

# A protocol for modelling generalised biological responses using latent variables in structural equation models

James Benjamin Grace<sup>‡</sup>, Magdalena Steiner<sup>§</sup>

<sup>‡</sup> U.S. Geological Survey, Lafayette, United States of America

<sup>§</sup> University of Fribourg, Fribourg, Switzerland

Corresponding author: James Benjamin Grace ([gracej@usgs.gov](mailto:gracej@usgs.gov))

Academic editor: Gbenga Akomolafe

## Abstract

In this paper, we consider the problem of how to quantitatively characterise the degree to which a study object exhibits a generalised response. By generalised response, we mean a multivariate response where numerous individual properties change in concerted fashion due to some internal integration. In latent variable structural equation modelling (LVSEM), we would typically approach this situation using a latent variable to represent a general property of interest (e.g. performance) and multiple observed indicator variables that reflect the specific features associated with that general property. While ecologists have used LVSEM in a number of cases, there is substantial potential for its wider application. One obstacle is that LV models can be complex and easily over-specified, degrading their value as a means of generalisation. It can also be challenging to diagnose causes of misspecification and understand which model modifications are sensible. In this paper, we present a protocol, consisting of a series of questions, designed to guide the researchers through the evaluation process. These questions address: (1) theoretical development, (2) data requirements, (3) whether responses to perturbation are general, (4) unique reactions by individual measures and (5) how far generality can be extended. For this illustration, we reference a recent study considering the potential consequences of maintaining biodiversity as part of agricultural management on the overall quality of grapes used for wine-making. We extend our presentation to include the complexities that occur when there are multiple species with unique reactions.

## Keywords

generalised responses, latent variables, multivariate responses, statistical modelling, structural equation modelling

## Introduction

The quest for generalisation in the ecological sciences is a fundamental challenge. One way that a general reaction by a system or organism can be detected is if there is a multivariate response where numerous individual properties change in concerted fashion. While such concerted reactions are often described using standard multivariate statistical analyses, causal investigations of the nature of integrated multivariate responses fall primarily into the purview of latent variable structural equation modelling (LVSEM, Joreskog 1970, Bollen 1989). LVSEM represents a foundational method in quantitative training in the fields of psychology and sociology, while other fields, including ecology, have historically focused on observed variable models (e.g. OVSEM, aka path analysis). While there has been an increased use of LVs in ecological models in the past 20 years (e.g. Arhonditsis et al. 2006, Shipley et al. 2006, Cubaynes et al. 2012, Liu et al. 2016, Souchay et al. 2018), only a few descriptions of the methods have emerged (Pugesek et al. 2003, Grace 2006, Shipley 2016). More importantly, the depth of coverage of the subject in ecological treatments is much less than in social science treatments, leaving untapped the full potential for LVSEM to advance ecological understanding. In this paper, we present a protocol designed to guide ecological researchers through the evaluation of hypotheses about generalised responses using LVSEM.

Fig. 1 provides a high-level perspective of the problem of interest in this paper. The particular form of representation here is referred to as a Structural Equation Meta-Model (SEMM), which represents the conceptual entities of interest in the study and the hypothesised causal connections amongst them (Grace et al. 2010). SEMMs are meant to solve a ubiquitous problem in statistical modelling, which is that our theories exist at a general level, while our fully-specified statistical models are highly specific. SEMMs are meant to facilitate an explicit linkage between our theoretical ideas and our specific findings. Typically, this step in the science process is implicit and, thus, the question of how our results generalise goes undescribed (and presumably only lightly considered).

The hypothesis represented in Fig. 1 is very general. It simply posits that the response of a system to perturbation can be understood through an evaluation of the mediating mechanisms. In this paper, a primary emphasis is on the nature of the response. We are especially interested in the degree to which the subjects of study exhibit a tightly integrated suite of changes. Several theoretical possibilities exist. In highly integrated responses, multivariate reactions will manifest strong and consistent correlations amongst measured indicator variables. At the other extreme, non-integrated responses will manifest reactions, such that each measured property acts independently from the others. Between those extremes are cases where there are mixtures of general and individual responses. All of these details are of interest if we are to develop an understanding of the study system and if we wish to properly represent hypotheses that can explain the observed data. That said, a substantial degree of intercorrelation amongst measured indicator variables is required if general responses are expected to conform to the requirements for latent variables.

LVSEM allows us to empirically test hypotheses about the latent responses so as to arrive at a model with both theoretical and empirical support.

Fig. 2 illustrates the LVSEM approach to representing a general multivariate response. Conceptually, Fig. 2 describes the case where there is some underlying cause for the various specific, yet correlated, responses. It is characteristic that the underlying cause is not directly observed. Instead, we learn about the latent properties of the underlying cause through the patterns of correlations amongst specific manifestations. Statistically, Fig. 2 represents a very particular hypothesis. For those not accustomed to causal modelling, an important shift in thinking is required to differentiate one's logic from that associated with descriptive statistical models. For example, we might employ principal components analysis to estimate a set of parameters consistent with Fig. 2. Such parameters would be purely descriptive, a summary of the data and fail to correspond to any particular causal explanation. In LVSEM, the model evaluated is treated as a causal hypothesis, which, in this case, is that the intercorrelations amongst indicator variables can be adequately explained by a single common causal process. All other sources of variation for the indicators are assumed to be independent. There are many ways in which data may fail to confirm this hypothesis, leading us to construct alternative causal hypotheses in order to explain what is going on.

Numerous complexities can be encountered when analysing models containing latent variables with multiple indicators. It is probably safe to say that the available literature may be inadequate for the beginning user of SEM to navigate the various diagnostics and decisions required for such models. Our primary objective in this paper is to provide a series of questions that can guide the investigator through the process. Our advice is targeted for the general objective outlined in Fig. 1 and may have to be supplemented for models with more complex purposes.

## **The Basic Analytic Machinery for LVSEM**

There exist many technical descriptions of the analytical machinery used to implement LVSEM. Here, we provide a non-technical summary and refer the reader to Bollen 1989 for a detailed treatment. A concise presentation of the equations and notation corresponding to our presentation can be found in Suppl. material 1.

Fig. 3 provides a more complete representation of the kind of model we seek to evaluate in this paper. Here, we see that, in addition to a latent variable for the general response and its four indicator variables, our hypothesis includes latent variables for a specific perturbation and a hypothesised mediator variable that explains the effect of perturbation on the general response. The model also proposes that the Perturbation can affect the General Response independently from the Mediator and the Mediator can affect particular responses (e.g. response4) to some degree differently from the other responses.

The classical approach to implementing SEM involves the analysis of covariances. For this, the rows of raw data are converted into a variance-covariance square matrix.

Hypothesised models represent a set of expectations about the patterns of covariances that should be found in data. Typically, covariance modelling estimates the parameters of the causal diagram via maximum likelihood while respecting the assumed causal relationships specified in the causal graph. Covariance SEM also produces a statistic that summarises the differences between the observed covariances and those predicted while agreeing with the model structure and tests the null hypothesis that the observed and predicted covariances are equal, except for random sampling variation. Failure to reject this null hypothesis is evidence that the assumed causal structure is correct.

For LVSEM, model structure is described using equations representing the relationships between latent variables and their indicators and equations describing relationships amongst latent variables (Suppl. material 1). For the model in Fig. 3, software would implement the variables Specific Perturbation and Specific Mediator as latent variables with single indicators in order to be consistent in the use of separate matrices for relationships amongst latent variables and relationships between latent and observed variables (Suppl. material 1, Table S1.1). Variables are described as *exogenous* if they serve only as predictors of other variables (no arrows pointing to them) and *endogenous* if predicted by other variables (possess incoming arrows). In Fig. 3, Perturbation is the only exogenous variable, all others are endogenous. The Greek symbols used to denote various parameters correspond to the matrices used to implement the models. The lambda matrix contains the model-implied weights between latent variables and their indicators, the beta matrix contains the effects of endogenous variables on other endogenous variables and the gamma matrix contains the effects of exogenous variables on endogenous variables. Two additional matrices are the psi matrix, which contains the latent error terms (zeta 1 and 2), as well as their intercorrelations, if such are specified and the theta-epsilon matrix, which includes the errors for indicators (epsilon 1-4) and their intercorrelations, if any. The practical value of these matrices is that they permit a great flexibility in model specification, allowing for error correlations, reciprocal effects and a great many customised specifications.

## **An Ecological Example: The Responses of Grape Nutritive Qualities to Intensity of Agricultural Management**

Steiner et al. 2021 conducted a study of Swiss vineyards managed under different intensities of weed management, which resulted in different levels of non-crop plant biodiversity. Grape qualities of known importance to wine-making (nitrogen, sugars, tartric acid and malic acid) were measured and served as indicators of the general property "Grape Qualities". Features of the spontaneous vegetation that were measured included total cover, total species richness and abundance of nitrogen fixing plants. The empirical measurements corresponding with the concepts of theoretical interest are summarised in Table 1.

The overall study objectives are summarised in Fig. 4. Here, it can be seen that the overall problem was to determine whether management intensity has an influence on grape

qualities and whether such an effect can be explained by hypothesised mediators. One question of interest in this study was whether the measured properties of grapes show an integrated pattern of response. This question was evaluated first before examining effects of management intensity, spontaneous non-crop vegetation and soil nitrogen on grape qualities.

## Illustration of a Protocol for Evaluating Hypotheses about General Responses

### **Question #1: What are the Anticipated Characteristics of the Theoretical Construct(s) of Interest?**

We learn about latent variables indirectly. More specifically, we learn about them through theorising and empirical investigations, rather than direct measurement. It is important, therefore, that we consider the theoretical meaning of constructs carefully and explicitly. Most ecologists are accustomed to using descriptive procedures, such as principal components analysis (PCA), when faced with a set of related measurements. PCA seeks to reduce a set of variables to some smaller number of composite variables (aka components) that contain most of the information in the set. PCA is purely a data-reduction method and there is no basis for drawing causal interpretations of the resulting components (McCune et al. 2002, Chapter 30).

With LVSEM, we might pose a hypothesis, such as the one shown in Fig. 5A for the grape study. Here, we are hypothesising that the chemical composition of grapes is tightly regulated within plants and, thus, there should be strong integration and tight correlations (positive or negative) amongst indicator variables. However, the simple structure shown in Fig. 5A is not necessarily the only model to initially consider. For example, we might wish to also consider a model like the one in Fig. 5B, which anticipates that certain chemical properties will be inherently more tightly constrained. In that case, we might expect an additional effect, which can be represented by including an error correlation. The error correlation is actually an implied latent variable. We might want to represent this effect using a double-headed arrow rather than by including an explicit latent variable for several reasons. First, it allows us to retain focus on our primary hypothesis. Second, it only involves a single parameter, thereby saving model degrees of freedom and maintaining statistical power.

In thinking about our theoretical constructs, of fundamental importance is whether we think the concept is unidimensional (behaves like it is one thing) or multidimensional (behaves like a collection of different things). Taken literally (which software estimation will do), the hypothesis being evaluated in Fig. 5A is that there is a single common cause influencing all the response indicators. The effects ( $\lambda$ s) do not have to be equal, but their relative strengths have to be consistent if we are to believe they are correlated due to a single common cause. All other causes contributing variations to response indicators are independent (the  $\epsilon$ s). If we think there are latent factors causing some response

indicators to be more strongly correlated than could be explained by the single cause of theoretical interest, we might expect the need to include correlations amongst errors.

There are many other possibilities that might be supported by theory. The most common alternative is that a theoretical “construct” or concept may be a collection of independent or semi-independent causes. The details that accompany this situation are beyond our purpose in this paper and the reader is referred to Grace et al. (2010) for a discussion of this larger topic.

## **Question #2: Are there Appropriate Measured Variables that can Serve as Indicators of the General Theoretical Constructs?**

When interested in a general property of a study system, it is recommended that one gives careful consideration to the previous question about expected attributes when designing the sampling scheme. This is one of those interesting differences between science practice in the social sciences versus the ecological sciences. In the social sciences, particularly when studies involve human behaviour, the default assumption is that the latent properties are of primary interest. Studies may involve human attitudes and motivations, which are assumed from the outset to be “deeply latent” and only discernible indirectly. This has led to the development of a process for careful consideration of the development of proper measures for the constructs of interest. For example, the American Association of Psychology Dictionary (VandenBos 2007) provides the following description for *scale development*:

“The process of creating a new instrument [a set of specific measurements] for measuring an unobserved or latent construct, such as depression, sociability, or fourth-grade mathematics ability. The process includes defining the construct and test specifications, generating items and response scales, piloting the items in a large sample, conducting analyses to fine-tune the measure, and then readministering the refined measure to develop norms (if applicable) and to assess aspects of reliability and validity.”

Our purpose here is to raise awareness of the fact that there has been substantial development of methodologies in other scientific disciplines that could be of interest to natural scientists, but that has been systematically ignored to the detriment of our scientific studies. It is beyond the scope of the present paper to consider this body of knowledge in detail, though the expected requirements for a set of indicators to represent a theoretical construct will be illustrated via our presentation. For a more general introduction to scale development, one can refer to DeVellis (2016).

When one wishes to develop a latent variable SE model, it is possible to proceed by having one or more indicator measurements. Having only a single measure provides limited opportunities. The most commonly adopted approach is to simply assume that the measured variable is a perfect representation of the latent property. The main accomplishment achieved in such a model is to make a conceptual distinction between the concept of interest and the observed measure. When we have some estimate for the reliability (repeatability) of a measurement process, we can insert that information into our

model and remove bias due to measurement error. Once we have two or more indicators, it is possible to confirm or not the presence of a latent cause. This is the example situation we address in the current paper.

*Indicator validity* refers to the requirement that measured variables are interpretable as measures of the concept of interest. This is a theoretical requirement, but one to not forget to address in a paper. We recommend the construction of a table such as Table 1 as a formal means of defining explicitly the basis for explaining the logic connecting indicators to latent variables.

### **Question #3: What do the Patterns of Intercorrelations Amongst Indicator Variables Suggest?**

It is one thing to conceptualise a set of observed variables as reflections of a concept of interest, but it is another thing for the data to agree with one's conceptualisation. A simple first approach to this problem is to construct a correlation matrix to see if the patterns of correlations amongst indicator variables are roughly consistent with theoretical expectations. For this exercise, we focus on the sub-model shown in Fig. 5A rather than the entire model for simplicity. Data for this paper were simulated from the results found by Steiner et al. (2021), as our goal is not to revisit the analyses performed by Steiner et al. (2021), but instead to use that study as a tangible example of the methods demonstration in this paper.

Fig. 5A implies two main things about our expectations of the data. The first expectation is very general; we expect the indicators to be correlated with one another. If we fail to find significant correlations (based on standard tests and null hypothesis testing,  $p < 0.05$ , then the indicators are not varying in concert and the data will not support the claim of a latent cause. Beyond that primary expectation, we might expect the standardised lambdas to be similar to each other in magnitude if they are simply controlled by a single common cause. There are many reasons this expectation might not be met. For example, if the true model is as shown in Fig. 5B, we would expect a somewhat different pattern of correlations. If that model were the true model, we would expect sugars and malic acids to be more strongly correlated with each other than with the other measures. Going one step further, the pattern of correlations may tell us which of several a priori theoretical alternatives are likely to be supported.

When one starts working with LVSEM, it is found that there are many ways that data may deviate from showing equal correlation strengths amongst indicators, aside from error correlations, some of which are suggested in Fig. 3. In the case of the data presented in Table 2, the correlations amongst indicators vary in strength, ranging from 0.61 to 0.27. The first thing to note is that sugars are negatively correlated with the other properties, suggesting some metabolic trade-off. A second observation is that tartaric and malic acids are only weakly correlated, which argues against the idea that acid production is a general property.

### **Question #4: Do Analyses Support There Being a Generalised Response?**

It is customary in SEM practice to analyse latent variable models in two stages, first evaluating the fit between latent variables and indicators (Fig. 5A) and then second, evaluating the full model (e.g. Fig. 6). Results obtained from the analysis of sub-models are only provisional, but this provides an opportunity to isolate certain kinds of misspecifications without working through models containing multiple misspecifications. SE models that include only latent variables and their indicators are often referred to as confirmatory factor analysis (CFA) models. Grace (2020) describes an integrated approach to comparing SE models that is consistent with current views from the field of statistics. The approach described involves an assessment of the weight of evidence supporting each model in the theoretically-defensible set of models being compared and considers the use of  $p$ -values along with information criteria, such as Akaike Information Criteria. Some readers may find that treatment a useful complement to the presentation in this paper.

Table 3 presents the code used to conduct a CFA examination of the model shown in Fig. 5 A. As can be seen in Table 3, lavaan allows for the use of covariance matrices as input data, which is very helpful in methods illustrations. The CFA command “GrapeQual =~ N +Sugars +Tart +Malic” is read as, “The latent property Grape quality is measured by four indicators, nitrogen, sugars, tartaric acid and malic acid.”

Tables of results for all models run in the paper are provided in Suppl. material 2. The reader may find it useful to download that file to follow along if they wish to see more of the raw results. References to specific tables of results (e.g. Table S2.1) are provided in the text that follows. Suppl. material 3 contains the R script as a separate document.

Examination of results focuses initially on overall model fit (Suppl. material 2, Table S2.1). We do not want to interpret results related to parameter estimates until we are confident there are no major model-data discrepancies; therefore, initial focus is placed on model fit evaluation.

Results show strong support for our initial model (Table S2.1). A test statistic (Model Chi-square) value of 0.808 with an associated  $p$ -value of 0.668 was found. This  $p$ -value is well above the 0.05 criterion, providing strong support for there not being major model-data discrepancies. A Comparative Fit Index value of 1.000 further indicates a near-perfect explanation of the observed covariances by the model. Thus, it is extremely unlikely that additions to our model, such as shown in Fig. 5B, would be justified. That said, we do not use a 0.05 criterion as an absolute cutoff for adequate fit, since  $p$ -values above 0.05 can hide important discrepancies.

Having assessed the global model fit, we turn attention to the parameter estimates (Table S2.1). Again, we do not treat  $p$ -values as absolute cutoffs, but instead as continuous measures of evidence that a parameter or model deviates from the default expectation (Grace 2020). For our assessment of global model fit (Table S2.1), the default expectation is our hypothesised model (Fig. 5A). For parameter estimates (Table S2.1), however, the default expectation is a value of zero. The  $p$ -values for parameters, which range from < 0.001 to 0.003, provide further support for the model in Fig. 5A. Note that the estimate for N (lambda1) is fixed to a defined value of 1.0 as a lavaan default. This is done to set the



scale for the latent variable, which would otherwise be unidentified. Finally, *R-square* values are provided in Table S2.1 and indicate the degree to which our hypothesised latent cause explains the total observed variation in the indicator variables. Values returned were 0.78 for N, 0.38 for Sugars, 0.21 for Tart and 0.46 for Malic. Thus, in this case, variation in N appears to be less influenced by factors outside the model compared to the other indicators.

#### **Question #5: Does the Generalised Response Exhibit a Concerted Reaction to Perturbation? and**

#### **Question #6: Are there Unique Reactions by Specific Indicators?**

The complexity of SE models and the variety of inferences we typically wish to make lead us to move through the evaluation of our overall hypothesis in stages. It is important to keep in mind that conclusions one might draw, based on the analysis of sub-models, may need to be reconsidered once the full model is examined. Having examined the latent response sub-model, we now move to a pair of competing models shown in Fig. 6. Here, we use single measures for management intensity and non-crop vegetation.

In Fig. 6, we address a pair of questions; “Do grape qualities vary as a function of management intensity?” and “Does the cover of non-crop N-fixing plant species explain some or all of the effects of management intensity.” The first of these questions is represented in Fig. 6A and the lavaan code is provided in Table 4. Lavaan code is used to specify two latent variables using the  $\approx$  operator, then to represent the hypothesis that Grape Qualities depend on Management Intensity. We first request overall fit measures using the “show” command, then modification indices to see if there are meaningful suggestions for model improvement.

Results for the initial model (Fig. 6A) revealed substantial model-data discrepancy (Table S2.3). Not only is the *p*-value for the test statistic  $< 0.001$ , the CFI value of 0.789 is well below typical recommendations for a value of 0.95 or greater (Hu and Bentler 1999, Grace 2020). Modification indices suggest some sort of a unmodeled relationship between management intensity and tartaric acid. Since modification indices are best thought of as uninformed suggestions, we must use theoretical knowledge to decide what alternative model would be appropriate to consider.

As illustrated in Fig. 3, one of the theoretical possibilities we might anticipate for this model is specific effects of management on particular grape qualities. Here, we considered a reasonable alternative to be the addition of a direct effect of management intensity on tartaric acid (Table 4, Revised Net Effect Model). Results showed this revised model to have near-perfect correspondence with the data (Table S2.2).

Our second question, represented in Fig. 6B, involves a mediator that might explain why management intensity has an observed effect on grape qualities. Steiner et al. 2021 considered a number of possibilities. We do not revisit the full variety of possible mediator models considered in the original study, but focus on the possibility that the cover of nitrogen-fixing non-crop plants might explain all or part of the effects of management

intensity on grape qualities. While we might expect to again find a specific effect of management on tartaric acid, we nonetheless begin with the most general version of the mediator model shown in Fig. 6B (which we refer to as LVMed1 in Table 5.) Results again indicate that management intensity has a specific effect on tartaric acid separate from its general effect. Once that link is included in the Revised Mediated Effect Model (LVMed2, Table 5), fit was found to be very close (CFI = 1.0, Table S2.3).

### **Question #7: Can we Simplify the Model, Thereby Increasing Generality?**

Since SE models are used for explanatory representations of scientist's understanding of systems (Grace and Irvine 2020), there are many cases where that purpose may lead investigators to use a light hand in pruning weak effects from their models. However, when the goal is to make general inferences, model simplicity is preferred. In the context of the current paper, simplicity can be approached by fixing parameters to set values (such as zero) so that model structure is maintained while the number of estimated parameters is reduced. Fixed values of parameters represent general statements. In addition, the fewer the number of estimated parameters, the greater the statistical power per parameter (number of observations/number of estimated parameters).

Regarding our example, we next turn to an examination of individual parameter estimates to determine whether model simplification of model LVmed2 is possible (Table S2.4). *P*-values provide strong support for all estimated lambdas (all < 0.001), as well as all other estimated parameters, except beta1 ( $p = 0.713$ ), which is the effect of the mediator Non-Crop Vegetation on Grape Qualities. We estimated a simplified model (not shown) with beta1 set to zero (beta1 == 0) and determined that model fit was improved, as discrepancy increased very slightly while the number of estimated parameters was reduced by one (and fit is a measure of the amount of discrepancy prorata to the number of estimated parameters). We continue discussing ways to minimise the number of estimated parameters in the next section where we address the complexity that arises when there is more than one variety of grape being modelled.

### **Question #8: What About Generality Across Groups?**

LVSEM has the capacity to formally evaluate parameter equality across groups. Referred to as *multi-group analysis*, the investigator can test hypotheses by asking whether models of the same general form apply beyond single groups. With regard to the Swiss grape study, the investigators sampled vineyards that cultivated two different varieties of grapes, *Chasselas* and *Pinot noir*. Suppl. material 3 includes simulated data for the two varieties of grape, as well as code for multi-group analysis.

If multigroup models are specified without constraints, all parameters will be independently estimated for each group by default. One way to set equality constraints across groups is to add labels to the code. In this case, one first uses the format `c("label1", "label2")` to create names for the parameters where there are two groups. This example will generate two independent parameter estimates, one for each group, since the labels are unique. If we specify `c("lambda1", "lambda1")`, the repeated use of a common label means a single

value will be estimated for both groups (Table 6). The effects of this constraint will be reflected in the model discrepancy and we can judge whether this equality constraint has a small or large effect (Table S2.5). Adding the three constraints caused Model Degrees of Freedom to increase by 3 (going from 4 to 7) in the constrained model. This reduction in the number of estimated parameters is very helpful for small sample studies because multigroup models can contain twice as many parameters.

Using the approach in Table 6, the results obtained show that even with all lambdas constrained to be equal across groups, global model fit is very close (CFI = 1.0, Table S2.6). Based on this finding, we proceeded to estimate the full model, first allowing all gamma and beta parameters to be independent, then adding equality constraints and revisiting model fit. Proceeding in this fashion, we arrive at the model in Table 7 and the results are shown in Fig. 7.

## Discussion

It is important to be able to judge whether a system exhibits a generalised multivariate response to environmental change rather than an independent collection of uncoordinated responses. This paper presents an approach to addressing that question. A particular aspect of the approach demonstrated is that it invokes causal reasoning. We ask if suites of observed properties behave as if they are jointly influenced by a “hidden hand” or integrative cause.

Studying generalised responses is inherently challenging. Our objective is to focus our attention on the general, while moving the specifics to the background – at least initially. The sequence of operations described support a “general first, specifics second” perspective. Ultimately, SEM forces us to address both. Along the way, we must confront the large number of possible explanations that can exist for the actual functioning of the system being studied. This complexity means one cannot take a rigid approach, but must follow clues along a path to selecting a final model to use for interpretation. We suggest a series of questions that can guide investigators through several critical steps in model evaluation. In addition, we recognise that the research context matters, so the list may need to be modified for particular applications.

Success in applying a flexible, adaptive approach requires a solid understanding of how the analytical system ‘thinks’ about things. Within LVSEM, latent variables represent the common variance or overlapping information for a set of measures. They represent, in essence, the consensus opinion about the latent factor that functions as their common causal connection. There will, of course, be unique information associated with the individual measures, particularly if they are selected to represent multiple facets of a theoretical construct. Our core challenge is to capture the general opinions of the data without becoming overly distracted by the unique responses.

Fig. 7 provides us with a vehicle for making some main points about the evaluation of general responses. By including latent variables and their indicators, an explicit distinction

between concepts of interest and the measures used to quantify those concepts is made. This approach means that in future studies, one may retain their general hypothesis while adapting the details of model to the particulars of the measured indicators. It also means that we have a model that can adjust for measurement error if we choose to incorporate that information (e.g. Grace and Keeley 2006). While Fig. 7 presents standardised results, Table S2.7 shows that the unstandardised lambdas for Grape Qualities are equal for the different grape varieties. This suggests the observed indicators have physiological meaning at a fundamental level. Future studies may wish to further examine the physiological properties of grapes to develop a deeper understanding of the role of external conditions on their expression.

A number of mysteries are exposed in our multigroup model (Fig. 7). For some reason, the *Chasselas* variety does not respond to the abundance of nitrogen-fixing plants, while *Pinot noir* is quite responsive. Beyond that, the total effect of management intensity on grape qualities for *Pinot noir* is partly dependent on management through its influence on non-crop vegetation (indirect effect:  $-0.47 \times 0.59 = -0.28$ ), but is also largely impacted through other mechanisms (direct effect: 0.63). We might speculate that these other mechanisms have to do with a reduction in the competitive effects of non-crop vegetation on grape qualities, but further studies could explore that relationship in greater detail. Finally, the *Chasselas* variety exhibits a differential response of tartaric acid to management compared to the other grape properties, which suggests the need for further examination.

It is our hope that this paper demonstrates both how to approach using LVSEM to investigate multivariate responses and also to hint at the variety of scientific insights that can be gleaned from the effort. We believe there is an important opportunity for LVSEM to play a greater role in our quantitative understanding of ecological responses to environmental change.

## Acknowledgements

We thank two anonymous reviewers for helpful comments and suggestions. This work was supported by the USGS Ecosystems and Land Change Science Climate Research and Development Programs. Any use of trade, firm or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

## Conflicts of interest

## References

- Arhonditsis GB, Stow CA, Steinberg LJ, Kenney MA, Lathrop RC, McBride SJ, Reckhow KH (2006) Exploring ecological patterns with structural equation modeling and Bayesian analysis. *Ecological Modelling* 192: 385-409. <https://doi.org/10.1016/j.ecolmodel.2005.07.028>

- Bollen KA (1989) Structural equations with latent variables. John Wiley & Sons, New York. <https://doi.org/10.1002/9781118619179>
- Cubaynes S, Doutrelant C, Grgoire A, Perret P, Faivre B, Gimenez O (2012) Testing hypotheses in evolutionary ecology with imperfect detection: capture-recapture structural equation modeling. *Ecology* 93: 248-255. <https://doi.org/10.1890/11-0258.1>
- DeVellis RF (2016) Scale development: Theory and applications. 26. Sage publications
- Grace JB (2006) Structural equation modeling and natural systems. Cambridge University Press <https://doi.org/10.1017/CBO9780511617799>
- Grace JB, Keeley JE (2006) A structural equation model analysis of postfire plant diversity in California shrublands. *Ecological Applications* 16: 503-514. [https://doi.org/10.1890/1051-0761\(2006\)016\[0503:ASEMAO\]2.0.CO;2](https://doi.org/10.1890/1051-0761(2006)016[0503:ASEMAO]2.0.CO;2)
- Grace JB, Anderson TM, Olff H, Scheiner SM (2010) On the specification of structural equation models for ecological systems. *Ecological Monographs* 80: 67-87. <https://doi.org/10.1890/09-0464.1>
- Grace JB (2020) A 'Weight of Evidence' approach to evaluating structural equation models. *One Ecosystem* 5 (e50452). <https://doi.org/10.3897/oneeco.5.e50452>
- Grace JB, Irvine KM (2020) Scientists guide to developing explanatory statistical models using causal analysis principles. *Ecology* 101 (e02962). <https://doi.org/10.1002/ecy.2962>
- Hu LT, Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal* 6: 1-55. <https://doi.org/10.1080/10705519909540118>
- Joreskog KG (1970) A general method for analysis of covariance structures. *Biometrika* 57: 239-251. <https://doi.org/10.1093/biomet/57.2.239>
- Liu X, Swenson NG, Lin D, Mi X, Umana MN, Schmid B, Ma K (2016) Linking individual level functional traits to tree growth in a subtropical forest. *Ecology* 97: 2396-2405. <https://doi.org/10.1002/ecy.1445>
- McCune B, Grace JB, Urban D (2002) Analysis of ecological communities: Chapter 30. MjM Software, Gleneden Beach, OR, USA.
- Pugesek BH, Tomer A, von Eye A (Eds) (2003) Structural Equation Modeling: Applications in Ecological and Evolutionary Biology. Cambridge University Press <https://doi.org/10.1017/CBO9780511542138.005>
- R Core Team (2019) R: A language and environment for statistical computing. R Foundation for Statistical Computing URL: <https://www.R-project.org/>
- Rosseel Y (2012) lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software* 48: 1-36. URL: <http://www.jstatsoft.org/v48/i02/>
- Shipley B, Lechowicz MJ, Wright I, Reich PB (2006) Fundamental tradeoffs generating the worldwide leaf economics spectrum. *Ecology* 87: 535-541. <https://doi.org/10.1890/05-1051>
- Shipley B (2016) Cause and correlation in biology. Second. Cambridge University Press <https://doi.org/10.1017/CBO9781139979573>
- Souchay G, Wijk RE, Schaub M, Bauer S (2018) Identifying drivers of breeding success in a long distance migrant using structural equation modelling. *Oikos* 127: 125-133. <https://doi.org/10.1890/05-1051>
- Steiner M, Grace JB, Bacher S (2021) Biodiversity effects on grape Quality depend on variety and management intensity. *Journal of Applied Ecology* In Press.

- VandenBos GR (2007) APA dictionary of psychology. American Psychological Association URL: <https://dictionary.apa.org/scale-development>



Figure 1.

Structural equation meta-model representing the general modelling goal. Note that a meta-model is a generalisation that defines a finite set of possible fully-specified models. Dotted outlines are used to convey that the entities represented are general concepts rather than specific variables.

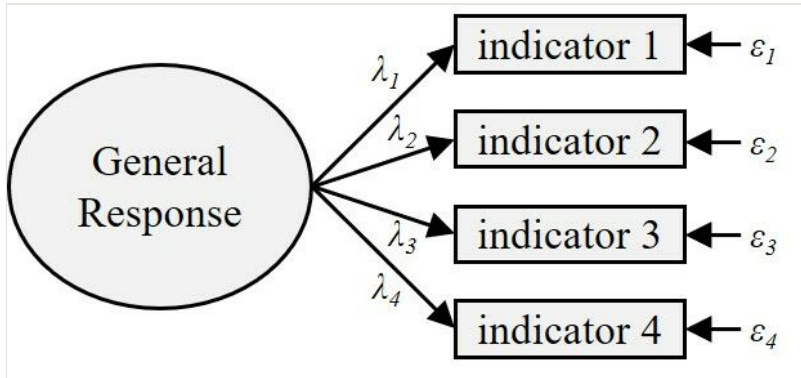


Figure 2.

Structural equation model representing the hypothesis that observed intercorrelations amongst response indicator variables (1 – 4) can be explained by a common cause, the generalised response. In contrast to Fig. 1, here we are dealing with specific variables, not concepts and solid outlines are used to represent that difference. Ellipses or circles are characteristically used to denote latent variables and rectangles used to denote observed variables (Grace 2006, Chapter 4).



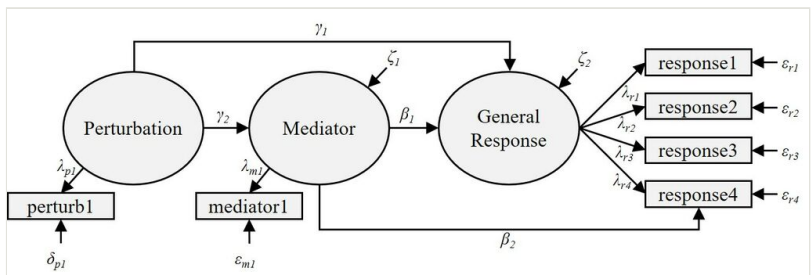


Figure 3.

Example SE model for the study of a general response to a specific perturbation and the role of a specific hypothesised mediator variable. See text for a discussion of notation.

### Meta-Model for Study of Grape Responses to Agricultural Management

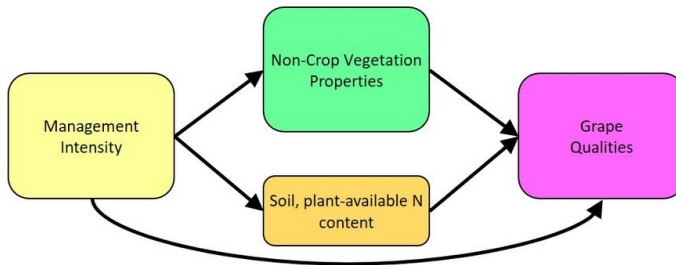


Figure 4.

Meta-model for the ecological example referenced in this paper (modified from Steiner et al. 2021).

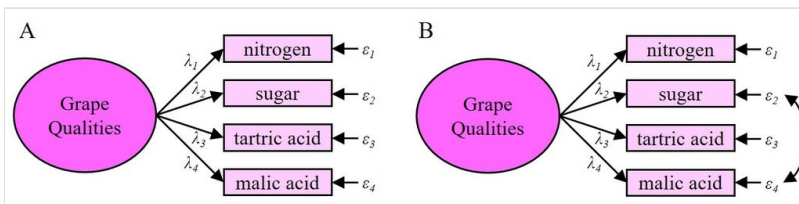


Figure 5.

(A) The initial hypothesis evaluated by Steiner et al. 2021 for the Grape Qualities construct.  
 (B) A slightly modified hypothesis that predicts sugar and malic acids will be more closely associated with each other than with the other measured properties.

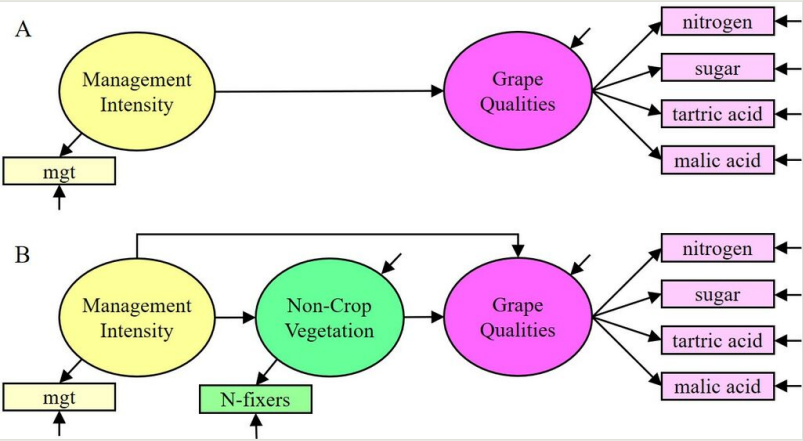


Figure 6.  
(A) The Net Effect Model. (B) The Mediated Effect Model.

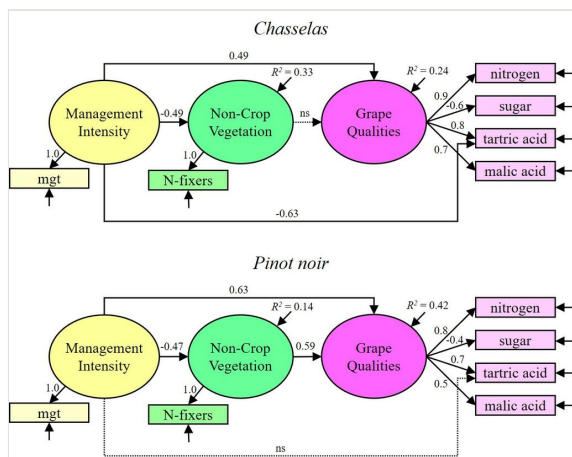


Figure 7.

Standardised results for the final full model showing both groups.

Table 1.

Concepts related to the structural equation meta-model in Figure 4 and their relationships to measured variables (from Steiner et al. 2021).

Concept of interest	Measurements	Scientific rationale
Management intensity	Intensity is a three-level index {1,2,3}. 1 = minimal control of inter-row vegetation, 2 = vegetation removal in every other row between grape plants, 3 = vegetation removal in all rows between grape plants.	The primary purpose of management is to reduce competitive effects of non-crop plants on grape plants. It is assumed that competition primarily acts through reductions in soil water and nutrients, but other forms of interference could be possible.
Non-crop vegetation properties	Plant species richness (numbers), abundance of N-fixing plants (% cover)	One possibility we wished to consider was a general beneficial effect of plant richness on grape qualities due to complementarity.  Another possibility of interest was a specific effect of the abundance of N-fixing plants on grape properties due to facilitation.
Soil nitrogen	Total soil N content (%)	We considered it possible that variations in total soil N might help explain variations in grape N. Such an effect either might or might not be indirectly related to management intensity.
Grape qualities	Nitrogen concentration Sugar concentration Tartaric acid Malic acid	We measured a suite of standard grape chemical parameters of importance for wine-making. While all of these parameters determine the character of wine, N concentration is perhaps of primary concern because of its critical role in the fermentation process (Bell and Henschke 2005)

Table 2.

Correlations amongst simulated indicators of Grape Qualities.

	Nitrogen	Sugars	Tartric Acid	Malic Acid
Nitrogen	1.00			
Sugars	-0.53	1.00		
Tartric Acid	0.40	-0.36	1.00	
Malic Acid	0.61	-0.41	0.27	1.00

Table 3.

R code for the Latent Response Model (Fig. 5A) using the lavaan software (R Core Team 2019, Rosseel 2012)

library(lavaan)					
input.cov <- '					
2.602					
-1.187	1.896				
1.038	-0.781	2.536			
1.270	-0.726	0.559	1.688		
-0.592	0.451	0.147	-0.219	1.670	
0.821	-0.364	-0.455	0.578	-0.864	1.366 '
cov.dat <- getCov(input.com, names = c("N", "Sugars", "Tart", "Malic", "Nfixers", "Mgt"))					
cfa1 <- 'GrapeQual =~ lambda1*N +lambda2*Sugars +lambda3*Tart +lambda4*Malic '					
cfa1.fit <- sem(cfa1, sample.cov = cfa.cov.dat, sample.nobs = 50)					



Table 4.

R code for examining the Net Effect Model (Fig. 6A).

**## Initial Net Effect Model**

```
LVNet1 <- '  
GrapeQual =~ lambda1*N +lambda2*Sugars +lambda3*Tart +lambda4*Malic  
ManInten  =~ lambda5*Mgt  
GrapeQual ~ gamma1*ManInten'
```

```
LVNet1.fit <- sem(LVNet1, sample.cov=cov.dat, sample.nobs=50)
```

```
show(LVNet1.fit); fitMeasures(LVNet1.fit, "cfi")  
subset(modindices(LVNet1.fit), mi>3)
```

**## Revised Net Effect Model**

```
LVNet2 <- '  
GrapeQual =~ lambda1*N +lambda2*Sugars +lambda3*Tart +lambda4*Malic  
ManInten  =~ lambda5*Mgt  
GrapeQual ~ gamma1*ManInten  
Tart      ~ gamma3*ManInten'
```

Table 5.

R code for examining the Mediated Effect Model (Fig. 6B).

**### Initial Mediated Effect Model**

```
LVmed1 <- '  
GrapeQual =~ lambda1*N +lambda2*Sugars +lambda3*Tart +lambda4*Malic  
ManInten  =~ lambda5*Mgt  
NonCrop   =~ lambda6*Nfixers  
GrapeQual ~ gamma1*ManInten + beta1*NonCrop  
NonCrop   ~ gamma2*ManInten'
```

**### Revised Mediated Effect Model**

```
LVmed2 <- '  
GrapeQual =~ lambda1*N +lambda2*Sugars +lambda3*Tart +lambda4*Malic  
ManInten  =~ lambda5*Mgt  
NonCrop   =~ lambda6*Nfixers  
GrapeQual ~ gamma1*ManInten + beta1*NonCrop  
NonCrop   ~ gamma2*ManInten  
Tart       ~ gamma3*ManInten' #added direct effect
```

Table 6.

Example R code for multigroup analysis of full model.

**## CFA independence model with distinct labels for each group**

```
mg.mod0 <- '  
GrapeQual =~ c("lambda1a","lambda1b")*N  
+ c("lambda2a","lambda2b")*Sugars  
+ c("lambda3a","lambda3b")*Tart  
+ c("lambda4a","lambda4b")*Malic'
```

**## CFA model with parameters equal across groups (using repeat labels)**

```
mg.mod1 <- '  
GrapeQual =~ c("lambda1","lambda1")*N  
+ c("lambda2","lambda2")*Sugars  
+ c("lambda3","lambda3")*Tart  
+ c("lambda4","lambda4")*Malic'
```

Table 7.

R code for final multigroup analysis of full model.

```
mg.mod4 <- '  
# declare latent variables  
GrapeQual =~ c("lambda1","lambda1")*N + c("lambda2","lambda2")*Sugars  
            + c("lambda3","lambda3")*Tart  
            + c("lambda4","lambda4")*Malic  
ManInten =~ c("lambda5","lambda5")*Mgt  
NonCrop  =~ c("lambda6","lambda6")*Nfixers  
# regressions  
GrapeQual ~ c("gamma1a","gamma1b")*ManInten  
            + c("beta1a","beta1b")*NonCrop  
NonCrop ~ c("gamma2a","gamma2b")*ManInten  
Tart ~ c("gamma3a","gamma3b")*ManInten  
# set constraints  
beta1a == 0  
gamma3b == 0  
gamma2a == gamma2b'
```

## Supplementary materials

### **Suppl. material 1: A protocol for modelling generalised biological responses using latent variables in structural equation models**

**Authors:** Grace JB, Steiner M

**Data type:** Mathematical equations and notation for latent variable structural equation modelling.

**Brief description:** This text file contains the equations and notation mentioned in Grace JB, Steiner M (2021) A protocol for modelling generalised biological responses using latent variables in structural equation models. One Ecosystem

[Download file](#) (134.48 kb)

### **Suppl. material 2: A protocol for modelling generalised biological responses using latent variables in structural equation models**

**Authors:** Grace JB, Steiner M

**Data type:** Results Tables

**Brief description:** This file contains the results tables for the demonstrations included in Grace JB, Steiner M (2021) A protocol for modelling generalised biological responses using latent variables in structural equation models. One Ecosystem.

[Download file](#) (117.55 kb)

### **Suppl. material 3: A protocol for modelling generalised biological responses using latent variables in structural equation models**

**Authors:** Grace JB, Steiner, M

**Data type:** R code

**Brief description:** : This text file contains the R code used to develop the demonstrations included in Grace JB, Steiner M (2021) A protocol for modelling generalised biological responses using latent variables in structural equation models. One Ecosystem.

[Download file](#) (8.34 kb)