Nonresponse Issue in Noncognitive Measures: Validity Approach Using Explanatory Item Response Modeling

Jiaying Xiao & Okan Bulut
Centre for Research in Applied Measurement and Evaluation
University of Alberta, Canada

Associated with the
Minnesota Youth Development Research Group
www.mnydrg.com

April 2019

Paper presented at the annual meeting of the
National Council on Measurement in Education, Toronto, Canada.

Citation:

Xiao, J., & Bulut, O. (2019, April). *Nonresponse issues in noncognitive measures: Validity approach using explanatory item response modeling.* Paper presented at the annual meeting of the National Council on Measurement in Education, Toronto, Canada.

Nonresponse Issue in Noncognitive Measures: Validity Approach Using Explanatory Item Response Modeling

Abstract

Missing data always exist in empirical studies, which could be harmful to validity results. Using explanatory item response modeling, we examine validation concerns due to nonresponse issue in a noncognitive measure of bullying. A total of 10,681 U.S. students in grades 8, 9 and 11 reported their experiences of bullying and skipped at least one item in the bullying scale. Results showed that female students were more willing to answer bullying items. Students in higher grade levels were more likely to answer physical and verbal bullying items than indirect bullying items. These item-level and respondent-level factors lead to different nonresponse patterns and thus potentially influence score interpretation. Practical implications are also discussed.

Background

Realizing that cognitive tests may not properly reflect student's school performance, researchers have shifted their attention to the effects of noncognitive factors (e.g., social behaviors, attitudes) (Farrington et al., 2012). Then, the next step was to develop sound measures to understand such constructs. Validity evidence is critical to interpret scores when using noncognitive measures (Kane, 2013). Based on the *Testing Standards* (AERA, 2014), there are several sources of validity evidence, including content, internal structure, the relationships with other variables, and response processes. 'Evidence based on response processes generally comes from analyses of individual responses" (p. 15). Previous studies only focus on the interpretation of answered items, however, nonresponse issue is inevitable in the data collection and the existence of missing data may be harmful to the validity results (McKnight et al., 2007). For example, when researchers use complete case analysis to delete respondents who omit some items, it is possible to get biased results and reduce the study power (Schafer, 1999). Besides, if missing data mechanism is missing at random (MAR) (missing data could be explained by other observable variables) or not missing at random (NMAR) (missing values on a variable X are dependent on the values of X itself, or some other variables that are related to it, but which have not been measured) (Graham, 2012), it means that nonresponse items themselves have valuable information and deserves to be explored more. Therefore, it is necessary to investigate nonresponse patterns in noncognitive measures.

In this study, we used a measure of bullying as an example to examine item-level and respondent-level factors contributing to missing responses via an explanatory item response model (EIRM). Researchers often categorized bullying as indirect (i.e., relational), physical, and verbal (Rivers & Smith, 1994). Previous studies indicated that nonresponse behaviours in bullying-related items might differ depending on bullying types measured by the items (Birkinshaw & Eslea, 1998). Missingness could also be accounted for by respondent-related predictors (Wayman, 2003). Some researchers have shown significant gender differences in bullying. For example, more males reported more bullying involvements compared with females (Seals & Young, 2003), and males were more willing to admit to bullying others (Whitney & Smith, 1993). One UK survey found the frequency of bullying others was steady through the junior/middle school year and fell during the first two years at secondary school, only to rise again in the third, fourth and fifth year (13-15 years) before falling again in the sixth form to its lowest (but not zero) levels (Whitney & Smith, 1993). Due to their effects on bullying behaviors, this study used gender and grade levels to as predictors of missingness. Furthermore, previous research suggests that physical bullying decreased whereas indirect bullying increased in higher grades (Crick, Grotpeter, & Bigbee, 2002), implying interaction effects between bullying type and grade. Therefore, the purpose of the study is to examine the effects of bullying type, gender, grade-level, and their interactions on students' nonresponse behaviours in a bullying scale.

In this study, we aim to examine whether nonresponse patterns for the bullying scale are consistent across items, gender, and grade levels. We hypothesize there are significant differences in nonresponse patterns under different bullying types, gender, and grade levels conditions.

Methods

Data

The data came from the 2016 Minnesota Student Survey (MSS; MN Department of Education, 2017). The MSS is designed by an interagency team from the MN Departments of Education, Health, Human Services, and Public Safety. The survey is administered every three years to students in grades 5, 8, 9, and 11. Approximately 85% of the MN public school districts participated in the 2016 MSS administration. The sample closely matched the state population in terms of race and ethnicity (67% White only, 9% Latino, 5% American Indian, 5% Black non-Somali, 2% Somali, 4% Asian non-Hmong, 3% Hmong), as well as participation in special education (10%) and free and reduced-price lunch (28%). Because some bullying items were not included for students in grade 5, this study only focused on students in grades 8, 9 and 11 who omitted at least one item in the bullying scale. Also, this study deleted 124 participants who did not respond to the gender question. So, the final sample size was 10,681. Table 1 shows the basic demographic information.

Table 1

Demographic Information of the 2016 Minnesota Student Survey (MSS) Sample

Variable	*N*	%
Gender		
Male	6,724	62.95%
Female	3,957	37.05%
Grade		
8	3,296	30.86%
9	4,157	38.92%
11	3,228	30.22%
Total	10,681	100%

The bullying scale includes eight items asking students to report the frequency of each form of bullying (see Rodriguez, 2017, for psychometric properties of the scale). Based on the item content, three items were recorded as physical bullying, two items were recorded as verbal bullying and other two were recorded as indirect bullying. Table 2 shows the missing proportion, item content, and bullying type for each item. The dependent variable in the current study was binary responses where “1” refers to a bullying item for which students chose one of the available response options, and “0” refers to an omitted item for which students did not choose any response option. For example, one of the questions in the bullying scale is “During the last 30 days, how many times at school have you made sexual jokes, comments or gestures towards someone else?” and the response options are “Never”, “Once or twice”, “About once a week”, “Several times a week”, and “Every day”. If a student chose one of the response options available, then the student’s response was recoded as “1”. If, however, the student omitted the question, then the missing response was recoded as “0”.

Table 2
Specifications of the Bullying Items

Item Content	Bullying Type	Missing Proportion
1-Pushed, shoved, slapped, hit or kicked someone when you weren't kidding around	Physical	21.59%
2- Threatened to beat someone up	Verbal	21.98%
3- Spread mean rumors or lies about someone else	Indirect	22.23%
4- Made sexual jokes, comments or gestures towards someone else	Indirect	22.56%
5- Excluded someone from friends, other students or activities	Indirect	22.83%
6- Was a bully or threaten other people	Verbal	67.83%
7- Started fights with other people	Physical	68.25%
8- Hit or beat up another person	Physical	84.40%

Analytical Model

Under the EIRM framework, traditional item response model could be considered as generalized linear mixed models (GLMM) where item covariates, person covariates, and item-by-person covariates could be included (De Boeck & Wilson, 2004). Using the binary dependent variable described above, respondent j’s $(j = 1, 2, \ldots, J)$ probability of choosing one of the response options available for question i follows a Bernoulli distribution with mean, π_{ij}. To place π_{ij} onto a continuous scale between $-\infty$ and $+\infty$, a logit link function can be applied as

$$\eta_{ij} = \log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right). \tag{1}$$

In order to estimate the probability of choosing one of the response options available for a set of survey questions, the following GLMM[1] can be used:

$$\eta_{ij} = \beta_i X_{ij} + \theta_j, \quad (2)$$

where β_i refers to the answerability of question i $(i = 1, 2, \ldots, K)$ where higher values indicate a higher likelihood of obtaining a valid response (rather than the item being omitted), X_{ij} is respondent j's binary response to the question i, (i.e., 1=answered the question, 0=omitted the question), and θ_j refers to respondent j's level of responsiveness to answering the survey questions. As the level of θ_j increases, respondent j becomes more likely to choose a valid response option rather than omitting the question.

The GLMM in Equation 2 provides information about the answerability of the questions and respondents' level of responsiveness, but does not explain what factors contributes to respondents' decision of answering or omitting the questions. To explain the probability of valid response based on a set of predictors related to questions (e.g., content characteristics of the questions) and respondents (e.g., gender), the GLMM in Equation 2 can be altered as follows:

$$\eta_{ij} = \sum_{m=1}^{M} \beta_m X_{im} + \sum_{p}^{P} \vartheta_p Z_{jp} + \theta_j, \quad (3)$$

where X_{im} is the value of question i on a question-related variable m $(m = 1, 2, \ldots, M)$, β_m is the fixed effect of the variable m, Z_{jp} is the value of respondent j on a respondent-related variable p $(p = 1, 2, \ldots, P)$, ϑ_p is the fixed effect of the variable p, and θ_j again refers to respondent j's level of responsiveness to answering the survey questions.

To describe the item-level (bullying types) and respondent-level (gender and grade-level) factors contributing to the students' nonresponse behaviours in the bullying scale, four EIRMs were specified and analyzed. In each model, the dependent variable was binary responses and the respondents (i.e., students) were the random effects. The items were nested within the students in a multilevel structure. The first model (M1) was based on Equation 2, where the questions were the only predictors of the dependent variable (i.e., binary responses). The remaining three models (i.e., M2 to M4) were designed to evaluate the effects of bullying type, gender, and grade levels based on Equation 3. M_2 examined the answerability of the items with regard to the bullying type addressed in the items. M_3 examined whether gender and grade levels significantly affected the answerability. Lastly, M_4 examined the interaction effects between bullying types and grade levels. Table 3 provides a summary of the models. The models were estimated with *lme4* (Bates et al., 2015) in R (R Core Team, 2018).

[1] For the sake brevity, the model does not involve a constant intercept. That is, the estimated regression lines are assumed to run through the origin.

Results

Table 4 contains the fit indices of the four models. AIC and BIC results showed that M_3 fits better than M1. Compared with M2, M4 fits better. The fixed effects for the four models are shown in Table 5 and all of them are significant.

M1 results showed that students were least willing to answer item 8 (-3.10) which describes a physical bullying behaviour (hit or beat up another person), and they were most likely to answer item 1 (2.35) which also mentions a physical bullying behaviour (pushed, shoved, slapped, hit or kicked someone when you weren't kidding around).

Table 3

Model Formulations of the Four EIRMS Used to Predict Nonresponse Behaviours

Model	Fixed Effects	Random Effects
M_1	$\beta_i X_{ij}$	$\theta_j \sim N(0, \sigma_\theta^2)$
M_2	$\beta_1 Physical_{i1} + \beta_2 Verbal_{i2}$	$\theta_j \sim N(0, \sigma_\theta^2)$
M_3	$\beta_i X_{ij} + \vartheta_1 Gender_{j1} + \vartheta_2 Grade_{j2}$	$\theta_j \sim N(0, \sigma_\theta^2)$
M_4	$\beta_1 Physical_{i1} + \beta_1 Verbal_{i1} + \vartheta_1 Gender_{j1} + \vartheta_2 Grade_{j2} + \delta_1 Physical_{i1} * Grade_{j2} + \delta_2 Verbal_{i1} * Grade_{j2}$	$\theta_j \sim N(0, \sigma_\theta^2)$

Note. In M_2 and M_4, "indirect bullying" was the reference category for the type of bullying. In M_3 and M_4, male was the reference category for gender.

In M2, "indirect bullying" was the reference category for the type of bullying. Based on M2 results, both physical bullying (-2.37) and verbal bullying (-1.60) items were more likely to be omitted compared with indirect bullying type items. M3 chose male as the reference category for gender. Results showed that compared with females (0.31), males were more likely to omit bullying items. In addition, students in higher grades preferred to omit bullying items.

In M4, "indirect bullying" was the reference category for the type of bullying and male was the reference category for gender. We could also get the same conclusions about bullying types and gender as the ones in M2 and M3. In addition, the significant interaction indicates that for students from different grade levels, their nonresponse behaviors for various bullying items were different. In the MSS data, students in higher grades were more likely to answer physical and verbal bullying items (0.08) than students in lower grades were.

Table 4

Model Fit Indices for the Explanatory Item Response Models

Model	AIC	BIC	Log-likelihood	Deviance
M_1	69434	69518.7	-34708.3	69416.5
M_2	92978	93015.4	-46485	92970
M_3	69385	69488.4	-34681.8	69363.5
M_4	92914	92988.9	-46449	92898.1

Table 5

Fixed Effects Parameters of Four Explanatory IRT Models

Model	Predictors	Fixed Effect	Standard Error
M_1	Item 1	2.35	0.05
	Item 2	2.29	0.05
	Item 3	2.26	0.05
	Item 4	2.22	0.05
	Item 5	2.18	0.05
	Item 6	-1.63	0.04
	Item 7	-1.66	0.04
	Item 8	-3.10	0.04
M_2	(Intercept)	1.75	0.02
	Physical	-2.37	0.02
	Verbal	-1.60	0.02
M_3	Item 1	3.25	0.22
	Item 2	3.20	0.22
	Item 3	3.17	0.22
	Item 4	3.12	0.22
	Item 5	3.09	0.22
	Item 6	-0.73	0.22
	Item 7	-0.76	0.22
	Item 8	-2.20	0.22
	Female	0.31	0.06
	Grade	-0.11	0.02
M_4	(Intercept)	2.88	0.19
	Physical	-3.08	0.18
	Verbal	-2.30	0.19
	Grade	-0.13	0.02
	Female	0.21	0.04
	Physical*grade	0.08	0.02
	Verbal*grade	0.08	0.02

Note. All fixed effects were significant, p-value < .001.

Conclusion

Currently, bullying is a common issue in school and attracts widespread attention in academia. Most large-scale studies rely on self-reports when identifying victimization and bullying groups (Juvonen, Graham, & Schuster, 2003). Unfortunately, they often have to face data with a large number of nonresponse items and figure out adequate methods to handle missing data (i.e., Glew et al., 2008). If we can know which groups are declined to omit questions and which items are more likely to be skipped before the data collection process, it would contribute to improving item designs and optimizing sampling selections; thereby enhancing the overall quality of data obtained from such measures.

Also, understanding nonresponse pattern could provide more accurate score interpretation as validity evidence for noncognitive measures on sensitive topics. To be specific, if we have a clear view about the missing pattern, we can add some useful auxiliary variables to missing data imputation procedure, which results in increased calculation efficiency and reduced estimation bias (Collins, Schafer, & Kam, 2001).

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–7. 2014.

Birkinshaw, S., & Eslea, M. (1998, September). *Teachers' attitudes and actions toward boy v girl and girl v boy bullying*. Paper presented at the annual conference of the developmental section of the British Psychological Society, Lancaster University, Lancaster, England.

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological methods, 6*(4), 330.

Crick, N. R., Grotpeter, J. K., & Bigbee, M. A. (2002). Relationally and physically aggressive children's intent attributions and feelings of distress for relational and instrumental peer provocations. *Child development, 73*(4), 1134-1142.

De Boeck, P., & Wilson, M. (2004). A framework for item response models. In P. De Boeck & M. Wilson (Eds.), *Explanatory item response models* (pp. 3-41). New York, NY: Springer.

Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., & Beechum, N. O. (2012). *Teaching adolescents to become learners: The role of noncognitive factors in shaping school performance--A critical literature review*. Chicago, IL: Consortium on Chicago School Research.

Glew, G. M., Fan, M. Y., Katon, W., & Rivara, F. P. (2008). Bullying and school safety. *The Journal of Pediatrics, 152*(1), 123-128.

Graham, J. W. (2012). Missing data theory. In *Missing Data* (pp. 3-46). New York, NY: Springer.

Juvonen, J., Graham, S., & Schuster, M. A. (2003). Bullying among young adolescents: The strong, the weak, and the troubled. *Pediatrics, 112*(6), 1231-1237.

Kane, M. (2013). Articulating a validity argument. In *The Routledge handbook of language testing* (pp. 48-61). Routledge.

McKnight, P. E., McKnight, K. M., Sidani, S., & Figueredo, A. J. (2007). *Missing data: A gentle introduction.* New York, NY: Guilford Press.

MN Department of Education. (2017). *Minnesota Student Survey*. Roseville, MN: Author. Retrieved from https://education.mn.gov/MDE/dse/health/mss/

Rivers, I., & Smith, P. K. (1994). Types of bullying behaviour and their correlates. *Aggressive behavior, 20*(5), 359-368.

Seals, D., & Young, J. (2003). Bullying and victimization: prevalence and relationship to gender, grade level, ethnicity, self-esteem, and depression. *Adolescence, 38*(152), 735.

Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical methods in medical research, 8*(1), 3-15.

R Core Team (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2016.

Rivers, I., & Smith, P. K. (1994). Types of bullying behaviour and their correlates. *Aggressive behavior, 20*(5), 359-368.

Rodriguez, M.C. (2017). Technical report on developmental skills, supports, and challenges from the 2013-2016 Minnesota Student Survey. Minnesota, Minneapolis, MN: University of Minnesota, Minnesota Youth Development Research Group. Retrieved from https://sites.google.com/view/mnydrg/research

Wayman, J. C. (2003, April). *Multiple imputation for missing data: What is it and how can I use it*? Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Whitney, I., & Smith, P. K. (1993). A survey of the nature and extent of bullying in junior/middle and secondary schools. *Educational research, 35*(1), 3-25.