Conservation Biology

Special Section Conservation Methods

Tools for assessing the psychometric adequacy of latent variables in conservation research

Gerard Kyle ,[1 *] Adam Landon ,[2] Jerry Vaske ,[3] and Kenneth Wallen[4]

[1]Department of Rangeland, Wildlife, & Fisheries Management, Texas A&M University, 2261 TAMU, College Station, TX 77843-2261, U.S.A.
[2]Division of Fish & Wildlife, Minnesota Department of Natural Resources, 500 Lafayette Road, St. Paul, MN 55155, U.S.A.
[3]Human Dimensions of Natural Resources, Colorado State University, Fort Collins, CO 80523, U.S.A.
[4]Department of Natural Resources and Society, University of Idaho, 875 Perimeter Dr, Moscow, ID, 83844-1139, U.S.A.

Abstract: Conservation psychology has a history of measuring variables that cannot be seen (e.g., values, attitudes, norms). Such latent variables are critical drivers of human action and are often measured using responses to survey questions. Tools for establishing the psychometric adequacy of unobservable, latent variables has been a century-long pursuit and challenge for quantitative psychologists and statisticians. Fundamental questions at the heart of this challenge include is what is claimed to be measured (validity) being measured and is measurement consistent (reliability)? We examined common methods used to establish the validity and reliability of psychometric instruments. Through a case study of anglers in Texas, we investigated the protocols and metrics used to evaluate the measurement of latent variables. The indicators we tested (identity, awareness of consequences, ascription of responsibility, and personal norms) validly and reliably assessed latent variables. Our findings also illustrated decision protocols (e.g., discriminant validity, convergent validity, internal consistency) involved in assessing the psychometric adequacy of latent variable indicators. The ability to correctly identify significant relationships among unobserved variables and their influence on human action is directly tied to the adequacy of measurement. In an era of instability and change that threatens social-ecological systems worldwide, the need for accuracy and precision in conservation social science has never been greater. Research that employs flawed measures has potential to lead to erroneous conclusions and undermine conservation and biodiversity protection.

Keywords: conservation psychology, latent variable modeling, measurement, reliability, validity

Herramientas para Evaluar la Idoneidad Psicométrica de las Variables Latentes en la Investigación de la Conservación

Resumen: La psicología de la conservación tiene la reputación de medir variables que no pueden ser vistas (p. ej.: valores, actitudes, normas). Dichas variables latentes son impulsores importantes de la acción humana y con frecuencia se miden usando las respuestas dadas en una encuesta. Las herramientas para establecer la idoneidad psicométrica de las variables inobservables y latentes ha sido una búsqueda y un desafío de todo un siglo para los psicólogos cuantitativos y los estadistas. Las cuestiones fundamentales en el núcleo de este desafío son: si es medido lo que se dice está siendo medido (validez) y si la medición es uniforme (confiabilidad). Examinamos los métodos comunes usados para establecer la validez y la confiabilidad de los instrumentos psicométricos. Mediante un estudio de caso de pescadores en Texas, investigamos los protocolos y las medidas usadas para evaluar la medida de las variables latentes. Los indicadores que analizamos (identidad, noción de las consecuencias, adscripción de la responsabilidad y normas personales) evaluaron a las latentes variables en validez y confianza. Nuestros hallazgos también reflejaron los protocolos de decisión (p. ej.: x, y, z) involucrados en la evaluación de la idoneidad psicométrica de los indicadores de variables latentes. La habilidad para identificar correctamente las relaciones significativas entre las variables no observadas y su influencia sobre las acciones humanas está vinculada directamente a la idoneidad de la medición. Hoy en día, en esta época de inestabilidad y cambio que

*email: gkyle@tamu.edu
Article impact statement: *Correctly identifying significant relationships among unobserved variables and their influence on human action is tied to adequacy of measurement.*
Paper submitted November 10, 2019; revised manuscript accepted April 10, 2020.

Conservation Biology, Volume 34, No. 6, 1353–1363

DOI: 10.1111/cobi.13625

amenaza a los sistemas socio-ecológicos en todo el mundo, la necesidad de tener certeza y precisión en las ciencias sociales de la conservación nunca ha sido tan grande. Las investigaciones que emplean medidas imperfectas tienen el potencial de derivar en conclusiones erróneas y perjudicar a la conservación y a la protección de la biodiversidad.

Palabras Clave: confiabilidad, medida, modelación de variables latentes, psicología de la conservación, validez

摘要: 保护心理学一向需要测量看不见的变量 (如价值观、态度、规范) 。这些潜在变量是人类行为的关键驱动因素, 通常使用问卷调查来进行测定。为难以观察的潜在变量的心理测量适当性开发工具, 是定量心理学家和统计学家一个世纪以来的追求和挑战。这一挑战的核心基本问题包括研究是否测量了其声称要测量的内容 (效度) 以及测量是否一致 (信度) 。本研究对确定心理测量工具的效度和信度的常用方法进行了检验。我们通过对德克萨斯州垂钓者的案例研究, 分析了用于评估潜在变量测量的方案和指标。本研究分析的指标 (身份、对后果的认识、责任的归属和个人规范) 可以有效且可靠地评估潜在变量。我们的结果还展现了用于评估潜在变量指标的心理测量适当性的决策方案 (如 x, y, z) 。正确识别难以观察的变量之间的重要关系及其对人类行为的影响的能力与测量适当性直接相关。在一个全球社会生态系统受到不稳定和变化威胁的时代, 人们对保护科学的准确性和精确性具有前所未有的需求。方法存在缺陷的研究有可能得出错误的结论, 以至于危害保护和生物多样性。【**翻译: 胡怡思; 审校: 聂永刚**】

关键词:潜在变量建模, 效度, 信度, 测量, 保护心理学

Introduction

Social science is increasingly found in *Conservation Biology* (de Snoo et al. 2013; Sandbrook et al. 2013; Bennet et al. 2017). This growth presents a challenge for researchers. Some published work, for example, could be deemed methodologically questionable given the normative standards of parent disciplines (e.g., psychology) from which the work draws (St. John et al. 2014). This is especially apparent in research that adopts variables such as values, beliefs, norms, and attitudes. Because these variables cannot be directly observed like distance or temperature in biophysics, psychologists have developed tools to capture their manifestation. To embrace the theories and methods of parent disciplines, requires adherence to the norms of the discipline regarding conceptualization, measurement, analysis, and reporting (Teel et al. 2018).

We devised a primer on a set of methods conservation scholars can use to establish the validity and reliability of latent variables. Although the concepts presented here are not new to psychology, their application in conservation science is relatively recent. We compiled an overview of latent variables and their utility for conservation science and a step-by-step presentation of the analyses and decisions analysts make. We used a case study to illustrate how these tools are applied in practice and identify problems that scholars may encounter.

Measuring the Unobservable

Establishing the adequacy of psychometric measures involves the following sequential steps: identify theoretical framework and relevant past empirical evidence; select manifest indicators; test the measurement model evaluate validity and reliability; and then test the hypothesized structural model (Table 1). Although generally linear, some iteration may be necessary if data anomalies are encountered. The successful implementation of each step depends on the quality of the output emerging from the previous step, and, fundamentally, from the research and sampling design. The importance of rigor in the data collection process should not be downplayed. Because surveys are commonly used for data collection, readers should consider Dillman et al. (2014) and Vaske (2019) as informative guides for survey design and protocols for minimizing error and bias. Throughout the article, we reference terms from quantitative psychology (defined in Table 2). Data for the case study were drawn from Landon et al.'s (2018) investigation of the psychological drivers of anglers' adoption of conservation behaviors in the United States. Data were collected in 2015 ($n = 948$) through a tailored design method (Dillman et al. 2014). We ran our analyses in STATA (version 16) and R (version 3.5.1) (Rosseel 2012). The Stata and R syntax from the case study are in the Supporting Information.

Theory and Empirical Evidence

The first step in measuring latent variables requires identifying a theoretical framework and relevant past empirical evidence. This approach guides which measures will be selected and analyzed and ensures the latent variables are consistent with the researcher's intentions. The psychological processes that shape human attitudes and behaviors are sometimes viewed within a broad class of research called the cognitive hierarchy (Homer & Kahle 1988). Specific behaviors and intentions are hypothesized to be a function of attitudes and norms which, in turn, are influenced by value orientations and values (Fulton et al. 1996; Fishbein & Ajzen 2010). In the context of Landon et al. (2018), latent variables hypothesized to influence behavioral intent were drawn from identity theory (Stryker & Burke 2000; Burke & Stets 2009) and the norm activation model (NAM) (Schwartz 1977). These types of theoretical frameworks are a crucial first step for any systematic inquiry of the psychological basis of conservation-related behavior. Because latent variables exist only as concepts, it is critical that their measurement be rooted in theory and past evidence to support the hypothesized relationships. For other empirical examples in the conservation literature, see Mastrangelo et al. (2014), Lute et al. (2016), and De Groot & Steg (2009).

Indicators of Latent Variables

Step 2 is the operational foundation for measuring unobservable variables. In psychology theoretical concepts cannot be directly observed and are referred to as latent variables or constructs (Kline 2016). Researchers must operationally define a construct in terms of phenomena it is thought to represent. The unobserved latent variable is then linked, via theory-based instrumentation, to measures that are observable (e.g., a rating scale), thereby making measurement possible (Brown 2015).

Observed measures of latent variables often use responses to instruments such as questionnaires or interviews. These measures are termed observed or manifest variables and serve as indicators of the underlying latent variable they are presumed to represent. For example, to measure angler identity (AI), respondents expressed their level of agreement on rating scale with the item "being an angler is an important part of who I am" (X1 in Table 3). Although we cannot see the concept of identity, it is indirectly manifested in subjects' responses to this and other items in the AI scale.

Psychometrically sound instruments provide the foundation for examining causal relationships between latent variables in conservation psychology. Based on their conceptual foundations, Landon et al. (2018) hypothesized that variables related to AI, awareness of the consequences (AC), ascription of responsibility (AR), and personal norms (PN) would influence respondents' conservation behaviors. Items used to measure these latent variables were drawn from past work that had demonstrated good validity and reliability (Steg & DeGroot 2010; Landon et al. 2017) (Table 3). It is generally recommended that at least 3 manifest indicators should be associated with each latent construct. This recommendation is based on algebraic identification in models used to estimate latent variables (Bollen 1989) and scale reliability. In our case study, respondents indicated their agreement with each statement on a 5-point rating scale from 1, strongly disagree, to 5, strongly agree. This specification bounds the variable on hypothesized extremities of response. Most statistical programs default to the maximum likelihood estimator when a variable is continuous or ordinal and the rating scale has 5 or more response categories. When the data are nominal or ordinal with fewer than 5 response categories, other estimators (e.g., robust maximum likelihood, weighted least squares) or correction protocols are advised (Brown 2015).

Common Factors Model

The third step involves linking observed measures to latent variables through factor analysis (Thurstone 1947).

Table 1. Steps for testing and evaluating models of latent variables.

Step	*Action*	*Objective*	*Outcome*
1	identification of theoretical framework and past empiricalevidence	inform construct definition and the nomologicalnetwork	model identification
2	selection of indicators used to measure latentconstructs	measures are a manifestation (reflective) of the latent construct	items venture beyond bounds of the defining construct
3*	test measurement model (confirmatory factoranalysis)	test the hypothesized factor structure (dimensionality)	show congruence between the model implied covariance matrix (σ) and the sample covariancematrix (s)
4	empirically evaluate validity and reliability	evaluate the psychometric adequacy of the manifestindicators	determination of construct validity (convergent and discriminant) and reliability (internal consistency andcomposite)
Step 5	test structural associations	test hypotheses of causal inference among latent constructs derived from theory or past empiricalevidence	identification of processes and driversof influence

**Following collection of data.*

Table 2. Glossary of basic terms associated with psychometric evaluation.

Term	*Definition*	*References*
Psychometrics	field concerned with the quantification and measurement of mental attributes, behavior, and design, analysis, and improvement of the tests, questionnaires, and other instruments used in such measurement	American Psychological Association 2020
Latent variable (unobserved variable)	variable with a theoretical basis presumed to reflect (explain) a concept that cannot be directly observed or measured	Kline 2016
Manifest variable (observed variable)	observed measure that underlies a latent variable it is presumed to represent (i.e., the data collected)	Kline 2016
Common factors model	formal proposition that assumes each indicator in a set of observed measures is a linear function of ≥1 common factor and 1 unique factor	Thurstone 1947; Brown 2015
Exploratory factor analysis (EFA)	data-driven approach that does not specify the number of factors or the pattern of relationships between the common factors and the indicators	Brown 2015
Confirmatory factor analysis (CFA)	theory-driven approach that specifies (hypothesizes) the number of factors and the pattern of relationships in advance	Brown 2015
Indicator (measure)	observed variable used as an indirect statistical measure of a construct; term used within the context of CFA but is generally synonymous with *manifest* variable	Kline 2016
Factor (construct)	unobservable variable that statistically influences ≥1 observed measure and accounts for correlations among observed measures (indicators); term used within the context of CFA but generally synonymous with *latent* variable.	Brown 2015; Kline 2016
Measurement model	statistical model that defines the relationships between observed indicators and unobserved factors and evaluates how well those factors are measured by the indicators (i.e., provides empirical estimates psychometric properties)	Kline 2016
Goodness of fit	extent to which observed data are predicted (reproduced) by hypothesized model	Anderson & Gerbing 1988; Hu & Bentler 1999; Kline 2016
Reliability	extent to which instrument yields scores that are consistently repeatable	Cronbach 1951; Raykov 1997
Validity	extent to which instrument measures what it claims to measure	Cronbach & Meehl 1955; Campbell & Fiske 1959; Brown 2015
Construct validity	overarching principle of validity that refers to the extent to which a psychological measure, test, or instrument in fact measures the concept it purports to measure	Brown 2015
Convergent validity	differentmeasures of theoretically similar or overlapping constructs being strongly interrelated	Brown 2015
Discriminant validity	measures of theoretically distinct constructs not highly intercorrelated	Brown 2015
Cross sectional data	observations made at 1 point in time	Babbie 2016
Nomological network	hypothesized pattern of relationships among variables that is inclusive of antecedents, mediators, and outcomes	Cronbach & Meehl 1955; Hagger et al. 2017

Factor analysis identifies the number and nature of latent variables (i.e., factors or constructs) that account for the variation among a set of observed measures (i.e., indicators). A factor is an unobservable latent variable that influences ≥1 observed measure and accounts for the correlations among the observed measures. Factor analysis assumes that the observed measures are manifestation of a theoretically defined latent variable. Observed measures are intercorrelated because they share a common cause in their latent variable. The goal of factor analysis is to develop a parsimonious understanding of the covariation among a set of indicators because the number of factors is often substantially less than the number of observed variables (Brown 2015).

There are 2 types of factor analyses: exploratory (EFA) and confirmatory (CFA). Exploratory factor analysis is used when the links between the observed and latent variable are unknown, a priori. As an inductive approach,

Table 3. Confirmatory factor analysis of recreational angler identity, awareness of the consequences, ascription of responsibility, and personal norms.[a].

Measure		*b (SE)*[b]	Λ[c]	*SMCs*[d]	*z*	*p*	*ρ*	α	*Mean*
Angler identity							0.852	0.847	0.542
X1	Being an angler is an important part of who I am.	1.00	0.80	0.64	47.36	<0.001			
X2	Angling is something that I rarely even think about.	0.92 (0.04)	0.74	0.58	36.12	<0.001			
X3	I would be at a loss if I were forced to give up angling.	0.97 (0.05)	0.66	0.44	28.24	<0.001			
X4	For me, being an angler is about more than just going fishing.	0.80 (0.04)	0.70	0.50	32.40	<0.001			
X5	I really don't have any clear feelings about being an angler.	0.97 (0.05)	0.77	0.59	40.43	<0.001			
Awareness of consequence							0.818	0.817	0.627
X6	Human activities have a negative impact on fisheries resources and aquatic ecosystems.	1.00	0.64	0.41	26.45	<0.001			
X7	Fisheries resources and aquatic ecosystems are threatened by human activities.	1.15 (0.06)	0.88	0.77	47.89	<0.001			
X8	Human impacts on fisheries resources and aquatic ecosystems are a serious problem.	1.13 (0.06)	0.83	0.69	42.80	<0.001			
Ascription of responsibility							0.795	0.784	0.598
X9	Conserving fisheries resources and ecosystems is my responsibility.	1.00	0.85	0.72	71.48	<0.001			
X10	I feel responsible to do my part to conserve fisheries resources and aquatic ecosystems.	1.03 (0.03)	0.88	0.77	84.97	<0.001			
X11	I am not responsible for conserving fisheries resources and aquatic ecosystems.	0.75 (0.05)	0.55	0.30	21.01	<0.001			
Personal norms							0.828	0.841	0.632
X12	People like me should do whatever they can to conserve fisheries resources and aquatic ecosystems.	1.00	0.88	0.77	88.29	<0.001			
X13	I would feel guilty if I didn't do my part to conserve fisheries resources and aquatic ecosystems.	1.03 (0.04)	0.75	0.56	43.63	<0.001			
X14	I feel morally obliged to try to conserve fisheries resources and aquatic ecosystems.	0.97 (0.04)	0.75	0.56	42.82	<0.001			

[a] *The first item in each variable is constrained to 1 to set the scale.*
[b] *Unstandardized factor loading.*
[c] *Standardized factor loading.*
[d] *Squared multiple correlations.*

EFA reveals associations between the observed measures and underlying factors based on empirical evidence. The goal is to identify the minimum number of factors that account for covariation among the observed measures. Alternately, CFA is used when the researcher has an a priori understanding of the underlying factor structure, which stems from theory, past empirical evidence, or both. Confirmatory factor analysis examines the associations between the observed measures and their underlying latent factors (i.e., CFA tests a hypothesized factor structure). When conducting CFA, the researcher specifies both the number of latent factors and their associations with the observed measures. For both EFA and CFA, relationships between observed measures and latent variables are represented by factor loadings.

Although EFA and CFA are both based on the common factors model, CFA is driven by theory, empirical evidence, or both. Unlike EFA, CFA tests a priori hypotheses about the number of factors, the pattern of factor loadings, and errors associated with the latent

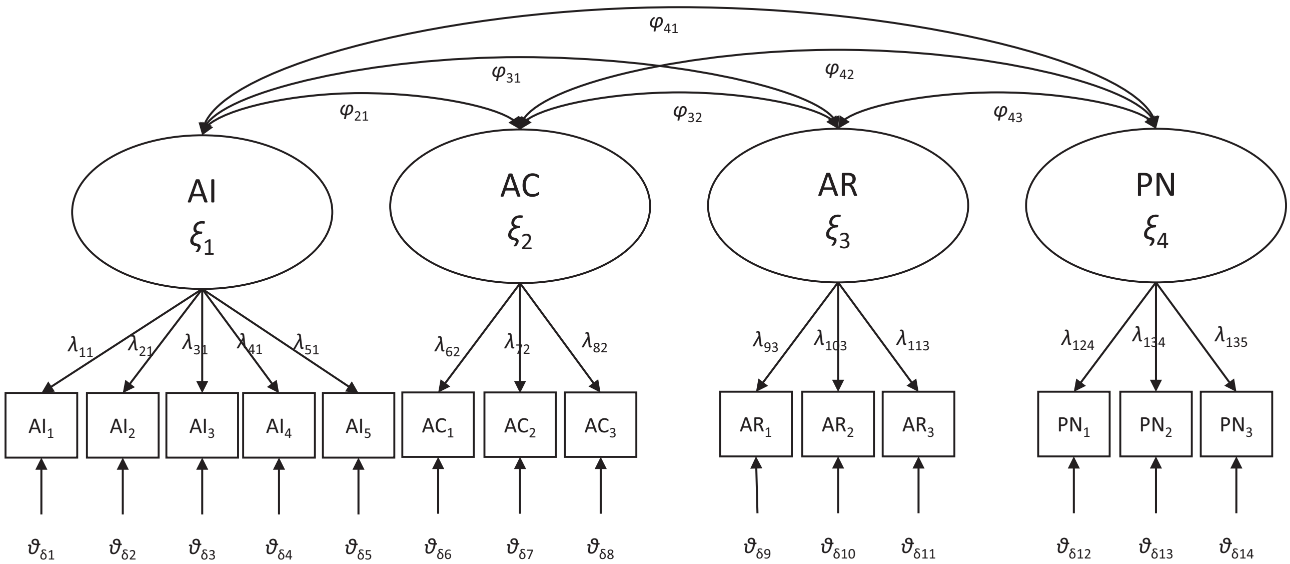

Figure 1. Hypothesized measurement model of recreational angler identity (AI), awareness of the consequences (AC), ascription of responsibility (AR), and personal norms (PN).

constructs. The acceptability of the specified model is evaluated relative to goodness of fit and the strength of the resulting parameter estimates. Goodness of fit refers to the extent to which the hypothesized model reproduces the observed data (i.e., the variance or covariance matrix). Model fit is a necessary but insufficient condition for establishing the psychometric fitness of latent measures. The output associated with CFA provides information for preliminary evaluating the adequacy of the measures and guidance on how to improve model.

In our case study, the factor structure of our constructs and measures was known, a priori, so CFA was considered appropriate (as opposed to EFA). Goodness of fit indices include χ^2, comparative fit index (CFI) (acceptable CFI value is often >0.95), normed fit index (NFI) (an acceptable NFI value is often >0.90), and the root mean square error of approximation (RMSEA) (an acceptable RMSEA value is <0.08) (Hu & Bentler 1999). Although $\chi^2$is the most basic indicator of model fit, it is sensitive to sample size ($n > 400$ cases) and the strength of variable correlations (i.e., high values produce an inflated χ^2) (Kenny 2015). Consequently, the array of alternate-fit indices are more often used to assess model fit. For a more extensive discussion of model fit see Kline (2016). Confirmatory factor analysis is also often used as a precursor to test structural equation models (SEM) that specify relationships (e.g., regressions) among latent variables (Anderson & Gerbing 1988). Within SEM, the CFA model is also referred to as the measurement model and provides foundational data for estimating a scale's psychometric properties. A full account of the steps for estimating and interpreting SEMs is outside the scope of this paper, but Brown (2015), Kline (2016), and Schrieber et al. (2006) offer detailed accounts. Applications in the context of conservation research are also presented by Kaltenborn et al. (2012), Sakuri et al. (2017), and Manfredo et al. (2020). In our case study, factor CFA revealed that the hypothesized model adequately fit the data ($\chi^2 = 379.64$, df $= 71$, RMSEA $= 0.073$, CFI $= 0.949$, NFI $= 0.935$) (Fig. 1).

Validity and Reliability

In step 4, validity and reliability of the latent variables are established, which is central to determining a scale's psychometric properties. Validity refers to the extent to which a scale measures what it claims to measure; reliability represents the extent to which the scale scores are repeatable. In the context of reliability, ensuring that a measure contains no random error is equally challenging. These measurement concerns must be addressed prior to examining the latent variables' associations with other variables of interest (Hagger et al. 2017). Although several metrics to establish construct validity and reliability have appeared in the literature over the past 50 years, the advent of CFA has provided researchers with an array of tools (e.g., Rosseel 2012; Jorgensen et al. 2019). Which tests to select are driven by the overriding theoretical framework, its associated research questions and hypothesis, and the study design (Kenny 2019).

Beginning with the assessment of validity, the measurement model is indispensable for evaluating the psychometric adequacy of test instrumentation and for construct validation. Validity tests determine whether the measure looks and behaves like a measure of the target variable. We considered 2 forms of construct validity: convergent and discriminant validity. Convergent validity is the extent to which different indicators of the same latent variable are correlated, whereas discriminant validity refers to analysts' ability to differentiate between latent constructs.

Landon et al. (2018) used several tests of convergent and discriminant validity that are commonly reported (Fornell & Larker 1981; Bagozzi & Phillips 1982; Bagozzi & Yi 1988). First, for convergent validity, we used 3 metrics to assess the extent to which each of the indicators shared a common factor: strength and statistical significance of factor loadings and estimated average variance extracted (AVE). Fornell & Larcker (1981) suggest that standardized factor loadings of 0.707 are desired given that values less than this indicate that the latent factor is capturing <50% of the variation in the indicator. Percent variance in the indicator explained by the latent factor is measured using the squared multiple correlation (SMC), which is the square of the standardized factor loading (see SMCs in Table 3). When the SMC is <0.5, the variance due to error is greater than the variance being captured by the latent variable. Three loadings fell below this threshold: X3, X6, and X11 (defined in Table 1).

Anderson & Gerbing (1988:416) suggest that "[c]onvergent validity can be assessed… by determining whether each indicator's estimated pattern coefficient on its posited underlying factor is significant." All items in our analyses had loadings that were statistically significant ($z \geq 1.96$), indicating the rejection of the null hypothesis and suggesting that the factor loadings were 0 (Table 3).

Estimates of the AVE (Fornell & Larcker 1981) for each latent variable provides an estimate of the variance captured by the variable in relation to the amount of variance due to measurement error. Fornell & Larcker (1981) suggest that values <0.5 infer that the validity of the indicators and the construct is questionable. All AVEs in our analyses were above Fornell and Larcker's recommendation (Table 3).

Collectively, these 3 metrics suggest the indicators in Landon et al. (2018) support the measures' convergent validity. Although 3 items fell below the 0.707 cutoff, they were not confined to a single variable; rather, they were spread across 3 factors. In applied research, the 0.707 standard can be quite demanding of the data. Other authors have offered alternative thresholds, such as 0.45, fair; 0.55, good; and 0.71, excellent (Tabachnick & Fidell 2007). When evaluating the acceptability of a parameter estimate, all available evidence should be considered. Beyond the convergent validity, reluctance to drop items with factor loadings slightly <0.707 was driven by concerns that the data could be used in the structural model. Because these variables were included in Schwartz's norm activation theory (1977), removing the items could potentially undermine the predictive validity of the model (Cronbach & Meehle 1955; Hagger et al. 2017). In testing the structural model, Landon et al. (2018) used these variables to predict respondents' stewardship behavior.

Tests of discriminant validity included confidence intervals around latent variable correlation estimates, AVEs greater than the squared correlation among latent variables, and constraining latent factor correlations. Anderson & Gerbing (1988:416) suggest that a complimentary way to assess "discriminant validity is to determine whether the confidence interval (CI with 2 SEs) around the correlation estimate between the 2 factors includes 1.0." Intervals that include 1.0, could suggest that the measures reflect the same variable. In our analyses, all but 1 of the CIs did not include 1.0 (Table 4). The CI around the correlation estimate between the latent variables AR and PN included 1.0, suggesting that these 2 variables do not possess discriminant validity. Fornell & Larcker (1981) suggest that the AVE calculated for each latent variable should be greater than the squared correlations between each of the variables. This test compares the relative amount of variance explained by the latent variable and the latent variable's relationship with other variables in the measurement model. When a latent variable's association with other variables in the measurement model is stronger than its relationships with its manifest indicators, it points to a lack of discrimination between the 2 variables (i.e., they are measuring the same underlying construct). In our analyses, all but 1 of the squared latent factor correlations were below each of the variable AVEs. The squared correlation between AR and PN ($r^2 = 1.033$) (Table 4) was larger than each variable's AVE, suggesting a lack of discriminant validity.

We individually fixed each of the correlations between the latent variables (i.e., 6 pairs of correlations) to equal 1.0. We then used the chi-squared difference test (Byrne 1998) with 1 df to provide an empirical indication of whether this constraint affected model fit (Bagozzi & Phillips 1982). The results illustrated that the difference between the fixed and free solutions were all statistically significant (Table 4), providing evidence of discriminant validity.

Unlike the tests for convergent validity that revealed relatively minor concerns (i.e., 3 factor loadings below 0.707), testing for discriminant validity revealed more serious problems. Of the 3 tests, 2 revealed problems relating to the empirical distinction between AR and PN. Conceptually, these variables each link an individual's compulsion to act in an environmentally responsible manner to personal obligation. The variable's operationalization reflects this sentiment (Table 3). The testing illustrated redundancy. In this situation, there were 2 choices: remove 1 of the variables from the model

Table 4. Squared correlations (r^2), CIs of latent variable correlations, and change in model fit ($\Delta\chi^2$) after constraining latent variable correlations to unity with 1 df in a study of recreational.

	Angler identity (AI)	*Awareness of the consequences (AC)*	*Ascription of responsibility (AR)*	*Personal norms (PN)*
AI	1.00			
AC	$r = 0.04$, CI $= -0.04$, 0.12 latent factor r constraint $= \Delta\chi^2 = 366.68$ $r^2 = 0.001$	1.00		
AR	$r = 0.53$, CI $= 0.46$, 0.59 latent factor r constraint $= \Delta\chi^2 = 271.54$ $r^2 = 0.277$	$r = 0.33$, CI $= 0.26$, 0.40 latent factor r constraint $= \Delta\chi^2 = 329.08$ $r^2 = 0.108$	1.00	
PN	$r = 0.49$, CI $= 0.43$, 0.56 latent factor r constraint $= \Delta\chi^2 = 274.39$ $r^2 = 0.243$	$r = 0.37$, CI $= 0.30$, 0.45 latent factor r constraint $= \Delta\chi^2 = 299.89$ $r^2 = 0.139$	$r = 1.02$, CI $= 1.00$, 1.03 latent factor r constraint $= \Delta\chi^2 = 208.73$ $r^2 = 1.033$	1.00

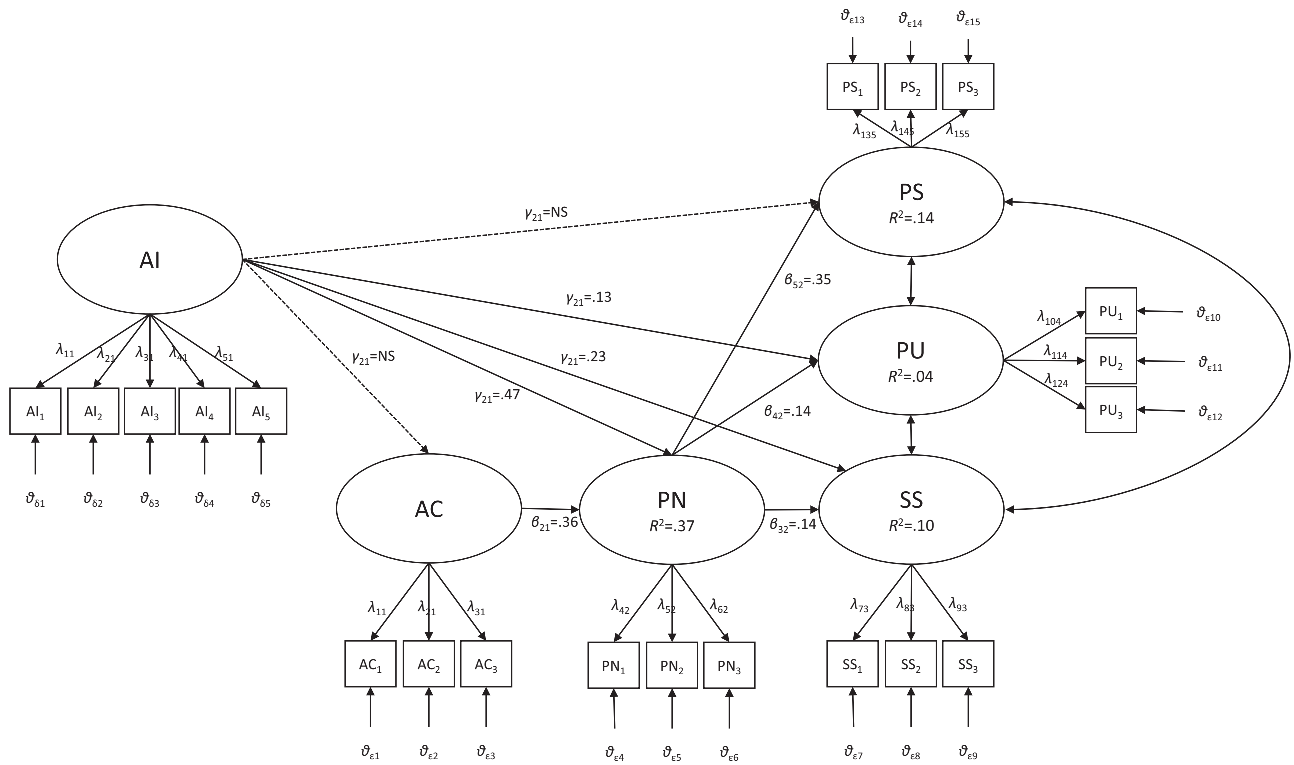

Figure 2. Final structural model ecosystem stewardship reported by Landon et al. (2018) (AI, angler identity; AC, awareness of consequence; AR, ascription of responsibility; PN, personal norms; PS, private-sphere behavior; PU, public-sphere behavior; SS, social stewardship behavior).

or merge the 2 variables to form a single latent variable. Unfortunately, the latter alternative violates theory and past empirical evidence. Each of these variables has a rich history in the literature that has demonstrated their distinction both conceptually and empirically. Other authors (MacCallum 1986; MacCallum et al. 1992; Brown 2015) argue against post hoc model modifications that are atheoretical and are potential idiosyncrasies of the data and context in which the data were collected that are unlikely to be replicated. Retaining both constructs in the model could generate inconsistent estimates, as a function of their multicollinearity. Therefore, we removed AR from the model and reran the measurement model ($\chi^2 = 225.32$, df $= 41$, RMSEA $= 0.075$, CFI $= 0.953$, NFI $= 0.937$) (Fig. 2). Although conservative, it does not contradict the literature. Ascription of responsibility was chosen for removal ahead of PN owing to its stronger convergent validity (e.g., strength of factor loadings), reliability (discussed below), and position as an indirect antecedent of behavior in the underlying theory of the model being tested. Both statistical evidence and theory informed the decision.

Finally, the reliability of each of the variables was assessed by examining the internal consistency (i.e., Cronbach's alpha) and the composite reliability (aka Jöreskog's rho; Raykov 1997) of the latent variables. These measures of reliability bear conceptual similarity to convergent validity because they provide an assessment of how consistently the manifest indicators measure the latent variable (Nunnally 1951). Indicators that are correlated with one another are assumed to be consistent in their assessment of that variable and deemed reliable. Departure from convergent validity relates to the examination of the indicators' relationship with other items and variables in the measurement model. Research shows that coefficient alpha is a good estimate of reliability in conditions of tau equivalence (i.e., factor loadings are equal) and error terms are not correlated (Novick & Lewis 1967; Raykov 1997; McDonald 1999). In applied research, however, this assumption often does not hold, leading to an alpha that provides a lower bound estimate of reliability (Sijtsma 2009). In the context of congeneric measures (i.e., inequality among the factor loadings for a unidimensional latent factor), an alternate estimator such as composite reliability is recommended. Both alpha (α) and composite reliability (ρ) are reported in Table 3. The lower bound cutoff for alpha is 0.70 (Nunnally 1951) and 0.60 for composite reliability (Bagozzi & Yi 1988).

Structural Associations

In step 5, following establishment that the measurement model is psychometrically adequate the (steps 1–4), the structural model is examined (Fig. 2). The parameters of specific interest in this step are the structural paths that lead from the exogenous (independent) to endogenous (dependent) latent variables (i.e., regression coefficients γ and β). Beyond model fit, these parameters are evaluated in terms of their strength, valence, and the variance they explain in the endogenous latent variables. The testing specified in Landon et al.'s (2018) examination of the Schwartz's (1977) NAM also provides evidence of predictive validity of the theory's tenets (Hagger et al. 2017). For a detailed account of step 5, see Kline (2016) and Schrieber et al. (2006).

Additional Tools for Establishing Reliability and Validity

Although we report several techniques to assess the validity and reliability of measures of latent variables, other approaches exist. For construct validity, a common limitation relates to the use of single measurement scales and cross-sectional data. Within CFA questions remain concerning the extent to which the solution is, in part, an artifact of method effects (e.g., convergent validity may be influenced by similarly worded items). When each variable is assessed by the same measurement approach (e.g., self-reports), it cannot be determined how much of the shared variance among factors is due to the common method effect as opposed to the true covariance among factors. In response Campbell & Fiske (1959) developed the multitrait-multimethod matrix (MTMM) to establish construct validity. The approach requires that variables (referred to as *traits* by Campbell and Fiske [1959], i.e., attitudes, norms, personality characteristics) be assessed using different methods. The approach is commonly used in clinical assessment and education evaluations. Despite its potential, human dimensions researchers have been reluctant to utilize the technique. A notable exception was Corral-Verdugo & Figueredo's (1999) assessment of the convergent and discriminant validity of 3 measures of conservation-related behavior (recycling). In correspondence with Campbell & Fiske (1959), their analysis identified method biases (e.g., self-report vs. independent observation) that undermined the construct validity of their measures.

For reliability other tests exist. One is the test-retest approach, which involves taking a measurement with the same indicators and sample at 2 different times. The correlation between the 2 sets of scores is calculated to provide an estimate of reliability (Pedhazur & Schmelkin 1991). Unlike the MTMM approach, test-retest reliability is reported commonly by human dimensions researchers across a variety of conservation-related contexts (Kaiser & Wilson 2000; Kim et al. 2007; Markle 2013; Al Menhali et al. 2018). There are assumptions, however, that can make the application of this test-retest method difficult in applied contexts. First, test–retest assumes measures are stable over time, which can be an onerous requirement in a time of rapid social and environmental instability. Second, temporal stability can be an artifact of a carry-over effect stemming from the first administration of the test (Pedhazur & Schmelkin 1991). Third, social desirability biases can also result in the appearance of stability. For example, there is an inherent desire to act in environmentally responsible ways (Oerke & Bogner 2013), such that scores on these measures tends to appear stable over time. Finally, the applied nature of conservation research can make accessing the same respondents on multiple occasions infeasible.

A second method to examine scale and indicator reliability is alternative forms. This method is like test–retest; however, it involves 2 different measures of the same variable at time T1 and T2. The extent to which 2 different indicators are correlated at T1 and T2 provides insight on their reliability. Although the method has the same practical limitations of test–retest, it is less susceptible to carry-over effects because of the use of different indicators. Also, because it uses different items there is less likelihood of covariation in error terms (Bollen 1989). Unlike test–retest, we were not able to identify any published work in the human dimensions of conservation that applied this method.

A third method is the split-half approach, which assumes a number of manifest indicators are available. Half of the items are combined to form 1 variable and the other half are combined to form a second variable with the same conceptual and empirical meaning. The result is 2 tests and 2 new measures that examine the same variable. The correlation between the halves provides a reliability coefficient for the whole (Nunnally 1951; Bollen 1989). The split-half approach avoids the carry-over effect of test–retest, but, like alternative forms, it has not been applied in the context of conservation.

Discussion

We illustrated the steps and decisions that underlie the assessment of psychometric adequacy for latent variables. Our testing revealed that 1 variable should be removed due to the lack of discriminant validity. That decision was undertaken with consideration of all the available metrics that facilitate an evaluation of instrumentation. Untangling why there was a multicollinearity issue with AR and PN required returning to theory and operational definitions. Drawing from Schwartz's (1977) NAM, Landon et al. (2018) considered AR an artifact of the individual's awareness of the consequences (AC) stemming from inaction. Respondents' understanding of human's impact on aquatic ecosystems shaped the extent to which they considered their actions' impact on the resource. Awareness of consequence was also antecedent to PN. A belief that people ought to act to protect the environment stimulated a moral imperative to act. Beyond the temporal distinction, PN addressed a compulsion to act to protect fisheries resources and ecosystems relative to a moral obligation, whereas their AR provided insight on the individual's sense of ownership (or not) related to the protection of the resource. Although these variables' measures were adapted from past work (Steg & DeGroot 2010; Landon et al. 2017) in which no issues of collinearity were reported, findings illustrate ex situ operationalization warrants further empirical consideration.

Commenting on the increasing volume of social-science-informed research appearing in outlets that have traditionally published work from the life sciences, St. John et al. (2014:2) noted the increasing incidence (and failure) of "ecologically trained scientists adding social science research to their mainly ecological studies, with greater or lesser success." They contend that low-quality science may be published under these circumstances given editors and reviewers training in the life sciences. They warn of the emergence of a potentially sinister problem: "because researchers read publications in high-impact journals which validate lower quality social science… journals risk institutionalizing poor social science" (p. 2). Teel et al. (2018) echo this sentiment, noting the need for social science to be evaluated by the disciplinary norms from which the work draws. Although the full impact of mismeasurement remains unknown, its pervasiveness and persistence may erode trust and perpetuate misconception. In response, we provided an overview of tools and the normative standards researchers working in quantitative psychology utilize to evaluate psychometric adequacy. As demonstrated, the application of these tools in applied contexts can lead to difficult decisions (e.g., the loss of data). Given many conservation issues emerge within social-ecological systems, the inclusion of a psychometrician within a team to assist with the design and analyses is prudent.. If our measures are not what we claim them to be, then how can claims about the psychological drivers of conservation behavior be accepted? As emerging data illustrates a rapidly warming climate along with myriad other anthropogenic stressors to ecosystems and populations of Earth's biodiversity, has the need for precision been greater?

Literature Cited

American Psychological Association (APA). 2020. APA dictionary of psychology. APA, Washington, D.C. Available from https://dictionary.apa.org/psychometrics (accessed August 2020).

Al Menhali EA, Khalid K, Iyanna S. 2018. Testing the psychometric properties of the environmental attitudes inventory on undergraduate students in the Arab context: a test-retest approach. PLOS ONE **13:**e0195250.

Anderson JC, Gerbing DW. 1988. Structural equation modeling in practice: a review and recommended two-step approach. Psychological Bulletin **103:**411–423.

Babbie E. 2016. The practice of social research. 14th edition. Carnegie Learning, New York.

Bagozzi RP, Phillips LW. 1982. Representing and testing organizational theories: a holistic construal. Administrative Science Quarterly **27**(3)**:**459–489.

Bagozzi RP, Yi Y. 1988. On the evaluation of structural equation models. Journal of the Academy of Marketing Science **16:**74–94.

Bennett N., et al. 2017. Conservation social sciences: understanding and integrating human dimensions to improve conservation. Biological Conservation **205:**93–108.

Bollen KA. 1989. Structural equations with latent variables. Wiley, New York.

Brown TA. 2015. Confirmatory factor analysis for applied research. (2nd ed.). Guildford Press, New York.

Burke P, Stets J. 2009. Identity theory. Oxford University Press, New York.

Byrne B. 1998. Structural equation modeling with LISREL, PRELIS, and SIMPLIS:basic concepts, applications, and programming. Erlbaum Associates, Mahwah, New Jersey.

Campbell DT, Fiske DW. 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin **56:**81–105.

Corral-Verdugo V, Figueredo A. 1999. Convergent and divergent validity of three measures of conservation behavior: the multitrait-multimethod approach. Environment and Behavior **31:**805–820.

Cronbach LJ. 1951. Coefficient alpha and the internal structure of tests. Psychometrika **16:**297–324.

Cronbach LJ, Meehl PE. 1955. Construct validity in psychological tests. Psychological Bulletin **52:**281–302.

De Groot J, Steg L. 2009. Mean or green: which values can promote stable proenvironmental behavior? Conservation Letters **2:**61–66.

de Snoo G., et al. 2013. Toward effective nature conservation on farmland: making farmers matter. Conservation Letters **6:**66–72.

Dillman D, Smyth D, Christian L. 2014. Internet, phone, mail and mixed-mode surveys: The tailored design method. Wiley, Hoboken, New Jersey.

Fishbein M, Ajzen I. 2010. Predicting and changing behavior: The reasoned action approach. Psychology Press, New York.

Fornell C, Larcker DF. 1981. Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research **18:**39–50.

Fulton DC, Manfredo MJ, Lipscomb, J. 1996. Wildlife value orientation: A conceptual and measurement approach. Human Dimensions of Wildlife **1**(2):24–47.

Hagger MS, Gucciardi DF, Chatzisarantis LD. 2017. On nomological validity and auxiliary assumption: the importance of simultaneously testing effects in social cognitive theories applied to health behavior and some guidelines. Frontiers in Psychology **8:** 1933.

Homer PM, Kahle LR. 1988. A structural equation test of the value-attitude-behavior hierarchy. Journal of Personality and Social Psychology **54:**638–646.

Hu L, Bentler PM. 1999. Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal **6**(1): 1–55.

Jorgensen TD, Pornprasertmanit S, Schoemann AM, Rosseel Y. 2019. semTools: useful tools for structural equation modeling. R package version 0.5-2. Available from https://CRAN.R-project.org/package=semTools.

Kaiser FG, Wilson M. 2000. Assessing people's general ecological behavior: a cross cultural measure. Journal of Applied Social Psychology **30:**952–978.

Kaltenborn BP, Andersen O, Vittersø J, Bjerke TK. 2012. Attitudes of Norwegian ptarmigan hunters towards hunting goals and harvest regulations: the effects of environmental orientation. Biodiversity Conservation **21:**3369–3384.

Kenny DA. 2015. Measuring model fit. Available from http://www.davidakenny.net/cm/fit.htm (accessed on January 7, 2019).

Kenny DA. 2019. Enhancing validity in psychological research. American Psychologist **74:**1018–1028.

Kim SM, Zeman C, Kostareva O. 2007. Reliability of the children's attitudes toward the environment scale. Psychological Reports **101:**117–118.

Kline R. 2016. Principles and practice of structural equation modeling. Guilford Press, New York.

Landon AC, Kyle GT, Kaiser RA. 2017. An augmented norm activation model: the case of residential outdoor water use. Society and Natural Resources **30:**903–918.

Landon AC, Kyle GT, van Riper C, Schuett MA, Park J. 2018. Exploring the psychological dimensions of stewardship in recreational fisheries. North American Journal of Fisheries Management **38:**579–591.

Lute ML, Navarrete CD, Nelson MP, Gore ML. 2016. Moral dimensions of human-wildlife conflict. Conservation Biology **30:**1200–1211.

MacCullum RC. 1986. Specification searches in covariance structure modeling. Psychological Bulletin **100:**107–120.

MacCullum RC, Roznowski M, Necowitz LB. 1992. Model modifications in covariance structure analysis: the problem of capitalization on chance. Psychological Bulletin **111:**490–504.

Manfredo MJ, Urquiz-Haas EG, Don Carlos AW, Bruskotter JT, Dietsch AM. 2020. How anthropomorphism is changing the social context of modern wildlife conservation. Biological Conservation, **241:** 108297. https://doi.org/10.1016/j.biocon.2019.108297.

Markle GL. 2013. Pro-environmental behavior: does it matter how it's measured? Development and validation of the pre-environmental behavior scale (PEBS). Human Ecology **41:**905–914.

Mastrangelo ME, Gavin MC, Laterra P, Linklatter WL, Milfont TL. 2014. Psycho-social factors influencing forest conservation intentions on the agricultural frontier. Conservation Letters **7:**103–110.

McDonald RP. 1999. Test theory: A unified treatment. Earlbaum Associates, Mahwah, New Jersey.

Novick MR, Lewis C. 1967. Coefficient alpha and the reliability of composite measurement. Psychometrika **32:**1–13.

Nunnally JC. 1951. Coefficient alpha and the internal structure of tests. Psychometrika **16:**297–334.

Nunnally JC. 1978. Psychometric theory. McGraw-Hill, New York.

Oerke B, Bogner FX. 2013. Social desirability, environmental attitudes, and general ecological behavior in children. International Journal of Science Education **35:**713–730.

Pedhazur EJ, Schmelkin LP. 1991. Measurement, design, and analysis: an integrated approach. Erlbaum Associates, Hillsdale, New Jersey.

Raykov T. 1997. Scale reliability, Cronbach's coefficient alpha, and violations of essential tau equivalence with fixed congeneric components. Multivariate Behavioral Research **32:**329–353.

Rosseel Y. 2012. Lavaan: an R package for structural equation modeling. Journal of Statistical Software **48:**1–36.

Sakurai R, Ota T, Uehara T. 2017. Sense of place and attitudes towards future generations for conservation of coastal areas in the Satoumi of Japan. Biological Conservation **209:**332–340.

Sandbrook C, Adams W, Büscher B, Vira B. 2013. Social research and biodiversity conservation. Conservation Biology **6:**1487–1490.

Schrieber JB, Nora A, Stage FK, Barlow EA, King J. 2006. Reporting structural equation modeling and confirmatory factor analysis results: a review. The Journal of Educational Research **99:**323–338.

Schwartz SH. 1977. Normative influences on altruism. Advances in Experimental Social Psychology **10:**221–279.

Sijtsma K. 2009. On the use, the misuse, and very limited usefulness of Cronbach's alpha. Psychometrika **74:**107–120.

St. John F, Keane A, Jones J, Milner-Gulland E. 2014. Robust study design is as important on the social as it is on the ecological side of applied ecological research. Journal of Applied Ecology **51:**1479–1485.

Steg L, De Groot J. 2010. Explaining prosocial intentions: testing causal relationships in the norm activation model. British Journal of Social Psychology **49:**725–743.

Stryker S, Burke PJ. 2000. The past, present, and future of an identity theory. Social Psychology Quarterly **63:**284–297.

Teel TL, et al. 2018. Publishing social science research in conservation biology to move beyond biology. Conservation Biology **32:**6–8.

Thurstone LL. 1947. Multiple-factor analysis. University of Chicago Press, Chicago.

Vaske JJ. 2019. Survey research and analysis. 2nd edition. Sagamore-Venture Publishing, Urbana, Illinois.