HEALTH AND ADDICTIONS/SALUD Y DROGAS
ISSN 1578-5319 - ISSNe 1988-205X

Vol. 26. Núm. 1, 35-48 2026
DOI: https://doi.org/10.21134/1067
RECIBIDO: 12/03/2025 ACEPTADO: 23/02/2026

ENHANCING ALCOHOL USE SELF-REPORT: AN EXPERIMENTAL STUDY ON THE ALCOHOL USE DISORDERS IDENTIFICATION TEST

MEJORANDO EL AUTORREPORTE DEL CONSUMO DE ALCOHOL: UN ESTUDIO EXPERIMENTAL SOBRE EL CUESTIONARIO DE IDENTIFICACIÓN DE LOS TRASTORNOS DEBIDOS AL CONSUMO DE ALCOHOL

Paula Victoria Gimenez
Institute of Basic, Applied Psychology and Technology (IPSIBAT), National Scientific and Technical Research Council (CONICET), National University of Mar del Plata (UNMdP), Mar del Plata, Argentina
paulagimenez@mdp.edu.ar
https://orcid.org/0000-0002-3090-5228

Tomás Salomón
Institute of Basic, Applied Psychology and Technology (IPSIBAT), National Scientific and Technical Research Council (CONICET), National University of Mar del Plata (UNMdP), Mar del Plata, Argentina
tomas.salomon.mdp@gmail.com
https://orcid.org/0000-0003-0797-8581

Raquel Inés Peltzer
Institute of Basic, Applied Psychology and Technology (IPSIBAT), National Scientific and Technical Research Council (CONICET), National University of Mar del Plata (UNMdP), Mar del Plata, Argentina
ripeltzer@gmail.com
https://orcid.org/0000-0002-9116-4478

Mariana Cremonte
Institute of Basic, Applied Psychology and Technology (IPSIBAT), National Scientific and Technical Research Council (CONICET), National University of Mar del Plata (UNMdP), Mar del Plata, Argentina
marianacremonte@gmail.com
https://orcid.org/0000-0001-5853-1873

***Karina Conde**
Institute of Basic, Applied Psychology and Technology (IPSIBAT), National Scientific and Technical Research Council (CONICET), National University of Mar del Plata (UNMdP), Mar del Plata, Argentina
kariknc@gmail.com
https://orcid.org/0000-0003-1913-8328

Cómo citar: Gimenez, P. V., Salomón, T., Peltzer, R. I., Cremonte, M., & Conde, K. (2026). Enhancing alcohol use self-report: An experimental study on the Alcohol Use Disorders Identification Test. Health and Addictions/Salud y Drogas, 26(1), 35-48. https://doi.org/10.21134/1067

*Autor de correspondencia

Resumen

Objetivo: Evaluamos si reorganizar el orden de los ítems y las opciones de respuesta en el Test de Identificación de los Trastornos por Uso de Alcohol (AUDIT) podría mejorar su desempeño psicométrico. **Método:** Se realizó un experimento factorial 2x2 con una muestra no probabilística de la población general argentina (n=1128; M=37,29 años, SD=11,93; 75,9% mujeres cis). Los participantes fueron asignados aleatoriamente a una de cuatro versiones del AUDIT, variando el orden de presentación de los ítems (consumo de alcohol primero vs. problemas relacionados con el alcohol primero) y el orden de las opciones de respuesta ("nunca" a "diariamente/casi diariamente" vs. orden inverso). **Resultados:** Las versiones que presentaban primero los problemas mostraron mayor fiabilidad. La versión con estos ítems primero y con opciones de respuesta en orden inverso tuvo la mayor consistencia interna (α=0,88). La tendencia a responder afirmativamente a los problemas fue mayor en estas versiones, independientemente del orden de respuesta. Además, los ítems tuvieron mejor desempeño cuando se presentaban primero los problemas. **Conclusiones:** El orden de las opciones de respuesta parece irrelevante, pero presentar primero los problemas pareciera mejorar el desempeño del AUDIT. Este cambio podría facilitar una adecuada identificación de las personas que necesitan asistencia de los servicios de salud, lo que deberían profundizar futuras investigaciones.

Abstract

Objective: We examined whether rearranging the order of items and response options in the Alcohol Use Disorders Identification Test (AUDIT) could improve its psychometric properties. **Method:** We conducted a 2x2 factorial experiment with an Argentinian general population sample (n=1128; M=37.29 years, SD=11.93; 75.9% cis-women). We randomly assigned participants to one of four AUDIT versions, varying the order of item presentation (alcohol consumption first vs. alcohol-related problems first) and response options ("never" to "daily/almost daily" vs. reverse order). **Results:** Versions presenting alcohol-related problems first showed higher reliability. The version with alcohol-related problems first and response options in reverse order had the highest internal consistency (α=.88). Endorsement for alcohol-related problems was also higher in these versions, regardless of response option order. The versions presenting items about alcohol-related problems first also showed better item performance, regardless of the response options' order. **Conclusions:** the order of response options seems irrelevant, but presenting items about alcohol-related problems first appears to enhance the AUDIT's performance. This change could facilitate the adequate identification of people needing assistance from health services, which future studies should further explore.

Palabras clave

Auto-reporte del consumo de alcohol; consumo perjudicial; tamizaje; Test de Identificación de los Trastornos por Uso de Alcohol; problemas relacionados con el consumo de alcohol.

Key Words

Alcohol self-report; harmful drinking; screening; Alcohol Use Disorders Identification Test; alcohol related-problems.

1. Introduction

People often answer self-report questionnaires considering what they perceive as socially appropriate, leading to a bias (Vergés, 2022). Bias, such as social desirability, are present when collecting sensitive information that compromises privacy and what is culturally accepted. It has been found that people interested in presenting a favorable image of themselves tend to report less drinking, even in confidential and anonymous surveys (Davis et al., 2010). Underreporting alcohol consumption can transfer errors to epidemiology studies, for example, those that estimate the prevalence of harmful use (Ay et al., 2025; Harford, 1994), alcohol use problems (Steinweg & Worth, 1993), or alcohol-attributable deaths (Esser et al., 2022), significantly impacting public health policy decisions.

The Alcohol Use Disorders Identification Test [AUDIT] is a screening questionnaire developed by the World Health Organization [WHO] (Babor et al., 2001). It has the advantage of brevity, with only ten items. The first three items are AUDIT-C (Bradley et al., 2003; Higgings-Biddle & Babor, 2018) and measure the habitual quantity of standard units consumed per occasion, frequency of alcohol consumption, and frequency of heavy episodic drinking (having six or more standard units per occasion). The remaining seven items measure problems related to alcohol consumption (e.g., not being able to stop drinking once started). Following Selin et al. 2006, from now on we will refer to these seven items as AUDIT-P. The AUDIT has been translated into different languages (e.g., Cremonte et al., 2010; Neufeld et al., 2023) and adapted to diverse populations (e.g., Schøler et al., Pereira et al., 2025). Compared with similar screening questionnaires, the AUDIT presents better psychometric properties (Cremonte et al., 2010), and its use is highly recommended (Ohtani et al., 2023).

The AUDIT is employed in healthcare settings to identify people at risk and help them change their alcohol use through brief interventions (Babor & Higgins-Biddle, 2001). If the answers given to the AUDIT are insincere, appropriate recommendations will not be given. The difficulties of the AUDIT to adequately detect an alcohol use disorder have been noted, especially among people with greater stigmatization, for example, women in specific cultural contexts (Lange et al., 2019).

In addition, lower AUDIT scores were found when the level of social desirability increased (Adong et al., 2019; Latkin et al., 2017). A straightforward way to reduce bias could be to alter the order of the items. Traditionally, the AUDIT consists of the first three questions about quantity and frequency of alcohol consumption in the last 12 months and then the seven questions about alcohol problems. Considering that alcohol consumption is highly accepted in many cultures, questions regarding quantity and frequency of consumption might be less likely to be affected by bias than questions regarding alcohol-related problems. Bischof et al. (2005), in a study with a sample of general practitioner patients, found higher scores in quantity and frequency items with the traditional AUDIT and higher scores in alcohol problems questions when these were introduced first. However, other studies did not find differences in the scores varying from the original presentation of the instrument (Kypri et al., 2016; Ivis et al., 2000)

Another simple way to reduce the AUDIT's bias could be to reverse the AUDIT's answer options. Answer options go from the least risky (e.g., "never") to the most severe (e.g., "daily/almost daily to never"). Presenting the least socially desirable option (e.g., "daily/almost daily" to "never") first could facilitate its endorsement. In addition, other studies have observed that the response order affects the Likert scale's scores, inflating appropriate behaviors (Chyung et al., 2018; Hays et al., 1994).

This study aimed to determine if the AUDIT performance improves by alternating the order of presentation of a) the items about alcohol consumption compared to the items about alcohol problems and b) its answer options: presenting first the less socially desirable option compared to the most socially desirable option (i.e., high quantity of alcohol consumption vs low quantity). We hypothesize that presenting the least socially desirable items and answer options first will enhance the accuracy of AUDIT performance.

2. Method

2.1. Design

We used a between-subjects 2x2 factorial experimental design. Participants were randomized to one of four possible AUDIT versions. The two independent variables were a) the order of the AUDIT items (either consumption or problems presented first) and b) the order in which the response options were presented (the most or least socially desirable presented first). The four resulting conditions, each administering a different AUDIT version, are described in Figure 1.

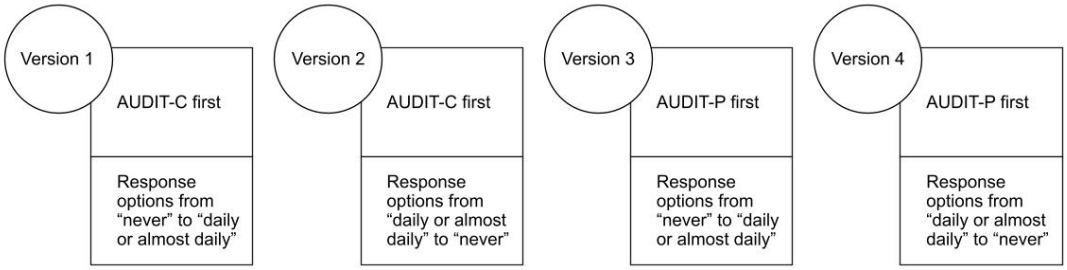

Figure 1

Four experimental conditions, each with different versions of the AUDIT

Note. The four versions resulted from combining four conditions: AUDIT's consumption or problem items presented first, and AUDIT's most or least socially desirable response options presented first. Version 1= items about alcohol consumption first and responses options from "never" to "daily or almost daily"; Version 2= items about alcohol consumption first and responses options from "daily or almost daily" to "never"; Version 3= items about alcohol problems first and responses options from "never" to "daily or almost daily"; Version 4= items about alcohol problems first and responses options from "daily or almost daily" to "never".

2.2. Participants

We collected data from a non-probabilistic convenience sample of Argentina's general population. We included residents of Argentina who were 18 years of age or older and consented to participate. We collected data through a web survey, and n=1940 eligible participants entered the survey link. The final sample was composed of n=1128 participants. Figure 2 shows the flow diagram of all eligible and excluded participants.

Participants were from all the Argentinian provinces. The majority were from Buenos Aires (48.7%), followed by Córdoba (6.4%), the Autonomous City of Buenos Aires (4.4%) and Jujuy (4.1%). There were no significant differences between experimental conditions (see the Design section below) regarding age, gender, and educational level. Table 1 shows the sociodemographic characteristics of the sample.

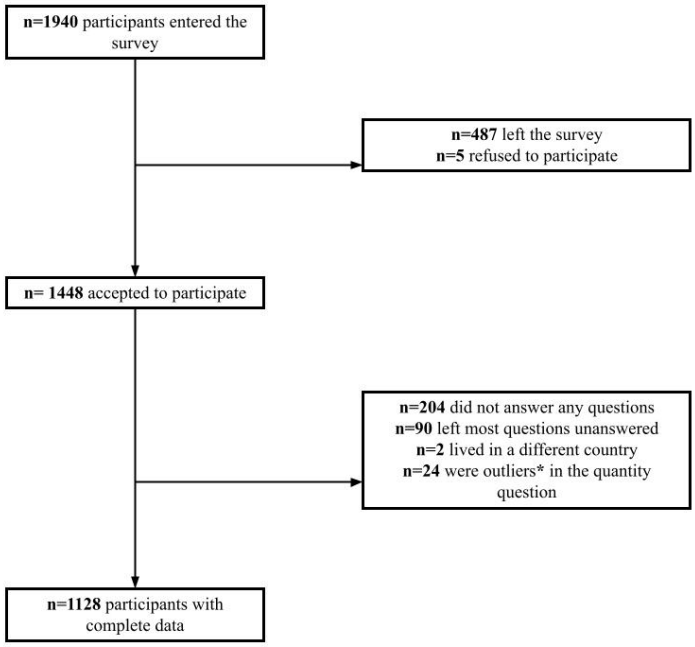

Figure 2

Flow diagram

*****Outliers were identified by examining dispersion plots; values reflecting physiologically implausible levels of consumption (e.g., reports of 300 standard units on a single occasion) were considered indicative of response or data-entry errors and were excluded from the analyses, in line with common practices in alcohol research.

Table 1

Sociodemographic characteristics, general population, Argentina, N=1125

	All	AUDIT[a] Version 1[b] (n=273)	AUDIT Version 2[c] (n=258)	AUDIT Version 3[d] (n=302)	AUDIT Version 4[e] (n=292)	X^2
Age (M, SD)	37.29(11.93)	37.11(11.16)	37.17(11.91)	37.63(12.39)	37.22(12.22)	–
Gender identity %						10.01
Cis women	75.9	75.5	74.8	77.5	75.7	
Trans women	0.2	0.4	–	0.3	–	
Cis male	22.8	22.7	23.6	21.9	23.3	
Trans male	0.4	0.7	–	–	0.7	
Other	0.7	0.7	1.6	0.3	0.3	
Educational level %						16.46
Primary	0.4	0.4	0.8	–	0.3	
Incomplete secondary	2.4	4.4	1.6	2	1.7	
Complete secondary	12	10.3	13.2	12.9	11.6	
Incomplete tertiary or university	34.7	35.5	34.1	34.8	34.2	
Complete tertiary or university	36.5	36.3	34.5	36.1	39	
Postgraduate studies	14	13.2	15.9	14.2	13	

Note. No statistically significant differences were found between groups for any sociodemographic variables (all p > .05).
[a]AUDIT= Alcohol Use Disorders Identification Test. [b]Version 1= items about alcohol consumption first and response options from "never" to "daily or almost daily". [c]Version 2= items about alcohol consumption first and response options from "daily or almost daily" to "never". [d]Version 3= items about alcohol problems first and response options from "never" to "daily or almost daily". [e]Version 4= items about alcohol problems first and response options from "daily or almost daily" to "never".

2.3. Instruments

Alcohol Use Disorders Identification Test (AUDIT). The AUDIT contains ten items aimed at assessing the level of risk of alcohol consumption in the last 12 months. The items inquire about 1) habitual quantity of alcohol consumption, measured in standard units (1 standard unit=12 grams of alcohol), 2) frequency of alcohol consumption, 3) frequency of heavy episodic drinking, measured as six or more standard units consumed per occasion, 4) losing control over drinking (not being able to stop), 5) not being able to do what is expected due to consumption, 6) morning consumption after an occasion of excessive consumption, 7) guilt or remorse after consumption, 8) alcohol blackouts, 9) injuries to oneself or others due to drinking and 10) other people's concerns about one's drinking. The answer options range from 0 to 4, and the total score from 0 to 40 (Babor et al., 2001).

AUDIT-C. The AUDIT-C comprises the three items of the original test regarding quantity and frequency of consumption and frequency of heavy episodic drinking (items 1 to 3 as listed above) (Bradley et al., 2003).

AUDIT-P. The AUDIT-P comprises the seven items of the original test regarding alcohol-related problems (items 4 to 10 as listed above) (Selin, 2006).

Sociodemographic characteristics. We asked about the province of residence, age in years, gender identity (cis-women, trans-women, cis-male, trans-male or other), and educational level (primary, secondary, uncompleted tertiary or university, completed tertiary or university, postgraduate studies).

2.4. Procedure

Data was gathered between September 2023 and July 2024. We invited participants to the study through our research group's official websites and online social networks (Facebook, Instagram, and Twitter). We used the LimeSurvey platform for data collection, which automatically randomized participants. First, participants received the information sheet containing the study's detailed characteristics. Then, those who accepted to participate had access to the self-administered questionnaire.

2.5. Data Analysis

This study evaluated the performance of different AUDIT versions by altering the order of the items and the answer options. First, we compared the reliability through the internal consistency of each version. To this end, we used Cronbach's alpha and, given that the AUDIT may present some multidimensionality and that the response options are ordinal, the McDonald's omega coefficient. Since the alpha and omega coefficients were almost identical, Cronbach's alpha values were compared across the four AUDIT versions using AlphaTest chi-square (χ^2) (Lautenschlager & Meade, 2008). Second, we described the scores for the complete AUDIT, the AUDIT-C, and the AUDIT-P and examined which of the four AUDIT versions resulted in greater endorsement, thus shielding higher scores. For this purpose, we used single negative binomial regressions, including the total AUDIT scores, the AUDIT-C scores, and the AUDIT-P scores as the outcome, and compared them between versions by pairs. To minimize type I error due to multiple comparisons, pairwise comparisons between the four groups were adjusted using the Bonferroni correction ($\alpha = .05/6 = .008$). We also evaluated differences among missing responses (yes/no) and harmful drinking (i.e., AUDIT total score of 7 or more, yes/no) with χ^2. Lastly, we examined each item endorsement (dichotomous items, any score above zero as endorsed) to estimate its sensibility and specificity to detect harmful drinking using a proxy measure of a criterion of a positive score (≥ 7) (Conde et al., 2018) for the whole AUDIT. We also estimated item-total correlation and the differences in the response categories by versions in two subsets, presence or absence of harmful drinking (i.e., AUDIT total score ≥ 7), with

χ^2 for the original ordinal items. We used the Fisher r-to-z transformation to test the significance of the difference between two independent r values (http://vassarstats.net/rdiff.html). Data was analyzed using SPSS version 21, R version 4.3.0, and the McDonald's Omega module from JAMOVI version 2.6 (https://www.jamovi.org).

2.6. Ethical Considerations

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000. Informed consent was obtained from all participants included in the study. The data gathered was confidential and anonymous. No compensation was offered for participating in the study. The Bioethics Interdisciplinary Thematic Program of the Universidad Nacional de Mar del Plata, Argentina, approved the study.

3. Results

Reliability through internal consistency of each AUDIT version

All the AUDIT versions showed good to excellent reliability. The version that starts with items about alcohol problems first and answer options from "daily/almost daily" to "never", showed the highest internal consistency. Internal consistency for each version was: version 1 -the original AUDIT-: α=.79; ω=.81, version 2: α=.83; ω=84, version 3: α=.87; ω=88, and version 4: α=.88; ω=.88. However, the versions presenting the items about alcohol problems first had higher internal consistencies: significant differences were found between versions 3 and 1 (X^2=15.49; p<.01), between versions 4 and 1 (X^2=15.46; p<.01) and between versions 2 and 4 (X^2=4.64; p<.05).

Level of endorsement for each AUDIT version comparing their AUDIT, AUDIT-C, and AUDIT-P scores

Table 2 shows the range scores, their mean and standard deviations in the AUDIT, AUDIT-C, and AUDIT-P for each AUDIT version.

Table 2

Score results for the AUDIT, AUDIT-C, and AUDIT-P by AUDIT version, N=1104

		AUDIT[a]		AUDIT-C[b]		AUDIT-P[c]	
AUDIT version	n	Min-Max	M(SD)	Min-Max	M(SD)	Min-Max	M(SD)
1[d]	273	0-23	5.7(4.49)	0-10	3.72(2.3)	0-15	1.98(2.83)
2[e]	258	0-30	6.51(5.85)	0-11	3.99(2.58)	0-21	2.52(4)
3[f]	287	0-32	6.47(6.26)	0-11	3.67(2.33)	0-24	2.72(4,36)
4[g]	286	0-39	6.7(6.55)	0-12	3.71(2.53)	0-27	2.97(4.67)

Note. [a]AUDIT= Alcohol Use Disorders Identification Test. Scores on the AUDIT range 0-40. Higher scores indicate higher alcohol consumption and related problems. [b]The three items of the original test regarding quantity and frequency of consumption and frequency of heavy episodic drinking. [c]The seven items of the original test regarding alcohol-related problems. [d]Version 1= items about alcohol consumption first and response options from "never" to "daily or almost daily". [e]Version 2= items about alcohol consumption first and response options from "daily or almost daily" to "never". [f]Version 3= items about alcohol problems first and response options from "never" to "daily or almost daily". [g]Version 4= items about alcohol problems first and response options from "daily or almost daily" to "never".

Although weak, we initially observed a tendency toward higher endorsement (i.e., higher AUDIT total scores) when participants received version 4 (items about alcohol problems presented first, with response options ranging from “daily or almost daily” to “never”) compared with the original AUDIT (version 1); however, this difference did not remain statistically significant after applying the Bonferroni correction. In contrast, endorsement to items assessing alcohol-related problems (AUDIT-P), and consequently higher scores, were

significantly greater for version 4 compared with the original AUDIT, even after the Bonferroni correction. Table 3 shows the differences among versions for the AUDIT, AUDIT-C, and AUDIT-P scores.

There was no difference among the four versions in missing responses by item, nor by the presence of harmful drinking (i.e., participants scoring below and above 7) (not shown in Tables).

Table 3

Scores' comparison for the AUDIT, Audit-C, and AUDIT-P between AUDIT versions, N=1128

Predictor	AUDIT[a]			AUDIT-C[b]			AUDIT-P[c]		
	Wald X^2	B(SE)	CI 95%	Wald X^2	B(SE)	CI 95%	Wald X^2	B(SE)	CI 95%
Version 1[d] (Ref. 2)	3.26	0.13(0.07)	[-0.01, 0.28]	1.63	0.07(0.06)	[-0.04, 0.18]	3.35	0.24(0.13)	[-0.02, 0.5]
Version 1 (Ref. 3)	2.89	0.13(0.07)	[-0.02, 0.27]	0.05	-0.01(0.05)	[-0.12, 0.09]	6.25	0.32(0.13)	[0.07, 0.56]
Version 1 (Ref. 4[g])	4.64	0.16(0.08)	[0.01, 0.31]	0.01	-0.01(0.05)	[-0.11, 0.11]	10.19	0.40(0.13)	[0.16, 0.65] **
Version 2[e] (Ref. 3)	0.01	-0.01(0.08)	[-0.16, 0.15]	2.27	-0.08(0.06)	[-0.19, 0.03]	0.32	0.08(0.14)	[-0.19, 0.34]
Version 2 (Ref. 4)	0.13	0.03(0.08)	[-0.13, 0.19]	1.61	-0.07(0.06)	[-0.18, 0.04]	1.45	0.16(0.13)	[-0.1, 0.43]
Version 3[f] (Ref. 4)	0.18	0.04(0.08)	[-0.12, 0.19]	0.04	0.01(0.06)	[-0.1, 0.12]	.442	0.09(0.13)	[-0.17, 0.34]

Note. Negative binomial regressions. Significance levels were adjusted using the Bonferroni correction for six pairwise comparisons (α = .05/6 = .008) ***p*<.008. [a]AUDIT= Alcohol Use Disorders Identification Test. [b]The three items of the original test regarding quantity and frequency of consumption and frequency of heavy episodic drinking. [c]The seven items of the original test regarding alcohol-related problems. [d]Version 1= items about alcohol consumption first and response options from "never" to "daily or almost daily". [e]Version 2= items about alcohol consumption first and response options from "daily or almost daily" to "never". [f]Version 3= items about alcohol problems first and response options from "never" to "daily or almost daily". [g]Version 4= items about alcohol problems first and response options from "daily or almost daily" to "never".

Item level analysis for each AUDIT version

Table 4 shows each item's performance (sensibility, specificity, and item-total correlation) for each AUDIT version and the strength of the association between endorsing each item and the AUDIT version received, separately for those with and without harmful drinking. While sensibility and specificity indexes did not seem to vary across versions, we observed significantly better performance in item-total correlation (p<.01) for items 4 (losing control), 7 (guilt), and 8 (blackout) in versions 2, 3, and 4, item 6 (morning consumption) and 9 (injuries to oneself or others) in version 3 and version 4, item 10 (other people concerns for one's drinking) in version 3, and item 5 (not being able to do what is expected) in version 4. Among participants with harmful drinking, we found a tendency for low endorsement of the never category in item 3 (frequency of heavy episodic drinking) and 9 (injuries to oneself or others) in AUDIT versions 3 and 4, and more endorsement of higher frequencies response options for item 5 (not being able to do what is expected) in version 3 and 4, and item 7 (guilt) in versions 2, 3 and 4 (not shown in Tables).

Table 4

Item level analysis for each AUDIT version: sensibility, specificity, and item-total correlation; and the strength of the association between endorsing each item and the AUDIT version received, separately for those with and without harmful drinking, N=1128

Items	AUDIT[a] Version 1[b]			AUDIT Version 2[c]			AUDIT Version 3[d]			AUDIT Version 4[e]			Harmful drinking[f]	
													No	Yes
	Se[g]	Sp[h]	r	Se	Sp	r	Se	Sp	r	Se	Sp	r	phi	phi
Frequency (item 1)	.30	1	.42	.33	1	.48	.29	1	.38	.34	1	.45	.11	.21
Quantity (item 2)	.59	.94	.44	.57	.88	.41	.51	.94	.53	.61	.91	.47	.11	.15
Frequency of heavy episodic drinking (item 3)	.40	.97	.68	.44	.97	.66	.40	1	.75	.49	1	.73	.17**	.26*
Losing control (item 4)	.78	.87	.58	.79	.85	.70*	.73	.88	.72*	.76	.87	.73*	.10	.21
Not being able to do what is expected (item 5)	.86	.83	.57	.76	.76	.47	.73	.88	.62	.76	.87	.70*	.09	.27*
Morning consumption (item 6)	.76	.77	.36	.93	.73	.38	.72	.78	.54*	.80	.75	.57*	.08	.24
Guilt or remorse after consumption (item 7)	.54	.87	.45	.64	.87	.68*	.52	.90	.58*	.59	.88	.70*	.10	.26*
Blackout (item 8)	.61	.87	.53	.63	.85	.66*	.63	.91	.68*	.64	.86	.68*	.02	.24
Injuries to oneself or others (item 9)	.79	.78	.24	.77	.75	.32	.75	.82	.58*	.77	.78	.53*	.08	.20*
Other people's concerns (item 10)	.91	.80	.50	.88	.80	.59	.80	.85	.62*	.76	.79	.57	.10	.17

Note. *p<.05; **p<.001. We used the Fisher r-to-z transformation to test the significance of the difference between two independent r values. [a]AUDIT= Alcohol Use Disorders Identification Test. [b]Version 1= items about alcohol consumption first and response options from "never" to "daily or almost daily". [c]Version 2= items about alcohol consumption first and response options from "daily or almost daily" to "never". [d]Version 3= items about alcohol problems first and response options from "never" to "daily or almost daily". [e]Version 4= items about alcohol problems first and response options from "daily or almost daily" to "never". [f]Harmful drinking= AUDIT total score ≥ 7. [g]Se= sensibility [h]Sp= specificity.

4. Discussion

In this article, we compare four different versions of the AUDIT. These versions contain the same items and response options but diverge in the order in which they are presented to participants. Thus, we aimed to explore whether the order of questions and response options alters AUDIT outcomes by presenting more socially desirable items and response options at the beginning or the end of the test.

On the one hand, we found that introducing questions about alcohol problems first seems to increase the AUDIT's performance. When AUDIT-P was introduced before AUDIT-C, reliability increased, and the items' performance improved. In addition, participants reported more alcohol problems. This finding supports our hypothesis that presenting the least socially desirable choice first reduces social desirability bias. Garbarski et al. (2015) reported similar results, finding that people rate their health positively when the most socially desirable option is presented first. Also in line with our findings, Bendtsen et al. (2023) showed that presenting questions on alcohol-related problems first influences subsequent consumption measures, leading participants to report higher levels of drinking.

Another not mutually exclusive reason for this finding is that presenting questions about alcohol problems first facilitates people's understanding of the purpose of the instrument. When participants are asked about the negative consequences of alcohol consumption at the beginning, they may recognize that the tool is assessing the severity of their alcohol use. When the AUDIT-P is presented first, people with harmful drinking reported low endorsement to "never" categories and more positive answers to problem questions, as expected. Accordingly, they answered the AUDIT-C questions in line with their AUDIT-P responses. We believe inverting the order of the questions increased accuracy because individuals used their previous responses about problems to answer how much and with what frequency they drank (Hays et al., 1994). However, future studies should test this conjecture since the literature is contradictory about the influence of previous answers (Bendtsen et al., 2023; Bischof et al., 2005; Kypri et al., 2016).

On the other hand, and contrary to our expectations, altering the answer options from the least socially desirable to the most did not appear to affect the AUDIT's performance. This evidence is against different order-response effects, such as the extreme response bias that assumes people tend to choose first or last options (Costa & Filho, 2019). In addition, our findings are contrary to the primacy effect and descending bias identified in written surveys with Likert response options. Other studies have shown that individuals select the first answer option, maybe because most writing systems read left to right (Chyung et al., 2018). Our results also disagree with studies that found higher scores in self-report questionnaires when the answer options were ordered from most frequent to least frequent (Chyung et al., 2018; Hays et al., 1994). The discrepancy between our findings and the literature underscores the need to replicate and extend our study ideally including other measures such as participants' history of alcohol use, other substance use or prior treatment, and social desirability, perception of judgment, stigma, or manipulation checks.Interestingly, the AUDIT-C scores did not significantly differ when comparing the four instrument versions. Although this finding may show the robustness of the AUDIT-C, it is essential to highlight that the sensibility of its three items was the worst, and the reliability of the whole instrument increased significantly when the AUDIT-P was presented first, suggesting people's responses to the consumption items were more consistent when they had first answered the problem items. Other studies in our context found similar results respecting AUDIT-C's capability to detect people with harmful drinking (Conde et al., 2018; Cremonte et al., 2010). The AUDIT-C has the advantage of being short and easy to use where time is scarce; therefore, more research is necessary to improve its psychometric performance (Toner et al., 2019). Wardell et al. (2020) recommended including the losing control item (item 4) in AUDIT-C to increase accuracy. Another possibility that future studies should test is dropping the frequency of consumption question (item 2) in countries with Mediterranean drinking cultures where individuals may drink frequently but not heavily (Savic et al., 2016).

One outcome that deserves to be highlighted is that the differences between the AUDIT versions were slight and may seem limited to the AUDIT-P items. However, and importantly, the reliability of the whole instrument improved significantly when the AUDIT-P was administered first. Validity is often considered the most essential psychometric property; since we did not alter the nature or content of the AUDIT items, we cannot conclude about the validity of the AUDIT. However, high reliability is a necessary condition for measurement validity (Thurstone, 1931), and it can be argued that increasing reliability without narrowing the scales' content (i.e., the construct-related information the test provides) may indirectly strengthen structural validity (Clark, & Watson, 2019).

While more categorical evidence is required to justify modifying the structure of a widely used and robust screening test, our findings should not be dismissed. The differences between the AUDIT versions could have a significant impact, considering that essential clinical and policy decisions (e.g., brief interventions, referral to treatment, population-level interventions) are made based on a cut-off score. Thus, a more accurate test could determine whether some people receive appropriate and much-needed care, whether resources are allocated, and whether policies are set in place.

4.1. Limitations

Finally, some limitations of this study should be addressed. First, the AUDIT was self-administered with an online survey. Although our previous study shows that the AUDIT performs equally well when self-administered online or administered in interviews (Conde & Cremonte, 2015), it could be possible that social desirability bias operates differently according to the administration method. Second, different people and groups may have diverse conceptions of what is or is not socially desirable. For example, it could be desirable for adolescents and young people to show themselves as individuals who drink heavily. Nevertheless, randomization to each version neutralized this possible bias, as we found non a priori differences among conditions. Additionally**,** future studies could incorporate a measure of social desirability to test the hypothesis that the observed changes resulting from alterations in the order of items or response options are attributable to a reduction in social desirability bias. Third, the sample was non-probabilistic, and predominantly composed by women so it may not represent the general population in Argentina, limiting the generalization of results. Future research should include a more significant proportion of male participants, especially given that men typically exhibit higher alcohol consumption than women. A fourth limitation that should be addressed in future studies is the lack of data on participants' history of alcohol or other drug use, as well as prior treatment experiences. In addition, other relevant measures—such as social desirability bias and manipulation checks—were not included. Although random assignment may help balance some of these unmeasured variables across groups, future studies would benefit from a more comprehensive assessment incorporating this information. Lastly, item-level sensitivity and specificity were estimated using the AUDIT total score as the criterion, which implies a lack of independence between the predictors and the criterion and may have partially inflated the estimates.

4.2. Conclusions

In sum, presenting the items about alcohol-related problems first, regardless of the response order options, appears to increased reliability, endorsement about alcohol-related problems, and improved item performance. Our results suggest that implementing this question order modification may be recommended. Even though the traditional AUDIT has adequate performance, this change could enhance its psychometric performance. It is important to replicate this study to ensure that healthcare and surveillance settings utilize the AUDIT version that optimizes the identification of those needing assistance and make well-informed decisions.

4.3. Acknowledgements

The authors would like to thank all participants for their time and contribution to this study.

4.4. Conflict of interest

None to declare.

4.5. Funding

Funding for this study was provided by the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina.

4.6. Author contributions

PVG: conceptualization, methodology, investigation, Writing – Original Draft. TS: conceptualization, methodology, investigation, Writing – Review & Editing. RIP: conceptualization, methodology, investigation, supervision. MC: conceptualization, methodology, investigation, supervision. KC: conceptualization, methodology, investigation, Data Curation, Formal analysis, supervision, funding acquisition.

4.7. Data availability

The datasets generated during and/or analyzed during the current study are not publicly available due confidentiality reasons but are available from the corresponding author on reasonable request.

5. References

Adong, J., Fatch, R., Emenyonu, N. I., Cheng, D. M., Muyindike, W. R., Ngabirano, C., ... & Hahn, J. A. (2019). Social desirability bias impacts self-reported alcohol use among persons with HIV in Uganda. *Alcoholism: Clinical and Experimental Research, 43*(12), 2591-2598. https://doi.org/10.1111/acer.14218

Ay, A., Çam, C., Kilinc, A., Fatih Önsüz, M., & Metintaş, S. (2025). Prevalence of hazardous alcohol consumption and evaluation of associated factors in university students. *Social psychiatry and psychiatric epidemiology, 60*(1), 223-233. https://doi.org/10.1007/s00127-024-02680-8

Babor, T. F., & Higgins-Biddle, J. C. (2001). *Brief Intervention. For Hazardous and Harmful Drinking. A Manual for Use in Primary Care.* World Health Organization. https://apps.who.int/iris/handle/10665/67210

Babor, T. F., Higgins-Biddle, J. C., Saunders, J. B., & Monteiro, M. G. (2001). *The alcohol use disorders identification test.* World Health Organization. https://iris.who.int/bitstream/handle/10665/67205/W?sequence=1

Bendtsen, M., Garnett, C., Toner, P., & Shorter, G. W. (2023). The effect of question order on outcomes in the core outcome set for brief alcohol interventions among online help-seekers. *Addiction Science and Clinical Practice, 18*(Suppl 1), O5.

Bischof, G., Reinhardt, S., Grothues, J., Dybek, I., Meyer, C., Hapke, U., ... & Rumpf, H. J. (2005). Effects of item sequence on the performance of the AUDIT in general practices. *Drug and Alcohol dependence*, *79*(3), 373-377. https://doi.org/10.1016/j.drugalcdep.2005.03.002

Bradley, K. A., DeBenedetti, A. F., Volk, R. J., Williams, E. C., Frank, D., & Kivlahan, D. R. (2007). AUDIT-C as a brief screen for alcohol misuse in primary care. *Alcoholism: Clinical and Experimental Research, 31*(7), 1208-1217. https://doi.org/10.1111/j.1530-0277.2007.00403.x

Chyung, S. Y., Kennedy, M., & Campbell, I. (2018). Evidence-based survey design: The use of ascending or descending order of Likert-type response options. *Performance Improvement, 57*(9), 9-16. https://doi.org/10.1002/pfi.21800

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological assessment*, *31*(12), 1412. https://doi.org/10.1037/pas0000626

Conde, K., & Cremonte, M. (2015). Calidad de los datos de encuestas sobre consumo de alcohol en estudiantes universitarios. *Cadernos de Saúde Pública, 31*, 39-47. https://doi.org/10.1590/0102-311X00061114

Conde, K., Gimenez, P. V., & Cremonte, M. (2018). Características de los instrumentos de tamizaje del trastorno por uso de alcohol en jóvenes universitarios de Argentina. *Clínica y Salud, 29*(1), 27-33. https://dx.doi.org/10.5093/clysa2018a5

Costa, A. R. L., & Hauck Filho, N. (2019). Methods for the control of extreme response styles in self-report instruments: A review. *Trends in Psychology, 27*(2), 309-323. https://doi.org/10.9788/TP2019.2-02

Cremonte, M., Ledesma, R. D., Cherpitel, C. J., & Borges, G. (2010). Psychometric properties of alcohol screening tests in the emergency department in Argentina, Mexico and the United States. *Addictive Behaviors, 35*(9), 818- 825. https://doi.org/10.1016/j.addbeh.2010.03.021

Davis, C. G., Thake, J., & Vilhena, N. (2010). Social desirability biases in self-reported alcohol consumption and harms. *Addictive behaviors*, *35*(4), 302-311. https://doi.org/10.1016/j.addbeh.2009.11.001

Esser, M. B., Sherk, A., Subbaraman, M. S., Martinez, P., Karriker-Jaffe, K. J., Sacks, J. J., & Naimi, T. S. (2022). Improving estimates of alcohol-attributable deaths in the United States: impact of adjusting for the underreporting of alcohol consumption. *Journal of studies on alcohol and drugs*, *83*(1), 134-144. https://doi.org/10.15288/jsad.2022.83.134

Garbarski, D., Schaeffer, N. C., & Dykema, J. (2015). The effects of response option order and question order on self-rated health. *Quality of Life Research*, *24*, 1443-1453. https://doi.org/10.1007/s11136-014-0861-y

Harford, T. C. (1994). The effects of order of questions on reported alcohol consumption. *Addiction, 89*(4), 421-4.

Hays, R. D., Bell, R. M., Damush, T., Hill, L., Dimatteo, M. R., & Marshall, G. N. (1994). Do response options influence self-reports of alcohol use? *International journal of the addictions*, *29*(14), 1909-1920. http://dx.doi.org/10.3109/10826089409128263

Higgins-Biddle, J. C., & Babor, T. F. (2018). A review of the Alcohol Use Disorders Identification Test (AUDIT), AUDIT-C, and USAUDIT for screening in the United States: Past issues and future directions. *The American journal of drug and alcohol abuse, 44*(6), 578-586. https://doi.org/10.1080/00952990.2018.1456545

Ivis, F. J., Adlaf, E. M., & Rehm, J. (2000). Incorporating the AUDIT into a general population telephone survey: a methodological experiment. *Drug Alcohol Depend, 60*, 97-104. https://doi.org/10.1016/S0376-8716(00)80012-5

Kypri, K., Wilson, A., Attia, J., Sheeran, P., Miller, P., & McCambridge, J. (2016). Social desirability bias in the reporting of alcohol consumption: a randomized trial. *Journal of studies on alcohol and drugs, 77*(3), 526-531. https://doi.org/10.15288/jsad.2016.77.526

Lange, S., Shield, K., Monteiro, M., & Rehm, J. (2019). Facilitating screening and brief interventions in primary care: A systematic review and meta-analysis of the AUDIT as an indicator of alcohol use disorders. *Alcoholism: Clinical and Experimental Research*, *43*(10), 2028-2037. https://doi.org/10.1111/acer.14171

Latkin, C. A., Edwards, C., Davey-Rothwell, M. A., & Tobin, K. E. (2017). The relationship between social desirability bias and self-reports of health, substance use, and social network factors among urban substance users in Baltimore, Maryland. *Addictive behaviors*, *73*, 133-136. https://doi.org/10.1016/j.addbeh.2017.05.005

Neufeld, M., Bunova, A., Fadeeva, E., Nadezhdin, A., Tetenova, E., Vyshinsky, K., ... & Rehm, J. (2023). Translating and adapting the Alcohol Use Disorders Identification Test (AUDIT) for use in the Russian Federation: A multicentre pilot study to inform validation procedures. *Nordic Studies on Alcohol and Drugs*, *40*(5), 536-553. https://doi.org/10.1177/14550725231183236

Ohtani, Y., Ueno, F., Kimura, M., Matsushita, S., Mimura, M., & Uchida, H. (2023). Highly endorsed screening and assessment scales for alcohol problems: A systematic review. *Neuropsychopharmacology Reports*, *43*(4), 470-481. https://doi.org/10.1002/npr2.12363

Pereira, F. A., de Oliveira, L. G., da Silva, G. T., Scatena, A., Kim, H. S., & Andrade, A. L. M. (2025). Validation of alcohol use disorders identification test (AUDIT) in Brazilian colleges: Network analysis, measurement invariance and screening efficiency. *International Journal of Mental Health and Addiction, 23*(4), 2907-2925. https://doi.org/10.1007/s11469-024-01268-2

Savic, M., Room, R., Mugavin, J., Pennay, A., & Livingston, M. (2016). Defining "drinking culture": A critical review of its meaning and connotation in social research on alcohol problems. *Drugs: Education, Prevention and Policy, 23*(4), 270- 282. https://doi.org/10.3109/09687637.2016.1153602

Selin, K. H. (2006). Alcohol Use Disorder Identification Test (AUDIT): what does it screen? Performance of the AUDIT against four different criteria in a Swedish population sample. *Substance use & misuse*, *41*(14), 1881-1899. https://doi.org/10.1080/10826080601025532

Schøler, P. N., Andersen, M. B., Andersen, K., Becker, U., Thiele, M., & Nielsen, A. S. (2025). Validation of the alcohol use disorders identification test in a Danish hospital setting. *Substance Abuse Treatment, Prevention, and Policy, 20*(1), 7. https://doi.org/10.1186/s13011-025-00638-w

Steinweg, D. L., & Worth, H. (1993). Alcoholism: the keys to the CAGE. *The American journal of medicine*, *94*(5), 520-523. https://doi.org/10.1016/0002-9343(93)90088-7

Toner, P., Böhnke, J. R., Andersen, P., & McCambridge, J. (2019). Alcohol screening and assessment measures for young people: A systematic review and meta-analysis of validation studies. *Drug and alcohol dependence*, *202*, 39-49. https://doi.org/10.1016/j.drugalcdep.2019.01.030

Thurstone, L. L. (1931). Relation between reliability and validity of a test. En L. L. Thurstone *The reliability and validity of tests: Derivation and interpretation of fundamental formulae concerned with reliability and validity of tests and illustrative problems* (pp. 50–51). Edwards Brothers. https://doi.org/10.1037/11418-014

Vergés, A. (2022). On the desirability of social desirability measures in substance use research. *Journal of Studies on Alcohol and Drugs*, *83*(4), 582-587. https://doi.org/10.15288/jsad.2022.83.582

Wardell, J. D., Cunningham, J. A., Quilty, L. C., Carter, S., & Hendershot, C. S. (2020). Can the AUDIT consumption items distinguish lower severity from high severity patients seeking treatment for alcohol use disorder? *Journal of Substance Abuse Treatment*, *114*, 108001. https://doi.org/10.1016/j.jsat.2020.108001