Increasing Levels of Inference, based on McGill et al. (2006); also see Platt (1964)

A. Single theory test. Are empirical data consistent with or contradictory to a theory?
    1. Correct shape approach. The model displays the same general shape or relation as empirical data.
    2. Curve-fitting approach.—The model fits the empirical data well when the parameters are chosen via curve fitting.
B. Null hypothesis test. Does the theory fit empirical data better than a null hypothesis (H_0) representing a simple scenario.
    1. Hypothetico-deductive approach.—The model fits the empirical data significantly better than H_0 after penalization for number of parameters.
    2. *A priori* parameters approach. The model fits the empirical data significantly better than H_0 after penalization for number of parameters and when the model parameters are chosen independently of the empirical data.
C. Multiple, complex predictions approach.—A single model is tested using multiple *a priori* predictions. These predictions are more complex than the data-fitting predictions tested in levels A and B. Two examples of complex predictions are predictions of correlations and predictions about dynamic processes. Each of the complex predictions is then tested at least at level B (i.e., against an appropriate H_0).
D. Model comparison test.—Realistic alternate models (as opposed to null hypotheses) are contrasted against each other, consistent with Platt (1964).
    1. Best theory approach.—The alternative models are ranked according to their match to empirical reality according to some score (such as r^2), and the best model is selected.
    2. Last standing approach.—Rigorous attempts are made to falsify all models in a Popperian fashion until only one model remains unfalsified, which is then accepted as the best model (Platt 1964).
    3. Model weighting approach.—This level not only involves multiple realistic theories, but assigns weights to them according to their explanatory or predictive power. The classical analysis of variance and partitioning of sums of squares is a linear example of such a technique. Akaike weights achieve a similar but not identical result (Burnham and Anderson 1998).
    4. Conditional weights approach.—This approach identifies how the model weights of D3 depend on the scale or context. This level of model comparison answers the question, ‘‘under what conditions is one model better than another in explaining the data?’’

McGill, BJ, BA Maurer & MD Weiser. 2006. Empirical evaluation of neutral theory. Ecology 87:1411–1423.

Platt, JR. 1964. Strong inference. Science 146:347–353.