

Running Head: BULLY-VICTIMIZATION SCALE RASCH ANALYSIS

Bully-Victimization Scale: Using Rasch Modeling in the Analysis of a Qualitative Scale

Marybeth Lehto, M.Ed.

University of Denver

July 1, 2009

2010 Annual Meeting of the American Educational Research Association

"Understanding Complex Ecologies in a Changing World"

April 30 – May 4

Denver, Colorado

Bully-Victimization Scale: Using Rasch Modeling in the Analysis of a Qualitative Bully-Victimization Scale

Bully/victim relationships are commonplace and recurrent occurrences in childhood and adolescence and can contribute to a child's feeling of safety in the school environment (Olweus, 2001; Smith, 2000). Olweus (1995) identified negative actions, as making faces or dirty gestures, intentional exclusion from a group, hurtful words, and physical contact. As defined by Craig, Henderson, and Murphy (2000) and consistent with the perspective of Elinoff, Chafouleas, and Sassu (2004), bullying behaviors may be physical or verbal, and include social exclusion. Both direct behaviors (physical attack, name-calling) as well as indirect behaviors (spreading rumors) constitute acts of bullying. In a nationwide sample of U.S. public school students, 25% of middle and junior high school students reported deliberately avoiding specific locations in the school (e.g., hallways, restrooms) to protect themselves. Approximately 10% of African American and Latino students indicated they stayed home from school due to worry over being targeted. (U.S. Department of Education, 1993)

The most commonly used instruments to measure bully/victim conflicts are the Olweus Bully/Victim Questionnaire (OBVQ) and the Revised Olweus Bully/Victim Questionnaire. Chan, Myron and Crawshaw (2005) noted few studies addressed the reliability and validity of the Olweus questionnaire, and reported the development of the non-anonymous School Life Survey (SLS) with improved reliability, validity and features designed to resolve the shortfalls of the Olweus questionnaire. Unfortunately, no fit statistics were reported for either instrument.

Kyriakides, Kaloyirou, & Lindsay (2006) conducted an analysis of the revised OBVQ using the Rasch model to measure construct validity, reliability and conceptual design on two separate aspects of bullying, i.e. Bullying Others and Being Victimized. Each construct measure

consisted of 8 items. Analysis revealed acceptable psychometric elements for each scale. Limitations of this measure were also addressed, specifically, the inclusion of additional difficult items to improve item targeting, and item phrasing modifications for more specificity to enable exploration of the causes of indirect bullying.

A 21-item multiple choice Questionnaire of Cyberbullying (QoCB) was developed to measure germane psychological and behavioral constructs. Unfortunately, nominal response categories were used, therefore, only content validity was investigated. (Aricak, Siyahhan, Uzunhasanoglu, Saribeyoglu, Ciplak, Yilmaz, & Memmedov, 2008)

Another concern is the appropriate number of items needed in analysis to assure unidimensionality. In an Exploratory Factor analysis (EFA) conducted by Georgiou (2008), 14 victimization items were used, whereas, 8 victimization items were used in the analysis done by Kyriakides, et. al. (2006). Confirmatory Factor Analysis (CFA) indicated a 10 item measure was preferable over a 47 items as reported by Cook, Kallen, and Amtmann (2009).

Clearly, there is a severely limited amount of extant literature in regard to comprehensive Rasch model analysis of bully-victim instruments, as well as measures with only a few items. Therefore, a Rasch analysis was performed on a six item bully victimization measure to address this gap in the literature. The instrument was analyzed to determine whether the data from the qualitative study fit Rasch model requirements for the definition of a measure.

Participants

The target population was ninth grade students who attended four-year high schools, with the accessible population being all ninth grade students that attended a comprehensive suburban high school located in the metropolitan Denver, Colorado area. Participants were a convenience sample of 670 ninth grade students that attended during the 2006-2007 school year. Participants

were assigned to English and “Freshman Seminar” classes by the school registrar. Participant grouping was pre-determined by the school’s student information system managed by the registrar. Ethnicity demographics distribution was 1.0% Native American, 8.1% Asian, 20.1% African American, 11.4% Hispanic American, 59.2% White, and 0.2% unclassified. Males comprised 50.8% of the sample population, females 49.2%.

Participation was anonymous with a total sample size of 670 ninth grade students; 601 students were administered the pretest with 525 students from the pretest group administered the post-test. Attrition of 145 students was due to absenteeism, expulsion, or transfer.

Instrument

The self-report victimization scale was developed and administered by a University of Denver Ph.D. candidate as part of an overall school engagement instrument. Four items were borrowed with permission from the Illinois Bully Scale (Espelage & Holt, 2001), while the remaining two items were developed by the researcher. Students were asked how often they had been picked on, made fun of, called names, been hit or pushed, and excluded from social cliques and activities in the past 30 days. Responses were recorded using a five-point frequency scale: (1) Never, (2) 1 or 2 times, (3) 3 or 4 times, (4) 5 or 6 times, (5) 7 or more times. The following constructs were assessed: (1) peer victimization, (2) type of victimization, and (3) frequency of victimization.

Procedure

A pretest was administered during week 3 of the 2006-2007 school year, to clustered groups of ninth grade participants in 26 "Freshmen Seminar" classes. The measure was repeated as a posttest during week 27 of the same school year to clustered groups of ninth grade participants in 30 English classes. To preserve participant anonymity, a research assistant

randomly assigned numbers to participants, then provided them to the primary researcher, who collected the survey data and released it to this paper's author for Rasch analysis. Only data collected from the final sample were used in the analysis.

Frequency of peer victimization: Degree of victimization was measured by frequency of bullying behaviors experienced by a participant while at school or school-related activities on each item, and was treated as an independent variable. Scores on the frequency scale were calculated by adding the responses on all items into a composite score.

Type of peer victimization: Quantification of the "victimization" variable included categorizing bullying behaviors as follows: verbal, physical, and exclusion.

Results

Use of the Response Scale

Rasch-Andrich thresholds were calculated and Linacre's (2002) criteria were applied for collapsing adjacent categories in the scale analysis. A five point rating scale was used: 1 (Never), 2 (1 or 2 times), 3 (3 or 4 times), 4 (5 or 6 times), 5 (7 or more times). Table 1 illustrates no category was underused (observed count less than 10). The dominant proportions of responses were in categories 1 and 2 as chosen by 47% and 38% of responders respectively, while the remaining 15% chose category 3, 4, or 5. Categories 4 and 5 were used the least frequently (<6% of the time each). The observed average of category structure was ordered, increasing in value from -2.74 to 0.68. Infit and outfit mean squares revealed acceptable values less than 2.0 for all categories. Threshold calibrations were satisfactory, increasing in value from -1.99 to 0.8. The category probabilities plot illustrated low probability of response values for categories 3 and 4. Category 5 (7 or more times) revealed infit and outfit mean

squares and probability of response values nearly equivalent to the values for category 1 (Never).

Dimensionality, Overall Fit, and Reliability of Person Separation

The data were analyzed using the entire sample ($N=670$) with all 6 items collectively. Valid sample size reduction ($N=525$) was due to the presence of null value responses and test-re-test attrition.

In Table 2, statistics revealed infit and outfit mean squares at approximately 1.0, with infit and outfit t -scores at approximately zero. However, person separation reliability for these data was low at 0.63.

A principal components analysis of residuals (PCAR) was conducted with highly contradictory values. Total variance explained was 55.1%. The eigenvalue for unexplained variance in the 1st contrast was 2.0 and percent variance was 15.1% which indicated a possible second dimension in these data. Moreover, the variance component scree plot illustrated more than one factor.

Attempts to Improve Reliability

An attempt to improve in reliability was tested by collapsing the three least frequently used categories of the rating scale, categories 3 (8%), 4 (4%) and 5 (3%). Category probability curves were cleaner, however, infit and outfit mean squares and reliability were unchanged.

Item fit statistics analysis prompted deletion of redundant items 1 and 3, followed by item 2 in the lowest logit position without improvement. Item 3 was deleted which resulted in a decline in reliability to 0.53. When item 2 then item 1 was deleted, reliability returned to 0.63. The analyst was reticent to delete more than one item at a time due to the small number of original items. Items 4, 5 and 6 with overfit, described latent traits of more severity,

therefore were not removed due to the importance of balanced traits on items. The remainder of the analysis was performed on the original data.

Item-fit Statistics:

Table 3 illustrates all items had infit and outfit mean squares within the acceptable range of 0.5-1.5 (Linacre, 2008, p.249). Item 4-HIT & PUSHED displayed the worst fit with more random variation than expected, whereas item 2-MADE FUN OF ME displayed the best fit. Point-measure correlations fell within the acceptable -1 to +1 range for all items. No evidence of contradictory use of responses was found in the Option/Distractor sub-table.

Targeting and Person-fit Statistics

Figure 1 illustrates the item-person map with item difficulties and student measures calibrated on the same scale. All 6 items clustered between 0.5 and 0.75 logits with the majority of the persons positioned between -1 and -4 logits. This revealed that most of the students responded they were never or rarely victimized, across all items. Redundancy is a notable possibility for two of the items; 1-PICKED ON and 3-CALLED ME NAMES. Item 2 was most frequently used and item 4 was the least frequently used. No significant differences were found for person fit in the bully-victimization scale both on gender ($F = 115, p > .05$) and age ($F = .986, p > .05$).

Differential Item Functioning (DIF)

A comparison of items across gender was conducted. Significantly different meanings was indicated between items for the two genders. Mantel-Haentzel calibration differences were used as a test of invariance. Items 1, 2, 3, 5 and 6 were invariant with calibration differences less than 0.5 logits across groups. Calibration differences for items 1, 2, and 3

were 0.00 logits, item 5 was 0.20 logits, and item 6 was 0.22 logits. Item 4 failed invariance with a calibration difference of 0.60 logits.

Construct and Content Validity

The intention of the instrument was to measure frequency of minor to severe bullying behaviors. The expectation of responses to the measure's item hierarchy would result in the majority of responses clustering at categories 1 (Never) or 2 (1 or 2 times) and fewer responses at category 5 (7 or more times). There is a logical expectation of proportionality when comparing the number of bully occurrences with the severity of bullying behaviors based on normal distribution. It is also reasonable to expect fewer people to have been hit or pushed (Item 6) than to have been picked on (Item 1). The typical student in the sample supports these expectations by indicating s/he has never or rarely experienced the bullying behaviors surveyed (Figure 1). Construct validity was estimated by calculating the correlation between the school engagement survey and bully-victimization scale resulting in a correlation of .274, $p < .001$. Further support for validity was provided by two content experts following analysis of item-person logit position in Figure 1.

Discussion

Reliability and unidimensionality were questionable for this instrument. Contradictory PCAR and overall fit statistics compromised unidimensionality. Person-fit statistics indicated the items fit well to the measurement model with responders' answers matching projected expectations on all items. Scale use indicated that students used the response format appropriately. Differential Item Functioning (DIF) indicated failure for invariance on 1 of the 6 items. Construct and content validity were established. Targeting tests showed item functionality similar for all membership of the target population.

The main concern was low person separation for reliability. One possible solution for improvement is to increase the number of items at the frequent and rare ends of the scale (Figure 1). The Spearman-Brown Prophecy formula was applied which determined the minimum increase in number of comparable items would be 3 to improve reliability to 0.72. An additional 6 items would improve reliability to 0.77.

Further suggestions for improvement in unidimensionality and reliability include an increase in the number of rating scale categories, re-phrase or re-design redundant items, test persons with more extreme experiences (high and low), and/or better sample-item targeting.

References

- Aricak, T., Siyahhan, S., Uzunhasanoglu, A., Saribeyoglu, S., Ciplak, S., Yilmaz, N., & Memmedov, C. (2008). Cyberbullying among Turkish adolescents. *Cyberpsychology & Behavior, 11*, 253-261.
- Cook, K.F., Kallen, M.A., & Amtmann, D. (2009). Having a fit: impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption. *Quality of Life Research, 18*, 447-460.
- Craig, W. M., Henderson, K., & Murphy, J.G. (2000). Prospective teachers' attitudes toward bullying and victimization. *School Psychology International, 21*, 5-21.
- Barber, B. L., Clements, P., Eccles, & J. S., Fuligni, A. J. (2001). Early adolescent peer orientation and adjustment during high school. *Developmental Psychology, 37*, 28-36.
- Elinoff, M. J., Chafouleas, S. M., & Sassu, K. A. (2004) *Psychology in the Schools, 41*, 887-897.
- Espelage, D. L., & Holt, M. L. (2001). Bullying and victimization during early adolescence: Peer influences and psychosocial correlates. *Journal of Emotional Abuse, 2*(3), 123-142.
- Espelage, D. L., Bosworth, K., & Simon, T. R. (2001). Short-term stability and prospective correlates of bullying in middle school students: An examination of potential demographic, psychosocial, and environmental influences. *Violence and Victims, 16*, 411- 426.
- Georgiou, S. (2008). Bullying and victimization at school: The role of mothers. *British Journal of Educational Psychology, 78*, 109-125.
- Hoover, J. H., Oliver, R., & Thomson, K. (1993). Perceived victimization by school bullies: New research and future directions. *Journal of Humanistic Education and Development, 32*, 76-84.
- Johnson, S. M., & Greenberg, L. S. (1985). Differential effects of experiential and problem-solving interventions in resolving marital conflict. *Journal of Consulting and Clinical Psychology, 53*, 175-184.
- Kupersmidt, J. B., & Coie, J. D. (1990). Preadolescent peer status, aggression and school adjustment as predictors of externalizing problems in adolescence. *Child Development, 61*, 1350-1362.
- Kyriakides, L., Kaloyirou, C., & Lindsay, G. (2006). An analysis of the Revised Olweus Bully/Victim Questionnaire Using the Rasch Measurement Model. *British Journal of Educational Psychology, 76*, 781-801.

- Nansel, T. R., Overpeck, M., Pilla, R. S., Ruan, W. J., Simons-Morton, B., & Scheidt, P. (2001). Bullying behaviors among US youth: Prevalence and association with psychosocial adjustment. *Journal of the American Medical Association*, 285, 2094-2100.
- Olweus, D. (1991). Bully/victim problems among school children: Basic facts and effects of a school based intervention program. In D. Pepler & K. Rubin (Eds.), *The development and treatment of childhood aggression*. Hillsdale, NJ: Erlbaum.
- Olweus, D. (1993). *Bullying at school: What we know and what we can do*. Oxford: Blackwell Publishers.
- Olweus, D. (1994). Annotation: Bullying at school: Basic facts and effects of a school based intervention program. *Journal of Child Psychology and Psychiatry*, 35, 1171-1190.
- Olweus, D. (1995). Bullying or peer abuse at school: Facts and intervention. *Current Directions in Psychological Science*, 4, 196 – 200.
- Olweus, D. (1996). *Revised Olweus Bully/Victim Questionnaire*. Mimeo. Bergen, Norway: Research Center for Health Promotion (HEMIL Center), University of Bergen.
- Olweus, D. (1997). Bully/victim problems in school: Facts and intervention. *European Journal of Psychology of Education*, 12, 495-510.
- Olweus, D. (1999). Sweden. In P. K. Smith, Y Morita, J. JungerTas, D. Olweus, R. Catalano, & P Slee (Eds.), *The nature of school bullying: A cross-national perspective* (pp. 7-27). London:Routledge.
- Olweus, D. (2001). *Olweus' core program against bullying and antisocial behavior: A teacher handbook*. Bergen, Norway: Research Center for Health Promotion (Hemil Center).
- Olweus, D. (2004). Olweus bullying prevention program: Design and implementation issues and a new national initiative in Norway. In F. K. Smith, D. Pepler, & K. Rigby (Eds.), *Bullying in schools: How successful can interventions be?* (pp. 13-36). Cambridge: Cambridge University Press.
- Olweus, D. (2005). A useful evaluation design, and effects of the Olweus bullying prevention program. *Psychology, Crime and Law*. 11. 389-402.
- Reid, K. (1989). Bullying and persistent school absenteeism. In D. P. Tattum & D. A. Lane (Eds.), *Bullying in school* (pp. 89-94). Stoke-on-Trent, England: Trentham Books.
- Reid, K. (2005). The causes, views and traits of school absenteeism and truancy: An analytical review. *Research in Education* 74, 59-82.
- Schwartz, D., Dodge, K. A., Pettit, G. S., & Bates, J. E. (1997). The early socialization of

- aggressive victims of bullying. *Child Development*, 68, 665-675.
- Schwartz, D. (2000). Subtypes of victims and aggressors in children's peer groups. *Journal of Abnormal Child Psychology*, 28, 181-192.
- Smith, P. K., & Brain, P (2000). Bullying in school: Lessons from two decades of research. *Aggressive Behavior*, 26 Issue 1, p, 1-9.
- Solberg, M. E., Olweus, D., & Endresen, I. M. (2007). Bullies and victims at school: Are they the same pupils? *British Journal of Educational Psychology*, 77, 441-464.
- Stevens, V., de Bourdeaudhuij, I., & Van Oost, P. (1994). Anti-bullying interventions at school: Aspects of programme adaptation and critical issues for further programme development. *Oxford University Press*, 16, 155-167.
- U. S. Department of Education. (1993). School engagement and students at risk. (NCES 1993–470). *Washington, DC: National Center for Education Statistics*.
- Vaillancourt, T., Hymel, S., & McDougall, P. (2003). Bullying is power: Implications for school-based intervention strategies. In M. J. Elias & J. E. Zins (Eds.), *Bullying, peer harassment, and victimization in the schools: The next generation of prevention* (pp. 157-176). New York: The Haworth Press, Inc.

Table 1

Summary of Category Structure (6 Items 5 Categories)

Category Label	Score	Observed Count	%	Observed ^a Average	Sample Expect	Infit MNSQ	Outfit MNSQ	Threshold Calibration
1	1	881	47	-2.74	-2.77	1.22	1.05	None
2	2	718	38	-1.68	-1.59	.82	1.02	- .94
3	3	157	8	- .29	- .51	.73	.71	.50
4	4	71	4	.33	.24	.88	.94	.68
5	5	53	3	.68	.77	1.12	1.40	.80
Missing		4	0	-2.18				

^aObserved Average is mean of measures in category. It is not a parameter estimate.

Table 2

Overall Model Fit

	Raw Score	Count	Measure	Model Error	Infit MNSQ	Infit <i>t</i> -score	Outfit MNSQ	Outfit <i>t</i> -score
Mean	10.6	6.0	-1.91	.72	.93	-.1	1.00	.0
SD	3.9	.1	1.32	.22	.71	1.1	.83	1.1
Max	25.0	6.0	1.39	1.08	4.67	3.4	6.01	3.7
Min	7.0	4.0	-3.71	.39	.03	-3.4	.03	-3.0

Table 3

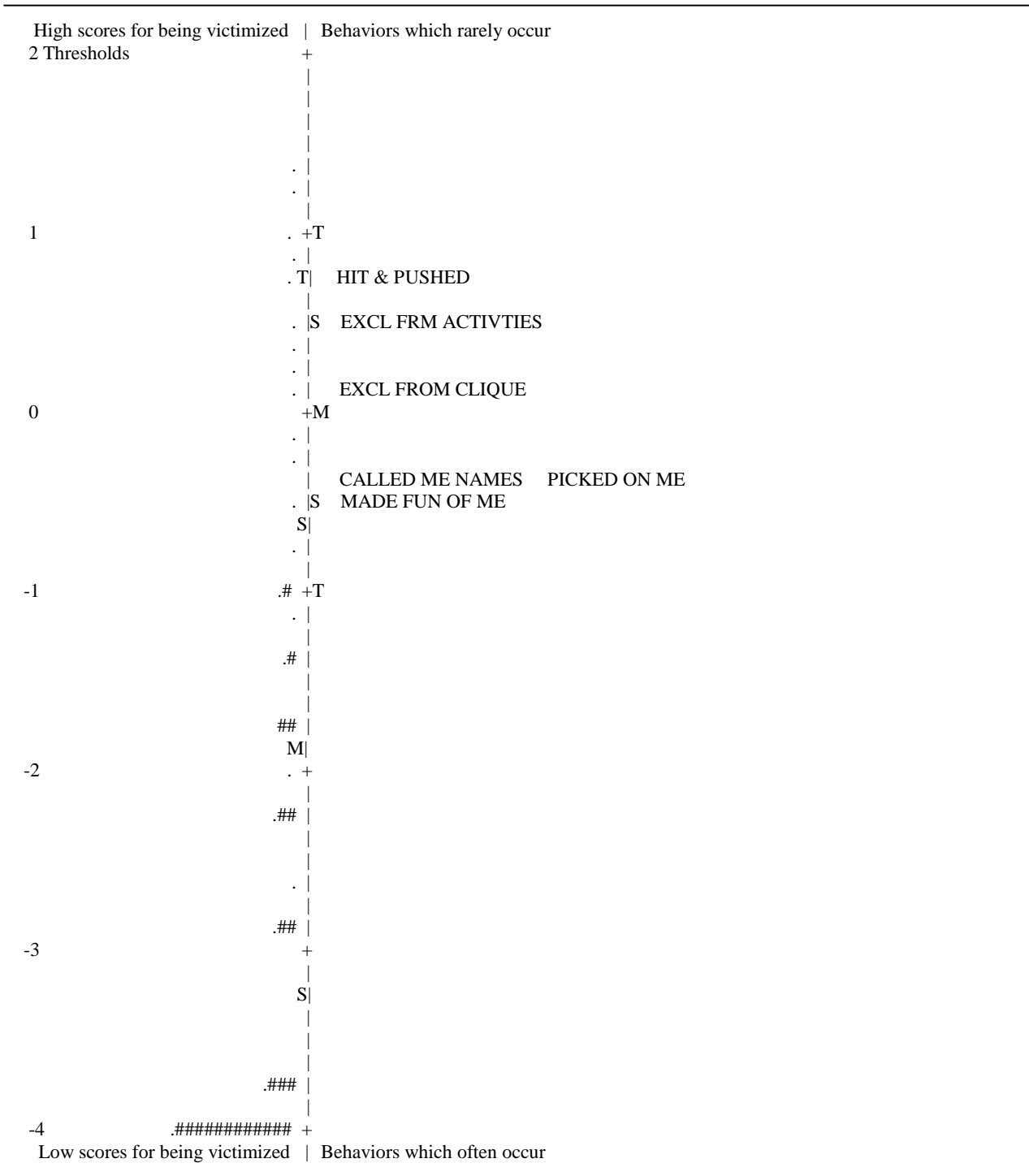
Item-fit Statistics

Item Number	Item	Infit MNSQ	Infit <i>t</i> -score	Outfit MNSQ	Outfit <i>t</i> -score	Point Measure Correlation
4	Hit and pushed	1.51	4.0	1.48	3.7	A .55
6	Excl from activities	1.38	3.2	1.19	1.7	B .61
5	Excl from clique	1.23	2.2	1.20	2.0	C .65
3	Called me names	.82	-2.0	.81	-2.2	c .77
1	Picked on me	.75	-2.9	.75	-3.0	b .79
2	Made fun of me	.57	-5.6	.59	-5.5	a .82
Mean		1.04	-.2	1.00	-.6	
SD		.35	3.6	.31	3.2	

Figure Caption

Figure 1. Item-person victimization map.

Figure 1.



Note: Each # represents 17 students. Each (.) represents 1 student.
 M represents the mean logit position for person or item.
 S represents 1 standard deviation above or below the mean.
 T represents 2 standard deviations above or below the mean.

Abstract The primary purpose of this study was to determine whether the data from the qualitative study fit Rasch model requirements for the definition of a measure, as well as to address concern in the extant literature regarding the appropriate number of items needed in analysis to assure unidimensionality. The self-report victimization scale was developed and administered as part of an overall school engagement instrument in a repeated measures design methodology. Participants were a convenience sample of 670 ninth grade students. Grouping was pre-determined by the school's student information system managed by the registrar. Results indicated validity was established and appropriate scale format use. However, reliability and unidimensionality were questionable. Recommendations are included. (Contains 3 tables and Item Map)