

ESS as an Index of Decision Consistency

Paul R. Yarnold, Ph.D.

Optimal Data Analysis, LLC

Decision consistency is "...appropriate for describing the degree of consistency that is realized when educational and psychological measures are used to make pass/fail decisions about examinees".¹ In this method a test or parallel form is administered twice and the pass/fail decision at each testing is cross-classified. The fraction (or percent) of overall agreement, and kappa, are used to index decision consistency, but both of these statistics have limitations overcome by the chance- and maximum-corrected *ESS* statistic.

Statistical limitations of the fraction of overall agreement and of kappa as indices of inter-rater, inter-device, and test-retest reliability, and of the comparative efficacy of *ESS* in this context, are discussed elsewhere.² Table 1 gives an example previously used¹ to illustrate these indices.

Table 1: Example for Computation of Decision Consistency¹

	<u>Second Test</u>	
<u>First Test</u>	<u>Fail</u>	<u>Pass</u>
Pass	5	75
Fail	5	15

For this example the fraction of overall agreement = kappa = 0.80, suggesting that the two test results agree very strongly. However, it seems inappropriate to consider the consistency of the two tests to be very strong since there was disagreement for half of the failure decisions made by the second test, and for three-fourths of the failure decisions made by the first test.

Treating the result of the second test as the class variable and the result of the first test

as the attribute, the associated sensitivity for fail (50.0) and pass (83.3) correspond to a moderate level of agreement, *ESS* = 0.33, which indicates that the observed consistency between the two tests corresponds to one-third of the possible increase in consistency (agreement) beyond what is expected by chance in this application. Switching the roles of the first and second test here yields a lower estimate of decision consistency: *ESS* = 0.19.

References

¹Traub RE (1994). *Reliability for the social sciences: Theory and applications* (Vol. 3). Thousand Oaks, CA: Sage (pp. 140-146).

²Yarnold PR, Soltysik RC (In Review). *Maximizing predictive accuracy*. Chicago, IL: ODA Books.

Author Notes

The study analyzed de-individuated data and was exempt from Institutional Review Board review. No conflict of interest was reported.