



## Rules of validation

There are two points of decision during the validation process. First, the admin of the CAP Babel webpage decides whether the uploader had previously given a proof of his/her knowledge regarding CAP coding. If the admin's decision is negative, then the uploader has to fill a tests coding of 200 sentences. In the case of at least 150 correct answer, the uploader's validation can be accepted. If the uploader fell short to reach this threshold, we suggest him/her to try the test again after further investigation and learning of the [CAP Policy Topics Codebook](#).

After a successful test coding, the next steps depend on whether the uploader has prepared and uploaded a proper pre-validated dataset or not. If the answer is yes, it is automatically compared to the automatically validated full dataset, resulting the intercoder reliability for the uploader. The uploader can also download immediately after the machine coding process the whole dataset.

If no pre-validation has been done (or the number pre-validated observations was lower than the required amount), the uploader can receive only a randomly selected sample equal to the one required during the pre-validation process of the dataset for a further post-validation. If the user uploads the post-validated sample, he/she gets the full dataset with the intercoder reliability index.

## Requirements for pre-validation

The pre-validated dataset should be randomly selected from the whole dataset. It should have identical structure to the uploaded original dataset, except that the variable major topic should be filled. The pre-validated dataset should include at least 10% of the full dataset's observation, but at least 500 ones.