Inter-coder Reliability Scores
Fukushima Case

Francesco Visconti (University of Leicester)[1]

This document[2] reports the reliability scores of coders who participated in the coding of the Fukushima policy juncture on nuclear energy. It provides an outline of the various steps of the coding process and of components of inter-coder reliability. It starts by explaining the tasks coders are supposed to do and continues with a description of the selection and training. It then presents in details the four steps of the India pre-test and what each of them is trying to measure in terms of inter-coder reliability. On the whole the inter-coder reliability test on the Fukushima case coders yielded satisfactory results.

ResponsiveGov coding description

In order to collect data to investigate governments' responsiveness to citizens between elections, the ResponsiveGov project employed a number of coders that coded a specific policy juncture for each country. Each coder worked independently from other coders and under the close supervision of one member of the ResponsiveGov core team based at the University of Leicester.[3] They were either a native speaker of the language(s) required to code her/his country of reference, or in a few cases a proficient speaker. The codebook and related appendices provided detailed instructions with regards to the identification of the policy junctures, the selection criteria of outlets and events to be coded, and general rules for coding missing data or non-applicable variables.

In ResponsiveGov the coding of events is carried out by policy juncture. At any one point in time, a coder codes the events of only one policy juncture in one country at a time. **Appendix 1** lists the policy junctures to be coded and the time period(s) (e.g. Jan 01, 1996 – Dec 31, 1999) they have to be coded for. In some cases (e.g., GMOs and Internet policy junctures) coders have been asked to help identify the time periods to be coded. In those cases they produced a **Coding Period Identification** file following the instructions contained in a document specific to each policy juncture. In the produced file coders had to list all governments and ministers in charge of the policy during the specific time period defined in **Appendix 1**. For each of these governments, coders needed to search for relevant policy

[1] The inter-coder reliability tests for the ResponsiveGov project were partly designed by Pietro Castelli-Gattinara, while a researcher in the project. His contribution to the results presented in this document is hereby acknowledged.

2 How to cite this document: Francesco Visconti (2016). "Inter-coder Reliability Scores - Fukushima Case" Leicester: ResponsiveGov, http://responsivegov.eu.

3 At different time points the supervisors included: Laura Morales, Maarja Luhiste, Pietro Castelli-Gattinara, Luca Bernardi, Daniel Bischof, Oriol Sabaté, and Francesco Visconti.

pledges in their election/coalition manifestos, programmatic speech, etc. Finally, coders were asked to search for legislation related to the policy juncture that was proposed, introduced or approved during each government period. Once this document had been prepared the coder's supervisor verified the whole document and selected the government period(s) to be coded.

Once the government period had been selected, the main task coders performed prior to coding the policy juncture consisted in gathering information on the government's initial policy position and on the main government characteristics (e.g., government ideological position, type of majority coalition). Detailed instructions concerning the coding of governments' position and characteristics are presented in **Appendix 1** (*i.e.*, hierarchy of sources and documents where to search for initial policy position, definition of the ideological scale to code the initial policy position, etc.), while the data collected summarising governments' positions and characteristics is stored in a separate Excel file: **Appendix 1a**. Information stored in this document is then used to code the first row of the **Main Data Matrix** Excel file specific to each country and policy juncture. In the case of each policy juncture, the governmental position is coded as the first and the last 'event'.

The next task performed by coders consisted in coding the initial and final policy position of all parties relevant for the country. This means the coder had to code parties' positions on the relevant issue in the election before the juncture started and the election after the juncture ended (or the election that ended the juncture). **Appendix 5a** lists for each country election years all relevant parties and therefore whose policy position needed to be coded. **Appendix 5b** provides information on the selection criteria for parties, along with instructions on how to code parties' positions on policy junctures. The initial and final policy positions for parties are coded based on the party's election manifesto for the national elections prior to the start of the juncture, and elections ending or following the end of the policy juncture. Initial and final parties' policy positions are coded in the Excel file **Appendix 5c**.

In the next step the coder searched for events to be coded in national press agencies newswires related to the specific policy juncture. For each policy juncture a specific keyword dictionary (**Appendix 2**) provided to coders was used to search for news related to the policy juncture. Before coding, coders translated the keyword dictionary into the language of the country they were coding and added keywords specific to their country (*e.g.*, interest groups of their country, etc.), and a supervisor verified the consistency of the translation. National press agencies newswires used for each country are listed in the codebook and were accessed through Factiva and Nexis databases. In some cases (e.g., Denmark, Germany, Greece, Iceland, and Sweden) the main press agency was not available either on Factiva or on Nexis, and newswires were accessed directly through the online database of the news agency with a subscription. In Nexis or Factiva database environments the coder selects the national news agency as the only source to be used in the search. Next, the

coder set a geographical delimiter in order to retrieve news stories relevant to the country he/she was coding. In Nexis, the coder selected from the "Add index terms:" selecting the category "geography", under which the coder ticked the name of the country he/she was coding at that point in time. In Factiva, the coder ticked the name of the country he/she was coding at that point in time under the "Region" section. With the keyword search the coder identified the range of dates during which the policy juncture appeared in the news. Then, the coder set the identified time period as the "range of dates" in the database search builder. Then, by utilising the keyword dictionary, the coder identified the news stories and coded all reported events, following the specific coding rules stated in the codebook.

The units of coding are events defined as one of the following occurrences mentioned in a news story or in one of our separate sources of information (survey databases, legislation databases, editorials, etc.) in relation to the policy issue/juncture of interest:

- Legislative proposal;
- Any piece of policy or legislation announced or passed;
- Declaration of government's position;
- Declaration of opposition's position;
- Declaration of interest group's (any organization that can lobby: employers, trade unions, environmental, professional, civil society, etc.) or individual activists' position;
- Declaration of the position of any other relevant social and political actor;
- Parliamentary debate;
- Parliamentary act;
- Opinion poll published;
- Collective action (e.g. petition, protest, demonstration, etc.) either online or 'offline';
- Editorial;
- Any other relevant "real world" event depending on the policy juncture (e.g. an earthquake, a nuclear accident, a stock exchange crash, etc.).

Although we are interested in the reactions of national governments to different expressions of the public opinion, any 'event' that occurs in the country — regardless of the level of government or geographical area where it happens — was coded. Thus, if local or regional actors intervened, these events were coded. Also, simple statements (i.e. declarations, claims) by the government or any other actor were coded as events, even if they do not change the policy, but the claim / statement was directly linked to the policy juncture. In order to code an event, the event had to be explicitly mentioned or referred to in a news story or any of the other sources specified for coding. In one news story several events can occur (e.g., one news story can cover a protest action against/for a policy change as well as the government's position on the policy juncture). In this case, all events needed

to be coded individually as separate events. Only events (claims, declaration, protests, opinion polls published, etc.) that took place in the country of coding were coded.

After coding the content of newswires, the coder had to search for additional survey sources related to the policy juncture during the identified time period. **Appendix 3** lists the surveys and polls for each country that the coder needed to check when coding a policy juncture. Cross-national surveys covering the policy area were centrally collected and distributed to coders. All surveys and polls that cover public opinion on the given policy juncture and have not been published in the newswires (and thus have not been coded as yet) have to be coded. Responses to each individual relevant survey question, related to the policy juncture, are coded as individual events. Then, the coder coded the survey responses as separate events, following the specific coding rules stated in the codebook.

After coding the news content and survey sources, the coder searched for additional legislative sources and parliamentary questions related to the policy juncture during the time period identified. **Appendix 4** lists the legislative sources for each country that the coder needed to check when coding a policy juncture. All legislative acts / bills / proposals / and oral parliamentary questions on the given policy juncture, which had not been published in the newswires (and thus were not coded) had to be coded. As a general rule, each legislative act, bill, proposal, set of debates on parliamentary questions, etc., related to the policy juncture, was coded. Then, the coder coded the legislative acts / bills / proposals as separate events, following the specific coding rules stated in the codebook.

After coding the news content, survey sources, and legislative sources of a given policy juncture, the coder looked for newspaper editorials[4] within the period of the policy juncture that deals with the issue being coded. **Appendix 6** lists the newspapers whose editorials needed to be coded for each country. The given policy juncture's keyword dictionary was used to search for editorials related to the policy juncture. In each country, the editorials of one "progressive" and one "conservative" broadsheet newspaper were coded. The newspaper editorials were accessed by using Nexis and Factiva databases, university library subscriptions, and the newspapers' own digital archives. As a general rule, each newspaper editorial, related to the policy juncture, was coded as a single individual 'event'. The coder coded the editorials as separate events, following the specific coding rules stated in the codebook. Only those editorials in which a clear position on the specific policy area is stated (*i.e.*, a prescriptive position: the govt/govts should do this or do that, etc.) was coded.

Coding of events was conducted electronically using a **Main Data Matrix** Excel file specific for each policy juncture and country. In the file coders could enter their codes in a template

[4] An editorial, leading article (UK), or leader (UK) is an opinion piece written by the senior editorial staff or publisher of a newspaper. Editorials may be supposed to reflect the opinion of the newspaper. An editorial often has no specific author stated. In Australian and major US newspapers, editorials are often classified under the heading "opinion".

with data validation rules to ensure consistency across matrices and minimise typos. Each case was typically coded by one coder only (unless in the cases of unexpected early departures), but to ensure reliability a member of the core team was in charge of quality controlling all the documents and datasets produced by them. In order to ensure that all supervisors followed the same steps in the quality control process the instructions were gathered in the file **Guidelines for coding quality control**. If discrepancies between the codebook instructions and the coding produced were found then the coder was asked to justify her/his choices, and if he/she could not properly justify them to change the coding accordingly.

Coder selection

Coders were recruited online via advertisements on the ResponsiveGov website[5], but also on social networking platforms and through advertisements sent to country experts and to universities of the countries to be coded.

Essential criteria for the selection of coders were:

- Fluency in the respective language of coding;[6]
- Knowledge of the political system of the country to be coded;
- Interest in the policy areas to be coded;
- Ability to work with spreadsheets;
- Reliability and organizational skills;
- Time availability.

In most cases, the coders selected were national citizens of the country they were coding. In a few instances citizens from other countries were selected if they met specific requirements: being a fluent speaker of the language(s) required and particularly knowledgeable of the political system of the country in question. Prior experience in coding was not a requirement because all coders participated in a training session. A first selection was based on a screening of the application (CV and cover letter) aimed at selecting candidates for the individual formal job interviews. All interviews were conducted by two members of the core research team, and consisted of the same set of questions aimed at testing the essential requirements. Once all candidates were interviewed a final selection was made by core team members, and if candidates accepted the position, working procedures and payment were explained.

Coder training

Coding was conducted by trained and supervised coders that had to take part in an intensive training session. Once the practical and bureaucratic matters were completed successfully

5 Available at http://www.reponsivegov.eu.
6 For some countries (Belgium and Switzerland) knowledge of two languages was requested in order to cover news sources and policy documents in both languages.

selected coders were assigned to a supervisor, a member of the ResponsiveGov core team, and asked to read in detail the codebook and all the appendices made available on a file hosting service. In order to ensure homogeneity of training across coders and that all necessary steps were followed, precise instructions for supervisors were listed in the **Coder training overview** file. After the coder carefully went through all project documents, a briefing session was scheduled (online or face-to-face according to the location of the coder) between the supervisor and the coder. The briefing consisted of a session (of duration ranging between one and a half to two hours) during which the supervisor carefully explained at great length the codebook and appendices content and covered step by step all the coding process following a presentation sent in advance to the coder (**Briefing for coders** Powerpoint file). Eventually coders' questions were answered, ambiguities clarified, and doubts addressed.

India pre-test

The next step consisted in the India pre-test, a test divided in four steps aimed at testing the reliability of the coder in all different stages of the coding process (identifying the government initial policy position, identifying the relevant dates and newswires to be coded, and finally coding a sample of newswires). The post-Fukushima nuclear energy policy juncture in India was selected as the case to be coded. India was selected as the country for conducting the pre-test for two main reasons: availability of newswires and documents in English, a language required for all coders; and the assumption that none of the coders were particularly familiar with the politics of India and hence all equally disadvantaged in this regard. Coding for the India pre-test was conducted electronically using the same Excel spreadsheets used for the real coding. All coders participated in the India pre-test in order to assess the individual and overall reliability of the data collection process. Data from the India pre-test were also used to establish whether the coder needed more training, or in some (few) cases to be dismissed. The next sections provide a detailed description of each of the four steps.

<u>Step 1: Coding the policy position of the government of India</u>

The supervisor sent to the coder the instructions of Step 1 and a deadline. Step 1 consists in finding the initial policy position of the Indian government on the nuclear energy policy prior to the Fukushima nuclear disaster. Following the instructions in Appendix 1 and the specifications relating to the Fukushima case, coders had to code the Indian government's policy position on nuclear energy prior to Fukushima nuclear disaster. Coders had to search on the internet for a policy document specifying the position of government toward nuclear power policy published between the electoral campaign of the latest general elections prior to Fukushima disaster and before the actual disaster itself, following the hierarchy of sources described in Appendix 1. Then they had to fill out the initial policy position in the Excel file Appendix 1a and send it back to the supervisor.

The objective of Step 1 is to test the ability of the coder to find the initial government policy position and to correctly classify its policy direction and saliency. For this step we look at the reliability of four variables that are important when determining the initial policy position of a government: salience of policy; page of the manifesto/document in which the policy is addressed; government policy position; and document used as a source for the government policy position.

Step 2: Specifying the time period during which news covering the post-Fukushima nuclear energy policy juncture appeared in the Indo-Asian News Service database

Once coders successfully ended Step 1, they were sent the instructions for Step 2. Step 2 consists in specifying the time period during which newswires covering the post-Fukushima nuclear energy policy juncture appeared in the Indo-Asian News Service accessed through Factiva. Using the post-Fukushima related keywords – Appendix 2 - coders needed to build a search on Factiva search tool following the instruction and noting down the range of dates within which they found news stories describing relevant events. Through the University of Leicester library website they were asked to access Factiva[7] and do the following:

1. Select Search - "Search builder";
2. Specify "Indo-Asian News Service" as the source;
3. Specify "India" as the region;
4. Select "enter date range" and enter 11/03/2011 as the start date and 06/02/2013 (date of elections) as the end date;
5. Turn duplicates off.

Once set up the Factiva search builder as mentioned above, coders had to enter the search terms from the keyword dictionary (Appendix 2) to look for news stories that are related to nuclear energy in India (not in other countries). Finally they reported back to their supervisor the range of dates (starting date and ending date) during which relevant stories of the nuclear energy policy juncture in India appeared.

The objective of Step 2 is to test the ability of coders to access and search independently the newswires database, and to correctly identify the relevant time period for the coding of meaningful events. With this step we look at the reliability of two variables, the starting and the ending dates of relevant newswires.

Step 3: Identifying the news stories that ought to be used for coding the post-Fukushima nuclear energy policy in India

[7] Available at https://www2.le.ac.uk/library/find/databases/f .

Step 3 consisted instead in identifying the news stories that ought to be used for coding the post-Fukushima nuclear energy policy in India between 01/09/2011 and 31/10/2011. Following the instructions provided by the supervisor, coders had to access Factiva and build up a search in the Indo-Asian News Service using the Appendix 2 keywords and the specified range of dates. Then they had to select all newswires describing events that should be coded and discard others. Next, they had to write down the title and date of each news story they found that is relevant to the nuclear energy policy juncture in India in an Excel file and email it to their supervisor. Coders had to select any news wire related to nuclear energy policy as described in Appendix 1.

The objective of Step 3 is to test the ability of coders in identifying all newswires that might contain information on relevant events for the policy juncture. As a measure of reliability for this step we compare the number of news stories identified by the coder, with the number of "true" news stories determined by the ResponsiveGov core team. An Excel file listing all the newswires that need to be listed by coders because they pertain to the nuclear policy juncture in India was produced by the core team for this purpose through deliberation.

Step 4: Coding the post-Fukushima nuclear energy policy in India

Step 4 consisted in coding the post-Fukushima nuclear energy policy in India between 07/09/2011 and 16/10/2011. Before starting to code coders were instructed to read again carefully all project documents (codebook and appendices) and to start only when they felt comfortable finding their way around them. The supervisor sent to the coder a file with all the newswires in the range of dates specified above from the Indo-Asian News Service, and the template of the main data matrix Excel file to use for the coding. After coding the government's initial policy position as the first case and marking it as the "start of the juncture", then they had to continue by coding the events that they were able to identify in the news stories provided. Once they had coded the first 15 rows, they had to send back to their supervisor the data matrix file for an initial inspection of their coding and to verify that there were no major issues of understanding. If the coding was basically fine they were invited to go on with the coding. Instead, if major issues of understanding were found, the supervisor would have sent back to the coder detailed feedback on how to address such problems and invited coders to check again relevant parts of the codebook or its appendices before continuing with the coding. In a few cases the supervisor had to set up a meeting in order to better clarify uncertainties of coders. Once the coder finished coding Step 4, the supervisor was in charge of performing again an inspection on the whole data matrix to check for errors or misinterpretations of the codebook. In case mistakes were found and additional training required, the supervisor would have set up a meeting with the coder to give her/him additional and focused training on the variables that had shown to be problematic.

The objective of Step 4 is to test the ability of coders in performing the coding of the main data matrix for a policy juncture, and therefore we look at the reliability of all the following variables:[8]

- V5a: Day of the event
- V5b: Month of the event
- V5c: Year of the event
- V11: Type of the event
- V13a: Location of the event
- V13c1: First addressee of the event
- V13c2: Second addressee of the event
- V13d1: 1st level of government influenced
- V13d2: 2nd level of government influenced
- V13e1: Number of participants at protest
- V13e1Source: Source for number of participants
- V13f: Duration of protest
- V14a: Type of actor general
- V14b :Type of actor specific
- V14d: Position of actor relative to government
- V14d1: Direction and intensity of position
- V14e: Policy position made explicitly by the actor
- V14f: Certainty of policy position
- V18a: Event triggered change in saliency
- V18b: Intermediate change in government position

For Step 4 in order to compare results provided by coders with true values, an Excel file, **True Scores**, listing for each one of the above variables the true values for the twenty most important events to be coded for India between 07/09/2011 and 16/10/2011 was produced by the core team. The twenty most important events were chosen by the core team through deliberation in order to best capture the policy dynamics during the period of interest (see Table 1). To determine what the true scores are these rules were followed:

- If there was unanimous agreement on a specific code among the data matrices of the core team members in Step 4, that code was selected;
- If there was no unanimous agreement, the core team discussed the specific events and through deliberation the 'true value' was determined.

[8] A more detailed description of these variables can be found in the codebook.

Table 1. List of the twenty most important events to be coded in Step 4.

Nr	Event description	Nr	Event description
1	Nuclear Safety Regulatory Authority Bill in the Lok Sabha	**11**	National Green Tribunal claim (R.Dutta)
2	Protests (demonstration) for three days	**12**	Shiv Sena president (U.Thackeray) statement on Jaitapur Nuclear Project
3	Verbal statement during demo above	**13**	Government meeting with protesters and administrators pledge to constitute group of experts
4	Over 125 people fasting protest	**14**	Statement CPI-M Communist party D.Raja
5	Statement By MDMK leader Vaiko supporting protester	**15**	Restart of Hunger strike
6	Letter by Jayalalithaa to Prime Minister	**16**	PM Singh letter to Jayalalithaa
7	Narayanasamy meets protesters	**17**	Blockade
8	Udaykumar (People MovAgainst Nuclear Energy) statement: protests stops if demands are met	**18**	Interview by S.K. Jain, chairman of NPCI: safeguard employees and project site
9	S.C. Chetal, (Indira Gandhi Centre for Atomic Research) no danger	**19**	Narayanasamy, Minister of State statement urging protesters to negotiate
10	Resolution by Tamil Nadu government asking Prime Minister to halt work	**20**	Sivasubramanian: protest called off for local elections.

Fukushima case coding summary

Coding for the internet policy juncture was conducted by a total of 21 coders as summarised in Table 2. Some coders (ID number: 9, 13, and 18) only coded preparatory documents (government identification period, Appendix 1a, Appendix 5c, and Appendix 2) for their country and left the project before working on the main data matrix.

Table 2. List of coders involved

Coder ID	Fukushima Case(s) Coded
1	ES
2	IT; UK; US
3	CA; CH; DE
4	BE; FI; NL
6	SE
7	FR
21	EU
25	EU

Inter-coder reliability tests

In order to assess the reliability of the data collected we resolved to select a variety of measures given the nature of our data. For step 1 and 2 we assess reliability with a measurement of average pairwise percent agreement, the percentage of all coding decisions made by pairs of coders on which coders agree. Values for this measure range between *0* no agreement, to *1* perfect agreement. It is a simple and intuitive index whose major weakness is the failure to account for chance agreement (Lombard *et al.* 2002). We could not apply different metrics given that the initial three steps of the India Pre-test are based on only one observation for each coder.

In order to evaluate the reliability of the salience variable we used a binary variable coded 1 if salience (measured as the number of words about the relevant policy divided by the total number of words of the policy document - often an election manifesto) reported by coders is equal to 0.02 ± 0.01 (average salience across coders is 0.0241 with standard deviation 0.01). Five out of eight coders reported it correctly, with an average percent agreement of 46.4%. Given that the result is not satisfactory supervisors of coders systematically verify how coders computed the score and double check to provide a more precise measurement.

The second relevant variable tracks the manifesto/document page on which the issue topic is discussed. It is an ordinal variable going from *1* to *4* (1 – First Page; 2 – On pages 2-3; 3 – On a page in the first half of the election manifesto, but not on pages 1-3; 4 – On a page in the second half of the election manifesto), and returned a poor average percent agreement of 46.4%. Nevertheless, it is true that sometimes coders got confused about what is a first page (e.g., whether the document cover counts or not) and the same manifesto is made available in different formats. Therefore, we also assessed the inter-coder reliability for this variable when recoded into a binary category (1&2 = 0 vs 3&4 = 1), and in this way the level of agreement increases dramatically (*90.5%*) given that only one coder coded it differently from others. In any case, the results of the inter-coder reliability of this variable has led to instruct coder supervisors to devote special attention to it while checking the quality of Appendix 1a.

Regarding the variable registering governments' policy position, instead, we have an average pairwise percent agreement of 100%, being correctly reported by all coders. Finally, for step 1 of the coding process, also the source of the government's initial policy position has been reported correctly by all coders with an agreement of 100%, meaning that all coders have been able to correctly follow the instructions in Appendix 1 and find the correct source.

Step 2 consists in finding the dates of the first and the last relevant Indo-Asian news agency newswires for the period between the 11/03/2011 and the 06/02/2013. In the first case there is a 100% agreement among coders and 75% in the second case if we allow for an

uncertainty range of up to two days over a two years period. Instead if we allow an uncertainty range of only one day the percentage of agreement shrinks to 75% for the initial date while it remains the same for the final one. If instead we consider the original date submitted by coders without any uncertainty the percentages of agreement become meaningless given that the majority of coders got it wrong in both cases.

For step 3 the average pairwise agreement could not be used meaningfully, and therefore we relied on two measures: the percentage of true stories listed by the coder and the average of the percentages of true stories for all coders who coded the nuclear policy juncture. Coders identified on average 95% of the 74 news stories to be identified, and the coder who identified the lowest number of newswires related to the nuclear energy policy in India post Fukushima disaster had a score of 77% (60 newswires out of 74). All but one coder have identified at least 90% of the 'true' newswires they were expected to detect, and the coder with the lowest value has been specifically retrained in order to improve his reliability. Table 3 summarises the results for Steps 1, 2, and 3 of the India pre-test for all coders that participated in data collection for the nuclear policy juncture.

Table 3. Summary of steps 1, 2, and 3 for Fukushima case coders. The table reports for each step original values provided by coders.

		Fukushima Case Coders ID							
Step	**Variable**	**2**	**3**	**7**	**1**	**4**	**6**	**21**	**25**
1.0	Salience	0.023	0.023	0.040	0.008	0.025	0.016	0.019	0.039
1.1	Page	4	3	3	4	4	4	4	3
1.2	Govt Policy Position	2	2	2	2	2	2	2	2
1.3	Source	INC Manifesto	INC Manifesto	INC Manifesto	INC Manifesto	INC Manifesto	INC Manifesto	INC Manifesto	INC Manifesto
2.1	Initial Date	12/03/2011	12/03/2011	11/03/2011	12/03/2011	12/03/2011	12/03/2011	12/03/2011	14/03/2011
2.2	Final Date	05/02/2013	05/02/2013	06/02/2013	05/02/2013	05/02/2013	23/01/2013	05/02/2013	06/02/2013
3.0	Total news stories reported	89	83	97	87	85	60	113	85
3.1	PCT True Stories identified	100%	100%	97%	100%	100%	77%	98%	90%

1 N.A. for coders not considered for step 3 and 4 because they worked only on preparatory documents and did not code the main data matrix for their country.

Summary of Inter-coder reliability scores for Step 1 of the India coding test

1.1 Salience of policy issue

Coded 1 if salience = 0.02 ± 0.01

Average Pairwise Percent Agreement: 46.43%

Average salience across coders: 0.0241

Standard Deviation: 0.01

1.2 Manifesto Page:

Original codes: 1 – First Page; 2 – On pages 2-3; 3 – On a page in the first half of the election manifesto, but not on pages 1-3; 4 – On a page in the second half of the election manifesto

Average Pairwise Percent Agreement: 46.43%

If recoded considering 1&2 = 0 vs 3&4 = 1:

Average Pairwise Percent Agreement: 100%

1.3 Government Policy Position:

Original codes: -2 Very progressive, -1 Progressive, 0 Neutral, 1 Conservative, 2 Very Conservative.

Average Pairwise Percent Agreement: 100%

If recoded considering Progressive = -2 and -1; Neutral = 0; and Conservative = 1 and 2:

Average Pairwise Percent Agreement: 100%

1.4 Source of the government initial position:

Original codes (1 if 2009 election manifesto of India National Congress used)

Average Pairwise Percent Agreement: 100%

Summary of Inter-coder reliability scores for Step 2 of the India coding test

2.1 Initial Date of coding period: code 1 if initial date is 12/03/2011 ± 2 days

Average Pairwise Percent Agreement: 100%

2.2 Final Date of coding period: code 1 if final date is 05/02/2013 ± 2 days

Average Pairwise Percent Agreement: 75%

Summary of Inter-coder reliability scores for Step 3 of the India coding test

3.1 Average Percentage of 'True' Stories Identified

95% (Range = 77% - 100%)

For Step 4, the most comprehensive one, we instead resorted to Krippendorff's Alpha index (Krippendorff 1980). This index has become a sort of standard in reliability tests because of its features: it allows for multiple coders; can be used with variables at different levels of measurement (nominal, ordinal, and interval/ratio); it deals successfully with missing values; and it accounts for chance agreement (Krippendorff 2004a).

$$Krippendorff's\ Alpha = 1 - \frac{Observed\ Disagreement}{\text{Expected Disagreement}}$$

The observed disagreement indicates the mismatch between coders in the values attributed to the same observations. Dividing by the expected disagreement corrects by the chance that coders code a unit rightly just by chance. This is calculated based on the amount of different values coded in the reliability subsample.[9] The more different values are, the lower the chance to randomly code it rightly. To account for this, Krippendorff's Alpha takes into account the prevalence of the categories coded for the variable that is tested. Therefore a major problem of Krippendorff's Alpha index is that in the case of binary variables with one of the two values (1 or 0) being very rare it will return low scores even with few mistakes from coders (De Swert 2012).

Given that this is the case for some of the variables (e.g., variables V14e – Actor's position is EXPLICITLY made by the actor, V18a – Change in saliency of issue for government, and V18b – Change in government's position) we also computed the percent of true scores for each variable (for the twenty events we counted how many times a variable was coded correctly comparing coders' values with those in the **True scores** file and then dividing by the number of times the variable was coded) and coder (for the twenty events we counted how many times coder's results matched those in the **True scores** for all variables and then we divided by the number of times the variables were to be coded) in order to gain better insights on the internal consistency of both of them.

Following Krippendorff's (2004b) recommendations in case of unsatisfactory levels of agreement across coders we decided to discard unreliable distinctions (re-aggregation of categories), or to drop variables. Krippendorff suggests to require an alpha level >= 0.8 in order to be able to provide valid inferences from the data; and if tentative conclusions are still acceptable an alpha level ≥ 0.667.

Table 4 summarises the reliability scores for the most relevant and non-trivial variables. The third column reports the alpha values for the original scores provided by coders without doing any data manipulation or aggregation. We can easily notice that the most important variables – date of event, type of event, location of event, duration of protests, general type

[9] Hence, it is not corrected by the amount of all possible categories in the codebook, but considers only those actually used by coders.

of actor (v14a), policy position relative to the government's, and direction and intensity of policy position – return all a satisfactory degree of reliability. It is worth noting that both the type of event (v11) and the specific type of actor (v14b) improve significantly when we recode them into the broader categories at the one digit level instead of at the three-digit level. Moreover, the policy position relative to the government's improves when recoded and simplified from a 5-categories variable into a 3-categories ordinal variable.[10]

Variables with low reliability scores are those registering the addressee (v13c1 and v13c2), the level of government influenced (v13c1 and v13c2), and the level of uncertainty of the first actor's policy position (v14f). Still if we look at Table 5, which reports the percentage of 'true scores' values coded by coders for each variable, the first addressee of the event has a value of 79% (and the second addressee of 53%), while the first level of government influenced was coded correctly 88% of the times (the second level of government reaches 64%) and the certainty of the policy position (v14f) achieves 83%. These levels of intra-variable agreement with the true scores seem to justify the use of at least the first addressee, first level of government influenced, and level of uncertainty for exploratory purposes. Moreover, it is useful to remember that the inter-coder reliability testing is not subjected to the quality control process that is compulsory for real coding outputs, and during the quality control process coders' supervisors control each variable.

The most problematic variables seem to be v14e - Policy position made explicitly by actor, v18a - Event triggered a change in saliency, and v18b - Intermediate change in government position as they return very low Krippendorf alpha values. Here we are facing a structural problem due to the nature of the variables and the Krippendorff alpha statistic. Indeed these three variables are binary with one of the values (1 or 0) being very rare, and in this case Krippendorff's alpha returns low values even with few mistakes from coders (de Swert 2012). To get a better sense of this structural problem we can compare the fourth column of Table 4 reporting Krippendorff's alpha for the same data recoded into dummies (1 if the coder reported the true score, 0 otherwise). If we look at interval variables with a very high alpha score (due to lack of variance between coders) in the third column - such as for v5a, v5b, or v5c - and compare it with the alpha for the binary values in the next column of Table 4, we notice a very low level of the latter even though these variables were coded correctly more than 90% of the times (see Table 5). Therefore for variables v14e, v18a, and v18b it seems more informative to look at the percentages in Table 5. It is also worth noting that the intra-variable percent agreement with true scores considers missing values as mistakes, while Krippendorff's alpha excludes them from the analysis (this is clear when looking at the alpha of v5c, year of event, which is 1 while the percent agreement is 96%), making it a restrictive measurement even if it does not account for chance. In Table 5 the percentages of v14e, v18a, and v18b seem to provide a satisfactory level of reliability.

[10] The recoded variable takes values of -1 (No: actor 1 explicitly opposes governmen); 0 (actor 1 is a priori neutral OR actor 1 doesn't support or oppose government); and +1 (Yes – actor 1 supports government).

Also in Table 5 the last row reports the average percent of 'true' scores for each variable (excluding more trivial variables: month and year of the event) which is a good 81%, with a minimum of 53% for the second addressee of the event (v13c2), variable that should therefore be handled carefully.

Finally, Table 6 reports the percent of true scores for each coder, both including and excluding more trivial variables. Looking at the third column, the most restrictive one, we have an average of 81% of true values reported by coders with a range between 74% and 91%.

Here it is important to reiterate that these are the first results produced by coders, and according to their specific mistakes or misconceptions about the variables after the India pre-test they receive a customized training in order to improve their reliability. This implies that most likely the actual individual level of reliability for each coder is greater than the one measured during Step 4. In addition all documents and data matrices compiled by coders are quality controlled by one member of the ResponsiveGov core team. The quality control procedure (see Guidelines for coding quality control document) consists in checking the consistency of all variables to ensure comparability across policies and countries. In case the supervisor of the coder faces problematic instances, then the latter are discussed and resolved through deliberation by the core team.

Table 4

Krippendorff's Alpha scores computed for selected variables on the India pre-test 20 events for the 8 Fukushima case coders.

		Fukushima Results Step 4	
		20 events and 8 coders	
Var Number	**Description of variable**	**Krippendorff Alpha (method)**	**Krippendorff's Alpha on True Scores binary variable**
v5a; 1	**Day of event**	0.989 (interval)	0.066 (nominal)
v5b; 2	**Month of event**	1 (interval)	0.115 (nominal)
v5c; 3	**Year of event**	1 (interval)	0.115 (nominal)
v11; 4	**Type of event**	0.692 (nominal); 0.824 (nominal, using first digit)	0.254 (nominal)
v13a; 5	**Location of event**	0.710 (ordinal)	0.127 (nominal)
v13c1; 6	**1st Addressee**	0.351 (nominal)	0.311 (nominal)
v13c2; 7	**2nd Addressee**	0.135 (nominal)	0.017 (nominal)
v13d1; 8	**1st Level of govt influenced**	0.347 (ordinal)	0.010 (nominal)
v13d2; 9	**2nd Level of govt influence**	0.112 (ordinal)	0.278 (nominal)
v13e1; 10	**Number of participants**	0.954 (interval)	0.320 (nominal)
v13e1source	**Source for nr of participants**	0.825 (nominal)	0.381 (nominal)
v13f; 12	**Duration**	0.029 (interval)	0.362 (nominal)
v14a; 13	**Type of Actor General**	0.743 (nominal)	0.239 (nominal)
v14b; 14	**Type of Actor Specific**	0.559 (nominal); 0.806 (nominal) with first digit	0.223 (nominal)
v14d; 15, 21	**Policy position Relative to Government**	0.723 (nominal) 0.888 (if recoded as ordinal)	0.221 (nominal)
v14d1; 16	**Direction and intensity of Position**	0.812 (ordinal)	0.109 (nominal)
v14e; 17	**Policy position made explicitly by actor**	0.068 (nominal)	0.206 (nominal)
v14f; 18	**Certainty about policy position**	0.025 (ordinal)	0.373 (nominal)
v18a; 19	**Event triggered a change in saliency**	0.443 (nominal)	0.267 (nominal)
v18b; 20	**Intermediate change in government position**	-0.021 (nominal)	0.070 (nominal)

Table 5

Intra-variable percent agreement with True Scores for the 20 events of the India pre-test coded by the 8 Fukushima case coders.

Fukushima Results Step 4

Variable	Intra Variable Percent Agreement with True Scores
V5a Day of event	91%
V5b Month of event	96%
V5c Year of event	96%
V11 Type of event	81%
V13a Location of event	73%
V13c1 First addressee of event	79%
V13c2 Second addressee of event	53%
V13d1 1st level of government influenced	88%
V13d2 2nd level of government influenced	64%
V13e1 Number of participants at protest	93%
V13e1Source Source for number of participants	94%
V13f Duration of protest	91%
V14a Type of actor general	84%
V14b Type of actor specific	69%
V14d Position of actor relative to government	86%
V14d1 Direction and intensity of position	79%
V14e Policy position made explicitly by the actor	77%
V14f Certainty of policy position	83%
V18a Event triggered change in saliency	77%
V18b Intermediate change in government position	91%
Average	82%
Range	53% - 96%
Average agreement with True Scores within variables (excluding month, year)	81%

Table 6

Percent agreement with True Scores of the 20 events of the India pre-test for each of the 8 Fukushima case coders.

Fukushima Results Step 4

Coder ID	Percentage Agreement with True Scores	Percentage Agreement with True Scores (excluding month, year)
1	87%	86%
2	84%	83%
3	86%	84%
4	92%	91%
6	76%	74%
7	80%	77%
21	78%	76%
25	75%	74%
Average	**82%**	**81%**
Range	**75% - 92%**	**74% - 91%**

References

De Swert, K. (2012). Calculating inter-coder reliability in media content analysis using Krippendorff's Alpha. *Center for Politics and Communication*, 1-15.

Krippendorff, K. (2004a). *Content analysis: An introduction to its methodology. Second Edition*. Thousand Oaks, CA: Sage.

Krippendorff, K. (2004b). Reliability in Content Analysis: Some Common Misconceptions and Recommendations. *Human Communication Research, 30* (3), 411-433. Available at: http://dx.doi.org/10.1111/j.1468-2958.2004.tb00738.x

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication research: An assessment and reporting of intercoder reliability. *Human Communication Research, 28*, 587-604.