

Intercoder Reliability Report

Toronto Star 2011

Two research assistant coders coded 21 issues of the Toronto Star for the period January 4, 2011, to August 8, 2011. We excluded the weeks between March 17 and May 4 because a federal election took place during this period and we considered the news mix during this period to be unrepresentative of typical content.

Prior to running intercoder reliability tests, each coder verified the accuracy of the data entered by matching the entries with their first assessment, written on the hard copy of the newspaper. Coders 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. Coders repeated the process until a set of 10 randomly selected items were found to be without data entry error.

We measured intercoder reliability by randomly selecting three sample issues (of the total 21 issues in our data set) and testing all variables within those three issues. The test issues were (35, 41, 43).

Issue number	Date of publication
35	2/5/2011
41	5/28/2011
43	6/13/2011

There were 365 local items in our test sample, of the total 1871 local items in our study. Thus, given that our test sample represents more than 10 per cent of our data, there is no reason to believe that the test sample is not representative of the data set.

After about 35 hours of training, each coder coded the 3 issues independently, meeting only once to reach a consensus on the number of local items before proceeding to independently code the item details. 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 PRAM (Program for Reliability Assessment with Multiple Coders) 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 .725 for all variables, with the exception of race/ethnicity 2 and race/ethnicity 3. The low level of agreement for race/ethnicity was because there were only 3 entries for this variable. We considered this variable still valid, however, because the coders agreed no data would be entered for 362 items; they agreed on one entry; and disagreed on 2 entries. Overall, we believe this demonstrates a high degree of agreement. The criteria for determining race/ethnicity entries is the same for all race/ethnicity variables.

Variable	Lin's Concordance Test Result
NoSptlRef	.936

Variable	Cohen's Kappa Test Result
ItemOrigin	0.944
ItemForm	0.939
FotoStatus	0.983
GenFotoCont	0.924
Religion	0.940
MinorityCmnty	0.927
GenSubj	0.843
SubjDetail	0.738
R/E1	0.725
R/E2	0.469
R/E3	0.419
R/E4	1.000

Of the remaining 18 non-test issues, coder Emily Jin coded 14 issues and coder Harbit Natt coded 4 issues.