Validation statistics over multiple subsets (out-of-sample, lower is better).