Formative Study Process. We conduct the study using the 107 projects. Table 1 shows summary characteristics of these 107 projects.

For the first study question, we use TraceMOP to generate a map from unique traces to their corresponding tests. We then calculate how many tests produce each unique trace. For the second study question, we use the trace-to-test map from the first study question to find essential tests. To answer the third study question, we run the essential tests with JavaMOP and compare their end-to-end time with running all tests with JavaMOP, then we compute the ratio of the two. Lastly, for the fourth study question, multiple authors together inspect a sample of 500 traces collected with TraceMOP.

The raw data are available in our artifact repository (in data/formative-study directory).

Table 1. Summary statistics on 107 evaluated projects in our study: no. of test methods (#tests), test time w/o RV in seconds (t), lines of code (SLOC), % statement coverage (cov^s), % branch coverage (cov^b), no. of GitHub commits (#SHAs), years since first commit (age), and no. of stars (#★).

	#tests	t	SLOC	cov^s	cov^b	#SHAs	age	#★
Mean	84.3	10.9	7,883.7	59.2	50.6	326.3	10.0	140.1
Med	34	3.8	3,496	60.7	52.3	162	10	45
Min	1	1.6	82	0.0	0.0	3	3	0
Max	1,131	110.6	1.3×10^5	100.0	100.0	4,259	22	2,640
Sum	9,024	1,165.1	8.4×10^5	n/a	n/a	n/a	n/a	14,996