NIH-PA Author Manuscript

Computer adaptive testing of liability to addiction: Identifying individuals at risk.

Levent Kirisci, Ph.D.[1], **Ralph Tarter, PH.D.**[1], **Maureen Reynolds, Ph.D.**[1], **Ty Ridenour, Ph.D.**[1], **Clement Stone**[2], and **Michael Vanyukov, Ph.D.**[1]

[1]Department of Pharmaceutical Sciences, Center for Education and Drug Abuse Research (CEDAR), University of Pittsburgh, 3520 Forbes Avenue, Suite 203, Pittsburgh, PA 15213, USA

[2] Department of Psychology in Education, School of Education, University of Pittsburgh, WWPH 5920, Pittsburgh, PA 15260, USA

Abstract

Background—Employed as a quantitative measure of substance use disorder (SUD) risk, the transmissible liability index (TLI) can be useful for detecting youths requiring prevention intervention. This study was conducted to develop and evaluate a computer adaptive test (CAT) version of the TLI to identifying individuals at risk for SUD.

Methods—In the first sample (N=425) of male and female subjects were recruited under aegis of the Center for Education and Drug Abuse Research in Pittsburgh, Pennsylvania, having a mean age of 18.8 years. A provisional CAT version of the TLI was assessed using simulation procedures. In sample 2, twins were recruited at the 2010 Twinsburg Festival in Twinsburg, Ohio. The CAT and paper and pencil (P&P) versions of the TLI were administered to 276 twin pairs having a mean age of 19.94 years.

Results—The simulated CAT version of the TLI predicted cannabis use disorder two years after initial study with 4% less accuracy (72% vs. 68%) than P&P version but with 78% reduction of items. In the twin sample, the CAT version predicted alcohol and drug use (OR=1.7 [2.1], p<.001) with 64% and 65% accuracy (sensitivity=75% [75%] and specificity =64% [65%]).

Conclusions—This study demonstrated that the CAT version of the TLI is an accurate and efficient measure of risk for SUD. The CAT version of the TLI potentially affords the opportunity for efficient screening of risk so that timely interventions can be implemented to prevent occurrence of SUDs having frequently lifelong consequences.

Keywords

Transmissible liability index; computerized adaptive testing; cannabis use disorder; item response theory

1. INTRODUCTION

It has been long known that substance use disorder (SUD) commonly runs in families. The portion of variance in risk manifest across generations for SUD due to the combined influences of genetic and environment comprises *transmissible liability* (Rice, et al., 1980). Responding to the call by the Genetic Consortium of the National Institute on Drug Abuse for a straightforward assessment of intergenerational risk for SUD (Conway et al., 2010), investigators at the Center for Education and Drug Abuse Research (CEDAR) spearheaded development of the transmissible liability index (TLI) (Vanyukov et al., 2009; Kirisci et al., 2009) which is a continuous trait encompassing the psychological and psychiatric characteristics comprising heritable risk for substance use disorder. Psychometric analyses have shown that the TLI has excellent internal reliability, discriminative validity and predictive validity (Vanyukov et al., 2009; Kirisci et al., 2009). The TLI has IRT-based reliability coefficient of .93. The mean difference is about .5 SD in the TLI scores between high risk (sons of SUD+ fathers) and low risk (sons of SUD- fathers). The heritability coefficient is estimated as $h^2 = .79$ (95% CI: .73, .84). This high heritability supports the TLI's construct validity as an index of transmissible risk for SUD.

The TLI at age 10-12 predicts development of cannabis use disorder by age 22. A modified TLI based on items contained in the National Epidemiological Survey of Alcohol and Related Conditions predicts all SUD categories (Ridenour et al.,2011). Another version of the TLI adapted for college students distinguishes freshman who subsequently developed SUD in the ensuing four years from peers who do not develop SUD (Arria et al., 2009). Paralleling results obtained on the CEDAR sample (Vanyukov et al., 2009; Kirisci et al., 2009), Hicks et al (this issue) found that genetic factors in the Minnesota Family Twin Study sample account for over 80% of TLI variance.

Employed as a quantitative measure of SUD risk, the TLI may, therefore, be useful for detecting youths requiring prevention intervention. Adoption of an assessment tool for use in practical settings is, however, contingent on satisfying several important criteria. In particular, the time required for administration and scoring cannot be burdensome to staff and clients. Accordingly, lengthy questionnaires not only detract from treatment delivery, but may also incur unacceptable cost. Indeed, fixed length tests may not even be the optimum method of measurement. As discussed by Weiss (2004), some items in fixed length instruments may contribute to error because measurement precision declines at both the high and low level of the trait.

Computer adaptive testing (CAT) provides a solution to these problems. It mitigates measurement error while maximizing efficiency since only the items pertinent to accurately measuring trait level are administered. Moreover, cost is minimal because scoring the responses is conducted automatically and immediately after completion of the questionnaire. Privacy is also ensured because there is no record of the person's responses on paper and access to the information is protected by password. These advantages have led to adoption of the CAT format in research to evaluate mental health and psychopathology (Walter et al., 2007; Fliege et al., 2005; Gardner et al., 2002; Gibbons et al., 2008; Roper et al., 1991; Handel et al., 1999), quality of life (Peterson et al., 2006), and personality traits (Forbey et al, 2007; Waller and Reise, 1989). To date, CAT procedures have not been used to assess

either substance use behavior or risk for SUD even though this is the increasingly preferred administration format for evaluating personality and psychopathology.

This investigation determined whether a CAT version of the TLI accurately measures risk for cannabis use disorder. Efficiently monitoring risk for this disorder across time is not only valuable for fiscal reasons but also because it affords the opportunity to obtain better compliance from the respondent due to abbreviated time required for the assessment. Hence, for example, using a CAT version of the TLI is an efficient surveillance tool that can potentiate the likelihood of expeditious intervention when an increase in SUD risk severity is observed.

Cannabis is the most widely used illegal drug in the world. In the U.S. consumption has increased among high school students, college students and young adults during the past several decades. For example, 30-day prevalence has increased from 13.8% to 19.4% in 12th grade students, and from 14.1% to 17.0% in college students between 1991 and 2008 (Johnston et al.,2011). Notably, the finding that cannabis use disorder typically manifests by age 18-19 (Wagner and Anthony, 2007) indicates that prior development during childhood and adolescence is critical to this outcome.

Significantly, previous research has shown that the TLI developed youths predicts cannabis use disorder by age 22 (Kirisci et al., 2009). However, before undertaking a long-term prospective study to determine whether a CAT format of the TLI is appropriate and accurately predicts cannabis use disorder in young children, this study evaluated the efficiency and accuracy of a CAT version of the TLI in young adults and examined the similarity of scores obtained with the already validated paper and pencil version.

2. METHOD

2.1 Participants

Development of the CAT version of the TLI was conducted in two stages. First, a prototype was developed using simulation procedures on the sample longitudinally tracked by the Center for Education and Drug Abuse Research (CEDAR). This sample enabled determining whether the CAT prototype derived using simulation procedures has validity for predicting cannabis use disorder. In a second sample, data were collected using the CAT and paper and pencil formats so as to compare the scores obtained using each procedure.

2.1.1 CEDAR sample—The method of recruitment and ascertainment of the sample has been previously described (see Tarter et al., this issue). This sample consisted of 318 males and 107 females having a mean age of 18.8 years (SD = .49) who were evaluated two years later to determine transition to cannabis use disorder. Table 1 summarizes key personal and demographic variables in the retained and attrited (21.4%) segments of the sample. As can be seen, education level, socioeconomic status, and rate of psychiatric diagnosis do not differ between these two groups. Attrition rate was also similar in European-American and African-American subjects. However, greater attrition was observed in males. Significantly, the mean TLI score was almost identical in the retained and attrited segments of the sample Also, rate of paternal and maternal SUD was not different between these two groups. Overall, these results indicate that there is no systematic attrition bias.

At the follow-up evaluation, diagnosis of cannabis use disorder (abuse or dependence) was formulated using DSM-III-R criteria because the DSM-IV taxonomy was introduced five years after this longitudinal project was initiated. Diagnoses were assigned by a clinical committee consisting of a psychiatrist certified in addiction psychiatry, another psychiatrist or psychologist, and the master-level clinical associates who administered the Structured

Clinical Interview for Diagnosis (SCID) (Spitzer et al., 1987) and other instruments in the research protocol that could inform about psychiatric disorder. At the second evaluation at age 22, 23.7% of the sample were diagnosed with cannabis use disorder.

2.1.2 Twinsburg sample—The sample consisted of 276 twin pairs (200 females and 76 males) who were monozygotic and 84 pairs (68 females and 16 males) were dizygotic. The sample was 19.9 (sd=4.50) years old and 88% of fathers and 85.6% of the mothers of twins were white. The paper and pencil version of the TLI was administered followed immediately by the CAT version. The sample was recruited at the 2010 Twinsburg Festival in Twinsburg, Ohio. Alcohol and cannabis use were assessed by self-report

2.2 Instrumentation

2.2.1 Transmissible Liability Index (TLI)—The rationale and method of deriving the TLI have been described in prior reports (Vanyukov et al., 2003a, b). In brief, exploratory factor analysis was conducted on items from questionnaires and interviews to derive psychological constructs that have been reported in the empirical literature to be associated with SUD risk. The factors that discriminated offspring of SUD+ and SUD- fathers were retained for further analysis following pruning of items that had low loading (<0.4). The remaining items in the construct were submitted to confirmatory factor analysis for verification of unidimensionality. The constructs that discriminated sons of SUD+/- fathers, were subsequently submitted to exploratory factor analysis to derive a second order factor. Confirmatory factor analysis, verifying unidimensionality of the trait, comprises the *transmissible liability index* (TLI). Item response theory (IRT) analysis was performed to calibrate the discrimination and threshold parameters of the items and to derive latent trait scores. The 65 items comprising the TLI, shown in Table 2, has an IRT-based reliability coefficient of .93.

2.3 Procedure

2.3.1 CEDAR sample—Written informed consent was obtained from the participants prior to administering the protocols employing the procedure approved by the University of Pittsburgh Institutional Review Board. The participants were also informed that all of the findings from this research were protected from disclosure by a Certificate of Confidentiality issued by NIDA to the Center for Education and Drug Abuse Research (CEDAR). The participants underwent a urine drug screen to ensure that the results were not confounded by the acute effects of psychoactive compounds or drug withdrawal. A positive result required rescheduling the evaluation. The protocols were administered in fixed order by research assistants who were blind to the diagnostic status of their parents.

2.3.2 Twinsburg Sample—After obtaining informed consent, the research protocol was administered in fixed order: The paper and pencil version of the TLI was administered first followed by the CAT version. The participants individually completed the paper and pencil and the CAT versions of the TLI. After the CAT version of the TLIwas completed, the number of items that were administered, the latent trait score, and the standard error of the estimate were stored in a text file. A demographic and medical history questionnaires asked 15 questions related to major medical conditions. Participants reported on heart disease (.8%), high blood pressure (1.9%), diabetes (.6%), cancer (0.3%), arthritis (1.4%), asthma (16.4%), allergies (33.1%), severe headaches (11.4%), epilepsy (0.8%), mental health problems (6.7%), drug and alcohol use (2.2%), chronic ear infections (4.7%), hearing problems (1.7%), cleft lip/palate (0%), and other birth defects (3.9%).

2.4 Statistical Analysis

2.4.1 CEDAR sample—Unidimensionality of the 65-item paper and pencil version of the TLI was first demonstrated using exploratory factor analysis. The ratio of the first eigenvalue to the second eigenvalue was computed as an indicator of unidimensionality (Lord, 1980; Hattie, 1985) along with percentage of variance explained by the first and second factors (Reckase, 1979). Unidimensionality is documented if the ratio of the first eigenvalue to the second eigenvalue is greater than 3 or if the first eigenvalue explains more than 20% of the variance, Confirmatory factor analysis was then used to verify unidimensionality. Factor loadings in the model were estimated using Mplus (Muthen and Muthen, 2001). Mplus uses the weighted least square means and variance adjusted parameter estimation method. Four indices of model fit were used: the χ^2 goodness-of-fit index, root mean square error approximation (RMSEA), comparative fit index (CFI), and Tucker-Lewis index (TLI). A non-significant χ^2 value (p .05) indicates that the data are consistent with the model. RMSEA values greater than .08 reflect poor model-data fit, values between .05 - .08 indicate acceptable fit, and values of less than .05 reflect good fit (McCallum et al., 1996). For the CFI and TLI, values greater than .90 and .95 indicate good model fit (Loehlin, 2004).

Next, the items were calibrated and IRT-based TLI scores were obtained using MULTILOG 7 (Thissen, 2003) for the entire set of items. MULTILOG is the preferred method to analyze items having mixed item response formats using a graded response mode (GRM). The graded response model (GRM) was selected to calibrate the items and estimate the latent trait scores (Samejima, 1969). In this model, each response is characterized by an item discrimination parameter and item threshold parameters (one less than the number of response categories).in addition, MULTILOG was used to test nested IRT GRM models. A CAT simulation was conducted next using Firestar (Chou, 2009). The question closest to the median trait level was the first item administered. Administering items was terminated when the standard error of estimate reached .30 (determined to be the optimum value after conducting several simulation studies) or after it was determined that administering more items had negligible impact on the final standard error of estimate. The expected *a posteriori* method was employed to estimate the TLI trait level (Embretson and Reise, 2000). The Firestar simulation analysis utilized item parameters that were already calibrated in the CEDAR sample. The paper and pencil and CAT simulated versions of the TLI were then correlated. Lastly, logistic regression analysis was used to predict cannabis use disorder at age 22 to demonstrate predictive validity of the CAT and paper and pencil versions of the TLI.

2.4.2 Twinsburg Sample—The CAT protocol, written in JAVA language (Stone and Weisman, 2005) used item parameters already calibrated in the CEDAR sample. Next, the CAT and paper and pencil scores of the TLI were correlated. In addition, the number of items administered relative to TLI severity score was plotted. Finally, conditional logistic regression, which takes into account dependency between scores of twin pairs, and also provides a robust standard error of estimates, was conducted to predict alcohol and drug use using the TLI scores obtained from the CAT and paper and pencil versions.

3. RESULTS

3.1. CEDAR Sample: Provisional Development of the CAT Protocol

3.1.1. Unidimensionality—Exploratory factor analysis and confirmatory factor analysis were performed to determine unidimensionality of the 65-item TLI. The ratio between the first eigenvalue ($\ell_1 = 21.83$) and the second eigenvalue ($\ell_2 = 3.97$) was 5.50. The first factor accounted for 32% of the variance whereas the second factor explained only 5.7% of the

variance, thereby indicating the TLI's unidimensionality (Reckase, 1979). Confirmatory factor analysis for a one factor structure of the covariance matrix revealed acceptable fit (chi-square=124.40, df=110, p=.16, RMSEA=.02, CFI=.98, TLI=.99).

3.1.2 Calibration of Item Parameters—Item parameters were estimated using the marginal maximum likelihood method in MULTILOG (Thissen, 2003). Two models of the graded response model were tested: 1) A graded response model with item discrimination parameters equal for all items, and 2) A graded response model with unequal item discrimination parameters for all items. The likelihood ratio (LR) test showed that the item discrimination parameters could not be set equal across items (LR=501.7, df=64, p<.001). Table 2 presents the item discrimination (slope) and item location (threshold) parameters. Figure 1 depicts test information function and standard error of the TLI scores. Subjects at elevated risk for SUD indicated by moderate and high TLI scores were measured more precisely than subjects having low risk for SUD.

3.1.3 Simulation Analysis—The paper and pencil version of the TLI was used to simulate the CAT format. The Firestar-Computerized Adaptive Testing Simulation Program (Chou, 2009) generated CAT scores for the TLI utilizing the item parameters shown in Table 2. The minimum and the maximum number of items administered were set at 8 and 20. The standard error of estimate threshold for terminating the administration of the TLI was set at .30. Figure 2 illustrates the distribution of the TLI scores obtained from the paper and pencil and the CAT versions of the TLI. As can be seen, the two distributions are almost identical; the correlation between two TLI versions is .95.

The paper and pencil version (M=.28, SD=.11) of the TLI has significantly lower standard error of estimate than the CAT (M=.37, SD=.10) version (t=47.28, p<.001). Figure 3 depicts the standard error of estimate of the CAT version of the TLI. Subjects who are at moderate and high risk for SUD have a smaller standard error score.

Figure 4 presents the average number of items administered using the CAT protocol of the TLI. The average number of items administered was 16.8 (SD=4.70). Subjects whose TLI scores were between +1 and +2 SDs above the mean required 8.1 items whereas subjects whose TLI score ranged up to +1SD above the mean is 10.8. The average number of items administered to subjects whose TLI scores were between -2 and -1SD below the mean was 20. Subjects whose TLI score ranged up to 1SD below the mean required 18.4 items to be administered. As expected, the CAT required fewer items to estimate the TLI score in high risk subjects.

The 10 most frequently administered items were: item #41 (100%) which was chosen as an initial item to be administered to all subjects, item #16 (90%), item #3 (87%), item #21 (84%), item #49 (81%), item #12 (79%), item #2 (76%), item #8 (72%), item #40 (70%), and item #11 (68%) (see Table 2).

3.1.4 Predictive validity—The TLI scores obtained using the paper and pencil (OR=2.94, p<.001, 95% CI=1.87-4.62) and CAT (OR=2.23, p<.001, 95% CI=1.47-3.40) versions predict cannabis use disorder diagnosis at age 22 with overall accuracy of 72% and 68%. The two versions have sensitivity of 75% and 70% and specificity of 64% and 58%. The substantial reduction of administration time and number of items using the CAT format result in only a 4% decrease in prediction accuracy.

3.2 Twinsburg Sample: Cross Validation

Figure 5 presents the distributions of scores of the CAT and paper and pencil version. The obtained score by the individual using the two versions is strongly correlated (r=.87). In

addition, the CAT version on average required administering 18.6 (SD=3.07) items compared to 65 items comprising the paper and pencil version. The paper and pencil and CAT versions predicted alcohol and drug use [OR=1.7 (2.1), p<.001] with 64% and 65% accuracy [sensitivity=75% (75%) and specificity = 64% (65%)]. As shown in Figure 6, subjects 1SD or higher above the mean were measured with greater precision than participants at low risk for SUD. In addition, standard error of estimates obtained from the paper and pencil (M = .34, SD = .11) and CAT (M = .42, SD = .10) versions are significantly different (t=22.61, p<.001).

3. DISCUSSION

To briefly recapitulate, this study demonstrated that the CAT version of the TLI at age 19 is an accurate and efficient measure of transmissible risk for SUD. High correlations were observed between the paper and pencil and the CAT versions of the TLI using simulated and real data. The CAT version of the TLI required administering an average of 16.8 items in the simulation study and 18.6 items in the cross-validation sample. In effect, the CAT version of the TLI reduced the number of items from the full length paper and pencil version by 71% (simulation) and 74% (real data). The CAT version of TLI also predicted cannabis use disorder diagnosis at age 22 with only 4% reduction of accuracy compared to the paper and pencil version.

The observation that the ten most frequently administered items using the CAT format denote conjointly behavior dysregulation and propensity for social norms violation aligns with findings showing covariance between the variety of SUDs in the DSM-IV and childhood externalizing disorder (Krueger et al., 2002). In broad terms, SUD manifest by early adulthood is an outcome of deviant socialization (Tarter et al., 2011). In effect, cannabis use, the necessary prodrome to CUD, is but one facet of illegal behaviors which via social selection and contagion promotes habitual use culminating in a rather brief interval in the diagnosis of CUD. Whereas externalizing behavior is a salient component of the TLI, it is important to emphasize that other attributes are also encompassed in transmissible risk. Nevertheless, the results herein underscores the importance of implementing prevention interventions during early childhood while it is opportune to bias the developmental trajectory toward normative socialization.

This is the first study to show that it is feasible to use a CAT format in young adults to assess risk for substance use disorder. However, several limitations of this study deserve mention. First, it should be emphasized that neither the CEDAR nor Twinsburg samples were randomly recruited. Moreover, the order of administering the CAT and paper and pencil was fixed. Because the paper and pencil version of the TLI was administered first to all subjects, it may have produced a systematic bias on the CAT results. Furthermore, the CAT protocol was evaluated in only young adults. However, based on these results, it is recommended that the accuracy and utility of the CAT format should also be investigated in younger populations. A reduction of over 70% in the number of items that need to be administered using the CAT format attests to its potential as a practical screening instrument. In this regard, it should be recognized that the complement of items constituting the TLI may not be most ideally suited for quantifying risk. Other characteristics that are not represented in the initial item pool may be pertinent to risk for cannabis use disorder. Indeed, the internal consistency coefficient of .93 suggests that all of the items comprising the current TLI version may also not be needed. Hence, further research is required to determine the final set of items. Lastly, it should be noted that the outcome variable in this study was cannabis use disorder. Research thus needs to be conducted to validate the CAT version of TLI for other SUD categories. These analyses could not be conducted in the present study due to the low rate of SUDs consequent to use of illegal drugs other than

cannabis use disorder. Although the TLI is highly likely an accurate predictor of other types of SUD based on theory (Vanyukov et al., 2003a,b) and data (Ridenour et al., in press), empirical verification of the CAT protocol nevertheless remains to be documented. A gender comparison was not conducted in either sample because of sample size restrictions. A further research is warranted to assess gender differences. In addition each item's performance needs to be contrasted across gender using differential item functioning to detect gender bias items.

Most prevention programs implement a uniform intervention for all individuals even though there is large variation in severity of risk. This study points to the utility of CAT procedures for quantifying and monitoring the transmissible component of SUD risk at the individual level. Reducing evaluation time to about 5 minutes illustrates that the TLI may have practical application such as routine screening in a variety of settings (e.g. prior to a medical checkup, beginning of school year, while in the waiting room before a session with counselor). Taking into account individual differences in severity of SUD risk enables calibration of intervention intensity to risk severity. Although tailoring intervention intensity to severity of the individual's risk for disorder is established practice for prevention of many medical disorders, this has not yet been adopted for prevention of psychiatric disorders, including SUD. The CAT version of the transmissible liability index (TLI), potentially affords the opportunity for efficient screening of risk so that timely interventions can be implemented to prevent occurrence of SUDs having frequently lifelong consequences.

Acknowledgments

REFERENCES

Arria AM, Vincent KB, Caldeira KM. Measuring liability for substance use disorder among college students: Implications for screening and early intervention. Am J Drug Alc Abuse. 2009; 35:233–241.

Chou SW. Firestar: computerized adaptive testing simulation program for polytomous item response theory models. Applied Psychol Measure. 2009; 33:644–645.

Conway KP, Levy J, Vanyukov M, Chandler R, Rutter J, Swan GE, Neale M. Measuring addiction propensity and severity: The need for a new instrument. Drug Alc Depend. 2010; 111:4–12.

Embretson, SE.; Reise, SP. Item Response Theory for Psychologists. Erlbaum; Mahwah, NJ: 2000.

Fliege H, Becker J, Walter OB, Bjorner JB, Klapp BF, Rose M. Development of a computer-adaptive test for depression (D-CAT). Qual Life Res. 2005; 14:2277–2291. [PubMed: 16328907]

Forbey JD, Ben-Porath Y. Computerized adaptive personality testing: review and illustration with the MMPI-2 computerized adaptive version. Psychol Assess. 2007; 19:14–24. [PubMed: 17371120]

Gardner W, Kelleher KJ, Pajer KA. Multidimensional adaptive testing for mental health problems in primary care. Med Care. 2002; 40:812–823. [PubMed: 12218771]

Gibbons RD, Weiss DJ, Kupfer DJ, Frank E, Fagiolini A, Grochocinski DK, Stover A, Bock RD, Immekus J. Using Computerized Adaptive Testing to reduce the burden of Mental Health Assessment. Psych Serv. 2008; 59:361–369.

Handel RW, Ben-Porath YS, Watt M. Computerized assessment with the MMPI-2 in a clinical setting. Psychol Assess. 1999; 11:369–380.

Hattie J. Methodology Review: Assessing unidimensionality of tests and items. Applied Psychological Measurement. 1985; 9:13–164.

Johnston, L.; O'Malley, P.; Bachman, J.; Schulenberg, J. Monitoring the Future National Survey Results on Drug Use 2009, 1975-2008, Volume 1, Secondary school students (NIH Publication No. 09-7402). National Institute in Drug Abuse; Bethesda, MD: 2009.

Kirisci L, Tarter R, Mezzich A, Ridenour T, Reynolds M, Vanyukov M. Prediction of cannabis use disorder between boyhood and young adulthood: Clarifying the phenotype and environtype. Am J Addict. 2009; 18:36–47. [PubMed: 19219664]

Krueger R, Hicks B, Patrick CJ, carlson SR, Iacono WG, McCue M. Etiological connections among substance dependence, antisocial behavior, and personality. Modeling the externalizing spectrum. Journal of Abnormal Child Psychology. 2002; 111:411–424.

Loehlin, JC. Latent Variable Models. 4th ed.. Lawrence Erlbaum; Mahwah, NJ: 2004.

Lord, FM. Applications of item response theory to practical testing problems. Erlbaum Associates; New York: 1980.

McCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. Psychol Methods. 1996; 1:130–149.

Muthen, LK.; Muthen, BO. Mplus User's Guide. 4th edition. Muthen & Muthen; Los Angeles, CA: 2001.

Peterson MA, Groenvold M, Aaronson N, Fayers P, Sprangers M, Bjorner JB. Multidimensional computerized adaptive testing of the EORTC QLQ-C30: Basic developments and evaluations. Qual Life Res. 2006; 15:315–329. [PubMed: 16547770]

Reckase MD. Unifactor latent trait models applied to multifactor tests: Results and implications. J Educ Stat. 1979; 15:65–75.

Rice J, Cloninger C, Reich T. General causal models for sex differences and the familial transmission of the multi-factorial traits: An application to human spatial visualizing ability. Soc Biol. 1980; 26:36–47. [PubMed: 7292035]

Ridenour T, Kirisci L, tarter R, Vanyukov M. National Epidemiological Study of Alcoholism and Related Conditions of the general U.S. population. Drug Alc Depend. in press.

Roper BL, Ben-Porath YS, Butcher JN. Comparability and validity of computerized adaptive testing with the MMPI-2. J Person Assess. 1991; 65:358–371.

Samejima F. Estimation of latent trait ability using a response pattern of graded scores. Psychometrika Monograph. 1969; N6:17.

Spitzer, R.; Williams, BW.; Miriam, G. Instruction manual for the structured clinical interview for DSM-III-R. New York State Psychiatric Institute; New York, NY: 1987.

Stone, CA.; Weisman, A. Simulating computer adaptive test. Final report for University of Pittsburgh, Central Research Development Fund; Pittsburgh, PA: 2005.

Tarter R,E, Fishbein D, Kirisci L, Mezzich A, Ridenour T. Deviant socialization mediatest ransmissible and contextual risk on cannabis use disoder deelopment: a prospective study. Addiction. 2011; 106:1301–1308. [PubMed: 21320228]

Thissen, D. *Multilog for Windows* (version 7.0) [computer software].. Scientific Software International; Lincolnwood, IL: 2003.

Vanyukov MM, Kirisci L, Moss H, Tarter RE, Reynolds MD, Maher BS, Kirillova GP, Ridenour T, Clark DB. Measurement of the risk for substance use disorders: Phenotypic and genetic analysis of an index of common liability. Behav Genetics. 2009; 39:233–244.

Vanyukov MM, Kirisci L, Tarter RE, Simkevitz HF, Kirillova GP, Maher BS, Clark DB. Liability to substance use disorders: 2. A measurement approach. Neurosci Biobehav Rev. 2003b; 27:517–526. [PubMed: 14599433]

Vanyukov MM, Tarter RE, Kirisci L, Kirillova GP, Maher BS, Clark DB. Liability to substance use disorders: 1. Common mechanisms and manifestations. Neurosci Biobehav Rev. 2003a; 27:507–515. [PubMed: 14599432]

Wagner R, Anthony J. Male-female differences in the risk of progression from first use to dependence upon cannabis, cocaine and alcohol. Drug Alc Depend. 2007; 86:191–198.

Waller NG, Reise SP. Computerized adaptive personality assessment: An illustration with the Absorption scale. J Person Soc Psychol. 1989; 57:1071–1059.

Walter OB, Becker J, Bjorner JB, Fliege H, Klapp B, Rose M. Development and evaluation of a computer adaptive test. Qual Life Res. 2007; 16:143–155. [PubMed: 17342455]

Weiss DJ. Computerized adaptive testing for effective and efficient measurement in counseling and education. Measurement Eval Counsel Develop. 2004; 37:70–84.
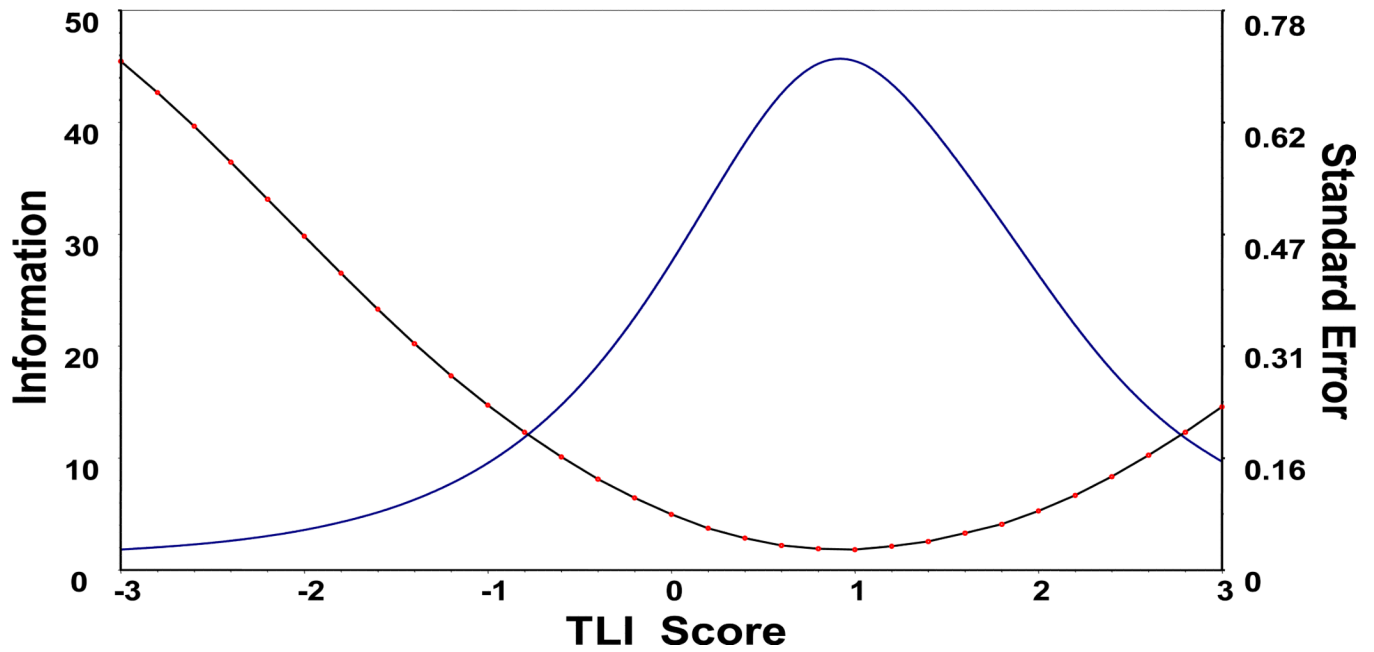
Figure 1.
Statistical information and standard error function of the paper and pencil TLI scores in the CEDAR sample of 425 subjects. Note: Information and standard error function indicate the total amount of information and precision the TLI scale at a given level of the IRT-based TLI score. Information=1/ SE.
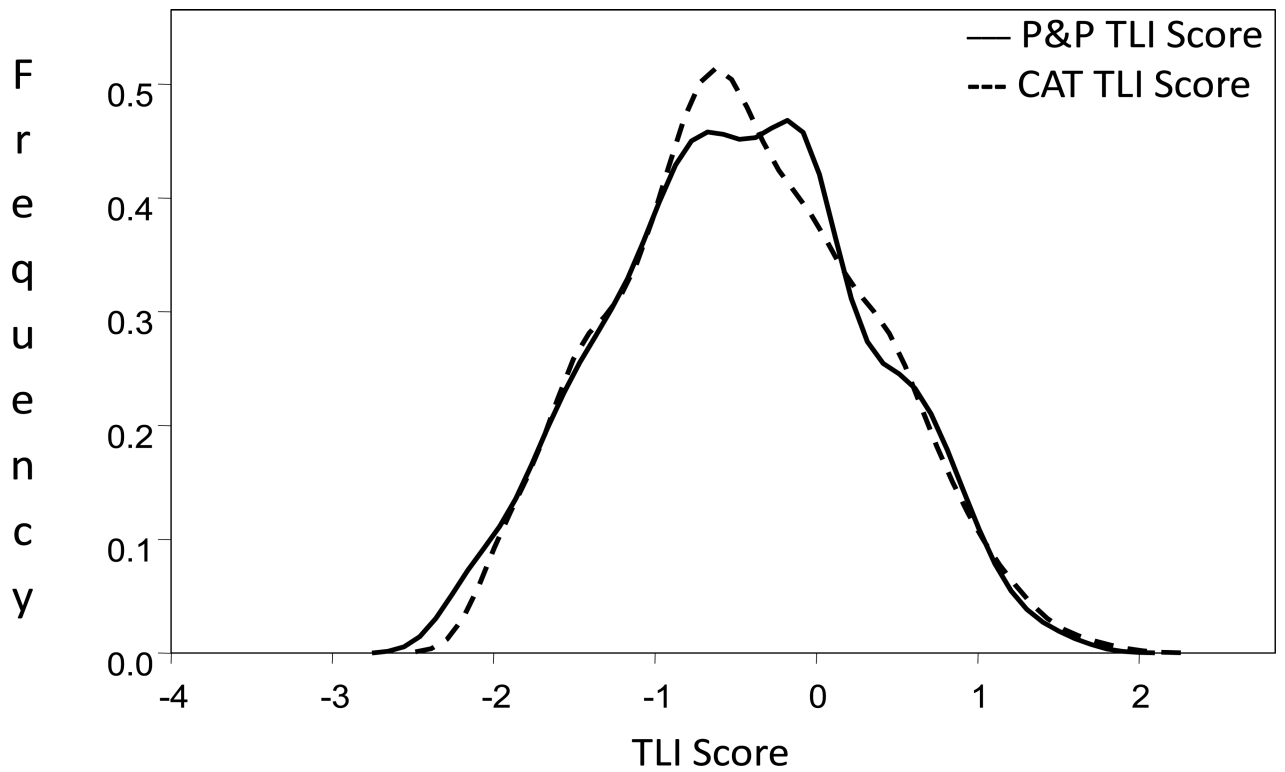
Figure 2.
Frequency distribution of paper and pencil and CAT TLI scale scores at a given level of the IRT-based TLI for 425 subjects in the CEDAR sample.
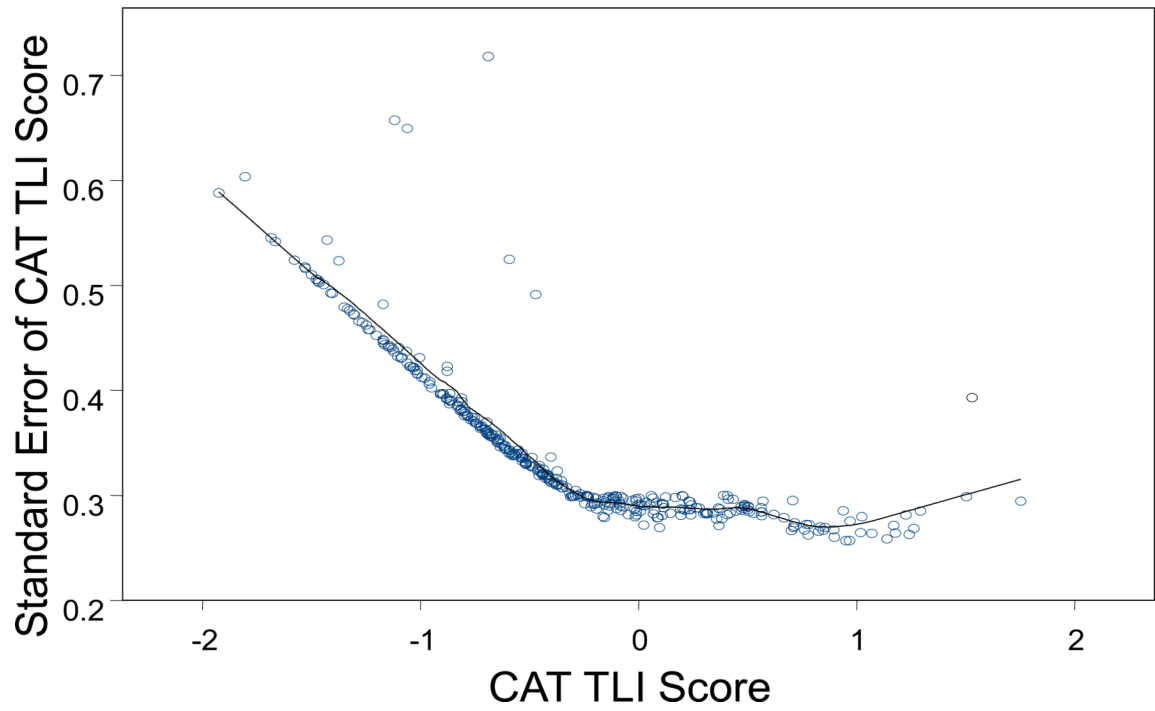
Figure 3.
Standard error of CAT TLI scale scores of 425 subjects at a given level of the CAT TLI in the CEDAR sample. Note: A plot suggests good measure precision for the majority of the TLI score range.
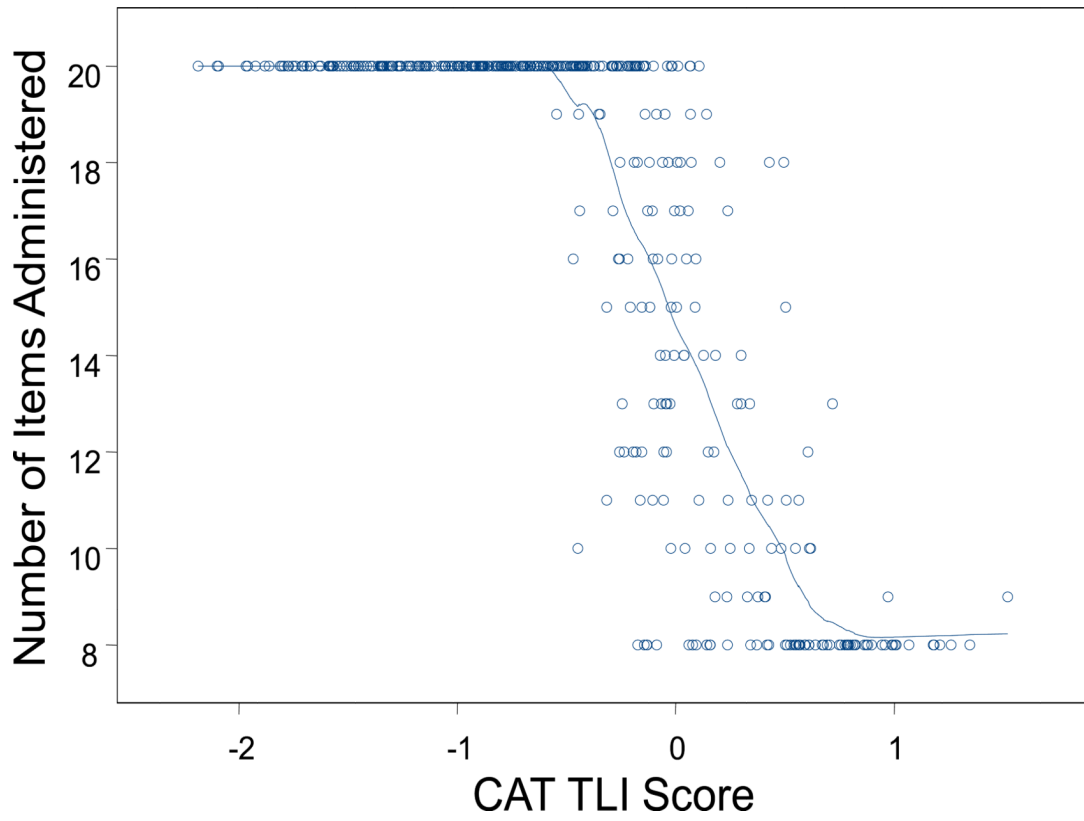
Figure 4.
Number of items administered by the CAT TLI at a given level of the IRT-based TLI in the
CEDAR sample of 425 subjects. Note: A maximum and minimum number of items
administered are set to 20 and 8, respectively.
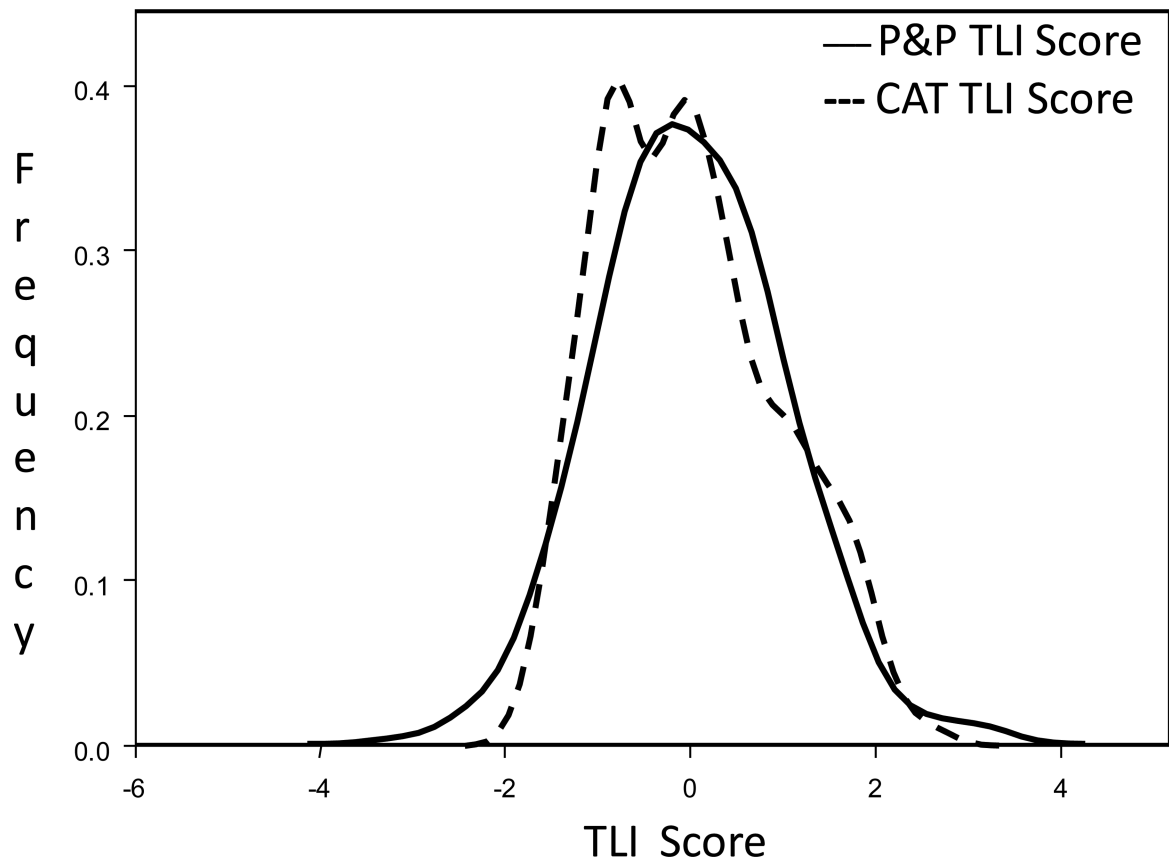
Figure 5.
Frequency distribution of paper and pencil and CAT TLI scale scores at a given level of
IRT-based TLI for 276 twin pairs in the Twinsburg sample.
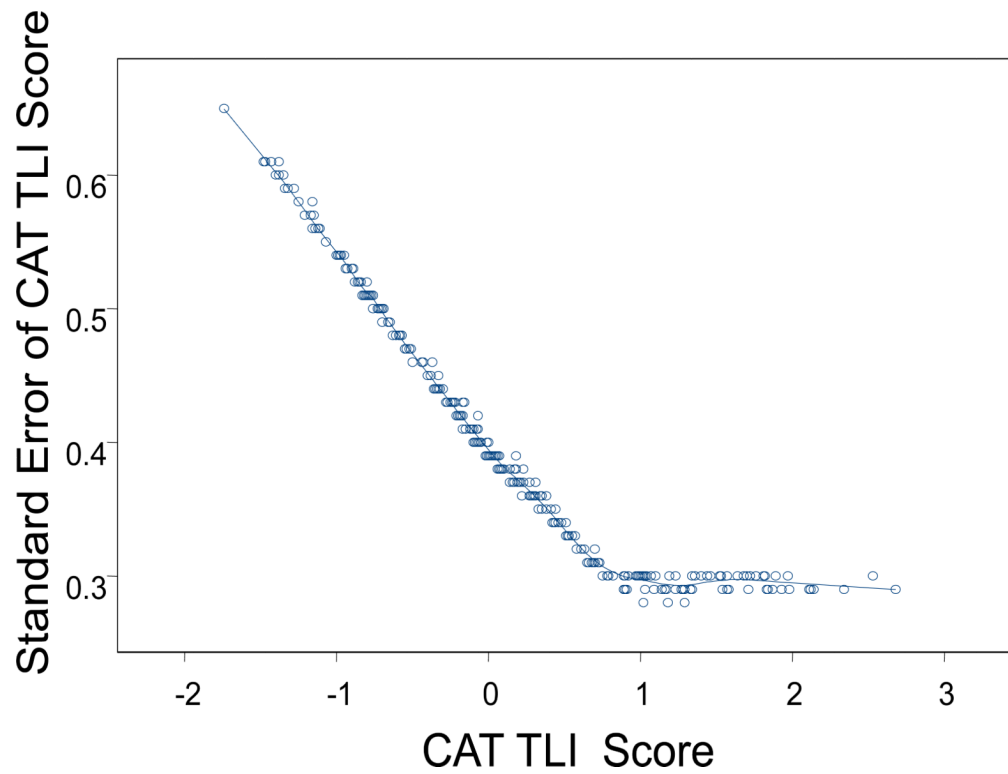
Figure 6.
Standard error of CAT TLI scale scores at a given level of IRT-based TLI of 276 twin pairs in the Twinsburg sample. Note: A plot suggests good measure precision for the whole of the TLI score range.

Table 1

Personal and demographic characteristics of attrited and retained segments of the sample

|	Attrited N = 91 Mean (sd)	Retained N=334 Mean (sd)	F	p
Socioeconomic Status[1]	39.82 (13.94)	41.59 (15.28)	.98	.323
Grade	12.30 (1.52)	12.07 (1.42)	1.67	.196
Transmissible Liability Index (TLI)[2]	-.05 (.78)	-.04 (.82)	1.78	.183
Female	16.5%	27.5%	χ^2=4.08	.043
Male	83.5%	72.5%		
European American	70.3%	73.4%		.566
African-American	29.7%	26.6%	χ^2=.33	
SUD	22.5%	23.1%	χ^2=.02	.900
Cannabis Use Disorder	17.7%	18.7%		.822
Depression	7.7%	12%	χ^2=.05	.248
Anxiety	9.9%	10.5%	χ^2=1.33	.870
Antisocial Personality	3.3%	3.3%		1.0
Disorder			χ^2=.03	
			χ^2=.0	.515
Father SUD	46.7%	50.5%		.963
Mother SUD	23.1%	23.3%		
			χ^2=.42	
			χ^2=.002	

[1] Hollingshead criteria

[2] IRT index score

Table 2

Items comprising the transmissible liability index (TLI)

	Item	Response category	Slope	Thresholds	Source
1	Over the past six months, have you destroyed things that belonged to others?	3	1.67	2.42/4.34	Youth Self Report[1]
2	Over the past six months, have you broken rules at school, work, or elsewhere?	3	1.32	0.97/3.24	
3	Over the past six months, have you been impulsive or acted without thinking?	3	1.54	.55/2.45	
4	Over the past six months, have you been biting your fingernails?	3	0.25	0.19/4.72	
5	Over the past six months, have you picked your skin or other parts of your body?	3	.60	2.82/6.05	
6	Over the past six months, have you stolen something?	3	1.87	1.88/3.42	
7	Over the past six months, have you had trouble finishing assignments?	3	.99	.20/2.66	
8	Over the past six months, have you had trouble concentrating or paying attention?	3	1.14	-.01/2.61	
9	Over the past six months, have you had trouble sitting still?	3	.89	.61/2.99	
10	Over the past six months, have you been pretty honest?	3	.93	1.13/4.44	
11	Over the past six months, have you been getting into a lot of fights?	3	2.29	1.73/2.73	
12	Over the past six months, have you had a hot temper?	3	1.31	0.83/2.48	
13	Over the past six months, have you threatened to hurt people?	3	2.13	1.63/2.95	
14	Over the past six months, have you failed to pay your debts or meet other financial responsibilities?	3	1.43	1.56/3.04	
15	Do you interrupt people when they are speaking?	4	0.64	-.65/4.73/7.66	Dysregulation Inventory Scale[2]
16	Within the past year, did you frequently do things without first thinking about the consequences?	2	1.78	1.12	Drug Use Screening Inventory[3]
17	Within the past year, did you steal things?	2	1.18	1.49	
18	Within the past year, did you have a bad temper?	2	1.32	1.33	
19	Within the past year, did you threaten to hurt people?	2	2.17	1.67	
20	Within the past year, did you frequently do risky or dangerous things?	2	1.68	1.41	
21	Within the past year, did you get into more fights than most people your age?	2	1.69	1.10	
22	Within the past year, did you have trouble concentrating?	2	.95	1.34	
23	Within the past year, did people stare at you?	2	1.53	1.69	
24	Within the past year, did you get tired very quickly when you exerted yourself?	2	1.25	1.72	
25	Did you get ever into trouble a lot for talking out of turn in school or talking without the teacher calling on you or for bothering people a lot?	2	1.28	2.30	Diagnostic Instrument[4]
26	Did you ever think about a specific plan to commit suicide?	2	1.02	5.07	

Item	Response category	Slope	Thresholds	Source
27 Did you ever do things to annoy people a lot like grabbing other childrens' hats?	2	1.64	2.18	
28 Did you ever frequently annoy people on purpose to get revenge?	2	1.47	2.21	

5. CEDAR (unpublished). Health Problem Checklist. School of Pharmacy, University of Pittsburgh, Pittsburgh, PA.

6. Andrew, J.M. (1974) Violent Crime Indices among community retained delinquents. Criminal Justice Behavior, 1, 123-130.

7. Tellegen, A. (1982). A Manual for the Differential Personality Questionnaire. Unpublished manuscript.

[1] Achenbach, T, & Edelbrock, C. (1991). Manual for the Youth Self-Report and 1991 Profile. Burlington, VT: University of Vermont, department of Psychiatry.

[2] Mezzich, A., Tarter, R., Giancola, P., & Kirisci, L. (2001).The Dysregulation Inventory: A new scale to assess the risk for substance use disorder. Journal of Child and Adolescent Substance Abuse, 10, 35-43.

[3] Tarter, R. (1990). Evaluation and treatment of adolescent substance abuse. A decision tree method. American Journal of Drug and Alcohol Abuse, 16, 1-46.

[4] Spitzer, R., Williams, J., & Gibbon, M. (1987). Instruction manual for the Structured Clinical Interview for DSM-III-R (SCID). New York, New York State Psychiatric Institute.