

An analysis of item non-response in a survey of law students, attorneys, and judges

Catherine Lau Crisp, University of Arkansas at Little Rock

Kaye McKinzie, University of Central Arkansas

Candace McCown, Arkansas Judges and Lawyers Assistance Program

Jennifer Donaldson, Arkansas Judges and Lawyers Assistance Program

Abstract

Missing data is an important issue in summed scales used in survey research. It can have a significant impact on the quality of the research by reducing the usable sample size, reducing the statistical power in small samples, limiting the generalizability of the results, and forcing the researcher to make decisions about whether to exclude responses from the analysis or to use a data replacement method. Excluding responses from the analysis reduces the usable size while replacing the missing items may result in an overestimation or underestimation of scale scores, affect the measure's reliability, and increase the likelihood of finding statistically significant results when there are none. Despite these challenges, missing data is rarely the focus of research studies. This article focuses on missing data in a sample of lawyers and law students who completed a survey that consisted of six different summed scales, each of which required that respondents answer all questions in the measure to compute an accurate score. The questions of interest in this study were 1) whether any demographic groups or combinations of groups were more or less likely to respond to all items in the summed scales and 2) whether there were statistically significant relationships between respondents' willingness to complete all items in the different summed scales. Implications for further research are discussed.

Keywords: missing data, item non-response, law students, summed scales, surveys, judges, attorneys

An analysis of item non-response in a survey of law students and attorneys

Introduction

The use of summed scales in research has a long history. Its onset is frequently attributed to Rensis Likert, who developed the widely used and accepted Likert scale in 1932 and is credited with its origins. Spector (1992) states that summed scales have four characteristics:

1. The scale must contain multiple items that will be combined or summed.
2. Each item must measure something that varies quantitatively.
3. Each item has no "right" answer.
4. Each item must be phrased as a statement in which respondents are asked to give a rating that best reflects their response to the item.

One issue that arises in research using summed scales is the issue of missing values as all questions in the scale must be answered in order to compute a score that reflects the respondents' true feelings about the issue of interest. When respondents neglect to respond to items in a summed scale (referred to as item non-response), the researcher has two choices: either delete the response using one of two methods (listwise or pairwise) or replace the data using one of several methods. Decisions about whether to delete or replace the data are often made based on a variety of factors including the extent of the missing data and whether the data can be characterized as Missing Completely at Random (MCAR), Missing at Random (MAR), and Missing Not at Random (MNAR). Deleting individual cases in which a single data point is missing (listwise deletion) can significantly reduce to pool of usable data while replacing the data can artificially increase scores and may not reflect respondents' actual feelings about the construct of interest.

Given the aforementioned issues, the questions of interest in this analysis were:

1. Are any demographic groups or combinations of groups more or less likely to respond to all items in the summed scales and
2. Are there statistically significant relationships between respondents' willingness to complete all items in the different summed scales?

Causes and implications of missing data

Item non-response is the failure of the respondent to answer individual items in a survey, despite being eligible to respond. According to de Leeuw, Hox, and Hussman (2003), item non-response has three different forms:

1. Information that is not provided by the respondent for certain question(s);
2. Information that is provided but is not usable to the researchers; and
3. Information that was provided but is lost and cannot be retrieved

Moreover, item non-response may occur when the items are missing by design, items do not apply to the respondent, the respondent has difficulty with the cognitive task involved in answering the question, the respondent refuses to respond, the respondent does not know the answer, and the respondent provides a response that does not fit the given response categories (Hussman, 1999). Item non-response may also be influenced by the sensitive nature of questions

such as questions about income, sexual behaviors, and drug and alcohol use (Blair, Sudman, Bradburn, & Stocking, 1977; Chen, 2019).

Item non-response can have a significant impact on the quality of the research. It may reduce the usable sample size, reduce the statistical power in small samples, limit the generalizability of the results, and force the researcher to make decisions about whether to exclude responses from the analysis or to use a data replacement method (McNeeley, 2012). Excluding responses from the analysis reduces the usable size while replacing the missing items may result in an overestimation or underestimation of scale scores, affect the reliability of the measure, and increase the likelihood of finding statistically significant results when there are none.

Methodology

Sample

This study sampled approximately 800 law students and 10,000 attorneys and judges licensed in a southern state in the United States. Its purpose was to examine mental health and well-being among this population and compare the findings to other studies conducted on this population.

IRB approval and survey distribution

Before beginning this study, approval was obtained from the first author's institutional review board (IRB). Once the approval was granted, the research team began data collection. Prior to beginning this research, the research team had concerns about asking law students, attorneys, and judges about sensitive issues such as their mental health and substance use. McNeeley (2012) claims that while there is no clear definition of sensitive issues, it is generally agreed that drug or alcohol use is considered a sensitive topic. Tourangeau and Smith (1996) state that a survey question is "sensitive if it raises concerns about disapproval or other consequences (such as legal sanctions) for reporting truthfully or if the question itself is seen as an invasion of privacy" (p. 276). McNeeley (2012) further states that the characteristics of the interest group must be considered as the sensitivity of the topic can vary by the target population. Because the questions inquired about licensed law students', lawyers', and judges' use of drugs and alcohol and their experience of depression, anxiety, and stress, several steps were taken to protect respondents' confidentiality. First, the recruiting email for the survey was sent from a member of the research team to a member of the licensing board who then sent the recruiting email to the approximately 10,000 licensed attorneys and judges. A separate email was sent to contacts in each of the law schools who then sent the survey to students enrolled in their respective programs. No member of the research team had access to any contact information for the law students, attorneys, and judges to whom the survey was sent. Second, all data was stored on an encrypted server owned by the Arkansas Judges and Lawyers Assistance Program. As such, the data is legally protected from being subpoenaed or requested by outside sources (Rule 10, 2017). Third, the above information was made explicit in the informed consent that respondents agreed to prior to beginning the survey. Fourth, because the IRB was concerned that requiring respondents to respond to any questions after agreeing to the informed consent could be seen as coercive, respondents simply had to agree to the informed consent and could submit the survey without answering any questions in the survey in order to receive the ability to earn

one Continuing Legal Education (CLE) credit at no cost. Finally, categories in the demographic variables with less than 10 respondents were combined into a single category and the research team only reported categories with 10 or more respondents. These measures were intended to obtain more truthful answers to the sensitive questions asked in this survey.

Following approval by the IRB, the state board that oversees licensed attorneys, and the deans of the law schools, emails were sent on April 9, 2024, inviting law students, attorneys, and judges to participate in the study. Reminder emails were sent via the licensing board and law schools every two weeks until June 18, 2024. In addition, information about the survey was distributed via social media and at conferences and judicial meetings during the same dates.

Instruments

Given that the focus of this study was on the well-being of law students and licensed attorneys and judges in a southern state in the United States, demographic questions and four summed scales that examined mental health and well-being were included in the survey packet. Three of the four measures (AUDIT, DASS-21, and DAST-10) and the demographic questions were asked to facilitate comparison to larger studies that had been conducted on well-being among attorneys and judges by Krill and colleagues (Anker & Krill, 2021; Krill, Johnson, & Albert, 2016). The fourth measure (MSC-SF) was included to examine the use of self-compassion in attorneys and judges and to expand the body of research on well-being among this population. Each of the instruments used in this study had been used in several other studies, had strong internal consistency, and undergone extensive validation studies.

Depression Anxiety Stress Scales (DASS-21)

The 21-item DASS-21 (Henry & Crawford, 2005) is a shorter version of the 42-item DASS-42 developed by Loviband and Loviband (1995) to assess depression, anxiety, and stress, each of which is measured with a separate subscale and is analyzed as a separate scale. Scores for each subscale range from 0 - 21, with higher scores indicating more severe symptoms.

Alcohol Use Disorders Identification Test (AUDIT)

The 10-item instrument was developed by the World Health Organization to screen for unhealthy alcohol use (Saunders, Aasland, Babor, De La Fuente, & Grant, 1993). Scores range from 0 - 40 with higher scores reflecting more hazardous alcohol consumption.

Drug Abuse Screening Test (DAST-10)

The 10-item DAST-10 was developed to assess drug use, not including alcohol and tobacco use, in the past 12 months. Scores range from 0 - 10 with higher scores reflecting a severe level of use (Skinner, 1982).

Mindful Self-Compassion Scale-Short Form

The 12-item Mindful Self-Compassion Scale-Short Form (MSC-SF) (Raes, Pommier, Neff, & Van Gucht, 2011) is a shorter version of the 26-item Mindful Self-Compassion Scale developed by Neff (2003) to assess self-compassion. Scores on the MSC-SF range from 1 - 5 with higher scores reflecting more self-compassion.

Statistical Analysis

The analysis was conducted using SAS Studio and Microsoft Excel 365. Responses to the demographic questions, each of the three subscales of the DASS-21 were examined individually along with the AUDIT, DAST-10, and SCS-SF. As noted earlier, to protect the confidentiality of the respondents as much as possible, demographic characteristics with less than 10 respondents were combined with other categories in the same variable so that only categories with 10 or more responses were reported. In examining the scale items, if a single item on one of the summed scales was missing, the score could not be calculated and the scale was thus coded as 0. When all the items in the scale were answered, a total score could be calculated and the scale was coded as 1. Chi-square, correlation tests, missing in common, and logistic regression were analyzed to assess which groups were more or less likely to answer all items on each of the six measures of interest and to examine relationships between the demographic variables and completion of the summed scales.

Results

Responses to summed scales

Of the 1,547 surveys returned, 37 (2.4%) were returned with only the informed consent completed. This resulted in 1,510 (97.6%) surveys in which the respondents answered at least one question beyond the informed consent and are the focus of our analysis. 1,149 (76%) respondents answered all 53 items in the summed scales while 1,425 answered all the demographic questions (94.37%). This resulted in 1,121 (74%) who answered all seven of the demographic measures.

Table 1

Missing data in summed scales

	Answered		Missing	
	<i>n</i>	%	<i>n</i>	%
<i>N</i> = 1,510				
DASS-21 (full scale)	1,439	95.3	71	4.7
Depression	1,483	98.2	27	1.8
Anxiety	1,490	98.7	20	1.3
Stress	1,479	98	31	2.1
AUDIT	1,288	85.3	222	15
DAST-10	1,427	94.5	83	5.5
SCS-SF	1,424	94.3	86	5.7

Demographics

Personal Characteristics

There was very little missing data in respondents' personal characteristics with a range of 0.27% (gender) to 1.06% (whether respondents had children). The sample consisted of more women than men and more Caucasian/white than other races. The most common age group was 41 - 50. Approximately two-thirds of the respondents had children and almost the same percent were married.

Table 2

Personal demographic characteristics

Characteristic (<i>N</i> = 1,510)	<i>n</i>	%
<i>Gender</i>		
Female	757	50.13
Male	737	48.8
Other	12	0.8
Did not answer	4	0.27
<i>Race</i>		
Black or African American	55	3.64
Caucasian/white	1,347	89.2
Multiracial	69	4.57
Single race other than African American or Caucasian	25	1.66
Did not answer	14	0.93
<i>Age</i>		
30 or younger	196	12.98
31 - 40	331	21.92
41 - 50	353	23.37
51 - 60	348	23.05
61 - 70	200	13.25
71 or older	68	4.5
Did not answer	14	0.93
<i>Did you have children?</i>		
Yes	997	66.03
No	497	32.91
Did not answer	16	1.06
<i>Marital status</i>		
Married	1011	66.96

Partnered but not married	79	5.23
Divorced	162	10.73
Separated	13	0.86
Single	213	14.1
Widowed	25	1.66
Did not answer	7	0.46

Notes:

1. This category includes respondents who identified as non-binary and others who identified in other ways. Because the N for each of these groups is less than 10, they are subsumed into a single category.
2. This category includes all respondents who checked more than one racial category.
3. This category includes all respondents who identified with a single race other than White/Caucasian or Black/African-American. Because the N for each of these groups is less than 10, they are subsumed into a single category.

Professional characteristics

There was less missing data for the professional characteristics than for the personal characteristics with just 1 (0.06%) respondent neglecting to answer their role and 3 (0.2%) respondents neglecting to answer the number of years they had worked in law. Almost 25% of respondents had worked in the legal field for 11 - 20 years while Managing Partner was the most frequently cited role with 26.76% of respondents identifying this as their role.

Table 3

Professional demographic characteristics

Characteristic (N = 1,510)	n	%
<i>Years in field</i>		
0 - 10	593	39.27
11 - 20	369	24.43
21 - 30	303	20.07
31 - 40	180	11.92
41 or more	62	4.11
Did not answer	3	0.2
<i>Role</i>		
Judge	55	3.64
Senior Partner	215	14.23
Senior Associate	97	6.42
Managing Partner	404	26.76
Junior Partner	43	2.85
Junior Associate	102	6.76

Law school faculty/staff	15	0.99
Law school student	76	5.04
Staff attorney/clerk/paralegal	215	14.24
None of the above	287	19.01
Did not answer	1	0.06

Likelihood of not responding to all items by demographic characteristic

Gender

There were significant differences in responses to five of the six measures as women and men were significantly more likely than “other” to complete all items in the three subscales of the DASS-21, AUDIT, and DAST-10 (Chi Sq: DASS-21 depression = 0.0042; DASS-21 anxiety = 0.001; DASS-21 stress = 0.0070; AUDIT = 0.0032; DAST-10 < 0.001). A small but insignificant difference with gender was found in the respondents’ completion of the SCS-SF.

Race

There were significant differences in responses to three of the six measures (Chi Sq: DASS-21 depression < 0.001; DASS-21 anxiety = 0.0108; DAST = 0.0078). Those who did not answer the question were significantly less likely than other racial groups to complete all items on the DASS-21 depression subscale and DAST-10; they were significantly less likely than groups other than those with single race other that was not Caucasian/White or African-American/Black to complete all items on the DASS-21 anxiety subscale. Those identified with a single race that was not Caucasian/White or African-American/Black were significantly less likely than all other groups to answer all items on the DASS-21 anxiety subscale.

Age

There were significant differences in the completion of the scales by age. Those who identified as 71 or older were significantly less likely to answer all items in five of the six scales (Chi Sq: DASS-21 depression = 0.0034; DASS-21 anxiety = 0.0326; AUDIT = 0.0143; DAST < 0.001; SCS-SF < 0.001) but were not significantly less likely to complete the DASS-21 stress subscale.

Marital status

Respondents who did not answer this question were the least likely to answer the DASS-21 anxiety subscale followed by those who were separated (Chi Sq: DASS-21 anxiety = 0.0128).

Have children

Respondents who did not answer if they had children were less likely to answer the DASS-21 Depression measure (Chi Sq: DASS-21 depression = 0.005).

Years as an attorney

Respondents who did not answer the number of years they have worked as an attorney and those who have worked for 41 years or more as an attorney were less likely to complete five of the six measures (Chi Sq: DASS-21 depression < 0.001; Anxiety < 0.001; Stress = 0.0021; DAST < 0.001; SCS < 0.001;).

Role

There were significant differences in respondents' roles in the completion of four of six measures. Law school faculty/staff were the least likely to answer all items in the DASS-21 depression and SCS-SF. Judges were the least likely to answer all items in the AUDIT and DAST-10 (Chi Sq: DASS-21 depression = 0.0182; AUDIT = 0.0007; DAST-10 = 0.015; SCS-SF = 0.0003).

Missing data in common

An analysis of the patterns of missing data was conducted to assess patterns of missingness. This is often referred to as a missing in common qualitative analysis. We looked for three different patterns of missing in common data:

1. Relationships between demographic questions: When respondents failed to answer a demographic question, were there patterns in their failure to answer other demographic questions?
2. Relationships between summed scales: When respondents failed to answer a single item on one summed scale, were there patterns in their failure to answer items on other scales in the survey?
3. Relationships between demographic questions and summed scales: When respondents failed to answer items on demographic questions, were there patterns in their failure to answer items on the summed scales? Correspondingly, when respondents failed to answer items on summed scales, were there patterns in their failure to answer demographic questions?

Demographic variables

An examination of missing in common demographic characteristics (failure to answer demographic questions) found 10 different combinations in which respondents failed to answer more than one demographic question and usually consisted of one respondent per combination. However, the largest instance was in only three cases (respondents failed to answer questions regarding race, age, marital status, and whether they had children). This small sample size precluded further analysis and/or conclusions about the pattern.

Summed scales

As previously noted, the failure to answer a single item on a summed scale renders the response to that scale unusable or leads the researcher to use a data replacement method. Given

this, we assessed whether there were patterns in respondents' willingness to respond to all items in different combinations of the summed scales. We found that:

1. 45 respondents failed to complete at least one item in both the AUDIT and the DAST-10;
2. 17 respondents failed to answer at least one item on both the AUDIT and SCS-SF.
3. All other combinations of scales in which respondents failed to complete at least one item in multiple scales had 7 respondents or less (small sample size).

Following the above analysis, further analysis was conducted to assess whether there was a relationship between respondents' completion of all items on the different summed scales. Weak but significant correlations exist between respondents' completion of the three subscales (depression, anxiety, and stress) of the DASS-21: respondents who completed all items in one measure were more likely to complete all items in the other two measures. This is to be expected since they are subscales of a single scale and are presented as a single measure in the survey instrument.

Weak but significant correlations were also found in respondents' completion of the AUDIT and SCS-SF and the DAST-10 and SCS-SF. The strongest correlation found was a moderate but significant relationship between the AUDIT and DAST-10. This is an interesting finding since the AUDIT asks about alcohol use and the DAST-10 asks about drug use, both of which are considered sensitive topics. This finding may indicate that those who are willing to answer questions about alcohol use are also willing to answer questions about drug use and that the converse is also true.

It is interesting to note and makes sense that the strongest relationship (AUDIT & DAST-10) is the one that had the most respondents (45) with this missing in common pair.

Table 4
Correlation of scale completion (N = 1,510)

Variable	DASS-21 depression	DASS-21 anxiety	DASS-21 stress	AUDIT	DAST-10	SCS-SF
	α	α	α	α	α	α
DASS-21 depression	-	0.1155*	0.1214*	-0.0136	0.0113	0.0746
DASS-21 anxiety	0.1155*	-	0.024	0.0173	0.0228	-0.0034
DASS-21 stress	0.1214*	0.024	-	0.0337	-0.0144	0.045
AUDIT	-0.0136	0.0173	0.0337	-	0.3429*	0.1077*
DAST-10	0.1131	0.0228	-0.0144	0.3429*	-	0.1162*
SCS-SF	0.0746	-0.0034	0.045	0.1077*	0.1162*	-

* $p < .001$.

Demographic variables and summed scales

An analysis of the relationship between missing values in demographic variables and summed scales found no significant combinations. There were only two combinations where more than one respondent had the same combinations of missing demographic variables and incomplete summed scales (age and the DAST-10 ($N = 3$); having children and the AUDIT ($N = 2$)). These small samples prevented further analysis.

Regression model

Logistic regression was used to assess whether any of the seven demographic variables could be used to predict the likelihood that someone would answer all items on a given scale. The dependent variable was whether respondents answered all items in the scale (yes or no) while the independent variables were the seven demographic characteristics discussed above. The model was not significant for predicting the completion of the DASS-21 depression and stress subscales. However, gender, race, and age were significant for predicting the completion of the DASS-21 anxiety subscale (Table 4).

Logistic regression models with a concordance (c) value below 0.5 are generally considered no better than chance models. Models above 0.7 are considered good while models above 0.8 are strong, with increased odds that the model can correctly predict the dependent variable. The concordance of the DASS-21 anxiety subscale model was 0.811, which is a strong model.

Table 5
Logistic Regression Anxiety = Gender, Race, Age

Predictor	β	SE β	Wald's χ^2	df	p	Odds Ratio (point estimate)
Constant	6.4701	136.4	0.0023	1	0.9622	
Gender			8.7179	2	0.0128	
Female	0.4545	0.5616	0.655	1	0.4183	5.975
Male	0.8785	0.5568	2.4892	1	0.1146	9.13
Other	0	.	.		.	
Race			9.1068	3	0.0279	
Black or African American	9.8976	409.1	0.0006	1	0.9807	>999.999
Caucasian/white	-2.4888	136.4	0.0003	1	0.9854	6.848
Multiracial	-2.9961	136.4	0.0005	1	0.9825	4.123
SRSC	0	.	.		.	
Age			14.137	5	0.0148	
30 or younger	0.1307	0.6629	0.0388	1	0.8438	5.095
31-40	1.4275	0.8943	2.5482	1	0.1104	18.635
41-50	0.3226	0.575	0.3148	1	0.5748	6.173
51-60	-1.0061	0.4221	5.6813	1	0.0171	1.635
61-70	0.6229	0.8795	0.5016	1	0.4788	8.335
71 or older	0	.	.		.	
Test			χ^2	df	p	
Overall model evaluation						
Likelihood ratio test			25.688	10	0.0042	
Score test			33.8453	10	0.0002	
Wald Test			22.1632	10	0.0143	
Goodness of fit test						
Hosmer & Lemeshow			3.6529	8	0.887	

Note: c = 0.811

In addition, both gender and age were significant in predicting the completion of the AUDIT. However, the concordance was only 0.601, and thus not a very good model.

Table 6
Logistic Regression AUDIT= Gender and Age

Predictor	β	SE β	Wald's χ^2	df	p	Odds Ratio (point estimate)
Constant	0.7385	- 0.6844	1.1642	1	0.2806	
Gender			12.9866	2	0.0015	
Female	2.0176	0.6264	10.3752	1	0.0013	7.52
Male	1.705	0.6253	7.4347	1	0.0064	5.501
Other	0
Age			14.2857	5	0.0139	
30 or younger	1.0149	0.3754	7.309	1	0.0069	2.759
31-40	0.9683	0.3373	8.2433	1	0.0041	2.634
41-50	0.6834	0.3274	4.3575	1	0.0368	1.981
51-60	0.6665	0.3247	4.2119	1	0.0401	1.947
61-70	0.3041	0.3375	0.8116	1	0.3676	1.355
71 or older	0
Test			χ^2	df	p	Test
Overall model evaluation						
Likelihood ratio test			26.8567	7	0.0004	
Score test			30.2023	7	<.0001	
Wald Test			27.7764	7	0.0002	
Goodness of fit test						
Hosmer & Lemeshow			4.2548	8	0.8334	

Note: c = 0.601

The only demographic variable that was significant in predicting completion of the DAST-10 was gender. With a concordance of 0.594, this model is not much better than a random chance of predicting the dependent variable.

Table 7
Logistic Regression DAST-10 = Gender

Predictor	β	SE β	Wald's χ^2	df	p	Odds Ratio (point estimate)
Constant	0.9808	0.677	2.0989	1	0.2806	
GenderMFO			14.742	2	0.0006	
Female	2.3684	0.7069	11.2244	1	0.0008	10.68
Male	1.7082	0.6939	6.0596	1	0.0138	5.519
Other	0
Test			χ^2	df	p	
Overall model evaluation						
Likelihood ratio test			13.0124	2	0.0015	
Score test			18.2159	2	0.0001	

Wald Test	14.742	2	0.0006
Goodness of fit test			
Hosmer & Lemeshow	0	1	1

Note: c = 0.594

Age and whether one had children were significant in predicting the completion of the SCS-SF. With a concordance of 0.641, this model is not much better than a random chance of predicting the dependent variable.

Table 8

Logistic Regression SCS-SF= Age and Whether the respondent had children

Predictor	β	SE β	Wald's χ^2	df	p	Odds Ratio (point estimate)
Constant	1.9235	0.3661	27.5997	1	<.0001	
Age			21.8155	5	0.0006	
30 or younger	1.604	0.5311	9.1194	1	0.0025	4.973
31-40	1.7005	0.4803	12.535	1	0.0004	5.477
41-50	1.5521	0.4663	11.0798	1	0.0009	4.721
51-60	0.8872	0.4231	4.3962	1	0.036	2.428
61-70	0.4887	0.4377	1.2467	1	0.2642	1.63
71 or older	0	.	.		.	
Children			4.4241	1	0.0354	
No	0.57	0.2671	4.4241	1	0.0354	0.57
Yes	
Test			χ^2	df	p	
Overall model evaluation						
Likelihood ratio test			22.0263	6	0.0012	
Score test			24.4786	6	0.0004	
Wald Test			22.573	6	0.001	
Goodness of fit test						
Hosmer & Lemeshow			1.7441	6	0.9417	

Note: c = 0.641

Neither marital status, time in the field, or profession were significant in predicting completion of any of the scales.

The Chi-Sq results for the individual demographics, in combination with other demographics, show a pattern with many of the demographic variables being statistically significant in many models. However, only the DASS-21 anxiety subscale Logistic Regression model was strong enough to consider that the model was not random.

Discussion

Item non-response is a significant issue in research using summed scales. In this study, 24% of respondents failed to complete all items on all six measures, thus rendering the answers they did answer unusable. Given this impact, more research needs to be conducted into 1)

methods that can increase respondents' completion of all items in summed measures and 2) why respondents neglect to respond to one or more items in a summed scale.

Analysis of missing data, while rarely conducted to this extent, can be used to gain insight into the limitations of generalizing the study results and provide direction for future research. For example, the finding that judges were less likely to respond to all items on the DASS-21 depression subscale, AUDIT, and DAST-10 suggests that caution should be taken in generalizing the findings in the study to judges. Moreover, researchers may want to take additional precautions, and steps need to be taken when surveying judges about sensitive topics such as depression and alcohol and drug use. Judges have often faced significant stress (Maroney, Swenson, Bibelhausen, and Marc, 2023) but more research is needed to fully understand the stress they experience and why they may be reluctant to answer questions that assess their mental health and well-being. Furthermore, additional research should be conducted to examine whether the patterns of missing data are unique to this sample or comparable to other studies of judges and lawyers as well as to other professional groups.

Researchers should be transparent about whether respondents are required to answer all items in a survey, all items in a particular section of the survey, or in a particular measure within the survey, or are not required to answer any items in a survey packet that contains summed scales. Regrettably, the limited research on judges' and attorneys' well-being often fails to provide information about what items must be answered to submit the survey. A 2020 study of 1,034 judges found that 1,026 answered the AUDIT (Swenson, Bibelhausen, Buchanan, Shaheed, & Yetter, 2020). However, the authors do not state whether respondents were required to answer some or all of the items in the survey packet. A subsequent article referencing the Swenson study states that while most respondents completed the AUDIT, they were “permitted to skip any portions of the survey” (Maroney et al., 2023, p. 28). Additional information about whether respondents were required to answer all items in the survey would have been helpful and given additional context for their findings.

The IRB's refusal to let the researchers require that respondents answer questions beyond the informed consent may have prevented a problem known as “straight-lining,” whereby respondents give similar or identical answers to all questions in order to finish the survey (Kim, Dykema, Stevenson, Black, & Moberg, 2019; Mirzaei, Carter, Patanwala, & Schneider, 2002) and, in this case, in order to receive the CLE. While the motivation for answering all items in the different summed scales cannot be known, one hopes and can reasonably assume that the respondents in this study answered the items truthfully, particularly given the precautions taken to protect their confidentiality and privacy.

At the same time, researchers and IRBs should be encouraged to require and perhaps incentivize responses to all demographic questions. Those who failed to respond to demographic items were less likely to answer all questions on many of the summed scales, thus suggesting a relationship between the two.

The results of this study also point to a need for more research on how the order of the measures in a survey affects the completion rate for each measure. The literature on the relationship between the location of scales in surveys and their completion rate is mixed with

some authors suggesting that measures assessing more difficult items be placed later in the survey while others suggest that they should be placed early in the scale when respondents are less prone to respondent fatigue. The DASS-21 was the first scale in the survey; the full 21-item scale and the three subscales had the highest completion rate of any of the measures. Although the 12-item SCS-SF was the last measure in the survey, it did not have the lowest completion rate. Although it has been well documented that respondent fatigue in surveys may lead to item non-response (Lavakras, 2008), this study does not support that concern. Researchers may want to consider putting scales that inquire about sensitive issues at the beginning of the scale. The 10-item AUDIT asks about alcohol use and was placed in the middle of the survey with 21 items asked before it and 22 items asked after it and had the lowest completion rate. Future research should examine how the order of the scales in the survey packet affects their completion rate. For example, studies could be conducted in which the same instruments are included in a survey packet but the order of the instruments varies. This would enable researchers to gain insight into the relationship between the location of the scale in the survey and the completion rate for that scale.

It is further important to examine relationships between measures and patterns of missing data among measures that are both theoretically and statistically related. The AUDIT and DAST-10 both ask about sensitive topics (alcohol and drug use) and there was a moderate relationship between failing to complete these two measures. Weak but significant relationships were found with the AUDIT and SCS-SF. Future research should explore both patterns of missing data among the measures and the nature of these relationships.

Another area of future research should be conducted on law students, attorneys, and judges' reluctance to answer questions about their age, race, and whether they have children. Although respondents answered 94.37% of demographic questions, there was a consistent non-response to these questions when they answered the six summed scales. One explanation is that this population is concerned that these characteristics could be used to identify them and tie them to their responses, despite attempts to limit the identification of respondents in the study.

In sum, missing data in the form of item non-response impacts the results of surveys using summed scales that require all questions in the measure to be answered in order to compute a score on the measure. A clear pattern was found with respondents who failed to answer demographics and also failed to answer all items in the summed scales. Analysis of missing data in surveys using summed scales should examine both missing data in demographic questions and in the summed measures to identify biases in the data caused by patterns of missing data among respondents. This will lead to more confidence in the results and increased awareness about the degree to which the findings may be limited to the specific population who responded or generalizable to broader populations.

References

- Anker, J., & Krill, P. R. (2021). Stress, drink, leave: An examination of gender-specific risk factors for mental health problems and attrition among licensed attorneys. *PLoS one*, 16(5), e0250563. <https://doi.org/10.1371/journal.pone.0250563>
- Blair, E., Sudman, S., Bradburn, N. M., & Stocking, C. (1977). How to ask questions about drinking and sex: Response effects in measuring consumer behavior. *Journal of Marketing Research*, 14, 316–321.
- Chen, S., and Haziza, D. (2019) Recent Developments in Dealing with Item Non-response in Surveys: A Critical Review. *International Statistical Review*, 87, S192–S218. <https://doi.org/10.1111/insr.12305>.
- de Leeuw, E., Hox, J., & Huisman, M. (2003). Prevention and treatment of item nonresponse. *Journal of Official Statistics*, 19, 153-176.
- Henry, J. D., & Crawford, J. R. (2005). The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*, 44(2), 227–239. <https://doi.org/10.1348/014466505X29657>
- Kim, Y., Dykema, J., Stevenson, J., Black, P., & Moberg, D. P. (2019). Straightlining: Overview of Measurement, Comparison of Indicators, and Effects in Mail–Web Mixed-Mode Surveys. *Social Science Computer Review*, 37(2), 214-233. <https://doi.org/10.1177/0894439317752406>
- Krill, P. R., Johnson, R., & Albert, L. (2016). The Prevalence of Substance Use and Other Mental Health Concerns Among American Attorneys. *Journal of Addiction Medicine*, 10(1), 46–52. <https://doi.org/10.1097/ADM.0000000000000182>
- Lavrakas, P. J. (2008). Respondent fatigue. In *Encyclopedia of Survey Research Methods* (pp. 743-743). SAGE Publications, Inc., <https://doi.org/10.4135/9781412963947>
- Lovibond, S.H., & Lovibond, P.F. (1995). *Manual for the Depression Anxiety Stress Scales*. (2nd. Ed.) Sydney: Psychology Foundation.
- Maroney, T., Swenson, D. X., Bibelhausen, J., & Marc, D. (2023). How Are You Holding Up? The State of Judges' Well-Being: A Report on the 2019 National Judicial Stress and Resiliency Survey. *Judicature*, 107, 22.
- McNeeley, S. (2012). Sensitive Issues in Surveys: Reducing Refusals While Increasing Reliability and Quality of Responses to Sensitive Survey Items. In L. Gideon (Ed.), *Handbook of Survey Methodology for the Social Sciences* (pp. 377-396). New York, NY: Springer New York.

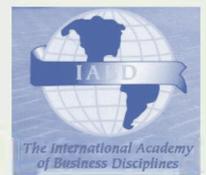
- Mirzaei, A., Carter, S. R., Patanwala, A. E., & Schneider, C. R. (2022). Missing data in surveys: Key concepts, approaches, and applications. *Research in Social and Administrative Pharmacy, 18*(2), 2308-2316. <https://doi.org/10.1016/j.sapharm.2021.03.009>
- Neff, K. D. (2003). Development and validation of a scale to measure self-compassion. *Self and Identity, 2*, 223-250
- Raes, F., Pommier, E., Neff, K. D., & Van Gucht, D. (2011). Construction and factorial validation of a short form of the Self-Compassion Scale. *Clinical Psychology & Psychotherapy, 18*, 250-255.
- Rule 10. (2017). *Rules of the Arkansas Judges and Lawyer Assistance Program (JLAP)*. https://opinions.arcourts.gov/ark/cr/en/item/1875/index.do#!fragment/zoupio-_Toc97817946/BQCwhgziBcwMYgK4DsDWszIQewE4BUBTADwBdoAvbRABwEtsBaAfX2zgE4B2ADgEYuHACwA2AJQAaZNIKEIARUSFcAT2gBydRIiEwuBIuVrN23fpABIPKQBCagEoBRADKOAagEEAcgGFHE0jAAI2hSdjExIA
- Saunders, J.B., Aasland, O.G., Babor, T.F., De La Fuente, J.R. & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption-II. *Addiction, 88*, 791-804. <https://doi.org/10.1111/j.1360-0443.1993.tb02093.x>
- Skinner H. A. (1982). The drug abuse screening test. *Addictive behaviors, 7*(4), 363–371. [https://doi.org/10.1016/0306-4603\(82\)90005-3](https://doi.org/10.1016/0306-4603(82)90005-3)
- Spector, P. E. (1992). *Summated rating scale construction: An introduction*. Sage Publications, Inc. <https://doi.org/10.4135/9781412986038>
- Swenson, D., Bibelhausen, J., Buchanan, B., Shaheed, D., & Yetter, K. (2020). Stress and resiliency in the US Judiciary. *2020 Journal of the Professional Lawyer, 1*.
- Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection, mode, question format, and question context. *Public Opinion Quarterly, 60*, 275–304.

QRBD

QUARTERLY REVIEW OF BUSINESS DISCIPLINES

February 2025

Volume 11
Number 3/4



A JOURNAL OF INTERNATIONAL ACADEMY OF BUSINESS DISCIPLINES
SPONSORED BY UNIVERSITY OF NORTH FLORIDA
ISSN 2334-0169 (print)
ISSN 2329-5163 (online)