

## **Intercoder Reliability Report**

### **Ming Pao 2008**

To be certain that the coding of Ming Pao is consistent with coding for the Toronto Star, the two Ming Pao coders received 80 hours of coder training for Toronto Star content. After random testing indicated they were coding in a similar/consistent manner with each other, the Ming Pao coders then coded three Toronto Star test issues (Issues 9, 16 and 19).

Once it was clear that the Ming Pao coders' decision-making vis a vis the Toronto Star was consistent, we moved on to coding Ming Pao. To be sure that the Ming Pao coders coded five test issues of Ming Pao (issues 11, 20, 21, 26, 28).

### **The Toronto Star Test Results**

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 selecting 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 coding three of the randomly selected sample issues (of the total 28 issues in our data set) and tested all variables within those three issues. There were 330 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 three 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 three test issues was coded by two research assistants. 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 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.

At the item level, intercoder reliability was at or above .757 for all variables, with the exception of minority community and religion. The low level of agreement for these variables is partially due to the small number of relevant cases. However, given the low level of agreement within these relevant cases, we will omit this variable from our detailed analysis and discussion.

#### Lin's Concordance Results

Variable	Lin's Concordance Test Result
NoSptlRef	.931

#### Cohen's Kappa Results

Variable	Cohen's Kappa Test Result
ItemOrigin	.804
ItemForm	.892
FotoStatus	.969
GenFotoCont	.945
GenSubj	.786
SubjDetail	.757

#### The Ming Pao Test Results

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 selecting 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 five sample issues (of the total 28 issues in our data set) and testing all variables within those five issues. There were 197 local items in our test sample, of the total 1157 local items in our study. Thus, given that our test sample represents over 10 per cent of our data and that the five 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 test issues was coded by two research assistants who are fluent in Chinese. 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 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 .754 for all variables, with the exception of minority community and religion. Our examination of the data suggests that the low level of agreement for this variable may be partially due to the small number of relevant cases where minority community/religion was coded for a news items. However, given the low level of agreement within these relevant cases, we will omit these variables from our detailed analysis and discussion.

#### **Lin's Concordance Results**

<b>Variable</b>	<b>Lin's Concordance Test Result</b>
NoSptlRef	.969

#### **Cohen's Kappa Results**

<b>Variable</b>	<b>Cohen's Kappa Test Result</b>
ItemOrigin	.934
ItemForm	1.00
FotoStatus	.993
GenFotoCont	.934
GenSubj	.881
SubjDetail	.754

After the coders passed the intercoder reliability test, the research assistants coded the remaining 24 non-test issues.