Supplemental Material for Online Publication

Study 4. A total of 682 mandated college students were randomly assigned to one of three groups: group-based motivational interviewing ($n = 228$), motivationally enhanced peer theater ($n = 230$), and an interactive alcohol education program ($n = 224$; Cimini et al., 2009).

Studies 8a, 8b, and 8c. Studies 8a, 8b, and 8c were part of the Motivating Campus Change Multisite Study (MC^2; Larimer et al., 2007), which was a large, multisite, multicohort, longitudinal prevention study. Participants were drawn from a random sample of enrolled students each year, and were followed annually until they graduated. The project evaluated three stepped-care interventions, with each intervention phase sequentially implemented annually within a multiple baseline design across three university campuses located in the Northwest region. Phase I tested campus-wide social norms marketing, providing accurate feedback with respect to descriptive norms for drinking by using mass-media techniques, which was implemented in successive years across three campus sites. Unpublished results indicated that the social norms campaign was associated with reductions in perceived descriptive drinking norms and increases in alcohol abstinence at the smallest campus site (Study 8c), but not at the two larger campus sites (Studies 8a and 8b).

Phase II was a randomized control trial (RCT) to evaluate effects of mailed motivational-enhancement feedback delivered to a random half of participants on each campus who were projected to remain enrolled a year from receiving the mailed feedback. The norms for feedback were derived separately for each campus because these campuses were distinctively different from one another in terms of student demographics and drinking. The results of the Phase II RCT from the first campus (Study 8a; Larimer et al., 2007) were published. Phase III was a RCT to evaluate the *added* effectiveness of a small-group, peer-delivered Alcohol Skills Training Program (ASTP) intervention delivered to a random half of those who continued to engage in heavy episodic drinking.

Data from the Phase II RCT participants are included in the Project INTEGRATE main data set. Of the 1486, 2155, and 600 participants at the three different universities, respectively, 736, 1094, and 303 received mailed feedback, respectively, for Studies 8a, 8b, and 8c. Given the mixed effects of the

social norms marketing (i.e., Phase I) and also in the context that many college campuses routinely engage in activities designed to increase student awareness about drinking, we deemed that the effect of the social norms campaign is equivalent to those from other campuses. In addition, the Phase III participants received two different types of interventions successively over two years. For these reasons, Phase III data were not included in the main data set. Note also that we counted Studies 8a, 8b, and 8c as three distinctive studies because these campuses had different drinking cultures as well as student demographics, and consequently students were given different drinking norms.

Study 9. Larimer and colleagues (Lee et al., 2009) completed a RCT with 604 students, randomized to one of six groups: in-person BASICS feedback ($n = 101$), Web BASICS ($n = 100$), Alcohol skills training program (ASTP; $n = 97$), Choices interactive journaling ($n = 103$), Alcohol education ($n = 102$), or assessment only control ($n = 101$). All participants were first-year students, who were screened for heavy episodic drinking via a web-based survey.

Study 10. Incoming first-year students were screened for their high-risk drinking based on heavy episodic drinking and alcohol-related negative problems during high school and were randomized into either the intervention ($n = 174$) or no-intervention control ($n = 174$) group (Baer et al., 2001; Marlatt et al., 1998). This study also recruited a random sample of 113 students to provide a natural history comparison, which included 25 high-risk students who were randomly assigned to either the intervention or control group (part of the randomized sample of 348). One of the 113 normative participants received the intervention due to heavy drinking. The remaining 87 of 113 individuals who did not receive any intervention and a mutually exclusive sample of 348 high-risk, randomized students were also included in the intervention data set.

Study 11. Walters, Vader, et al. (2007) conducted a RCT of the "electronic Check-Up to Go" (e-CHUG) for high-risk, first-year students ($n = 106$). However, participation in this study was open to all first-year students regardless of their drinking levels, and all participants were randomized into two groups at baseline: a personalized norm feedback ($n = 185$) and an assessment-only control ($n = 198$) group. A total of 383 first-year students were included in the Project INTEGRATE intervention data set.

Study 18. Martens et al. (2010) tested the efficacy of a targeted PFI among intercollegiate athletes from universities that are NCAA Division I and II schools with a specific focus on athletic performance-related feedback on their alcohol use. The sample included student athletes who participated in varsity sports ($n = 263$) and those who participated in club sports ($n = 64$). Martens et al. reported findings only for the former because the intervention was specifically designed for those who played varsity sports. The latter group was recruited because it was against the NCAA regulations to compensate varsity athletes if the study is open only to varsity athletes. Thus, we chose to include both groups in the intervention data set. The pooled data include a total of 329 students, including two participants who did not answer a question about their varsity affiliation, who were randomly assigned into one of three conditions: (1) a PFI condition that did not include athletic performance-related feedback (PFI-general; $n = 116$), (2) a PFI condition that included athletic performance-related feedback (PFI-sport; $n = 102$), and (3) a control condition ($n = 113$) that provided general educational information about alcohol use.

Study 19. LaBrie, Hummer, et al. (2008) implemented a clustered RCT in which campus organizations (fraternity, sorority, or service organization) were randomly assigned to either an intervention group ($n = 537$) or an assessment-only control group ($n = 641$).

Study 20. Larimer et al. (2001) reported findings from a combined individual and group motivational interviewing intervention provided to fraternity pledge classes, with random assignment of individuals to peer or professional interviewers. Twelve fraternity houses were randomized to either an intervention group or a treatment-as-usual, assessment-only control group. A total of 166 individual fraternity members agreed to participate and completed a baseline assessment. The reported data in Larimer et al. were a subset of a larger, multicohort, longitudinal study, which later included sorority members (ns = 136 and 39 for Cohorts 1 and 2, respectively) and members of the original fraternities from different cohorts in subsequent years (ns = 241 and 346 for Cohorts 3 and 4, respectively). Cohort 3 followed randomization of Cohort 1 but did not receive any actual intervention. Although it is possible that the effect of the original intervention that took place two years earlier can be sustained within fraternities to a certain degree, the treatment and control groups did not differ at baseline for Cohort 3.

Thus, both of the Cohort 3 groups were included as control cases, resulting in a total of 318 individuals who received the feedback intervention and 610 control students who did not receive any active intervention.

Study 21. Study 21 was a RCT dismantling the efficacy of in-person motivational interviewing interventions and stand-alone feedback interventions by creating the following four groups: (1) a web feedback only group ($n = 68$), (2) a single motivational interviewing session without feedback ($n = 72$), (3) a single motivational interviewing session with feedback ($n = 76$), and (4) an assessment only control group ($n = 72$) (Walters, Vader, Harris, Field, et al., 2009). At the 12-month follow-up, an additional delayed assessment group ($n = 75$) completed alcohol assessment and was compared against the immediate assessment-only group to examine the effects of assessment on drinking (Walters, Vader, Harris, & Jouriles, 2009). The delayed assessment group was included in the Project INTEGRATE larger data set (see the Combined Sample section) but not in the main, intervention data set as their first assessment occurred at 12 months post intervention for other participants.

Study 22. This RCT examined the efficacy of a brief motivational intervention (BMI) and a parent-based intervention (PBI) as universal preventive interventions to reduce alcohol use among incoming college students (Wood et al., 2010). Four intervention conditions were created by crossing the two interventions of interest (PBI only, BMI only, PBI + BMI, and Assessment-only Control). The PBI only group ($n = 256$) was not included in the Project INTEGRATE intervention data set because this intervention was very unique such that it was delivered to parents (who were, in turn, charged with communicating with their teens) and difficult to classify.

Table S1

Study Intervention Groups Included in Project INTEGRATE (N = 12,630)

Study	Original Group Label	New Group Label	Sample Size at Baseline
1	BMI	MI + PF	180
	WF	PF	168
2	Immediate WF	PF	111
	Delayed WF	Control	119
3	BMI	MI + PF	113
	Alcohol Education (Alcohol 101)	AE	112
4	Group MI	GMI	228
	Group Theatrical Presentation	AE*	230
	Alcohol Education	AE*	224
5	Group BMI	GMI	167
6	Group BMI	GMI	115
7.1	Professional/Peer LMC	GMI+*	100
	Control	Control	24
7.2	Professional/Peer LMC	GMI+*	317
	Control	Control	135
8a	Feedback	PF	736
	Control	Control	750
8b	Feedback	PF	1,094
	Control	Control	1,061
8c	Feedback	PF	303
	Control	Control	297
9	Alcohol Education (Alcohol 101)	AE	102
	ASTP	GMI	97
	Choices	GMI*	103
	BASICS	MI + PF	101
	Web BASICS	PF	100
	Control	Control	101
10.1	BMI	MI + PF	174
	Control	Control	174
10.2	Non-High-Risk Control	Control*	87
11	Feedback	PF	185
	Control	Control	198
12	BMI	MI + PF	84
	EC	AE*	81
	BMI + EC	MI + PF + AE**	87
	Control	Control	83
13	BMI	MI + PF	26
	Feedback	PF	28

14	BMI	MI + PF	30
	Alcohol Education	AE	29
	Control	Control	25
15	Group BMI	GMI	155
	Control	Control	108
16	Group BMI	GMI	161
	Control	Control	126
17	Group BMI	GMI	120
18	Targeted Feedback	PF*	114
	Standard Feedback	PF	102
	Alcohol Education	Control	113
19	Group-Specific Feedback	GMI*	537
	Control	Control*	641
20	BMI	MI + PF	318
	Control	Control	610
21	BMI with Feedback	MI + PF	76
	BMI without Feedback	MI without PF**	72
	Feedback	PF	68
	Control	Control	72
22	BMI	MI + PF	253
	BMI + PBI	MI + PF + PBI**	249
	Control	Control	256

Notes. BMI = Brief Motivational Interviewing; WF = Written Feedback; AE = Alcohol Education; LMC = Lifestyle Management Class; BASICS = Brief Alcohol Screening and Intervention for College Students; ASTP = Alcohol Skills Training Program; EC = Expectancy Challenge; PBI = Parent-based Intervention; MI + PF = Motivational Interview plus Personalized Feedback; PF = Stand-alone Personalized Feedback; GMI = Group Motivational Interview; AE = Alcohol Education. GMI+ in Study 7 provided personalized feedback, whereas other GMIs did not. * These groups are rather unique and difficult to classify, and thus excluded in some subsequent treatment-focused analyses. ** These groups are unique and thus omitted from the total number of 57 intervention groups reported in this paper (i.e., 10 MI + PFs, 11 PFs, 11 GMIs, 6 AEs, and 19 Controls). See Ray et al. (in press) for the detailed information about the content, personalization, and intervention fidelity of the brief motivational interventions included in Project INTEGRATE.

Table S2

Overview of Key Constructs, Approaches taken, and Overlap across Studies at Baseline

Construct	Approach	1	2	3	4	5	6	7	8a-8c	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Outcome																							
Alcohol use	Harmon.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Alcohol-related problems	IRT	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Sex cons.	Harmon.	X	X	X	X	X	X	---	X	X	---	---	X	---	---	X	X	X	X	---	X	---	X
Legal cons.	Harmon.	X	X	X	X	---	X	X	X	X	X	X	X	---	---	X	X	---	---	X	X	X	---
Tobacco & marijuana use	Harmon.	X	X	X	---	---	---	X	X	X	X	---	X	---	---	---	X	---	---	X	X	X	X
Mediator																							
Peer use norms	Harmon.	X	X	X	X	---	X	X	X	X	X	---	X	X	---	---	X	---	X	X	X	X	X
Alcohol exp. & drinking mot.	IRT	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	---	---
Protective beh.	IRT	X	X	X	X	---	---	---	X	X	---	---	X	---	---	---	X	---	X	---	---	X	X
Read. to change	No common metric	X	X	X	X	X	X	X	X	X	X	X	X	---	X	X	X	X	X	---	X	X	---
Moderator																							
Family hist. alcohol problems	Harmon.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	---	---	X	X	X
Demographic	Harmon.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Notes. An X indicates that the target construct was assessed. --- indicates that the target construct was not assessed. Harmon. = harmonization; IRT = IRT latent trait (θ) scores; Sex cons. = sex-related consequences due to drinking; Legal cons. = any arrest, DWI arrest, or driving after drinking; Alcohol exp. & drinking mot. = alcohol expectancies and/or drinking motives; Protective beh. = protective behavioral strategies; Read. to change = readiness to change; Family hist. alcohol problems = family history of alcohol problems.

Table S3

Overlap in Alcohol-related Problems Measures and Baseline Referent Time frames across Studies

Study	1 month	2 month	3 month	6 month	12 month	ADS
1			R		R	ADS
2		R			R	
3					Y, A	
4				R, Y		
5	R					
6	R					
7	P					
8a			R		Y	
8b			R		Y	
8c			R		Y	
9			R		Y	
10				R	A	ADS
11	R					
12				Y		ADS
13	R					
14		R				
15	R				A	
16	R, Y					
17	R					
18			B			
19	R					
20	R			R	R, Y	ADS
21			R		A	
22			Y			

Notes. R = the Rutgers Alcohol Problem Index (RAPI; White & Labouvie, 1989); Y = the Young Adult Alcohol Problems Screening Test (YAAPST; Hurlbut & Sher, 1992); B = the Brief Young Adult Alcohol Consequences Questionnaire (BYAACQ; Kahler, Strong, & Read, 2005); P = the Positive and Negative Consequences Experienced questionnaire (PNCE; D'Amico & Fromme, 1997); the Alcohol Dependence Scale (ADS; Skinner & Allen, 1982; Skinner & Horn, 1984); and A = the Alcohol Use Disorders Identification Test (AUDIT; Saunders et al., 1993). At follow-ups, alcohol-related problems were typically assessed for the past 1-3 months in almost all studies. The ADS was originally developed to assess alcohol use and problems for the past 12 months. No specific time frame was provided for the ADS in the original studies that are included in the Project INTEGRATE data set. For Studies 1, 2, and 20 that assessed the RAPI multiple times, we selected data from the shortest referent time frame for the IRT analyses.

Table S4

A Series of Actions Taken for IRT Analysis

| IRT Analysis	Steps Taken
1. Model considerations	a. Unidimensionality vs. multidimensionality
	b. Multi-unidimensional vs. higher-order IRT
	c. 2-PL IRT vs. Generalized Partial Credit Model
	d. Common mean vs. different means across studies
2. Data considerations	a. Link items across studies
	b. Check missing data (i.e., did not assess) at the item level across studies
	c. Identify items that can be collapsed across studies
	d. Harmonize response options within/across studies
	e. Harmonize referent time frames within/across studies
	f. Prune items, if necessary
3. Preliminary and final steps	a. EFA/CFA using raw data
	b. EFA/CFA using different covariance matrices
	c. EFA/CFA using the full information maximum likelihood estimation
	d. IRT analysis using BILOG, a software program
	e. Develop MCMC algorithms and run simulation studies
	f. Real data analysis (item calibration and scoring of theta score)
	g. Simulation study based on the results from the real data analysis to examine the impact of missingness
	h. Sensitivity analysis - Different sets of items and subsamples for item calibration for longitudinal data
4. Model checking	a. Visually examine item characteristic curves (ICCs) of all items
	b. Examine descriptive statistics of item parameters, structural parameters, and theta scores per study
	c. Examine IRT diagnostic plots
	d. Examine posterior predictive model check in the Bayesian analytic framework
	e. Model fit comparison using the deviance information criterion (DIC)
	f. Compare the distribution of observed scores against the distribution of the posterior predictive scores
	g. Plot individual and average growth curves per study and for the combined data
	h. Compute correlations between latent trait scores and original scale scores within studies
	i. Examine latent trait scores derived from IRT models that allowed some items to show different item

functioning (DIF)

j. Check equivalence of treatment and control groups at baseline

Notes. IRT = Item response theory, 2-PL IRT = 2-Parameter Logistic IRT, EFA = Exploratory factor analysis, CFA = Confirmatory factor analysis, MCMC = Markov chain Monte Carlo.

*Figure S1*. Total information curve of alcohol-related problems.
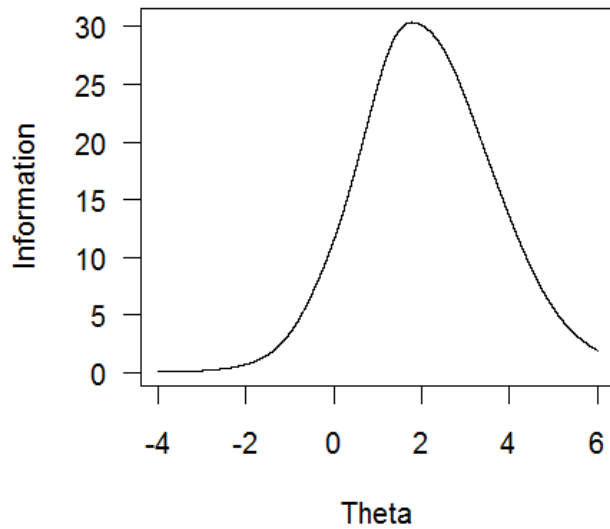
The total information curve summarizes the overall performance of the measure at each level of an underlying latent trait (Markon, 2013), and indicates how much information the scales yield across the range of latent traits and is computed as a function of the item location and slope parameters. The items assessing alcohol-related problems provided more reliable information for individuals whose traits were at the higher end of the spectrum (e.g., latent trait [θ] scores between 1 and 3) compared to those at the lower end of the spectrum.

Figure S2. Scatter plot of latent trait scores for alcohol-related problems from two IRT models.
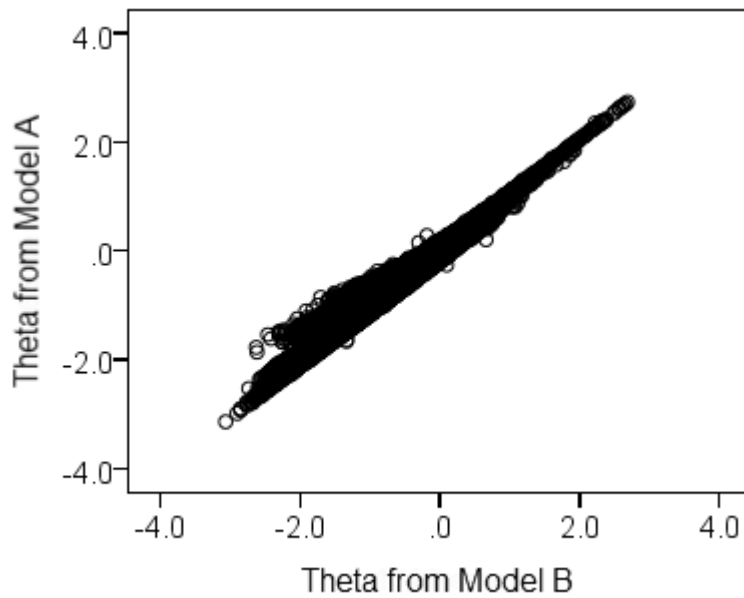
Model A = The IRT model explained in the present article (theta mean = −0.87, *SD* = 1.01); Model B = An alternative IRT model in which two common items were allowed to have DIF across studies (theta mean = −0.88, *SD* = 0.99). The correlation between Models A and B across studies were 0.99. The rank orders of individuals within studies were also extremely similar, with correlations of theta scores ranging from 0.97 to 1.00 between Models A and B. The deviance information criterion (DIC) values for Models A and B were 11,049 and 11,051, which favored the original, no-DIF IRT model.

Figure S3. Mean latent trait scores across studies from the original, no-DIF IRT model (A) and the DIF IRT model (B)
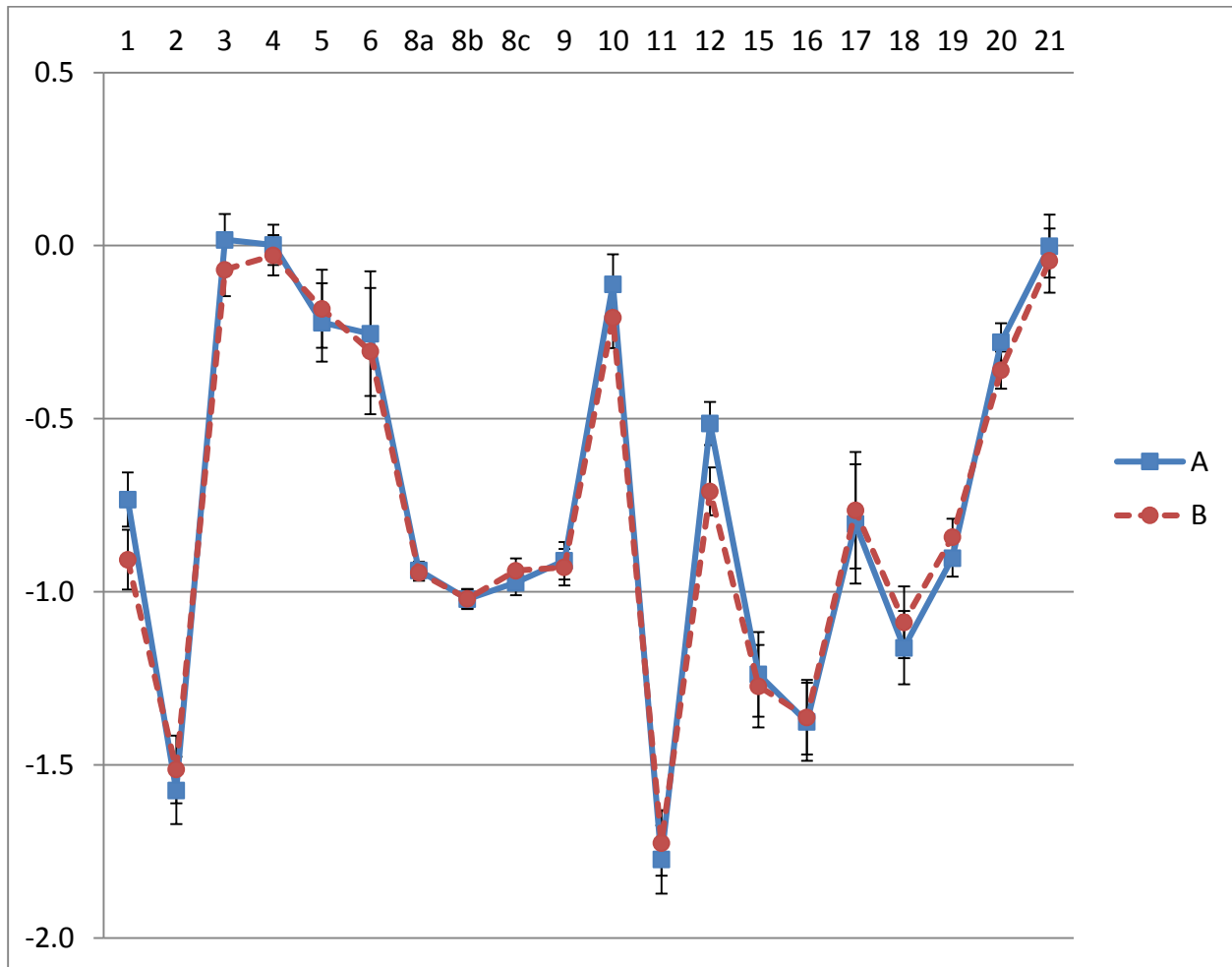
Model A = Theta scores for alcohol-related problems from the original, no-DIF IRT model explained in the present article. Model B = Theta scores from the alternative IRT model in which two common items were allowed to have DIF across studies. *Y*-axis indicates latent trait scores and *x*-axis indicates study. Error bars represent 2 times *SE* (standard error) in each direction, essentially representing the lower and upper bounds of 95% confidence intervals. Results shown in Figures S2 and S3 suggest that these two IRT models yielded essentially the same trait scores at the individual level and also at the study level.