

SESUG Paper 105-2019

## Examining Imputation effect by Using FACTOR and MI Procedures in SAS ® for a Bullying Scale for LGBTQ Youth in SC

Abbas S. Tavakoli, DrPH, MPH, ME<sup>1</sup> Laura C. Hein, PhD<sup>1</sup>, Mary F. Cox, PhD<sup>1</sup>,  
<sup>1</sup>University of South Carolina, College of Nursing

### ABSTRACT

Missing data presents a challenge to researchers. This study used LGBTQ youth data from SC to test the effect of an imputation using FACTOR and MI procedures in SAS. There are many reasons for missing data. Factor analyses were run to develop a Bullying Scale for LGBTQ youth. These runs included no imputation, single imputation, and multiple imputation (1000 times) for missing data. Two factors emerged - hearing bullying and experiencing bullying. The inter-factor correlation was .46 for hearing bullying and experiencing bullying. Similar results for factor extraction for no imputation, single imputation, and multiple imputation were found. Testing revealed all reliability coefficients exceeded .80 with no imputation and with imputation using the SAS ®<sup>6</sup> PROC FACTOR and STANDARD, and MI procedures for data analysis. All data analyses performed using SAS/STAT® version 9.4<sup>6</sup>.

**Keywords:** SAS, FACTOR, MI, Bullying, LGBTQ

**University of South Carolina, College of Nursing.**

### INTRODUCTION

Factor analysis is a statistical method to find a set of unobserved variables or factors from a larger set of observed variables. Factor analysis may be Exploratory or Confirmatory. Exploratory factor analysis examines the relationships among sets of observed variables without a prior fixed number of factors. The SAS ®<sup>6</sup> PROC FACTOR) procedure supports exploratory analysis. Observations with missing data for one or more variables called incomplete cases. Missing data presents a challenge to investigators, although it is common in most studies. There are many reasons for missing data such as poor research design, poorly structured questions, and attrition in longitudinal studies. Most SAS/STAT ® Statistical procedures eliminate the data point from data analysis if there is missing value.

### PURPOSE

The purpose of this paper is using FACTOR and MI procedures in SAS® 9.4<sup>6</sup> to examine imputation effect for a bullying scale for LGBTQ youth in South Carolina (SC).

### BACKGROUND

Students who identify themselves as lesbian, gay, bisexual, transgender, and/or questioning (LGBTQ) are at greater risk for bullying than those students who consider themselves heterosexual.<sup>4</sup> The Gay, Lesbian, & Straight Education Network (GLSEN) reports that the majority of LGBTQ students feel unsafe in their school environment.<sup>5</sup> Furthermore, four out of five students report being verbally harassed, and almost 40% reported being physically harassed due to their sexual orientation.<sup>3</sup> Despite these alarming statistics, researchers have found that teachers are uncomfortable intervening with bullying related to sexual orientation, as compared with other types of bullying.<sup>2</sup> The Bullying Scale developed from our study includes 11 items (see table 1).

**Table 1. Bullying Items, LGBTQ Youth study**

Items	
<b>b1</b>	hearing gay used in a negative way
<b>b2</b>	hearing other anti-LGBT remarks
<b>b3</b>	hearing sexist remarks

Items	
<b>b4</b>	hearing racist remarks
<b>b5</b>	hearing comments about not acting masculine enough
<b>b6</b>	hearing comments about not acting feminine enough
<b>b7</b>	hearing biased comments from teachers or other school staff
<b>b8</b>	experiencing fear of being physically harassed (e.g. verbal threats)
<b>b9</b>	experiencing actual physical harassment (e.g. pushed or shoved)
<b>b10</b>	experiencing actual physical violence (e.g. punched, kicked, or injured with a weapon)
<b>b11</b>	experiencing some form of electronic harassment (e.g. cyber bullying)

## METHODS

We examined data from a 2013 survey of LGBTQ South Carolinians, including self-identified transgender or genderqueer people. Single and multiple imputations used to replace the missing value for 11 items. Means of each item compared with and without imputation. Factor analyses run to develop a Bullying Scale for LGBTQ youth. These initial computations included no imputation, single imputation, and multiple imputation (1000 times) for missing data. Factor analysis used squared multiple correlations as prior communality estimates. The maximum Likelihood (ML) method used to extract factors. This followed by the promax (oblique) rotation. In interpreting the rotated factor pattern, an item was said to load on a given factor if the factor loading was .35 or greater for that factor, and was less than .35 for the other. Proc Means, Freq, Corr, Standard, MI, and Factor used to analyze data. All data analyses performed using SAS/STAT® version 9.4<sup>6</sup>.

## RESULTS

Table 2 contains the N and percentage of missing values for each item in the Bullying Scale. The results indicated there were missing values ranging from 6.33 percent to 7.45 percent among the bullying items.

**Table 2: N and percentage of missing value of bullying scales, LGBTQ Youth study (n=1154)**

Items	N	%
hearing gay used in a negative way	73	6.33
hearing other anti-LGBT remarks	75	6.50
hearing sexist remarks	75	6.50
hearing racist remarks	81	7.02
hearing comments about not acting masculine enough	83	7.19
hearing comments about not acting feminine enough	84	7.28
hearing biased comments from teachers or other school staff	82	7.11
experiencing fear of being physically harassed (e.g. verbal threats)	86	7.45
experiencing actual physical harassment (e.g. pushed or shoved)	82	7.11
experiencing actual physical violence (e.g. punched, kicked, or injured with a weapon)	80	6.93
experiencing some form of electronic harassment (e.g. cyber bullying)	81	7.02

Table 3 presents the mean and standard deviation of the bullying items without imputation, single and multiple imputation. The results indicate the mean for items are similar without imputation and with imputation.

**Table 3: N, Mean, and STD of Bullying Scales, LGBTQ Youth study (n=1154)**

Items	No Imputation			Imputation (Once) (n=1154)		Imputation (1000) (n=1154)	
	N	Mean	STD	Mean	STD	Mean	STD
b1	1081	3.98	1.12	3.98	1.08	3.98	1.08
b2	1079	3.83	1.13	3.83	1.09	3.83	1.09
b3	1079	3.80	1.09	3.80	1.05	3.80	1.05
b4	1073	3.72	1.11	3.72	1.08	3.72	1.08
b5	1071	3.18	1.33	3.19	1.28	3.19	1.28
b6	1070	2.68	1.40	2.68	1.35	2.68	1.35
b7	1072	2.51	1.28	2.52	1.24	2.52	1.24
b8	1068	2.53	1.38	2.53	1.33	2.53	1.33
b9	1072	2.00	1.23	2.00	1.19	2.00	1.19
b10	1074	1.61	1.08	1.61	1.04	1.61	1.04
b11	1073	1.56	1.06	1.56	1.02	1.56	1.02

The 11-item Bullying Scale was factor analyzed. Responses from the survey of LGBTQ South Carolinians subjected to exploratory factor analysis using squared multiple correlations as prior communality estimates. The maximum Likelihood (ML) method used to extract the factors. This followed by the promax (oblique) rotation. A scree test, Eigen values, and the proportion variance explained with each factor suggested two meaningful factors. In interpreting the rotated factor pattern, an item was said to load on a given factor if the factor loading was .35 or greater for that factor, and was less than .35 for the other. Using these criteria, two factors emerged Hearing (items b1-b7) and Experiencing (items b8-b11) (Table 4). Inter-factor correlation was .46 for Hearing and Experiencing. Table 4 depicts similar results for factor extraction for no imputation, single imputation, and multiple imputation.

**Table 4: Factor Loading (Standardized Regression Coefficients) of Bullying Scales, LGBTQ Youth study (n=1154)**

Items	No Imputation		Imputation (Once)		Imputation (1000)	
	Factor1	Factor2	Factor1	Factor2	Factor1	Factor2
b3	86	-3	86	-3	86	-3
b2	85	2	84	1	84	1
b1	81	-3	81	-4	81	-4
b4	75	-1	73	0	73	0
b5	54	18	53	18	53	18
b6	50	7	50	7	50	7
b7	43	34	42	35	42	35
b9	-2	91	-2	91	-2	91
b10	-8	87	-8	87	-8	87
b8	19	70	19	69	19	69
b11	6	51	7	50	7	50

Table 5 shows Coefficient Alpha Reliability without imputation and with imputation. Scale reliability assessed by calculating coefficient alpha. The result revealed all reliability coefficients exceeded .80 for no imputation, single imputation, and multiple imputation.

**Table 5: Alpha coefficient reliability of bullying scales, LGBTQ Youth study (n=1154)**

Items	No Imputation	Imputation (Once)	Imputation (1000)
<b>Total Bullying</b>	0.88	0.88	0.88
<b>Hearing (7 item)</b>	0.87	0.87	0.87
<b>Experiencing (4 item)</b>	0.85	0.85	0.85

## CONCLUSION

Our study shows that there were similar results when imputation used for replacing missing values compared to no imputation. Two factors emerged into the Hearing bullying and Experiencing Bullying. The inter-factor correlation was .46 for Hearing and Experiencing. Results showed similar results for factor extraction for no imputation, single imputation, and multiple imputation. The result revealed all reliability coefficients exceeded .80 with no imputation

and with imputation. There are several methods for replacing missing values when analyzing data. The single imputation method can be biased if the proportion of missing values is large (greater than 5%).<sup>1</sup> Multiple imputation (MI) is an alternative method to replace missing values. Researchers should consider using the imputation method to help resolve problems caused by missing data. As demonstrated in our study, the imputation method is equally reliable.

## References

1. Allison, P.D. (2000). Multiple Imputation for Missing Data: A Cautionary Tale. *Sociological Methods and Research*, 28,301-309.
2. Bradshaw, C. P., Waasdorp, T. E., O'Brennan, L. M., & Gulemetova, M. (2013). Teachers' and Education Support Professionals' Perspectives on Bullying and Prevention: Findings From a National Education Association Study. *School Psychology Review*, 42(3), 280–297.
3. Hein, L. C., & Scharer, K. M. (2013). Who cares if it is a hate crime? Lesbian, gay, bisexual, and transgender hate crimes—Mental health implications and interventions. *Perspectives In Psychiatric Care*, 49(2), 84-93. doi:10.1111/j.1744-6163.2012.00354.x
4. Robinson, J. P., & Espelage, D. L. (2012). Bullying explains only part of LGBTQ–heterosexual risk disparities: Implications for policy and practice. *Educational Researcher*, 41(8), 309-319. doi:10.3102/0013189X12457023
5. Perez, E. R., Schanding, G. J., & Dao, T. K. (2013). Educators' perceptions in addressing bullying of LGBTQ/gender nonconforming youth. *Journal of School Violence*, 12(1), 64-79. doi:10.1080/15388220.2012.731663
6. SAS Institute Incorporated. (2013). SAS for Windows 9.4. Cary, NC: SAS Institute Inc.

## Contact Information

Your comments and questions are valued and encouraged. Please contact the authors at:

Abbas S. Tavakoli, DrPH, MPH, ME  
 Clinical Associate Professor  
 College of Nursing  
 University of South Carolina  
 1601 Greene Street  
 Columbia, SC 29208-4001  
 Fax: (803) 777-5561

## ACKNOWLEDGEMENT

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies

## Appendix

### SAS Syntax

**\*\* Frequency and means for Items\*\*:**

```
Ods rtf; ods listing close; proc freq data=three;
  tables b1-b11 /missing;
  title ' frequency tables ';
  title2 'SCLGBT project'; run;
proc means data=three n mean std maxdec=2;
  var b1-b11;
  title ' means '; title2 'SCLGBT project'; run;
ods rtf close; ods listing; quit; run;
```

```

**** Single imputation ***;

ods rtf; ods listing close; title;

Proc STANDARD DATA=three OUT=stnd REPLACE noPRINT ;
  VAR b1-b11; RUN;
  title ' standard means/replacement of missing ';    title2 'SCLGBT project'; run;

proc means data=stnd  n mean std noprint maxdec=3;
  var b1-b11 ;
  title ' replacement of missing by means ';    title2 'SCLGBT project';
output out=meantest1 (drop=_type__freq_); run;
proc transpose data=meantest1
  out=meantest1 (rename=(col1=n col2=min col3=max col4=mean col5=std));
format mean 6.3 std 6.3; run;
proc print data=meantest1 noobs ;
  var _name_ _label_ n mean std ; run;
ods rtf close; ods listing; quit; run;

*** multiple imputation ****;

proc mi data=three seed=37851 out=outmi nimpute=1000 noprint;
var b1-b11;
title ' Multiple imputation ';    title2 'SCLGBT project'; run;
proc univariate data=outmi noprint;
var b1-b11;
output out=outuni mean =mb1-mb11;
by _imputation_; run;
data all;
  if _N_ = 1 then set outuni(keep = mb1-mb11); set three; run;
data final;
  set all;
array items b1-b11;
  array itemsb mb1-mb11;
  do over items;
    if items =. then items=itemsb;
  end; run;
*** Factor and reliability without imputation **;

Ods rtf; ods listing close; %macro fact (q,n,t);
proc factor data=three method=prin priors=smc scree rotate=promax reorder msa
  flag=.35 nfact=&n ; var &q ;
  title ' factor analysis ' &t;    title2 'SCLGBT project';
%mend fact;
%fact (b1-b11,2, Bullying scale); run;

%macro corr (q);
proc corr nocorr alpha nomiss data=three;
  var &q;
  title ' Reliability coeffcient';    title2 'SCLGBT project';
%mend corr;
%corr ( b1-b11);
%corr ( b1-b7 );
%corr ( b8-b11);
run;
ods rtf close; ods listing; quit; run;

```