Intercoder Reliability Report

TORONTO STAR 2008

Prior to running intercoder reliability tests, we verified the accuracy of the data entered by matching the entries with the coders' written responses. We randomly selected 10 items from each issue, and verified the data for all the elements within those items. If any errors were found within the set, the errors were corrected and another 10 items were selected and checked. We repeated the process until a set of 10 randomly selected items was found to be without data entry error.

We measured intercoder reliability by randomly selecting four sample issues (of the total 28 issues in our data set) and testing all variables within those four issues. The test issues were 9, 10, 16, and 19. There were 410 local items in our test sample, of the total 2731 local items in our study. Thus, given that our test sample represents over 10 per cent of our data, and that the four test issues fall on various days of the week, there is no reason to believe that the test sample is not representative of the data set.

Each of the four test issues was coded by the researcher and one research assistant. After 80 hours of training, each coder coded the issues independently, meeting only once to reach a consensus on the number of local items before proceeding to independently code the item details for local items. All other discrepancies were resolved through discussion after the intercoder reliability tests were completed.

We measured intercoder reliability for interval variables using Lin's concordance, with the aid of PRAM (Program for Reliability Assessment with Multiple Coders) software. Lin's concordance measures the correlation between coder responses, and takes into account systematic coding errors (coder bias), the possible range of responses, the magnitude of difference between coders' responses, and the agreement expected by chance. We considered reliability to be acceptable at or above .700 using Lin's concordance.

We measured intercoder reliability for nominal variables using Cohen's Kappa, with the aid of SPSS (Statistical Package for the Social Sciences) software. Cohen's kappa is a relatively conservative index that measures the extent to which coders make identical coding decisions, and takes into account the agreement expected by chance. We considered reliability to be acceptable at or above .700 using Cohen's kappa.

Intercoder reliability was at or above .766 for all variables, with the exception of minority community. The low level of agreement for this variable is partially due to the small number of relevant cases (four of the 410 local items). However, given the low level of agreement within these relevant cases, we will omit this variable from our analysis and discussion.

Lin's Concordance Results

Variable	Lin's Concordance Test Result
NoSptlRef	.950

Cohen's Kappa Results

Variable	Cohen's Kappa Test Result
Page	.977
Word_Count	.988
ItemOrigin	.900
ItemForm	.946
FotoStatus	.989
GenFotoCont	.981
Religion	.867
MinorityCmnty	.005
GenSubj	.864
SubjDetail	.766

Of the remaining 24 non-test issues, the researcher coded 10, and the research assistant coded 18.