

Preliminary Validation of the Young Adult Risky Drinking Measure using Rasch Analysis

Haley A. Wright

College of Public Health and Health Professions, University of Florida

While “binge drinking” is a highly sensitive indicator of major negative health outcomes, this single item yields poor measurement specificity resulting in a generic labeling of “at-risk” college students. Despite proven interventions, rates of binge-drinking have remained stagnant. To effectively identify students at greatest risk, a new measure of hazardous college drinking is needed. For the preliminary validation of young adult risky drinking, a convenience sample of college students from the health science program (n=241) completed a 40 item, voluntary, anonymous, online survey. Item fit, rating scale analysis, item hierarchy, person ability-item match, and person reliability and separation were examined. All 40 items fit the Rasch model ($MNSQ \leq 1.4$, $ZSTD \leq 2.0$) with 11 people (4.5%) misfitting. Person ability was approximately 2 logits lower than item difficulty. Person reliability was 0.91 and the sample separated into four distinct groups. Rating scale consistently violated the ‘minimum 10 observed counts in every category’ criteria. Rating scales were collapsed from a 9-point to a 4-point scale and re-examined, resulting in an improved rating scale. Preliminary evaluation of the young adult risky drinking suggests the measure reliably separates college students across a risk continuum. Further analysis is required to determine unidimensionality and reliability across other groups of college students.

BACKGROUND

College drinking accounts for nearly 744,000 student injuries, 124,000 reported cases of sexual assault, and 1,800 student deaths annually (Hingson, Zha, & Weitzman, 2009). With increased attention on college drinking, heavy episodic drinking has been named the number one public health threat to and the primary source of mortality and morbidity for college students (Wechsler et al., 1994). Wechsler and his colleagues found this style of drinking was linked to educational difficulties, psychosocial problems, antisocial behaviors, overdoses, high-risk sexual behaviors, and an increased likelihood to partake in other high-risk activities (Wechsler, Lee, Kuo, & Lee, 2000).

Currently two instruments are commonly used to assess hazardous drinking; Binge drinking and the Adult Use Disorder Identification Test-Consumption known as the AUDIT-C. The term “binge drinking” is defined by five or more drinks for a man and four or more drinks for a woman, within a two-hour period, at least once within the past two weeks (Wechsler & Nelson 2001). The term was developed to create a “cut-off point” in which the level of alcohol intake had harmful consequences. (Wechsler et al., 1998; Wechsler et al., 2000). Studies have repeatedly shown that approximately 40% of college students qualify as binge drinkers within the past two weeks (Wechsler et al., 2002; Wechsler et al., 1998; Wechsler et al., 2000). However, critics do suggest that a majority of binge drinking episodes occur without the student experiencing any major public health consequences (e.g. injuring oneself, having

unprotected sex, drinking and driving, etc.) (Dimeff et al., 1995; Weingardt et al., 1998; Dejong, 2003; Moorhouse et al., 2014).

The AUDIT-C is a three item-screening tool used to identify hazardous drinking (Babor, Higgins-Biddle, Saunders, & Monteiro, 2001). The three items, “How often do you have a drink containing alcohol?,” “How many standard drinks containing alcohol do you have on a typical day?,” and “How often do you have six or more drinks on one occasion?,” are based around quantity, frequency, and duration of drinking events. The AUDIT-C uses only the first three questions of the full 10 question AUDIT but is approximately equal in validity to that of the full AUDIT (Reinert & Allen, 2007).

Both binge drinking and the AUDIT-C dichotomize people into “at-risk” and “not at-risk” groups. However, such measures may not sufficiently capture the range of risky

college drinking behaviors, therefore failing to identify those students in need of alcohol-related intervention services (Moorhouse et al. 2014). We hypothesize this lack of contextual relevance and low drink threshold (e.g. 5 drinks for men, 4 drinks for women) capture all those students who are “at-risk”, but are not optimized as many of the students that are considered “at-risk” experience no major negative health outcomes. The purpose of this study is to examine the psychometric properties of a new college drinking measure, the Young Adult Risky Drinking (YARD) measure using Rasch Analysis.

METHODS

Participants

A convenience sample of 241 undergraduate health science college students from a major southern university were recruited for this study. Informed consent was presented to all participants through written email followed by the link to the survey. Inclusion criteria for participants consisted of being ages 18-24, enrolled in college, and agreeing to informed consent.

YARD Items

Prior to this study, the primary research team conducted seven focus groups with college students to identify drinking behaviors that span the risk continuum. Focus groups were facilitated using a semi-structured interview guide and data was coded. Sixty-three items were identified from the initial pool. Following researcher triangulation and multiple iterations of comparison, the item pool was reduced to 45 items. The research team then conducted cognitive interview sessions with 26 undergraduate students to ensure interpretability and item clarity. The final item pool consisted of 40 items with a 9-point rating scale ranging from 0 to 8 or more. Participants answered demographic questions and the 40 YARD items anonymously. All questions were administered via Qualtrics, a secured, online survey format.

Rating Scale

Linacre establishes 3 essential guidelines which should be used to optimize the rating scale. The first is that each rating category must have at least 10 observations. He argues that without at least 10 observations in the category, the category may be unnecessary to measure (Linacre 2002). Second, outfit mean-squares must be less than 2.0 for each rating category. Linacre states that any value over 2.0 indicates there is more misinformation than information in the observation (2002). The last of the essential criteria is that average measures advance monotonically with category, meaning a person with a strong ability should respond to higher categories. If a set of observations does not advance monotonically then the rating scale for that data set is uncertain (Linacre 2002). Depending upon these criteria, decisions will be made as to whether the rating scales should be collapsed.

Item Fit

The fit to the model was measured through infit statistics: looking at the difference in the observed and expected response for items that have a difficulty level near the person's ability level (Tennant.2007). For an item or person to be considered misfitting, the model must violate both the criterion for mean-squares (MNSD) and standardized t-

score (ZSTD). A mean-squares of 1.0 is the ideal fit. Since this was a survey with relatively low stakes, the accepted infit MNSD is 0.6-1.4 A mean-squared of 0.6 indicates that the item overfit the model and had 40% less variation than was predicted. Conversely a mean-square of 1.4 indicates the item underfit the model and had 40% more variation or randomness than was predicted (Wright. n.d). The criteria for ZSTD is ± 2.0 . The ZSTD determines the observed fit versus the expected fit and the probability that the data will fit the Rasch model (Tennant 2007).

Item Hierarchy, Person Ability-Item Match, and Reliability

Rasch analysis focuses on how closely the ability of the sample matches with the difficulty of the items. If the mean of the sample is lower than the mean of the items, there is a possibility for a floor effect. Similarly, if the mean of the sample is higher than the mean of the items, a ceiling effect is possible. Rasch also looks at individual ability compared with specific items. An item with a high difficulty is less likely to be answered correctly than an item with an easier difficulty and vice versa (Fraley, Waller, & Brennan. 2000). Person reliability is the Rasch equivalent to Cronbach's alpha. It estimates how well a measure can differentiate people on the construct. On the other hand, person separation determines how many strata, or distinct groups, the sample can be separated into.

RESULTS

Rating Scale Analysis

The stated criterion of having at least 10 observations per category was violated multiple times. The data was first analyzed using a 9-point scale (0 times- 8 or more times in the past 30 days). Since there were very few observations in categories from 6 times to 8 times or more, the decision was made to collapse some categories. The categories were recoded as 0 times, 1-2 times, 3-5 times, and 6 times or more. This increased the reliability, as well as improved the rating scale by having more observations in each category however, there were still several categories that did not reach 10 observations, as noted in Table 1.

Item Fit

All the items met the infit requirements. Elven people (4.5%) misfit, as they did not meet the infit criteria (MNSQ >1.4 and ZSTD $>\pm 2.0$).

Precision

Person reliability, which is the Rasch analysis equivalent of Cranach's alpha, was 0.91. Person separation was 2.98. When plugged into the formula for strata, $(4*2.98+1)/3$, the sample for YARD was broken down into 4 distinct groups.

Table 1: Item Response Categories

Item	0	1	2	3
1	231	4	4	2
2	231	6	3	1
3	171	35	29	6
4	80	57	62	42
5	183	31	18	8
6	225	7	5	2
7	158	42	35	6
8	155	45	21	18
9	149	58	27	7
10	96	69	49	27
11	85	73	57	24
12	166	41	23	9
13	141	63	22	9
14	163	50	19	5
15	200	25	11	3
16	102	72	46	18
17	122	68	32	14
18	192	37	5	4
19	168	54	9	8
20	210	23	3	3
21	78	69	56	36
22	163	54	15	6
23	118	63	39	18
24	173	41	19	5
25	215	15	6	2
26	226	7	4	1
27	201	32	4	1
28	225	11	2	1
29	234	1	1	2
30	179	37	18	4
31	206	24	7	1
32	230	4	3	0
33	232	2	2	2
34	198	30	8	1
35	212	21	4	1
36	226	9	3	0
37	180	39	16	2
38	157	62	15	4
39	180	42	10	5
40	207	21	9	1

Items are listed with their respective observation counts. All categories that violate essential criterion are noted by grey color. A '0' indicates an individual has never engaged in item, '1' represents engaging in item 1-2 times within past 30 days, '2' indicates engaged in item 3-5 times within past 30 days, and '3' indicates engaging in item 6 or more times in past 30 day. Item key can be found in appendix.

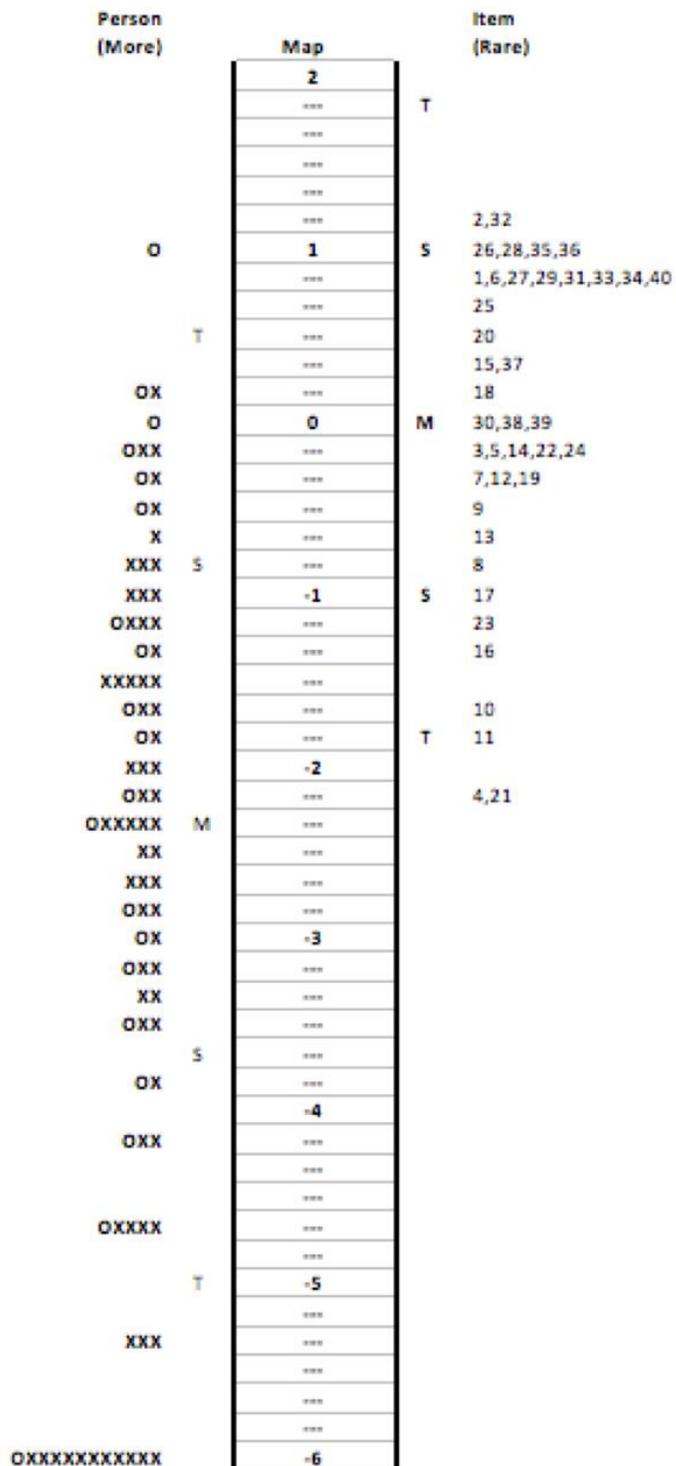


Figure 1. Person-ability item-ability sample match. *Note.* Each 'O' on the left side of the map represents 1-2 people, whereas each 'X' represents 3 people. The numbers on the right side of the map each correspond to an item. 'M' represents the mean of the sample or items. See endnotes for item key

Person-Ability Item-Difficulty Match

Figure 1 highlights the difference between the mean of the items and the mean of the people. The items, located on the right side of the map, are ranked from hardest to easiest starting at the top of the map. Person ability, located on the left side of the map, is tiered from persons with highest

ability at the top to those with lowest ability on the bottom. Though there was no ceiling effect, 34-35 individuals created a floor effect. A floor effect shows that the YARD was unable to capture those individuals with extremely low ability

DISCUSSION

Preliminary examination of the YARD shows it to be a reliable and consistent measure for the sample of college students that was used. There were no misfitting items and 11 people (4.5%) misfit. There was high person reliability of 0.91, though there was over a 2 logit difference between person and item means.

A floor effect was also observed in the sample. This is caused by the mean of the items being higher than the mean of the sample. 34-45 individuals created this effect and were not captured by the instrument. It is possible that some students that created the floor effect simply do not consume any alcohol. To improve this, items of lower ability could be added. Another improvement that could be made is to have a more heterogeneous sample. The current sample was relatively homogeneous, as almost all were students in a rigorous health science program; most were at a low ability level. By including a more diverse sample, we would be able to see how the instrument captures those with higher ability.

Some limitations in the validation of YARD were noted. One limitation, as mentioned above, was the homogenous sample used. Though it showed great reliability, more trials would be necessary to ensure the YARD would be useful screening tool when applied to the college population as a whole. Running a factor analysis would also be an improvement and recommendation for further studies on this instrument. Because this was a preliminary trial with a heterogeneous sample, the decision was made to keep all items. A factor analysis would help to determine if any of the items overlap and item bank could be condensed without compromising reliability. Examination of unidimensionality should also be included in future studies.

The preliminary evaluation of the YARD shows the sample of college students being broken down into 4 distinct groups based upon their risk. By being able to stratify the sample and target those that are at high or the highest risk, it would be possible to use interventions to target those individuals before a negative public health outcome occurs. Before this instrument could be used on a large scale, it would also be necessary to compare this measure to that of both to binge drinking and AUDIT-C in terms of its sensitivity and specificity relative to major negative public health outcomes. If the YARD proves to be a reliable and consistent instrument, capable of capturing and stratifying all college students, it is possible this measure could be used across all college campuses to identify those students who need targeted interventions the most.

ACKNOWLEDGMENTS

I would like to thank my mentor Dr. Michael Moorhouse for all his support and wisdom.

REFERENCES

- Babor, T. F., Higgins-Biddle, J. C., Saunders, J. B., & Monteiro, M. G. (2001). *Audit*. The Alcohol Use Disorders Identification Test (AUDIT): Guidelines for use in primary care.
- Carey, K. B., Scott-Sheldon, L. A. J., Carey, M. P., & DeMartini, K. S. (2007). Individual-Level Interventions to Reduce College Student Drinking: A Meta-Analytic Review. *Addictive Behaviors*, 32(11), 2469–2494. <http://doi.org.lp.hscl.ufl.edu/10.1016/j.addbeh.2007.05.004>
- Dawson, D. A., Grant, B. F., Stinson, F. S., & Zhou, Y. (2005). Effectiveness of the derived alcohol use disorders identification test (AUDIT-C) in screening for alcohol use disorders and risk drinking in the US general population. *Alcoholism: Clinical and Experimental Research*, 29(5), 844-854.
- DeJong, W., Definitions of binge drinking. *JAMA*, 2003. 289(13): p. 1635; author reply 1636.
- Demartini KS1, Carey KB (2012). Optimizing the use of the AUDIT for alcohol screening in college students. *2012 Psychol Assess*.Dec;24(4):954-63.
- Dimeff, L.A., et al., Binge drinking in college. *JAMA*, 1995. 273(24): p. 1903-4.
- Fraley RC, Waller NG, Brennan KA. An item response theory analysis of self-report measures of adult attachment. *J Pers Soc Psychol* 2000 Feb;78(2):350-365. [Medline: 10707340]
- Hays, R.D., L.S. Morales, and S.P. Reise, (2000) Item response theory and health outcomes measurement in the 21st century. *Medical Care*; 38(Suppl 9): p. 12.
- Hingson, R., W. Zha, and E. Weitzman, (2009) Magnitude of and trends in alcohol-related mortality and morbidity among US college students ages 18-24, 1998-2005. *Journal of Studies on Alcohol and Drugs*: p. 12-20.
- Linacre, J.M., Optimizing Rating Scale Category Effectiveness. *Journal of Applied Measurement*, 2002. 3(1): p. 22.
- Moorhouse, MD, Soule, EK, Hinson, WP, & Barnett, TE (2014). Assessing alcohol use in college: Is it time for a new approach to identify risky drinking behavior? *Journal of Substance Use*, 19, 262-267.
- Reinert, D. F. and Allen, J. P. (2007), The Alcohol Use Disorders Identification Test: An Update of Research Findings. *Alcoholism: Clinical and Experimental Research*, 31: 185–199. doi: 10.1111/j.1530-0277.2006.00295.x
- Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum* 2007 Dec 15;57(8):1358-1362 [FREE Full text] [doi: 10.1002/art.23108] [Medline: 18050173]
- Wechsler, H., et al.,(2002) Trends in college binge drinking during a period of increased prevention efforts: Findings from 4 Harvard School of Public Health College Alcohol Study Surveys: 1993-2001. *Journal of American College Health*. 50(5): p. 203-217.

Wechsler H, Davenport A, Dowdall G, Moeykens B, Castillo S (1994). Health and Behavioral Consequences of Binge Drinking in College: A National Survey of Students at 140 Campuses. *JAMA* ;272(21):1672-1677. doi:10.1001/jama.1994.03520210056032.

Wechsler H, Lee J, Kuo M, Lee H. (2000) College binge drinking in the 1990s: A continuing problem. Results of the Harvard School of Public Health 1999 College Alcohol Study. *J Am Coll Health*;48: 199-2 10.

Wechsler H, Nelson TF. (2001). Binge drinking and the American college students: What's five drinks? *Psychol Addict Behav.*;15:287–291.

Weingardt, K.R., et al., Episodic Heavy Drinking Among College Students: Methodological Issues and Longitudinal Perspectives. *Psychology of Addiction Behaviors*, 1998. 12(3): p. 13.

Wright B, Linacre J, Gustafson J, Martin-Löf P. Rasch Meas Trans. Reasonable mean-square fit values URL: <http://www.rasch.org/rmt/rmt83b.htm> [accessed 2017-01-20] [WebCite Cache ID 6guwvX0Of]

ENDNOTES

Item Key for Figure 1

1. Drank alcohol with pain medications (e.g., Lortab, Vicodin) to become more buzzed or intoxicated.
2. Drank alcohol with sedatives (e.g., Valium, Xanax) to become more buzzed or intoxicated.
3. Consumed an alcoholic beverage that was mixed with an energy drink (e.g., Red Bull and Vodka).
4. Consumed an alcoholic beverage that was mixed with a soda other than Sprite (e.g., Rum and Coke).
5. Drink alcohol and smoke marijuana within the same 2 hour time period.
6. Drink alcohol while also using other illegal substances such as cocaine or heroin.
7. Chugged or funneled alcohol (e.g., shot gunned a beer).
8. Ordered 2 or more drinks for yourself from a bartender or server at one time.
9. Drank liquor directly from the bottle.
10. Consumed 3 "Standard" drinks within the first hour of drinking.
11. Consumed 5 or more drinks in a 2 hour period (for males) / Consumed 4 or more drinks in a 2 hour period (for females)

12. Ordered multiple drinks from a bartender or server at one time for yourself.
13. Taken 5 or more shots in one night.
14. Consumed an alcoholic beverage but did not know what was in it (e.g., "hunch punch").
15. Accepted a drink from someone you just meant (not including bartenders or servers).
16. Participated in a drinking game (e.g., ring of fire, flip cup, or beer pong).
17. Participated in drink specials (e.g., free beer, beat the clock).
18. Participated in an organized bar/pub crawl (i.e., drank alcohol at 3 or more predetermined bars, clubs, or restaurants).
19. Drank 2 or more "standard" drinks when your were home alone.
20. Drank alcohol before noon (not including game days or in situations in which it may be socially acceptable such as mimosas during brunch).
21. Drank alcohol before going out (i.e., pre-drinking).
22. Continued to drink alcohol after 2 in the morning.
23. Drank alcohol regularly for longer than 3 consecutive hours.
24. Got drunk or buzzed on 3 consecutive nights.
25. Continued to drink after vomiting.
26. Continued to drink after being cut-off by a friend or bartender/server.
27. Consumed alcohol to the point where a friend had to physically help you walk or carry you so you could get home.
28. Drank alcohol within 2 hours of having to take an exam.
29. Consumed alcohol during class.
30. Drank alcohol without specifically knowing how you were going to get home.
31. Drank more than 1 drink despite being the designated driver.
32. Drank alcohol despite being diagnosed with an alcohol-related condition (e.g., fatty liver).
33. Drank alcohol despite current or pending alcohol-related legal issues.
34. Purposefully drank on an empty stomach to get more intoxicated.
35. Drank alcohol that you or someone else snuck into an event or place where alcohol was prohibited (e.g., class, football game).
36. Decided to go drink alcohol rather than attend class.
37. Consumed alcohol while a passenger in a moving vehicle.
38. Drank alcohol to cope with stress or anxiety.
39. Drank alcohol to cope with sadness.
40. Drank heavily to make ordinary activities such as playing intramural sports or going to the movies seem more fun..