Published in final edited form as:

Subst Use Misuse. 2009 ; 44(6): 835–847. doi:10.1080/10826080802484686.

Creating a Developmentally Sensitive Measure of Adolescent Alcohol Misuse: An Application of Item Response Theory

Maria Orlando Edelen, Ph.D., Daniel M. McCaffrey, Ph.D., Phyllis L. Ellickson, Ph.D., Joan S. Tucker, Ph.D., and David J. Klein

RAND Corporation, Santa Monica, California

Abstract

This study, funded by the US National Institute of Drug Abuse, evaluates the usefulness of item response theory (IRT) to create a developmental alcohol misuse scale. Data were collected during 1997–2006 from 5828 Midwestern US students who completed annual surveys at Grades 7 through 11 and two and four years after high school. Seventeen alcohol misuse items were calibrated with IRT and examined for differential item functioning (DIF) across five study waves. Eight items displayed DIF; in most cases, properties for items assessed two years after high school were different from those assessed in grades 8–11. Implications and suggestions for future research are discussed.

Keywords

Alcohol misuse; adolescents; young adults; longitudinal; item response theory

Recognition of differences between adolescents and adults in patterns of alcohol use and the development of alcohol use disorders has led to the development of numerous tools to identify and assess alcohol problems among adolescents (Winters, 2003). However, there has been little effort to create a developmental scale of alcohol misuse that reflects the changing manifestation of this construct from middle school years to the transition out of high school. Item response theory (IRT) lends itself very well to this type of unidimensional scaling. First, because of IRT's built-in linking mechanism (Embretson, 1996), a developmental scale can be constructed that results in comparable scores across the age span (Edelen & Reeve, 2007; Embretson & Reise, 2000). Second, IRT is ideally suited to the detection of differential item functioning (DIF; Holland & Wainer, 1993). DIF in this context refers to items that have different measurement properties for individuals at different age levels after controlling for the overall differences in alcohol misuse between age levels.

The present study used IRT to examine how the measurement properties of alcohol misuse items change over time as adolescents mature, and to evaluate the value-added aspects of the developmental IRT-generated scores relative to simple scores. To address these goals, we used data from a longitudinal study evaluating the effects of a school-based drug prevention program (Ellickson et al, 2003). Annual assessments were conducted from grade 7 to grade 11, as well as at age 19 and 21, with several items representing alcohol misuse included at each assessment.

Methods

Participants

Participants were part of the Project ALERT field trial conducted in the Midwestern US. A total of 5857 seventh-grade students (85% of all students) from 62 middle schools were enrolled in the study. Students completed annual surveys during school classes at Grades 7 through 11 and completed two additional surveys by mail or via the internet approximately two and four years after completing high school. Trained staff administered the in-school surveys. Parental consent for participants under age 18 was solicited via approved procedures. Students were asked to assent at each survey administration.

A small percentage of students (.5%, $n=29$) had missing values for the alcohol misuse items at baseline and were excluded from these analyses. The analytic sample of 5828 students was 48% female and 86% non-Hispanic white (9% Native American, 5% other race/ethnicity). At baseline, when participant age ranged from 11–16 ($M=12.4$, $SD=.52$), approximately 1 in 4 students reported grades of C or below, 67% were living in two-parent households, and 61% reported some use of alcohol in their lifetime.

Missing Data

Of the 5828 students who participated at Grade 7, 92% ($n = 5333$) participated at Grade 8, 87% ($n = 5084$) participated at Grade 9, 83% ($n = 4853$) participated at Grade 10, 84% ($n = 4921$) participated at Grade 11, 57% ($n = 3294$) participated in the first post high school survey, and 57% ($n = 3332$) completed the final survey. Within each wave, there was also a small percentage of item-level missing data. Observations with incomplete data were included in all analyses by using the maximum likelihood estimation for incomplete data methods provided in Mplus (Muthen & Muthen, 1998–2003) and Multilog (Thissen, 1991), which assume the data are missing at random (Little & Rubin, 2002). However, observations in each study wave were required to have at least one non-missing alcohol misuse item for inclusion.

Measures

A total of 17 items were included at one or more study waves to measure *alcohol misuse*. Six of the 17 items were administered at all waves (see Table 1). Aside from the two derived dichotomous indicators (items 6 and 8), all items were assessed with a 6-point likert scale asking the number of times in the past year that each situation occurred ($0=none$ to $5=20$ or *more times*). Because of small frequencies in the upper categories of the response scale, the six-point responses were collapsed to three ($0=none$, $1=1-2$ times, $3=3$ or *more times*) for IRT analysis.

Correlation analyses were conducted with scales measuring *mental health* (the 5-item MHI5 (Ware, Kosinski, and Keller, 1996) at waves 1–6, median $\alpha = .781$), *impulsivity* (5 items at waves 1,3,5, and 6, median $\alpha = .829$), *parental support* (3 items at waves 2–5, median $\alpha = .822$), *school satisfaction* (3 items at waves 2–5, median $\alpha = .716$), *sensation seeking* (4 items at waves 3–6, median $\alpha = .804$), and *lack of guilt* (3 items at waves 3–5, median $\alpha = .628$).

Alcohol dependence was assessed at wave 7 with the AUDIT (Allen et al, 1997; Bohn, Babor, & Kranzler, 1995; Chung et al, 2000). The scale score ranges from 0 to 36 with values of 8 and above indicative of alcohol abuse or dependence.

Analytic Approach

IRT Calibration and DIF Assessment—To meet the assumptions of basic unidimensional IRT analysis, we first conducted exploratory factor analyses to ensure that the alcohol misuse construct, at each study wave was essentially unidimensional (McLeod, Swygert, & Thissen,

2001; Stout, 1987) for IRT modeling. Next we used SAS v 9 (SAS Institute, 1990) to select non-overlapping subsets of respondents ($n=750$) from each wave to create a calibration sample which was used to simultaneously calibrate the 17 alcohol misuse items across the study waves.

The alcohol misuse items were examined using Samejima's graded IRT model (Samejima, 1997) which specifies the relationship between an individual's observed responses to a set of items and the unobserved (or latent) trait that is being measured by the item set. Each item characteristic curve generated by the graded IRT model specifies the probability of choosing a response as a function of one item slope parameter (a) and a location (b) parameter for each response category transition. The slope reflects the strength of the relationship of the item to the construct being measured. The location parameters reflect the spacing of the item responses across the construct continuum.

To test for DIF, it is necessary to first identify a subset of items that are the same across two or more assessment waves (anchors) to link the waves together (Embretson, 1996). The items not used as anchors are then examined for DIF. We accomplished this in two steps. First, we conducted a series of adjacent wave comparisons ($w1$ vs. $w2$, $w2$ vs. $w3$, etc.) with IRTLRDIF (Thissen, 2001) to identify anchor items to be used for linking across the administrations. The second step involved a simultaneous calibration of all assessments using MULTILOG (Thissen, 1991). In this calibration, we specified a version of the graded IRT model that allowed item parameters for the linking (anchor) items to be equal across the waves, whereas parameters of non-linking items were free to vary. From this basic model, we employed an iterative process of imposing constraints on items across waves, first by constraining the slopes, then the location parameters. The suitability of each set of constraints was evaluated using the likelihood-based model comparison approach (Edelen et al, 2006; Thissen, Steinberg, & Wainer, 1993). This process continued until no additional constraints were possible without impacting the fit of the model. Due to the large number of tests, we required a significance level of $<.01$ to indicate misfit. Items in the final model displayed DIF if they did not have equal parameters across all waves of data in which they were administered.

Generating and evaluating developmental scores—We developed two sets of developmental scores for the alcohol misuse items across the six waves of data. Scores were generated for the entire sample. The first set, the IRT scores (IRT_W1-IRT_W6), were generated using MULTILOG based on the final simultaneous item calibration and incorporated all identified DIF. The second set was calculated by transforming the observed scores at each wave (OBS_W1-OBS_W6). Specifically, all items were reduced to dichotomous indicators ($0=none$, $1=2$ or more times) and the scale score within each wave was calculated as the proportion of endorsed items. The IRT scores were calibrated relative to a mean of 0 and standard deviation of 1 for the wave 4 sample, whereas the OBS scores were proportions, ranging from 0–1.

We conducted several analyses to examine the added value of the IRT scores versus the derived observed scores. We first looked at the basic descriptive statistics for the two sets of developmental scores including the mean, standard deviation and reliability at each wave. Next we compared the correlations of the alcohol misuse scores with several psychosocial measures to determine whether the increased precision of the IRT scores resulted in stronger associations with other variables. Finally, we used growth modeling implemented in Mplus (Muthen & Muthen, 1998–2003) to compare the extent to which the six waves of IRT versus OBS scores predicted alcohol dependence (as measured by the AUDIT) in the final wave of the study.

Results

Item Calibration and DIF detection

Based on exploratory factor analyses, the alcohol misuse items at each wave were deemed sufficiently unidimensional to meet the assumptions of IRT analysis. However, preliminary analyses revealed that items from wave 1 did not have sufficient variance to be modeled with IRT; thus, wave 1 data were excluded from further analyses and the wave 2 parameters were applied for purposes of generating wave 1 scores.

Using a linked specification based on results from the adjacent wave calibrations as a starting point, and specifying the wave 4 mean to 0 to set the scale (all other wave means were freely estimated), item parameters were constrained to equality and evaluated for DIF in an iterative fashion. Slope parameters were constrained first, followed by location parameters if slopes were identified as DIF-free. This iterative process resulted in a final model with item parameters displayed in Table 2. Overall, binge drinking (item 7) was the item most strongly related to the underlying construct of alcohol misuse with a slope value of 4.12 across waves 2–5 (ages 13 to 16). Weekly drinking also displayed a high slope (3.73), but not for measurement at wave 6 (age 19). The location parameters for all items were positive, indicating that the items are most useful for measuring alcohol misuse at the higher end of the continuum. Items reflecting binge drinking, weekly drinking, and feeling really sick were endorsed at lower levels of misuse, whereas getting into a traffic accident and getting into trouble at school or work were endorsed at higher levels of alcohol misuse.

Seven of the fifteen items that were administered at two or more waves did not display any DIF (Table 2: Items 3, 7–8, 11–14). Of the eight items displaying DIF (Table 2: Items 1–2, 4–6, 9–10, 15), only weekly drinking had different slope parameters across waves. Weekly drinking was a much stronger indicator of alcohol misuse at the earlier waves with a higher slope value at waves 2–5 relative to wave 6. Four of the items (felt really sick, did something later felt sorry for, weekly drinking, drove after drinking) had higher location values at the earlier waves. Thus, relative to older respondents with the same level of overall alcohol misuse, endorsement of these four items by younger respondents was indicative of higher levels of alcohol misuse. In contrast, the four items showing the opposite pattern, with lower location values at the earlier study waves relative to later (got into a fight, got into trouble at home, consumed alcohol and some other drug on the same occasion, drank before school or work), are indicative of higher levels of alcohol misuse or more severe misuse at the later waves.

Developmental score evaluation

Table 3 displays summary statistics for the two types of developmental scores, those created based on the IRT calibration (IRT_W1-IRT_W6), and those based on a non-parametric transformation of the observed responses (OBS_W1-OBS_W6). Both sets of scores increase monotonically with age. The IRT scores have increasing variance whereas the OBS scores' variances increase for the first few waves then level out after wave 3. At waves 1 and 2 the IRT scores have poor reliability at the sample mean. However, the reliability of the wave 1 and wave 2 IRT scores reaches a maximum of .93 at theta values about 1 standard deviation above the mean. Reliability for waves 3–6 is somewhat better for the IRT scores relative to the OBS scores.

Table 4 shows the correlations of the alcohol misuse scores with several psychosocial scales. In this table, correlations in bold print indicate values that are larger in magnitude by at least .005 in comparing the IRT with the OBS. Seventeen of the 25 correlations (68%) favored the IRT scores, although the difference in magnitude was generally small. Additionally, the IRT

score correlations were not consistently higher than the OBS correlations: four cases favored the OBS scores and four cases showed associations of comparable magnitude.

A final set of analyses was conducted to examine the relative ability of the IRT and OBS scores to predict alcohol dependence at wave 7. For these analyses, we estimated growth models of alcohol misuse scores from waves 2–6 and specified both the intercept and slope of the misuse scores as predictors of the AUDIT, a continuous measure of alcohol dependence. Table 5 summarizes the results from the growth models, which were fairly similar across the two types of scores (the intercept estimates are different because of the difference in means for the two types of scores). In both cases, the alcohol misuse slope was strong and positive indicating an increase in alcohol misuse across the study waves. The correlations between the intercept and slope were both negative and relatively small in magnitude indicating that as intercept values increase slope values decrease moderately. The coefficients for the alcohol misuse growth parameters predicting wave 7 alcohol dependence were also similar. For both scores, the intercept and slope were both significant predictors of alcohol dependence. However, the slope coefficient for the IRT score model was slightly larger than that for the OBS score model, and the reverse was true for the intercept coefficients. Finally, the amount of variance in alcohol dependence predicted by the model was high in both cases, but slightly higher for the IRT score model.

Discussion

In this study we used IRT modeling to create a developmental scale of alcohol misuse from early adolescence into young adulthood. The IRT approach allows for the changing manifestation of this construct as individuals mature, incorporating new items into the scale where appropriate and dropping items that become less relevant. In addition, the IRT approach provides a straightforward way to identify and model differences in item properties over time with DIF analyses, which can provide an understanding of the way alcohol misuse develops. Finally, IRT scores generated based on the developmental model should be more precise than scores created with approaches from classical test theory.

Eight alcohol misuse items included in two or more study waves displayed non-equivalence over time. It is noteworthy that, with only one exception, differences in item parameters were found for wave 6 (two years after participants had completed high school) relative to assessments that occurred during middle and high school. The exception was the driving item which had higher location parameters at wave 4 relative to waves 5 and 6. This result makes sense given that driving after drinking is a much more serious event when one is just learning to drive (as would be the case for most participants in Grade 10, wave 4).

Four of the DIF items resulted in lower location parameters as individuals matured, whereas four others showed the opposite pattern, indicating more serious levels of alcohol misuse at later ages. The first pattern of DIF, lower locations as individuals mature, was observed for feeling really sick, doing something you later felt sorry for, weekly drinking, and driving after drinking. With the exception of driving after drinking, discussed above, these items were assessed at all waves and showed higher item location parameters for waves 2–5 relative to wave 6. This pattern of differences makes intuitive sense; the items reflect more serious alcohol misuse for school-aged participants relative to 19-year olds. The other pattern of DIF is more difficult to interpret. Two items with this pattern, getting into a fight and getting into trouble at home, were assessed at all waves. Perhaps these events reflect more serious alcohol misuse at age 19 because they involve more serious behavior or consequences at that age.

We also examined the precision of the IRT scores relative to scores calculated with a simple non-parametric transformation and found that the IRT scores displayed higher reliability

(especially at the later waves), tended to be more strongly correlated with other study variables, and were stronger predictors of subsequent alcohol abuse or dependence. However, the value added in all cases was modest. One reason for the limited value-added of the IRT scores is that the items did not show a great deal of variance and several had minimal endorsement, especially at the earlier waves, dampening the potential impact of the IRT scoring. Additionally, the amount of DIF was moderate, and the direction of the effect was split, making the scores relatively unaffected in the aggregate. Another pattern of DIF, or DIF of greater magnitude, would have rendered IRT scores that were more distinct from the OBS.

The IRT models used in this paper yield developmental scales with measurements at different ages on a single metric for which growth can be interpreted as changes in a common construct. In contrast, changes in scores created with simpler methods (e.g., averaging of item scores) might reflect changes in a common construct or they might reflect differential item functioning or other changes in the relationships between the items and the construct of interest. In our study, the difference between the developmental IRT scale and the wave-by-wave average are likely to have limited consequences on inferences made about adolescent alcohol misuse. However, even in this context the IRT measures had smaller measurement error and somewhat greater predictive power. Moreover, the IRT modeling exposed differences in the items over time and offered insights into the meaning of the items that simpler methods cannot provide.

Given time and resources, IRT modeling offers several benefits over other scaling methods. However, IRT scaling can be significantly more demanding and costly than creating wave-by-wave item average scores. In our example, these costs might have outweighed the benefits in terms of inferences we will make about the development of alcohol misuse. Future research is needed to demonstrate whether these findings hold in other studies of the development of adolescent substance use. The techniques presented here can serve as a foundation for those explorations.

Acknowledgments

The research reported in this article was funded by Grant 2R01DA11246 from the National Institute on Drug Abuse.

References

- Allen JP, Litten RZ, Fertig JB, Babor T. A review of research on the Alcohol Use Disorders Identification Test (AUDIT). *Alcoholism: Clinical and Experimental Research* 1997;21(4):613–619.
- Bohn MJ, Babor TF, Kranzler HR. The alcohol use disorders identification test (AUDIT): Validation of a screening instrument for use in medical settings. *Journal of Studies on Alcohol* 1995;56(4):423–432. [PubMed: 7674678]
- Chung T, Colby SM, Barnett NP, Rohsenow DJ, Spirito A, Monti PM. Screening adolescents for problem drinking: performance of brief screens against DSM-IV alcohol diagnoses. *J Stud Alcohol* 2000;61(4):579–587. [PubMed: 10928728]
- Edelen MO, Reeve B. Applying Item Response Theory (IRT) Modeling to Questionnaire Development, Evaluation, and Refinement. *Quality of Life Research* 2007;16:5–18. [PubMed: 17375372]
- Edelen MO, Thissen D, Teresi J, Kleinman M, Ocepek-Welikson K. Identification of differential item functioning using Item Response Theory and the likelihood-based model comparison approach: Application to the Mini-Mental Status Examination. *Medical Care* 2006;44:S134–S142. [PubMed: 17060820]
- Ellickson PL, McCaffrey DF, Ghosh-Dastidar B, Longshore DL. New Inroads in Preventing Adolescent Drug Use: Results From a Large-Scale Trial of Project ALERT in Middle Schools. *American Journal of Public Health* 2003;93(11):1830–1836. [PubMed: 14600049]
- Embretson SE. The new rules of measurement. *Psychol Assess* 1996;8:341–349.
- Embretson, SE.; Reise, SP. *Item response theory for psychologists*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers; 2000.

- Holland, PW.; Wainer, H. Differential Item Functioning. Hillsdale, NJ: Lawrence Erlbaum Associates; 1993.
- Little, RJA.; Rubin, DB. Statistical analysis with missing data. New York: Wiley; 2002.
- McLeod, LD.; Swygert, KA.; Thissen, D. Factor analysis for items scored in two categories. In: Thissen, D.; Wainer, H., editors. Test Scoring. Mahwah, New Jersey: Lawrence Earlbaum & Associates; 2001.
- Muthen, LK.; Muthen, BO. Mplus Users Guide. Vol. Third Edition. Los Angeles, CA: Muthen & Muthen; 1998–2004.
- Samejima, F. Graded response model. In: v. d. L. a., WJ.; Hambleton, RK., editors. Handbook of modern item response theory. Springer-Verlag: New York; 1997. p. 85-100.
- SAS Institute. SAS/STAT user's guide. Vol. Vol 1. Cary, NC: SAS Institute, Inc.; 1990.
- Stout W. A nonparametric approach for assessing latent trait unidimensionality. Psychometrika 1987;52(4):589–617.
- Thissen, D. MULTILOG user's guide: Multiple, categorical item analysis and test scoring using item response theory. Chicago, IL: Scientific Software; 1991.
- Thissen, D. IRTLRDIF v.2,0b: Software for the computation of the statistics involved in item response theory likelihood-ratio tests for differential item functioning. LL Thurstone Psychometric Laboratory, University of North Carolina at Chapel Hill; 2001.
- Thissen, D.; Steinberg, L.; Wainer, H. Detection of Differential Item Functioning Using the Parameters of Item Response Models. In: Holland, PW.; HW, editors. Differential Item Functioning. Hillsdale, NJ: Lawrence Erlbaum Associates; 1993. p. 67-114.
- Ware J Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996;34(3):220–233. [PubMed: 8628042]
- Winters, KC. Assessment of alcohol and other drug use behaviors among adolescents. In: Allen, JP.; Wilson, VB., editors. Assessing Alcohol Problems: A Guide for Clinicians and Researchers. Vol. 2nd edition. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism; 2003. p. 101-123. NIH Publication No. 03-3745

Table 1
Alcohol Misuse Item Content and Waves of Administration.

Item	Wave					
	1	2	3	4	5	6
1. Felt really sick due to drinking	x	x	x	x	x	x
2. Got into a physical fight due to drinking	x	x	x	x	x	x
3. Got into trouble at school/work due to drinking	x	x	x	x	x	x
4. Got into trouble at home due to drinking	x	x	x	x	x	x
5. Did something later felt sorry for due to drinking	x	x	x	x	x	x
6. Weekly drinking*	x	x	x	x	x	x
7. Binge drinking (3+ in one day)	x	x	x	x	x	x
8. Used alcohol and marijuana on same occasion*		x	x	x	x	x
9. Used alcohol and another drug on same occasion				x	x	x
10. Drove a car or motorcycle after drinking				x	x	x
11. Binge drinking (5+ in one day)					x	x
12. Got criticized by other kids due to drinking					x	x
13. Passed out due to drinking					x	x
14. Missed school/work due to drinking					x	x
15. Drank alcohol before school/work					x	x
16. Got criticized by an adult due to drinking					x	
17. Got into a traffic accident due to drinking						x

Note: Items with asterisk were derived dichotomous variables. All other items were assessed with a 6-point likert response scale.

Table 2
Item Parameters for Alcohol Misuse over Five Study Waves.

		a	b_1	b_2	χ^2
1. Felt really sick	w2-w5	2.72	0.55	1.45	
	w6	2.72	0.25	1.18	28.2
2. Got into a physical fight	w2-w5	2.64	1.56	2.20	
	w6	2.64	1.97	2.66	41.2
3. Got into trouble at school/work	w2-w6	2.57	2.04	2.68	
4. Got into trouble at home	w2-w5	2.32	1.27	2.27	
	w6	2.32	1.72	2.53	51.8
5. Did something later felt sorry for	w2-w5	2.52	0.97	1.81	
	w6	2.52	0.75	1.59	16.5
6. Weekly drinking	w2-w5	3.73	0.62		
	w6	2.31	0.57		26.0
7. Binge drinking (3+ in one day)	w2-w5	4.12	0.29	0.64	
8. Used alcohol and marijuana on same occasion	w2-w6	2.20	1.01		
9. Used alcohol and another drug	w4-w5	2.36	1.59	2.00	
(not marijuana) on same occasion	w6	2.36	2.00	2.36	28.6
10. Drove a car or motorcycle	w4	2.44	0.97	1.57	
after drinking	w5-w6	2.44	0.75	1.41	13.7
11. Binge drinking (5+ in one day)	w5-w6	3.19	0.45	0.80	
12. Got criticized by other kids	w5-w6	1.72	1.97	2.97	
13. Passed out	w5-w6	2.87	0.85	1.47	
14. Missed school/work	w5-w6	3.00	1.48	2.07	
15. Drank alcohol before	w5	1.54	1.47	2.06	
school/work	w6	1.54	2.02	2.72	30.2
16. Got criticized by an adult	w5	2.24	1.54	2.49	
17. Got into a traffic accident	w6	2.57	2.70	3.30	

Note: Chi-square values correspond to the likelihood ratio difference tests for the equality of item parameters (e.g., the value 28.2 for item 1 is from the test of equality for the b parameters for *felt really sick* from w2-w5 with those from w6). All tabulated chi-square values are significant at $p < .0001$.

Table 3
Summary Statistics for IRT and Observed Alcohol Misuse Scores at each Study Wave.

Reliability					
	N	N(items)	Mean	SD	Alpha
IRT_W1	5828	7	−0.646	0.479	.52
IRT_W2	5333	8	−0.443	0.676	.69
IRT_W3	5084	8	−0.084	0.697	.87
IRT_W4	4853	10	0.112	0.716	.91
IRT_W5	4921	16	0.186	0.757	.94
IRT_W6	3294	15	0.507	0.793	.93
Obs_W1	5828	7	0.049	0.158	.83
Obs_W2	5333	8	0.103	0.197	.81
Obs_W3	5084	8	0.164	0.235	.81
Obs_W4	4853	10	0.198	0.243	.83
Obs_W5	4921	16	0.214	0.232	.88
Obs_W6	3294	15	0.261	0.231	.85

* Note: θ refers to the latent alcohol misuse score. The ranges in parentheses are the latent score ranges within which the maximum reliability was obtained.

Table 4
Correlations* of Observed and IRT-Based Alcohol Misuse Scores with Psychosocial Measures across the Study Period.

	MHIS	Impulsivity	Parental Support	School Satisfaction	Sensation Seeking	Lack Guilt
Irt_W1	.266	.374				
Obs_W1	.254 (n=5828)	.343 (n=5828)				
Irt_W2	.192		.193	.289		
Obs_W2	.186 (n=5333)		.188 (n=5333)	.276 (n=5325)		
Irt_W3	.142	.300	.177	.254	.338	.274
Obs_W3	.145	.301	.164	.242	.319	.262
Irt_W4	(n=5001)	(n=5015)	(n=5028)	(n=5068)	(n=5014)	(n=4982)
Obs_W4	.114		.158	.264	.324	.229
	.122		.148	.255	.315	.219
Irt_W5	(n=4655)		(n=4789)	(n=4837)	(n=4702)	(n=4689)
Obs_W5	.110	.314	.144	.253	.359	.235
	.118	.325	.139	.238	.343	.229
Irt_W6	(n=4815)	(n=4904)	(n=4870)	(n=4907)	(n=4838)	(n=4808)
Obs_W6	.120	.330			.306	
	.120	.341			.309	
	(n=3249)	(n=3279)			(n=3280)	

* Note: All correlations are significantly different from 0 at $p < .0001$. Bold print indicates the correlation of larger magnitude (difference $> .005$) in comparing IRT with OBS scores.

Table 5

Results from IRT and OBS Alcohol Misuse Growth Models Predicting AUDIT Scores at Wave 7 (N=3332).

	Standardized Estimate [*]	
	IRT	OBS
Intercept (I)	-.858	.599
Slope (S)	1.602	.956
Correlation I with S	-.305	-.295
I predict AUDIT	.359	.361
S predict AUDIT	.680	.653
AUDIT R ²	.443	.417

* Note: All estimates are significant at $p < .0001$.