

Test Validity and Number of Response Categories: A Case of Bullying Scale

Satyendra Nath Chakrabarty

Former Director, Indian Maritime University
Kolkata Campus

Rumki Gupta

Indian Statistical Institute
Kolkata

About 1147 school students were selected based on a pre-designed sampling scheme and they underwent a Bullying test where each item had 5-response categories. Class teachers independently marked the students as bully or non-bully by giving a holistic rating in a qualitative fashion without any knowledge of students' test scores. Validity of the bullying scale was found by various approaches like Point bi-serial correlation, Chi-square measure of association and function of chi-square like Contingency Coefficient, Phi Coefficient, R^2 of Logistic regression separately for 2, 3, 4 and 5 response categories obtained by a collapsing data and Hit ratios. Fluctuations of validity indices were investigated to find the number of response categories, which performs best. Goodness of classification of students was done in two groups namely, bully and non-bully owing to the results of test scores and teachers' ratings. The two-point, three-point and four-point scales performed rather poorly and indices were higher for scales with 5-response categories. However, factor analysis showed that only one factor was measured by all the response categories implying unidimensionality of the Bullying Scale even if data are collapsed to have various response categories..

Keywords: Bullying, Validity, Point bi-serial correlation, Logistics regression, Contingency Coefficient, Hit ratios

Establishment of validity of tests developed to assess bullying behaviour in school students appears to be an important concern. Concurrent validity involves administration of two tests or measures of the same construct on the same sample with a minimum time gap. Both concurrent and predictive validity are subdivisions of criterion validity and the major difference is in terms of time gap. Thus, cross referencing the scores for every subject allows computation of a correlation, which is the numerical value of validity of the test and reflects the extent by which the test measures what it is supposed to. When the purpose of a test is to classify people at two ends of a construct say Bully and Non-bully, traditional techniques for estimating reliability and validity may not apply. Data and subsequent classifications from a test are valid when they are relevant and when they produce correct and consistent classifications of individuals.

Major issues of concurrent validity include:

- Inaccuracies associated with the benchmark test gets carried to the new test if the new test shows a high correlation
- Poor measurement in one variable may affect validity, expressed in terms of correlation
- Level of measurement of the two tests could be different and thus, validity requires a proper choice of type of correlation or association between the two tests or measures.
- If Bully is coded as 1 and Non-bully as 2, then it becomes a binary categorical variable. Appropriate correlation between such a dichotomous variable (Y) and a continuous variable, say test score (X) should be Point bi-serial correlation.

Pett (1997) observed that the criteria for evaluating the coefficient of determination with regards to standard correlation can be applied to

rpb^2 and suggested following intervals of values of coefficient of determination in terms of rpb^2 .

Very strong: ≥ 0.81 ; Strong: 0.49-0.80; Moderate: 0.25-0.48; Weak: .00-0.08

One can distinguish between two types of dichotomous variables viz. (i) purely dichotomous variable like gender and (ii) dichotomous variable with an underlying continuity but, measured discretely as two values like a Multiple Choice Questionnaire (MCQ) of achievement test where an individual's score is either zero or one. For the second case, appropriate correlation between such a dichotomous variable and a continuous variable is Bi-serial correlation, which assumes normal distribution for each of the variable. Computation of the Bi-serial correlation involves the ordinate of the standard normal distribution at the point, which divides the distribution into proportions of subjects with 0-score and 1-score in the dichotomous variable. The relationship between Bi-serial correlation and Point bi-serial correlation is well known. Validity is sample specific. Sample variances influence the validity and the power of significance tests of location. Zimmerman (2006) observed that for unequal sample sizes, the Type I error rate and power of test are repressed if a larger variance is associated with a larger sample size, and elevated if a larger variance is associated with a smaller sample size. This well-established effect is primarily due to unequal contributions of pooled estimates of error variance in the computation of test statistic.

Validity via test variance can be changed even after administration of the test by various ways like

- Ignoring successively top performers / low performers
- Ignoring items with low item - total correlation or low discriminating values etc.
- Changing number of response categories per item by collapsing data from a long scale to scales having a lesser number of response categories

As per Schutz and Rucker (1975), number of response categories does not materially affect the cognitive structure. Green and Rao (1970)

observed that six or seven response categories are appropriate and little extra information is gained by increasing the number of response categories beyond seven. This is in line with Miller (1956) who suggested that human mind has a span of apprehension capable of distinguishing about seven different items. Study of Hancock and Klockars (1991) establish superiority of higher number of response categories over 3-point or 4-point scales. Prestan and Colman (2004) observed that reliability, criterion validity, and discriminating power were much lower for scales with two, three or four response categories. Wakita et al (2012) found that number of options in a Likert type scale do influence the psychological distance between options particularly for seven-point scale without influencing descriptive statistics significantly, or reliability of the scale. Gupta and Dhara (2012) found that reliability in terms of Cronbach α declines with reduction in number of portions in Likert type scale. It may be kept in mind that scales with few response categories tends to result in lower value of test variances, which limit correlations with other scales (e.g., Chang, 1994; Martin, 1973, 1978; Nunnally, 1970). This restriction-of-range tends to lower the validity of scores from scales with few response categories and also limits their utilities.

Thus, a need is felt to explore answers to the following questions:

- i) How to find validity of bullying scale when teachers have rated the students as bully or non-bully and compare fluctuations of validity due to reduction in number of response category for each of the methods?
- ii) How many number of response category performs best?
- iii) What should be the best way to classify the students in two groups namely bully and non-bully?

Objectives

The present study attempts to find empirically the fluctuations of validity in terms of suitable measures of correlation or association when the number of responses per item is decreased from five to four and further.

Method

Participants

A multistage sampling design was adopted in this study. In the first stage, two districts of West Bengal were selected viz. Kolkata and Hooghly. From each district, schools were selected by Simple Random Sampling Without Replacement (SRSWOR). District-wise numbers of Government aided schools were available from a list (List of recognised X-class schools, 2005) published by the West Bengal Board of Secondary Education (WBBSE). From each selected school, students were also selected by SRSWOR. Thus, the sample in this study included 1147 students from the Government aided schools of West Bengal. The students were from class VII and class VIII from 44 different schools. The mean age of the students was 13.06 years and SD was 1.137 years.

Instrument

A Bengali version of The Peer Relation Questionnaire (PRQ) by Ken Rigby and Phillip Slee (1991) was administered on the selected students of the sampled schools. It contains 20 items and 3 subscales viz. bully scale, victim scale and pro-social scale. The scores pertaining to the Bully test was considered here to find the validity of the sub-test. The students are from the Bengali Medium schools but, the statements in the questionnaire are written in English. Considering this situation the statements were translated into Bengali. For reliability of the Bengali translation of the questionnaire, the Cronbach's α was calculated and it was .75.

Procedure

Before administration of the test, class-teachers of the selected schools were requested to mark each student selected from the class as a Bully or a Non-Bully. The teachers were well acquainted with their student's behaviors (at least class-room behaviors) and with rating guides, which was pre-prepared helped in their ratings. Not a single student was evaluated by more than one teacher. Thus, classification of students by teachers was done by holistic rating in a qualitative fashion without any knowledge of students' test scores. The situation is similar to Contrasting Groups Method as described

by Livingston & Zieky, 1982, which requires panellists to classify examinees according to whether they have or have not met the performance standard/s. Here, holistic rating was assigned simply on the basis of total impression of a student i.e. "General Impression Marking" as used by Charney (1984). Marking of the students as bully (1) or non-bully (0) by the teacher constituted the dichotomous variable (Y). The scores of students from the Bullying Scale constituted the continuous variable (X). The scale had five response categories for each item. After scoring the test, number of response categories was reduced progressively to four, three and two by merging successive categories. Thus, in addition to marking by teacher as a Bully (code-1) or a Non-Bully (code-0), each individual had four different scores viz. X_5 , X_4 , X_3 and X_2 corresponding to 5-response categories, 4-response categories, 3 response categories and 2-response categories, respectively.

To find the validity of the test, we are essentially looking for associations between test scores and ratings of performance in non-test settings. Validity of the test was evaluated in terms of a) Point bi-serial correlation between X and Y (since point bi-serial correlation is more appropriate for correlation between a binary variable and continuous variable) or function of point bi-serial correlation b) Chi-square measure of association or function of chi-square to ensure a range of association c) Coefficient of determination of Logistics regression d) Hit ratios.

Validity of test in the instant case were considered as

a) Point bi-serial correlation between test score (X) and ratings given by teachers (Y) for classifying students as a Bully or a Non-bully is given by

$$r_{pb} = \frac{M_1 - M_0}{S_n} \sqrt{\frac{n_1 n_2}{n}} \dots\dots\dots (1.1)$$

where M_1 = Mean of X for the group with Y=1 (Sample size - n_1)

M_0 = Mean of X for the group with Y=0 (Sample size - n_2)

$$n = n_1 + n_2 \text{ and } S_n = \sqrt{\frac{1}{n} \sum (X_i - X)^2}$$

a) Point bi-serial correlation lies between -1.0 to +1.0. For testing H0: Point bi-serial correlation = 0, the test statistics is

$$r_{pb} \sqrt{\frac{n_1 - n_2 - 2}{1 - r_{pb}^2}}$$

which follows t-distribution with (n1 + n2 – 2) degrees of freedom.

A high positive value of point bi-serial correlation in this context indicates that a student with high score on X is likely to be classified as a Bully by the class teacher.

b) Chi-square measure of association observes whether the two variables are associated. A contingency table of order r X c is a two way table containing frequencies by category and the association is given by

$$\chi^2 = \sum_{\text{all cells}} \frac{(f_o - f_e)^2}{f_e}$$

where $f_e = \frac{\text{Row Total} \times \text{Column Total}}{N}$

follows X²- distribution with(r-1) (c-1) degrees of freedom, where r denotes number of rows and c denotes number of columns of the table.

c) Pearson’s Contingency Coefficient (C) interpreted as a measure of the relative (strength) of an association between two variables and is given by

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}}$$

It can be proved that

$$0 \leq C \leq \sqrt{\frac{r-1}{r}}$$

where r denotes number of rows in the contingency table.

d) Phi coefficient – It is used only for 2x2 contingency table and is given by

$$\text{Phi} = \sqrt{\frac{\chi^2}{N}} \quad \text{In general, } -1 \leq \phi \leq 1$$

e) Cramer’s V Coefficient (V) – It is used to see whether a statistically, significant chi-square

value was the result of a large sample size instead of any substantive relationship between the variables. V ranges from 0 to 1 and is given by

$$V = \sqrt{\frac{\chi^2}{N(q-1)}}$$

Where q is the minimum (row, column). For 2X2 contingency table, V coincides with phi coefficient.

Here, the validity of the test was computed separately using r²_{pb}, x² measure of association, C-index between X and Y and phi coefficient for classifying students as a Bully or a Non-bully.

Results

Descriptive Statistics

Mean and variance of Test Score with different number of Response categories for bullies and Non-bullies are given in Table 1.

Table 1. Mean and Variance of Bullies and Non-bullies for Different Response Categories

| | 5- Response Categories | 4-Response Categories | 3-Response Categories | 2-Response categories |
|-----------------------------------------------|------------------------|-----------------------|-----------------------|-----------------------|
| Identified as Non-Bully by teachers (n = 751) | | | | |
| Mean | 6.49 | 6.27 | 5.03 | 4.68 |
| Variance | 6.60 | 5.12 | 1.79 | 0.77 |
| Identified as Bully by teachers (n = 396) | | | | |
| Mean | 9.94 | 7.22 | 5.49 | 5.09 |
| Variance | 6.09 | 6.18 | 2.79 | 1.66 |
| Combining two groups (n = 1147) | | | | |
| Mean | 7.68 | 6.96 | 5.19 | 4.68 |
| Variance | 9.12 | 5.69 | 2.18 | 1.11 |

Table 2. t - ratios and F - ratios between Bully and Non-bully for Different Response Categories

| | 5-response categories | 4-response categories | 3-response categories | 2-response categories |
|-----------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| t-ratio (Bully X Non-bully) | 1.71 | 29.19** | 0.57 | 1.79 |
| F-ratio (Bully X Non-bully) | 1.63 | 0.15 | 0.09 | 0.86 |

** Significant at 99% level

Table 3. Validity of the Test in Terms of Measures of Associations for Various Response Categories

| Measures | 5- Response Categories | 4- Response categories | 3- Response Categories | 2- Response categories |
|----------------------------------------------|------------------------|------------------------|------------------------|------------------------|
| Point bi-serial correlation (r_{pb}) | 0.54 | 0.18 | 0.15 | 0.18 |
| *Test statistic (t) for $H_0 : (r_{pb}) = 0$ | 32.57** | 27.87** | 27.67** | 27.83** |
| r_{pb}^2 | 0.29 | 0.03 | 0.02 | 0.03 |
| χ^2 | 217.06** | 38.23** | 14.54** | 21.07** |
| Contingency Coefficient(C) | 0.40 | 0.18 | 0.11 | 0.13 |
| Phi coefficient | 0.43 | 0.18 | 0.11 | 0.13 |

* Here, the test statistic $t = \frac{n_1 - n_2 - 2}{\sqrt{1 - r_{pb}^2}}$ follows t-distribution with 1145 degrees of freedom

** Significant at 99% level

The higher value of the mean test score for those identified as a Bully by teachers in comparison to those identified as a Non-bully for each Response category tends to imply face validity of the test.

t-ratios for testing equality of means and F-ratios for testing equality of variances between the bully and the non-bully groups for different response categories are given in Table – 2.

Observations

The mean test score for the combined group declined steadily with a reduction in number of response categories. Difference of mean scores between i-th and (i -1) th response categories (for i =5, 4, 3) were found to be significant. Similar trends were observed for the Bully group as well as for the Non-bully group. The mean for Bully group was more than the same for Non-Bully group for each response category.

ii) The variance of the test score for the combined group decreased in a non-linear fashion with a decrease in number of response

categories indicating a reduction in number of response categories, which tended to increase group homogeneity as can be seen from the Figure - 1.

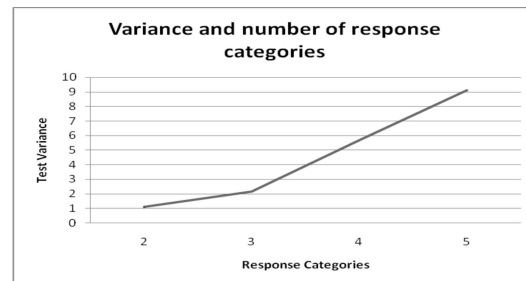


Figure 1. Variance and number of response categories

Variance of the i-th and (i - 1)-th response categories for i = 5, 4, 3 were found to be significant for all i. However, the extent of homogeneity that was achieved was not found to be linearly related with the number of response categories. A similar trend was observed for

the Bully group as well as for the Non-bully group. However, the null hypothesis of equality of variance of bully and non-bully groups could not be rejected for each response categories.

t-ratio for testing equality of means of Bully and Non-bully groups was found to be significant at 99% level only for 4-response categories.

Validity of test

In order to investigate the effect of number of response categories on validity of the test, all the measures of validity indicated in para 6 above were computed between the bully and non-bully groups identified by the teachers for each chosen category and are shown below:

The Point bi-serial correlation between the bully and the non-bully groups was found to be significant for each number of response categories. Thus, the Bullying Scale could be taken as valid for each number of response categories per item.

However, the value of Point bi-serial correlation decreased with a decrease in number of response categories implying a declining trend of test validity with a decline in variance as well due to the reduction in number of response categories as shown in Figure - 2 and Figure - 3

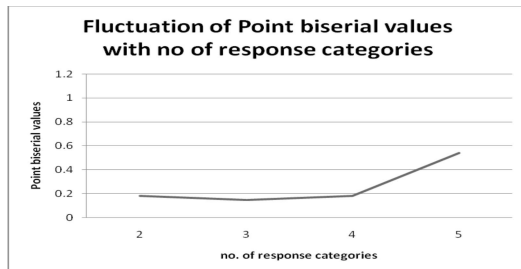


Figure 2. Fluctuation of Point Biserial values with no. of response categories

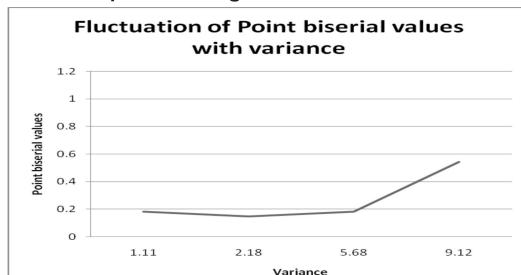


Figure 3. Fluctuation of Point biserial values with variance

Following the norms of Pett (1997), it can be said that the test has moderate validity when there are 5-response categories ($r_{pb}^2 = 0.2939$). The reduction in number of response categories resulted in a weak validity of the same test.

Values of r_{pb}^2 suggest that only 29.39% of variation of Y could be explained by the test score (X) when there were 5-response categories. The proportion of variance of Y explained by X was too poor with 4, 3 and 2-response categories. To measure the association between the score and ratings by teachers in terms of X^2 Pearson's Contingency Coefficient and Phi coefficient, students were classified into two classes. Those having scores greater than or equal to the average test score for a fixed number of response categories were considered in one class and the rest in the other class. For example, in the case of 5-response categories, students with scores ≥ 7.68 (mean score for the combined group with 5-response categories) constituted one class and the other class. X^2 values of the association fluctuated with changes in the number of response categories. Significant values of X^2 indicate existence of association between teachers' ratings and the test score. Thus, the test was valid for each number of response categories. However, values of X^2 declined steadily with the decrease in number of response categories. However, Pearson's Contingency Coefficient or Phi Coefficient declined for 4 and 3 - response categories but, increased marginally for 2-response categories.

When the number of response categories was 5, teachers' ratings and test scores were found to be well associated in terms of significant values of χ^2 and also in terms of higher values of Pearson's Contingency Coefficient (C) and Phi coefficient. Possibly the reduction in variance as a sequel to reduction in number of response categories makes the distribution of test score more flatter, which tends to increase the proportion of agreement in the main diagonal positions of contingency table with teachers' ratings, which in the present case was rather flat. The value of square of Phi Coefficient (ϕ^2) measures the proportion of one variable that is explained by the other variable. The situation with 5-response categories emerged as the best ($\phi^2 = 0.1892$). Poor percentage of explained

variance tends to indicate little overlapping of classification done by the test (X) and the teacher's ratings (Y) and the percentage of variance explained declined further with a reduction in number of response categories.

Classification

If the classification of students as bully or non-bully done by the teachers is good and the test on Bullying Behaviour is valid, then logistic regression of the dummy variable Y with the test score as the independent variable would give a significant value of beta co-efficient or R^2 or insignificant values of standard errors. A significant value of R^2 indicates that the test scores could be used to predict ratings given by the teachers. Thus, values of R^2 could be taken as a measure of validity of the test. Results of regression analysis are shown in the table below.

Logistic Regression resulted in value of R-Square as 0.296 for five response categories, which declined to 0.032, 0.023 and 0.033, respectively for 4, 3 and 2 response categories. However, F-ratios were found to be significant for all the response categories considered.

Poor value of R^2 even for 5-response categories indicates about 29.56% of overlapping between predicted value used for classification and classification done by the teachers. Thus, the test had moderate validity for 5-response categories with teachers' rating as criterion variable. Value of R^2 or adjusted R^2 was too low

when number of response categories was In other words, the regression model performed the best when the number of response category was ≤ 5 and failed to perform well when the number of response category was reduced.

An attempt was made to see correlations among various response categories. If the same construct is being measured even when response categories are collapsed, the correlation between i-th and (i-1)-th response categories should be high for $i=5, 4$ and 3. The matrix of measures of correlation for the entire group is shown below in Table 4.

The correlation for each pair of response categories was significantly high. However, relatively lower values of correlation found for few pairs despite a steady decline of mean and variance with a reduction in number of response categories could be probed further.

Strictly speaking, correlation between 2-number of response categories and scores with 2-response categories is not meaningful, since it boils down to correlation between two dichotomous variables.

Conclusions drawn from the above three measures are similar. Probably, collapsing of data different number of response categories results in different psychological distances between successive options. This is also indicated by the proximity matrix (a dissimilar matrix) obtained considering Euclidian distances as given below.

Table 4. Matrix of Correlation, Spearman's rho & Kendal's tau for Various Response Categories

| | 5-response categories | 4-response categories | 3-response categories | 2-response categories |
|-----------------------|-----------------------|-----------------------------------------|-----------------------------------------|-----------------------------------------|
| 5-response categories | 1 | 0.45 ** (Rho-0.48)** [Tau-0.44]** | 0.94 ** (Rho-0.92)** [Tau-0.87]** | 0.92 ** (Rho-0.91)** [Tau-0.83]** |
| 4-response categories | | 1 | 0.43 ** (Rho-0.45)** [Tau-0.40]** | 0.42 ** (Rho-0.44)** [Tau-0.40]** |
| 3-response categories | | | 1 | 0.91 ** (Rho-.97)** [Tau-0.93]** |
| 2-response categories | | | | 1 |

** Significant at 0.01 level (2-tailed).

Table 5. Proximity Matrix

| | 5-response categories | 4-response categories | 3-response categories | 2-response categories |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 5-response categories | 0.000 | 113.40 | 61.83 | 80.87 |
| 4-response categories | | 0.00 | 133.39 | 142.25 |
| 3-response categories | | | 0.00 | 26.29 |
| 2-response categories | | | | 0.00 |

The proximity matrix indicates i-th and j-th cells were much apart for ($i \neq j$ and $i, j = 2, 3, 4, 5$). An attempt was made to assess the goodness of classification done by the test through a chosen cut score from the test, which is normative as well as substantive. The cut score is the operationalization of the performance standard (Kane, 1994, p. 426). It defines the minimum test score required for an examinee to be classified as satisfying or not satisfying on the performance standard. In the present case, students with test score at or above the cut score may be treated as bully and those with below the cut score may be classified as non-bully. Accuracy of classification may be reflected by constructing a contingency table and computing Hit ratio given by

$$\frac{\text{Sum of frequencies in the diagonal}}{\text{Total no. of cases}}$$

or by (1- probability of misclassification). If probability of misclassification is defined as

$$\frac{\text{Sum of off-diagonal frequencies}}{N}$$

then Hit ratio + probability of misclassification = 1. Thus, Hit ratio can be taken as a measure of classificatory validity.

Considering the cut score as a mean test score for the entire group, students were classified in two classes namely those with score \geq mean and those with score $<$ mean, 2 X 2 Contingency table for 5-response categories was prepared with teachers' ratings as "Bully" & "Non-Bully" in columns and test scores in rows as $\text{Test Scores} \geq \bar{X}_{(s)}$ and $\text{Test Scores} \leq \bar{X}_{(s)}$ where $\bar{X}_{(s)}$ denotes mean score for 5-response categories.

Here, the probability of misclassification was

$$\frac{\text{Sum of off-diagonal frequencies}}{N} = \frac{90+190}{1147} = 0.244$$

and Hit ratio

$$\frac{\text{Sum of diagonal frequencies}}{N} = \frac{206+661}{1147} = 0.756$$

Values of Hit Ratio for various response categories were 0.642, 0.598 and 0.621, respectively for 4, 3 and 2 response categories.

Classification by the test scores in terms of mean score and classification done by the teachers agreed in 75.60% of cases for 5-response categories. Degree of agreement declined with fewer response categories.

Table 6. Total Variance Explained

| Factor | Initial | | | Extraction Sums of Squared Loadings | | |
|--------|-------------|---------------|--------------|-------------------------------------|---------------|--------------|
| | Eigen value | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 3.11 | 77.86 | 77.86 | 2.98 | 74.43 | 74.43 |
| 2 | .73 | 18.36 | 96.22 | | | |
| 3 | .09 | 2.35 | 98.60 | | | |
| 4 | .06 | 1.43 | 100.00 | | | |

Factor Analysis

However, to answer the major question on whether all these different number of response categories is measuring the same contrast, factor analysis was undertaken. Kaiser-Meyer-Olkin Measure of Sampling Adequacy was 0.811 and Chi – square value for Bartlett's Test of Sphericity was found to be significant at 0.01% level. Goodness of fit test gave chi-square = 6.465, which is significant at 3.9% level. Factors extracted by maximum likelihood method showed only one factor explaining 77.86% of variance as shown in Table - 6

Thus, only one factor was measured by all the response categories implying unidimensionality of the Bullying Scale even if the data are collapsed to have various response categories. To find a tally with the purpose of one dimension was tapped by the Bullying Scale.

Conclusions

It has been empirically found that

- Reduction in the number of response categories is associated with the reduction of test variance.
- Reduction of the test variance limits the magnitude of correlation and thus, results in lower value of convergent validity.
- Test variance and validity appear to have a positive relation.
- Among the various indices of validity considered in this study, the scales with few response categories performed the worst. Correlation among various response categories and classification using test scores tell similar stories.
- Degree of agreement between the classification done by teachers and the test in terms of Hit ratios was the maximum for 5-response categories.
- The test with 5-response categories was found to have moderate validity with teachers' rating as the criterion. Reduction in the number of response categories resulted in lower values of validity, which resulted in poor validity of the test.
- From the psychometric angle, the test with 2, 3 or 4 response categories, with poor validity coefficients, appeared to have limited usefulness.
- Intercorrelations among various response categories, however, measured the same, single factor indicating unidimensionality of the Bullying Scale even if data are collapsed to have various response categories.

References

- Charney, Davida (1984). The Validity of Using Holistic Scoring to Evaluate Writing: A Critical Overview. *Research in the Teaching of English*, 18(1), 65-81.
- Chang, L. (1994). A psychometric evaluation of four-point and six-point Likert-type scales in relation to reliability and validity. *Applied Psychological Measurement*, 18, 205-215.
- Green, P. E., & Rao, V. R. (1970). Rating scales and information recovery: How many scales and response categories to use? *Journal of Marketing*, 34, 33-39.
- Gupta, Rumki and Dhara, Jayeta (2012). *Fluctuation of Reliability with change of Response Categories in Questionnaire*. In Suhas Shetgovekar (Ed.), Proceedings of the National Conference of the Psychological Measurement: Strategies for the New Millennium, (pp. 126-132), New Delhi: IGNOU.
- Hancock, G. R. & Klockars, A. J. (1991). The effect of scale manipulations on validity: targeting frequency rating scales for anticipated performance levels. *Applied Ergonomics*, 22, 147-154.
- Kane, M. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research*, 64, 425-461.
- List of Recognised X-Class schools (2005). Published by West Bengal Board of Secondary Education.
- Livingston, S. A. & Zieky, M. J. (1982). *Passing scores: A manual for setting standards of performance on educational and occupational tests*. Princeton, NJ: Educational Testing Service.
- Martin, W. S. (1973). The effects of scaling on the correlation coefficient: a test of validity. *Journal of Marketing Research*, 10, 316-318.
- Martin, W. S. (1978). Effects of scaling on the correlation coefficient: additional considerations. *Journal of Marketing Research*, 15, 314-318.
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for

- processing information. *Psychological Review*, 63 (2), 81 -97.
- Nunnally, J.C. (1970). *Introduction to psychological measurement*. New York: McGraw-Hill.
- Pett, M. A. (1997). *Nonparametric Statistics for Health Care Research: Statistics for Small Samples and Unusual Distributions*. Thousand Oaks, CA: Sage Publications, Inc.
- Preston, Carolyn C. and Colman, Andrew M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104, 1 -15 (www.elsevier.com/locate/actpsy).
- Schutz, H. G., & Rucker, M. H. (1975). A comparison of variable configurations across scale lengths: an empirical study. *Educational and Psychological Measurement*, 35, 319-324.
- Wakita, T; Ueshima, N and Noguchi, H. (2012) – Psychological Distance Between Categories in the Likert Scale: Comparing Different Numbers of Options. *Educational and Psychological Measurement*, 72 (4), 533 – 546.
- Zimmerman, Donald W. (2009). Two separate effects of variance heterogeneity on the validity and power of significance tests of location, *Statistical Methodology*, 3 (4) 351-374.
- Manuscript submitted on July 14, 2015
Final revision received on December 07, 2015
Accepted on January 04, 2016.

Satyendra Nath Chakrabarty, M. Stat. with Psychometry as specialization from Indian Statistical Institute and Former Director, Indian Maritime University, Kolkata Campus.
E-mail: snc12@rediffmail.com

Rumki Gupta, Ph. D., Associate Scientist 'C', Psychology Research Unit, Indian Statistical Institute, Kolkata. Email: gupta.rumki@gmail.com

JIAAP Full Text Back Volumes (2005 to 2010) are available at www.medind.nic.in IndMED

A Bibliographics Data-base of Indian Biomedical Research

It is a matter of great pleasure that for appropriate publicity of Indian Biomedical Research, Indian MEDLARS Centre, under the National Informatics Centre, has designed and developed a database entitled IndMED meeting international standards. The database is accessible fulltext on internet at the website <http://medin.nic.in>. Fulltext of 38 journals taken up for the IndMED. Authors are requested to include abstracts with their papers. While sending their papers for publication in future.

For IndMED details please write to:
Bibliographics INformatics Division
National Informatics Centre
(Department of Informatics Technology)
A-Block, CGO Complex, Lodhi Road,
New Delhi - 110 003, India.

Telephone : 91-11-24362359, Fax : 91-11-24362628, Email : medinfo@nic.in