychol 120: 308–21.

10 Gomez R, Vance A, Gomez RM (2014) The factor structure of anxiety and depressive disorders in a sample of clinic-referred adolescents. J Abnorm Child Psychol 42: 321-332.

11 Embretson SE, Reise SP (2000) Item Response Theory for Psychologists. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

12 Steinberg L, Thissen D (1995) Item response theory in personality research. In: Shrout PE, Fiske ST, editors. Personal. Res. methods, theory A festschrift Honor. Donald W. Fisk., Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc 161-181.

13 de Ayala RJ (2009) The theory and practice of Item Response Theory. New York, NY, US: Guilford Press.

14 Silverman WK, Albano AM (1996) The Anxiety Disorders Interview Schedule for DSM-IV: Child Interview Schedule. San Antonio, Texas: Psychological Corporation.

15 Silverman WK, Saavedra LM, Pina AA (2001) Test-retest reliability of anxiety symptoms and diagnoses with the Anxiety Disorders Interview Schedule for DSM-IV: child and parent versions. J Am Acad Child Adolesc Psychiatry 40: 937-944.

16 Achenbach TM, Rescorla LA (2001) Manual for the ASEBA School-Age Forms & Profiles: An integrated system of multi-informant assessment. Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families.

17 Thissen D (1991) MULTILOG: multiple category item analysis and test scoring using item response theory.

18 Stark S, Chemyshenko S, Chua WL, Wadlington P (2003) Computing chi-square statistics and fit-plots using the MODFIT program.

19 Bock RD, Aitkin M (1981) Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. Psychometrika 46: 443-459.

20 Hemphill JF (2003) Interpreting the magnitudes of correlation coefficients. Am Psychol 58: 78-79.

21 Muthén LK, Muthén BO (2012) Mplus user's guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén.

22 Hu L, Bentler PM (1998) Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. Psychol Methods 3: 424-453.

23 Watson D (2005) Rethinking the mood and anxiety disorders: a quantitative hierarchical model for DSM-V. J Abnorm Psychol 114: 522-536.

24 Angold A, Costello EJ, Erkanli A (1999) Comorbidity. J Child Psychol Psychiatry 40: 57–87.

25 Krueger RF, Markon KE (2014) The role of the DSM-5 personality trait model in moving toward a quantitative and empirically based approach to classifying personality and psychopathology. Annu Rev Clin Psychol 10: 477-501.

26 Ehrenreich-May J, Bilek EL (2012) The development of a transdiagnostic, cognitive behavioral group intervention for childhood anxiety disorders and co-occurring depression symptoms. Cogn Behav Pract 19: 41–55.

27 McEvoy PM, Nathan P, Norton PJ (2009) Efficacy of transdiagnostic treatments: A review of published outcome studies and future research directions. J Cogn Psychother 23: 20–33.

28 Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE (2005) Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 62: 617-627.

29 Wittchen HU, Lieb R, Schuster P, Oldehinkel AJ. When is onset? Investigations into early developmental stages of anxiety and depressive disorders. In: Rapaport J, editor. Child. Onset "Adult" Psychopathol. Clin. Res. Adv., Washington, DC, US: American Psychiatric Press, Inc; 1999: 259-302.

30 Reise SP, Waller NG (2009) Item response theory and clinical measurement. Annu Rev Clin Psychol 5: 27-48.

31 Lucke JF. Unipolat item response models. In Reise SP, Revicki DA, editors. Handbook of item response theory modeling: Applications to typical performance assessment. NY, Routledge: 2015: 272-284.