Technological Forecasting & Social Change 173 (2021) 121142

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

journal homepage: www.elsevier.com/locate/techfore

A tailored-fit model evaluation strategy for better decisions about structural equation models

Robert Mai [a], Thomas Niemand [b], Sascha Kraus [c,*]

[a] Grenoble Ecole de Management, Department of Marketing, 12 rue Pierre Sémard, 38000 Grenoble, France
[b] Clausthal University of Technology, Department of Marketing Research, 38678 Clausthal-Zellerfeld, Germany
[c] Free University of Bozen-Bolzano, Department of Economics & Management, Piazza Università 1, 39100 Bolzano, Italy

ARTICLE INFO

Keywords:
Structural equation modeling
Fit indicators
Model evaluation
Review
Monte carlo simulation
Contingency paradigm

ABSTRACT

Proper measurement of technology knowledge and social change enables managers to advance strategies in technology management. Structural equation modeling is the ideal method in Technological Forecasting and Social Change (TFSC) and other leading journals to assess the measurement quality of the relevant decision variables and understand how they are related. However, a myriad of indicators are now available to judge how suitable these measurements are (i.e., how well they fit). Despite a consensus that fit indicators are highly context-dependent and no "one-fits-all approach" emerges, a more contingent perspective is surprisingly missing. To fill this gap, we advocate for a "tailored-fit model evaluation strategy" that is specific to the situation at hand to exploit the particular strengths of fit indicators. Motivated by a synthesis of structural equation modeling in TFSC, our simulation study finds that three critical distinctions regarding (a) model novelty, (b) focus on measurement or structural models, and (c) sample size are vital. The proposed strategy demonstrates that, in many contexts, only a few indicators are recommended to avoid artificially inflated Type I/II errors. We provide a decision tree to reach more accurate decisions in model evaluation in order to better theorize and forecast technological and social challenges.

1. Introduction

Emerging over the past 40 years, structural equation modeling (SEM) became an important tool to assess various issues in quantitative research, such as confirmatory factor analysis (CFA), discriminant validity, and common method bias, as well as complex theoretical models involving mediation and moderation analysis (Williams et al., 2009). What contributed to the success of covariance-based SEM is its ability to quantify empirical fit (Bentler, 1990), that is, the degree to which a model corresponds to the empirical data.[1] It is this model fit on which researchers base their decisions about whether the empirical data fit the theoretical model in order to accept or reject a model when testing a measurement, specific mechanism, or even whole theories.

The researcher's toolbox contains a myriad of fit indicators rooted in different principles (e.g., chi-squared distribution values or fit indices). For these indicators, thresholds are defined to separate correct from misspecified models. Recently, these cutoff values also account for the size of the model and sample (Niemand and Mai, 2018). Such "flexible" cutoffs can principally be derived for all fit indicators, which nearly doubles the number of candidates when evaluating a model. Scholars are likely puzzled about which indicators to consult because all fit indicators have specific strengths and weaknesses.

This research draws attention to a related problem that has largely been disregarded thus far. Common to most previous approaches is that they seek to identify recommendations which are effective under *all* conditions and *all* purposes; we label this a "one-size-fits-all-purposes" approach. Although the malleability of fit indicators is widely acknowledged, methodological research still fails to provide recommendations that are optimal in the researcher's situation.

This paper, to the best of our knowledge, is the first to suggest shifting our thinking from a "one-size-fits-all-purposes" approach to a "best-to-fit-a-specific purpose" paradigm of model testing. We propose

* Corresponding author.
E-mail addresses: robert.mai@grenoble-em.com (R. Mai), thomas.niemand@tu-clausthal.de (T. Niemand), sascha.kraus@zfke.de (S. Kraus).

[1] While partial least squares (PLS) is another important technique in structural equation modeling, PLS emphasizes prediction and does not rely on fit indicators (Ringle et al., 2020). The present research focuses on covariance-based SEM, but reports PLS in the review. Confirmatory Composite Analysis is another promising technique that can be used for CFA in the near future (Schuberth et al., 2018).

https://doi.org/10.1016/j.techfore.2021.121142
Received 26 February 2021; Received in revised form 17 August 2021; Accepted 18 August 2021
Available online 8 September 2021

the idea of a flexible evaluation strategy that relies on those indicators that are most effective in the given context. We seek to overcome established routines by advancing (and extending beyond) the recent suggestions of more flexible, contingency-based approaches (e.g., Niemand and Mai, 2018) that are still limited to improving cutoffs of more or less precise fit indicators. We propose, instead, to optimize the evaluation strategy systematically and rely on those fit indicators and cutoffs that are most precise for a given research purpose and model conditions. In updating current recommendations, this paradigm shift and the specific guidelines aims to assist scholars in making better decisions about their models.

Technology management research widely uses fit indicators to study questions about technology management and social change in order for the field to benefit from more accurate evaluations of the applied measures or theories. A review of research published in this journal will show that the conditions under which scholars assess their models are very different; fit indicators in the field are likely distorted by diverse sources of variation. The proposed flexible model evaluation strategy accounts for not only model and data characteristics, but also important conceptual considerations guiding the model estimation. Fit indicators are known to differ in the way they are immune to Type I errors (erroneously rejecting a correct model) and Type II errors (erroneously accepting a false model). Certain indicators may therefore be more suitable for testing a novel model for which Type II errors are especially harmful. Others may be better when estimating well-established models for which both error types are required to be more balanced. Likewise, fit indicators differ in their ability to identify the error in the structural model or errors in the measurement model (Hu and Bentler, 1999). A flexible model evaluation strategy exploits the fact that specific fit indicators are more sensitive for assessing structural paths, while others may be more appropriate for testing a measurement model (e.g., CFA). We suggest that a flexible model evaluation strategy should account, at least, for the following three key questions:

(a) The novelty of a model—estimating an established model *vs.* testing a novel model;
(b) Focus on the model—conducting CFA *vs.* testing a theoretically-derived structural model;
(c) Certainty of the evidence (i.e., sample size as an expression of certainty).

Our central premise is that—rather than a silver bullet—users of covariance-based SEM require a model evaluation strategy with clear guidelines about the optimal fit indicator(s) for the purpose of their research. In this paper, we first discuss current methodological advances about fit indicators and establish that they are highly context-dependent. Our review of SEM research in the Technological Forecasting and Social Change (TFSC) literature illustrates this context-dependency in the field, substantiating the need for a flexible evaluation strategy. To provide the basis for such a strategy, an extensive Monte Carlo simulation explores the relevant factors related to the research purpose and studies a comprehensive set of fit indicators under the various conditions. We conclude with guidelines and an easy-to-use decision tree to derive the optimal model evaluation strategy. Finally, we reflect on potential research propositions and limitations.

2. Recommendations from the methodological literature

2.1. Overview of fit indicators

At a very abstract level, two competing paradigms of model fit currently coexist: chi-squared distribution values of model fit (T statistics) and fit indices. Although T statistics and fit indices serve the same purpose, research on both is relatively independent. Covariance-based SEM seeks to overcome the drawbacks of traditional regressions, which disregard (a) the measurement error in the variables, (b) the complex relationships between these variables, and (c) the error due to omitted variables (Jöreskog and Sörbom, 1982, p. 404). This is achieved by comparing the observed (empirical) covariance matrix with an estimated (fitted) covariance matrix that rests on the conceptual model of latent variables (or factors) and items (Hu and Bentler, 1999, p. 426). By minimizing the discrepancy function of the estimated covariance matrix for a sample size n (Yuan et al., 2015, p. 380), the corresponding T statistic assesses the fit of both matrices. For a model to be identified as correctly specified, the empirical T value should not significantly differ from zero and should not exceed the theoretical chi-square distribution's critical value.

Unfortunately, T is notably sensitive to sample size (n-sensitivity). T values are inflated for small samples while being smaller than they should be for larger samples (e.g., Bentler and Bonett, 1980; Curran et al., 2002). Likewise, T increases with model size (p-sensitivity) because the estimated covariance matrix expands, such as when more manifest variables (p) are included (e.g., Kenny and McCoach, 2003; Moshagen, 2012). T is traditionally derived through Maximum Likelihood-estimators, so deviations from multivariate normality (hereafter termed mvn-sensitivity) can also affect the assumptions of T (e.g., Boomsma, 1985; Fouladi, 2000). These issues are not novel (e.g., Jalal and Bentler, 2018; McNeish, 2020), and different ways were proposed to address them.[2] Most of these suggestions focus on the corrections of the T statistic, which can be roughly clustered into two groups: corrections for the n- and p-sensitivity of the T statistic (Bartlett, 1951; Swain, 1975; Yuan, 2005; Yuan et al., 2015) and corrections for the mvn-sensitivity (Jiang and Yuan, 2017; McNeish, 2020; Satorra and Bentler, 2001).

However, research on these corrections is inconclusive because all corrections heavily depend on the study context (e.g., Fouladi, 2000; Herzog et al., 2007; Jackson et al., 2013; Savalei, 2010; Yang et al., 2018). Table 1 summarizes important empirical studies on these issues (a detailed discussion of the different correction procedures is provided in the Appendix). Overall, no correction type emerged as superior because, to some degree, all are "heuristics" (McNeish, 2020, p. 5). Prior approaches addressing the shortcomings of the T statistic assume that an unknown chi-square distribution is met by incorporating information regarding the model or data. However, this is often specific to certain (simulation) assumptions.

The second paradigm follows an approach that is not based on testing T itself. Instead, a comparison standard is taken, for which a "good" model scores high on an index ranging between 0 and 1 (Bentler and Bonett, 1980). These "fit indices" can be roughly grouped into "goodness-of-fit" indices (optimum: 1, e.g., GFI) and "badness-of-fit" indices (optimum: 0, e.g., RMSEA, SRMR; Appendix for further details). These are "absolute fit indices" because they are solely based on the estimated model. Additionally, it is reasonable to compare the estimated model with a baseline model that assumes all correlations and loadings to be zero. Indices that rely on such relative interpretation are termed "incremental fit indices" (e.g., NFI, TLI, CFI).

Compared to direct corrections of T, fit indices share a few important commonalities. First, fit indices are largely descriptive; all available indices (except for RMSEA) cannot be used to estimate a confidence interval for inference (Chen et al., 2008). Second, all fit indices are a function of T (except for SRMR), so they inherit some of T's sensitivity and can be subject to n-, p-, and mvn-sensitivities (Table 1). Consequently, fit indices are also heavily contingent on the many sources of variation other than model misspecification (e.g., Bagozzi and Yi, 2012; Fan and Sivo, 2005; Niemand and Mai, 2018; Sharma et al., 2005).

[2] We hereafter refer to them by "T" and the first letters of the surnames of authors, TB, TS, TY, TY1, TY2, TY3, TJY and TSB. Since the approach by McNeish (2020) is F-distributed, we term this FN.

Table 1
Important empirical and simulative investigations regarding fit indicators

Study	Type of investigation	Focus Indices	Cutoffs	Purpose	Model	Uncer-tainty	Major recommendations
Bagozzi and Yi (1988)	Example data focusing on Type I & Type II errors (NHST)	●				●	T should be non-significant and yield adequate power. Further, cutoffs for an incremental fit index (IFI ≥ .9) and adjusted goodness of fit index (AGFI ≥ .9) are recommended as additional criteria to indicate appropriate fit. Both fit indices are not recommended in later research.
Hu and Bentler (1998)	Simulation study focusing on Type I & Type II errors (NHST)	●				●	Comparing a wide set of indices, SRMR is found to be the most sensitive to structural model misspecification, while TLI, IFI, CFI/RNI are most sensitive to measurement model misspecification. SRMR and one of the other indices should be combined.
Hu and Bentler (1999)	Simulation study focusing on Type I & Type II errors (NHST)	●	●			●	Based on the recommended dual-index representation of Hu and Bentler (1998), cutoffs of ≤ .09 for SRMR and ≥ .95 for CFI/RNI, TLI, or ≤ .06 for RMSEA, or ≥ .90 for MC yield an acceptable balance of Type I and II errors.
Cheung and Rensvold, (2001)	Simulation study focusing on M and SD of chi-square values		●				Fit evaluation should account for sampling and parsimony errors that jointly affect the criteria's distribution. Hence, the probability distribution of fit indicators should be used, not arbitrary cutoffs.
Marsh et al. (2004)	Simulation study focusing on Type I & Type II errors (NHST)	●	●			●	Replicating Hu and Bentler (1998, 1999) and adding larger sample sizes showed that T and T/df are superior when the sample size is N = 2,500 or N = 5,000. Hence, dual-index representation cutoffs are not generalizable.
Fan and Sivo (2005)	Simulation study focusing on Type I & Type II errors (NHST)	●	●			●	Replicating Hu and Bentler (1998, 1999), the proposed dual-index representation is found to not be optimal, and more contingencies such as more complex model misspecifications are needed.
Sharma et al. (2005)	Simulation study focusing on Type I & Type II errors (NHST)	●				●	GFI, TLI/RNI, NNCP, and RMSEA are investigated regarding variations in sample size, the number of items, factor loadings, and factor correlation. TLI/RNI or NNCP should be coupled with RMSEA.
Nye and Drasgow (2011)	Simulation study focusing on Type I & Type II errors (NHST)	●				●	Non-normality, together with sample size, affects RMSEA, and SRMSR when using DWLS. Regressions for both indices are applied to obtain cutoffs given an acceptable level of Type I error.
Williams and O'Boyle, (2011)	Example data from six studies focusing on Type I & Type II errors (NHST)	●	●				Path model-based versions of NSCI and RMSEA outperformed traditional CFI and RMSEA indices and, thus, should be used for theoretical model evaluation.
Niemand and Mai (2018)	Simulation study focusing on Type I & Type II errors (NHST)		●			●	Flexible cutoffs yielded lower Type I & Type II error rates across various distorting patterns (sample size, model size, factor loadings, non-normality) and particularly for the 'gray area' of moderately misspecified models. A conservative level of uncertainty (α = .05) is recommended for flexible cutoffs.
McNeish (2020)	Simulation & example study focusing on the distribution of chi-square and F-values, Type II error	●	●			●	Using F-tests instead of chi-squared T-tests helped overcome rejection issues of T and fit indices (only in example data) with small samples.
This study	**Simulation study focusing on Type I & Type II errors (NHST)**	●	●	●	●	●	**No *one-fits-all* approach exists. Flexible cutoffs generally perform better than fixed cutoffs, but selecting appropriate indicators (esp. SRMR) and cutoff paradigms depend on the research purpose, model focus, and data certainty.**

Notes. Chronological order. Purpose: Established model vs. novel model. Model: CFI vs. SEM. Uncertainty: Sample size. Names of indices are used as in the studies—e.g., IFI from Bagozzi and Yi (1988) is equal to BL86 in Hu and Bentler (1998)—then standardized to ease understanding. NHST: Null Hypothesis Significance Testing.

2.2. Cutoffs to separate correct and misspecified models

Given that fit indices do not follow a testable theoretical distribution, methodologists recommend specific cutoff points to judge whether a model shows a "good" fit. For goodness-of-fit indices, scholars typically apply a "larger or equal to"-logic (e.g., CFI ≥ .95), while a "smaller or equal to "-notion is used for badness-of-fit indices (e.g., SRMR ≤ .08). To increase the accuracy of such cutoff points drawing a dividing line between correct and misspecified models, Hu and Bentler (1999) proposed combinations of indices in dual-index strategies (e.g., CFI/TLI/RMSEA and SRMR), which are currently the standard in model evaluation (Bagozzi and Yi, 2012). As such, cutoffs are universal, but fit indices are contingent on other factors; scholars argue that fixed, general cutoffs are often inappropriate (Fan and Sivo, 2005; Marsh et al., 2004). The distribution of fit indices vary with the size of the sample and model characteristics, along with the distribution of the data itself, which changes the implied levels of Type I and II errors. Remarkably, cutoff recommendations are either outdated (Bagozzi and Yi, 1988; Baumgartner and Homburg, 1996) or ignore this issue of contingency (Bagozzi and Yi, 2012).

To account for the fact that fit indices follow unique "sui generis" distributions specific to the model and sample (Cheung and Rensvold, 2001, p. 248), Niemand and Mai (2018) introduced "flexible" cutoffs that are determined through an array of contingency variables, including sample size, model size, reliability (magnitude of factor loadings), and non-normality. These cutoffs build on correctly specified models for which the unique fit distributions are individually determined, depending on the sample size (n = 100 to 1,000), model size (number of latent variables from 2 to 10, number of items from 2 to 10), factor loadings (.7, .8, .9), and the degree of non-normality (no, moderate and severe). These flexible cutoffs were derived from the quantiles of the left (goodness-of-fit indicators) or right tail (badness-of-fit), given a predetermined value (e.g., 5%) to allow for error. This flexible approach was only applied to fit indices (CFI, TLI, RMSEA, SRMR)

(Niemand and Mai, 2018), but the T statistic and its corrections are still ignored.

Academic and managerial research will substantially benefit from contingent cutoffs for the T statistic because of its known distribution and the potential to compare fit; for instance, when contrasting nested theoretical models (West et al., 2012). Another more important shortcoming of the current contingency-based approach is the limited focus on optimizing cutoffs. Certain fit indicators may be less sensitive or imprecise in specific situations. Advancing this contingency notion and extending beyond the empirical issues surrounding fit indices and cutoffs, the flexible paradigm developed in this article posits that the model evaluation should also take into account the conceptual considerations that guide a research project.

2.3. From "One-to-Fit-All-Purpose" to a "Best-to-Fit-a-Specific Purpose"

As reasoned above, different solutions have been proposed to account for the contingency of fit indicators (whether T corrections or fit indices) on data characteristics (e.g., sample size, non-normality issues) and model characteristics (e.g., the number of latent variables). Surprisingly, what has been ignored thus far is that these unique distributions of fit indicators also have unique implications from a conceptual perspective. For example, Bagozzi and Yi (2012) recommend combining CFI and SRMR with values of .93 and .07, respectively, which stands in contrast to earlier recommendations by Hu and Bentler (1999) of .95 and .09. Consequently, CFI became more lenient and SRMR more conservative, with notable implications at a conceptual level. Concretely, SRMR is particularly sensitive to misspecification in the structural model, whereas CFI is more sensitive to misspecification in the measurement model (Table 1). The updated recommendations for this index pair, therefore, increased the likelihood to accept models where the measurement model is more relevant, such as for CFA.

Evidently, prior recommendations in the literature were made with the intent to provide one-size-fits-all recommendations. Several scholars propose a combination of at least two specific indicators (Bagozzi and Yi, 2012; Baumgartner and Homburg, 1996; Nye and Drasgow, 2011; Niemand and Mai, 2018; Sharma et al., 2005); while other scholars advocate to rely on a single criterion (Cheung and Rensvold, 2001; McNeish, 2020). The present research proposes the novel idea of a tailored-fit evaluation approach that answers calls to be more contingent ("flexible") in the indicator selection (e.g., Cheung and Rensvold, 2001; Nye and Drasgow, 2011; Williams et al., 2009). Rather than focusing on a small subset of possible indicators, we apply a holistic approach and consider a large body of established and newly proposed fit indicators to identify those that are optimal for the purpose of a given research project.

As the malleability of the common fit indictors also has implications at an epistemological level, a flexible model evaluation strategy should extend beyond accounting for data and model characteristics to include the research purpose when selecting the appropriate fit indicator(s) and cutoff(s). We propose that the model evaluation strategy needs to be guided by considerations related to the research purpose and, specifically, by three key questions:

(a) whether an established model is estimated or a novel model is tested,
(b) whether CFA is conducted or primarily a theoretically derived structural model is tested, and
(c) the extent certainty that is present in the empirical data (e.g., sample size).

Novel vs. established model. The first fundamental question that needs to be answered is whether scholars seek to estimate a well-established model or whether they are testing a novel model that has not been specified before and is being tested for the first time. This question is relevant because fit indicators provoke different types of errors. When testing a model, two types of errors can occur: Type I errors ("erroneously rejecting a *correct* model") and Type II errors ("erroneously accepting a *misspecified* model"). Note that the pattern for hypothesis testing in covariance-based SEM is opposite to traditional null hypothesis testing with classic multivariate techniques.[3]

We argue that, depending on the research purpose, Type I and Type II errors may be differentially relevant when fitting a model. When testing a novel model that is specified for the first time, Type II errors are arguably more relevant than Type I errors. Following a critical rationalist principle, the consequences for science are much more harmful when erroneously accepting a misspecified model than erroneously rejecting a correct model. This notion rests on the principle that underlies hypothesis testing where both errors are weighted differently (e.g., often with a 4:1 ratio, such an accepted α of .05 and accepted β of .20). Bear in mind that both errors are mutually dependent and cannot be improved simultaneously: reducing Type I error in model estimation provokes inflated Type II errors. Not surprisingly, fit indicators with lower Type I error have inflated Type II errors, and vice versa. For this reason, balancing both types of errors is essential and should be aligned to the research purpose. When estimating a model that is well established in the literature and which was replicated numerous times (e.g., Theory of Planned Behavior, Ajzen, 1991), it is plausible to enhance the importance of Type I errors of erroneously rejecting a correct model, for example, adopting a 2:1 ratio or weighting both error types equally. Fit indicators vary in the likelihood of rejecting models and, therefore, are more or less efficient in accepting correct models or detecting misspecification. As a consequence of this variability, researchers may need to rely on different indicators, depending on the purpose of their research. In a similar fashion, flexibly derived cutoff points tend to be more conservative in the model assessment, which may also affect the balancing of Type I and Type II errors. A flexible model evaluation strategy makes use of these specific characteristics that fit indicators and cutoffs exhibit under the different conditions. Consequently, the selection of the appropriate fit indicator(s) and cutoffs are expected to be contingent on the research purpose.

CFA vs. structural models. Covariance-based SEM is applied for different tasks with two principal foci, namely, conducting CFA (also for establishing discriminant validity or assessing a common method bias) or confirming a theoretical model (Bagozzi and Yi, 2012). In terms of CFA, the loading patterns among manifest and latent variables are of primary interest. The error in the measurement model is particularly relevant here, whereas the correlations among latent variables are assumed to be complete, which makes errors in the structural model rather unlikely. By contrast, in research that focuses on estimating the structural paths, the measurement models (i.e., the constructs) were often validated beforehand (CFA's focus). The model's theoretically derived structure is in focus here, making structural model misspecification more important than misspecification of the measurement model. However, fit indicators are differentially effective in detecting misspecifications in measurement or structural parts of the model (Niemand and Mai, 2018). A flexible model evaluation strategy is therefore needed that may refer to different fit indicators for conducting CFA or testing a structural model.

Sample size. As a third major consideration, the flexible model evaluation strategy should take into account the certainty about the empirical evidence. We consider the size of the sample as a representation of certainty. The sample size is the most investigated source of variation in methodological research, directly affecting the indicator's accuracy (Steiger, 1990). Although most prior recommendations

[3] While many statistical tests seek to reject the null where a false rejection (Type I) often is more harmful than erroneous acceptance (Type II), covariance-based SEM seeks to accept the null hypothesis of perfect fit (= a correct model). Here, false acceptance of the null (Type II) would be more hurtful than an erroneous rejection of the null (Type I, Marsh et al., 2004).

acknowledge the importance of sample size, only Hu and Bentler (1999), and Niemand and Mai (2018) explicitly incorporate sample size into their recommendations. In general, the larger the sample size is, the greater the certainty is in model evaluation because both the ratio of the parameters to be estimated and sample size are relaxed (i.e., more data is available to assess the fit of the observed and estimated matrices). Sample size consequentially distorts the relevant fit indicators, leading to potentially biased decisions about the model. The model evaluation strategy should account for the fact that fit indicators differ in the degree to which they are sensitive to a distortion by sample size. For example, Niemand and Mai (2018) found that CFI, TLI and RMSEA are less sensitive to sample size than SRMR.

Apart from the research purpose (novel vs. established model) and the purpose of the model specification (CFA vs. theoretically derived model), a researcher's decision about which fit indicator(s) to consult and what cutoffs to use may consequently also be guided by the certainty related to the available data. This research, for the first time, incorporates these conceptual considerations into the model evaluation strategy. We next conduct a review of the TFSC literature to illustrate the strong context-dependency of SEM applications in this field, substantiating the need for a tailored-fit model evaluation strategy.

3. Review of structural equation modeling in technological forecasting and social change

To gain an overview of how SEM is applied to study technological forecasting and social change, we evaluated all papers published in this journal between 2016 and 2020. We chose a five-year period because recent methodological contributions (e.g., Bagozzi and Yi, 2012; Niemand and Mai, 2018) require some time to diffuse among applied researchers. Initially, an ISSN-specific keyword search for "structural equation", "structural equation model" (or "SEM") or "partial least squares" (or "PLS") in Google Scholar revealed 260 potential full-text papers that were then manually checked for applications of SEM-techniques, yielding a final sample of 72 papers (Appendix), of which 38 papers (52.8%) use PLS (e.g., Kraus et al., 2020), while 34 papers (47.2%) apply covariance-based SEM (e.g., Nadeem et al., 2020).

In the next step, we analyzed the literature regarding our three key questions to test our assumptions that research in TFSC varies considerably regarding the research purpose (established vs. novel model), the specification purpose (CFA vs. SEM), and the certainty (sample size).

Established vs. novel model. Regarding the research purpose, we coded as to whether SEM applications were based on an established model—a (modified) theoretical framework—compared to a novel model. A vast majority of 51 papers (70.8%) investigated a novel model, whereas 21 (29.2%) estimated an established model. Among these established models, the Technology Acceptance Model (TAM, Davis, 1989) was used eleven times (52.4%, e.g., Fotiadis and Stylos, 2017), followed by the related theories of Theory of Planned Behavior (Ajzen, 1991) with three (14.3%, e.g., Youssef et al., 2021) and the Theory of Reasoned Action (Ajzen and Fishbein, 1977) with two applications (9.5%, e.g., Barnes and Mattsson, 2017).

CFA vs. structural models. Next, we coded the purpose for which covariance-based SEM was used (multiple coding possible). Most of the applications tested a theoretical model (57, 79.2%), but about one-third of the papers also used the method for assessing the measurement model through a CFA (25, 34.7%). We also found other purposes, such as detecting common method bias (28, 38.9%) or establishing discriminant validity (40, 55.6%).

Sample size. As expected, our review shows a wide range of sample sizes. On average, SEM is conducted with a median sample size of 265.50 (SD = 264.37), while PLS is applied to smaller sample sizes (median = 245.50, SD = 259.59) than covariance-based SEM (median = 359.88, SD = 269.97), ranging from 47 (Blohmke et al., 2016) to 1,156 respondents (Gali et al., 2020). In these SEM studies, 68.1% collected larger samples (>200), and 31.9% used a small sample (< 200). Overall, this review confirms a wide variance in the TFSC literature regarding our three major factors, which underscores the need to account for them in a flexible model evaluation strategy.

Commonly applied fit indicators. We further identified the most frequently applied fit criteria, which comprise TLI (18, 52.9%), followed by RMSEA and CFI (both 15, 44.1%) and SRMR (4, 11.8%). T was applied in 11 cases (32.4%). It is noteworthy that the combinations proposed by Hu and Bentler (1999) and tested in other studies (e.g., Mai and Niemand, 2018) are rarely used (CFI & SRMR: 4, 11.8%, TLI & SRMR: 3, 8.8%, RMSEA, and SRMR: 4, 11.8%). To date, TFSC researchers do not take advantage of the unique benefits of fit indicators by paring SRMR with a second fit index (see, e.g., Gali et al., 2020 for an exception). We often found that scholars report extensive sets of fit criteria. For example, de Zubielqui et al. (2019) consulted T, T/df, CFI, TLI, and RMSEA. Remarkably, combinations of T with fit indices occurred frequently (with CFI or RMSEA: 9, 26.5%, with TLI: 8, 23.5%, with SRMR: 2, 5.9%). Evidently, combinations between the competing paradigms to determine model fit (T statistics and fit indices) were more common in the TFSC literature than combinations within the same paradigm.

We lastly investigated the occurrence of fit indices regarding these criteria via logit regressions. CFI and RMSEA (Odds ratio [OR] = 16.75, $z = 3.77$, $p < .001$ for both) as well as SRMR (OR = 6.22, $z = 2.28$, $p < .05$) were more likely to be applied for CFAs than for theoretical models. T instead is not related to any criteria covered in our sample. These results generated two insights. First, fit indices were more often used in TFSC studies that seek to validate the measurement model in a CFA. Second, T is neither related to sample size nor the purpose of the research (novel vs. established) and model specification (CFA vs. structural paths). This implies that T is used rather descriptively.

Overall, the widely used fit indicators do not appear specifically linked to the research purpose (established vs. novel) or sample size—two of our three major considerations for a flexible model evaluation strategy. This observation about the TFSC literature pinpoints that current model evaluations in the field do not appear to exploit the unique benefits of certain fit indicators. As technology management research may benefit from a more tailored-fit approach, we incorporated the three major considerations into a flexible model evaluation strategy and identified the appropriate fit indicators and cutoffs. To this end, we conducted an extensive simulation study that manipulates a broad range of relevant conditions.

4. Simulation study

4.1. Objective and procedure

The simulation study serves to develop the flexible model evaluation strategy and has two steps. First, data were simulated for different conditions to compare the fit indicators (corrections of T and fit indices) and their dependency on misspecification-unrelated factors. For these conditions, we secondly estimated the flexible cutoff values for T and its corrections as well as a comprehensive set of fit indices. Our data was obtained through a Monte Carlo simulation based on multivariate random number generators, using *R* and the *simulateData* function in package *lavaan*. A population model for CFA was initially defined by setting all latent variable correlations to .3 and all factor loadings between latent variables and items to equal, standardized values. The population model does not contain misspecifications and is thus a correct model.

For the first step, we systematically manipulated five model and sample factors in that population model: factor loadings (3 levels), the number of latent variables (4 levels), the number of items (4 levels), the degree of non-normality (3 levels), and the sample size (4 levels). The relevant levels of the manipulated factors were selected adopting the simulation procedure by Niemand and Mai (2018). Because of this paper's focus, we made a few changes, such as adding factor loadings of .6

to the levels of .7 and .8. We omitted the level of .9 because these loadings are likely to yield unrealistically high-reliability estimates for the latent variables, while loadings below .6 can provoke convergence issues, specifically for small samples (Gagne and Hancock, 2006). Additionally, sample size (100, 200, 500, and 1,000) is modified from the original sizes (125, 250, 500, 1,000) to avoid inconsistent use in the reference study (125 is only used in study 2, 1,000 only in study 1) and to obtain an equal ratio of smaller (100, 200) and larger (500, 1,000) samples. Finally, we followed the reference simulation procedure regarding the number of latent variables (2, 4, 6, 8), the number of items per latent variable (2, 3, 4, 5), and the manipulation of multivariate normality for skewness and kurtosis (no: 0 and 0; moderate: 1 and 3.5; severe: 2 and 7).[4] Overall, 576 conditions were created.

Based on these settings, each population model is then applied to estimate a robust Maximum Likelihood-estimator CFA, applying the standard options in *lavaan* (set free: intercepts of items, intercepts of latent variables, residual variances, variances of latent variables, covariances of latent variables; set to 1: first indicator per latent variable). For each condition, 1,000 replications are estimated. Further, the individual models were varied in four ways to manipulate the (mis)specifications in the measurement model and the structural equation model: (a) a correctly-specified model with no errors in the measurement model (MME = 0) and the structural model (SME = 0); (b) a model with one error in the measurement model (by switching the first items of latent variable A and latent variable B); (c) a model with one error in the structural model (by restraining the correlation between A and B to 0 instead of .3); and (d) a model version including both error types. Given the results from Niemand and Mai (2018), we opted not to introduce more errors as these are easily recognized by common fit indicators and are not relevant for our investigation. To illustrate the manipulation procedure, Fig. 1 depicts a model of four latent variables with three items each. This procedure led to a dataset of 2,304,000 cases. Additionally, fourteen relevant fit indicators are extracted directly or calculated.

To derive the flexible cutoff values for the second step, the respective 576 conditions were simulated with 1,000 replications per model type (no errors in the measurement and structural model). Empirical quantiles of type 8 (Hyndman and Fan, 1996) from definition 8 were extracted for a width of p = .05.

4.2. Descriptive findings

Sensitivity of fit indicators. We initially explored the degree to which the fit indicators are distorted. To this end, we investigated the basic patterns of sample size (n), number of items (i), number of latent variables (m), factor loadings (fl), and non-normality (non) within the simulated data. Table 2 illustrates these patterns for the fit indicators based on η^2 from ANOVAs (Sum of squares: Type II). According to Cohen (1988), we classified this sensitivity of the fit indicators as “small” ($\eta^2 \geq .01$), “medium” ($\eta^2 \geq .06$), or “large” ($\eta^2 \geq .14$). This first step in our analysis serves to identify whether a fit indicator exists that is primarily sensitive to misspecifications in the measurement model and the structural model, while simultaneously being immune to the distortions that are induced through the misspecification-unrelated factors related to the data (e.g., sample size, non-normality) and the model (e.g., the number of the latent variables or items).

As shown in Table 2, the T statistic is not only substantively sensitive to the number of items and latent variables; it is moderately sensitive to measurement model misspecification, but also slightly sensitive to sample size. Notably, this pattern persists for all corrections of T, which

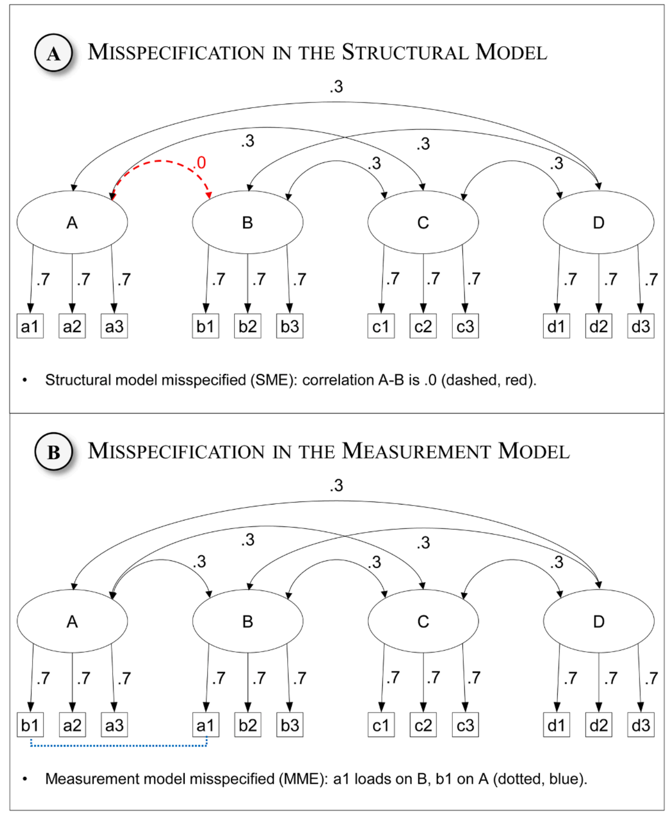

Fig. 1. Examples for different types of model misspecification
Notes. Example specification for four latent variables with three items each. A-D: latent variables (ovals), a1-d3: items (rectangles); two-sided arrows above the latent variables indicate latent variable correlations; one-sided arrows below the latent variables indicate standardized loadings.

implies that the suggested corrections do not help fully remedy the sensitivity of the T statistic to misspecification-unrelated aspects. The Jiang-Yuan-T correction (TJY) is additionally sensitive to sample size (Jiang and Yuan, 2017). It is noteworthy that the recently proposed use of F-tests (FN) shows only moderate sensitivity to the number of latent variables (McNeish, 2020).

With regard to the fit indices, this research replicates the patterns identified in previous research. SRMR is the sole fit indicator that is at least moderately sensitive to misspecification in the structural model ($\eta^2 = .11$, $p < .001$) while also being sensitive to errors in the measurement model ($\eta^2 = .24$, $p < .001$). Still, this fit index responds substantially to sample size and the number of latent variables (Table 2). This observation corresponds with the finding in our review of the TFSC literature that scholars do not follow a consistent evaluation strategy in covariance-based SEM. Our review further pinpointed that no superior fit indicator emerged in the field. These observations, coupled with the results in Table 2, provide a first indication of the need for a model evaluation strategy that is tailored-fit to the situation at hand. RMSEA, CFI, and TLI share high sensitivity to errors in the measurement model but are also moderately (CFI) or strongly (RMSEA, TLI) affected by the number of latent variables. Owing to the fact that CFI responds to interactions between the misspecification of the structural model and that of the measurement model, the previously proposed combination of CFI and SRMR appears to work best for disentangling the two misspecification types (measurement model and structural model) from the unintended nuisance variance introduced by misspecification-unrelated aspects. The following analyses, therefore, focus on the most relevant indicators CFI, TLI, RMSEA, SRMR, and TSB.

Fixed vs. flexible cutoffs. We now turn our attention to the cutoffs that draw a dividing line between correctly specified models and

[4] A Vale-Maurelli approach (VM) is used to derive non-normal data (*simulateData*-function in *lavaan*). As only one estimator is applied, differences in estimator performance as described by Foldnes and Grønneberg (2015) are irrelevant.

Table 2
Sensitivity of fit indicators for manipulated factors (η^2 in %)

Indicator	Sample size (n)	Number of items (i)	Number of latent variables (m)	Average factor loading (fl)	Non-normality (non)	Structural misspecification (SME)	Measurement misspecification (MME)	SME * MME
T statistic	*4.2*	**32.7**	**31.0**	(.8)	(.2)	(.5)	6.9	(.1)
TB	6.4	**31.9**	**29.0**	(.9)	(.2)	(.5)	7.5	(.2)
TS	*5.7*	**32.0**	**29.7**	(.8)	(.2)	(.5)	7.4	(.2)
TY	6.1	**31.9**	**29.4**	(.8)	(.2)	(.5)	7.4	(.2)
TSB	*3.4*	**33.2**	**33.9**	(.5)	(.0)	(.2)	6.7	(.2)
TJY	**19.8**	**20.1**	**17.2**	(.8)	(.1)	(.4)	9.4	(.3)
TY1	*5.9*	**31.9**	**29.6**	(.8)	(.2)	(.5)	7.4	(.2)
TY2	*5.9*	**31.9**	**29.6**	(.8)	(.2)	(.5)	7.4	(.2)
TY3	*5.9*	**31.9**	**29.6**	(.8)	(.2)	(.5)	7.4	(.2)
FN	*4.2*	*3.6*	11.9	(.7)	(.0)	(.6)	*5.5*	(.2)
CFI	*2.7*	(.5)	9.9	(.0)	(.2)	(.7)	**41.2**	*1.8*
TLI	*1.2*	*3.0*	**16.0**	(.0)	(.1)	(.6)	**34.2**	(.6)
RMSEA	*1.0*	*1.8*	**23.0**	*1.4*	(.1)	(.7)	**30.4**	(.1)
SRMR	**17.8**	*3.9*	10.5	*1.9*	(.1)	11.1	**23.6**	(.4)

Notes. η^2 (in %) from Type II-sum of squares ANOVA (dependent variable: indicator, independent variables: n, i, k, fl, non, SME, MME, and interaction SME * MME). TB = Bartlett-T (Bartlett, 1951), TS = Swain-T (Swain, 1975), TY = Yuan-T (Yuan, 2005), TSB = Satorra-Bentler-T (Satorra and Bentler, 2001), TJY = Jiang-Yuan-T (Jiang and Yuan, 2017), TY1 = Yuan-Empirical-T1 (Yuan et al., 2015), TY2 = Yuan-Empirical-T2 (Yuan et al., 2015), TY3 = Yuan-Empirical-T3 (Yuan et al., 2015), FN = F-test (McNeish, 2020), CFI = Comparative Fit Index, TLI = Tucker Lewis Index, RMSEA = Root Mean Square of Error Approximation, SRMR = Standardized Root Mean Residual. All fit indices (except SRMR) are based on TSB. Sensitivity: in parentheses "insensitive" ($\eta^2 < 1.0$), *italics* "small" ($\eta^2 \geq 1.0$), normal "medium" ($\eta^2 \geq 6.0$), or **bold** "large" ($\eta^2 \geq 14.0$).

misspecified models. Fixed cutoffs rely on a constant threshold (e.g., .95), regardless of the distortions mentioned above regarding the fit index score. Flexible cutoffs (parametrized as "flex") account for these distortions by determining a unique distribution for the given model and data characteristics. To demonstrate how fit indicators respond to the sources of distortion, Table 3 provides the mean values of the flexibly derived cutoffs for all manipulated factors. These results demonstrate that all cutoff values are not constant but do vary, as expected. Regardless of actual model misspecification, fit indices are generally more likely to indicate appropriate fit for larger samples, higher factor loadings, fewer items, and latent variables, as well as normal data. Note that this variation in the cutoff values occurred for correctly specified models and therefore arises regardless of whether actual misspecification is present in the model. A flexible cutoff captures the lowest empirical score that is to be expected for a fit indicator under the given model and data conditions (the maximum score for badness-of-fit indices). For example, when investigating a model based on a sample size of n = 500, 95% of same-sized correctly specified models would show a mean CFI of .958 or higher and an SRMR score of .035 or lower (across all manipulated factors).

4.3. Overall precision of the fit indicators under the different research conditions

Having determined their sensitivity to data and model characteristics, we next address the question of whether conceptual considerations should guide the researcher's choice of fit indicator(s) and cutoffs. Supported by our review of the TFSC literature and the great variety in the field, we propose that different indicators should be consulted. For developing feasible and simple guidelines, we detailed the three critical questions that scholars face when estimating a model and investigating its fit. Our notion of a flexible model evaluation strategy posits that the selection of the optimal fit indicator(s) is contingent on the research purpose (established vs. novel model), the specification purpose (CFA vs. SEM), and the certainty of the evidence (sample size). To compare the accuracy of the fit indicators and cutoffs (fixed and flexible), we calculated their hit rates in model evaluation (Hu and Bentler, 1999)

Table 3
Patterns of derived flexible cutoffs for manipulated factors

Factor	CFI	TLI	RMSEA	SRMR	TSB
Sample size					
n = 100	.769	.663	.077	.077	226.302
n = 200	.893	.831	.052	.055	210.690
n = 500	.958	.929	.032	.035	202.763
n = 1000	.979	.963	.023	.024	200.285
Number of items					
i = 2	.913	.784	.060	.038	48.290
i = 3	.895	.850	.046	.049	134.821
i = 4	.897	.872	.040	.051	252.426
i = 5	.895	.879	.036	.053	404.503
Number of latent variables					
m = 2	.907	.792	.071	.042	28.660
m = 4	.907	.865	.043	.048	112.284
m = 6	.899	.869	.036	.050	251.060
m = 8	.886	.859	.033	.050	448.036
Average factor loading					
fl = .6	.875	.810	.045	.049	209.612
fl = .7	.902	.850	.046	.048	210.125
fl = .8	.923	.880	.046	.046	210.292
Non-normality					
normal	.916	.869	.046	.046	209.958
moderate	.904	.852	.046	.047	209.915
severe	.880	.818	.046	.050	210.157

Notes. Mean values for flexible cutoffs (p = .05) dependent on manipulated factors, CFI = Comparative Fit Index, TLI = Tucker Lewis Index, RMSEA = Root Mean Square of Error Approximation, SRMR = Standardized Root Mean Residual, TSB = Satorra-Bentler-T (Satorra and Bentler, 2001), all fit indices calculated except SRMR based on TSB.

under different conditions.

A hit rate specifies the relative proportion of a positive outcome (i.e., confirming a correct model or identifying a falsely specified model). The T statistic (and all related corrections) indicate confirmation if $p \geq .05$, given a chi-square distribution with the model-implied df (or vice versa for rejecting a model). Hit rates for fit indices with fixed cutoffs apply the thresholds recommended by Hu and Bentler (1999) (CFI, TLI: .95; RMSEA: .06; SRMR for single indices: .08, SRMR in combination with CFI/TLI/RMSEA: .09). Hit rates for flexible cutoffs apply the quantiles derived to estimate the flexible cutoffs ($p = .05$).

Table 4 presents the hit rates for the three key questions (research purpose, specification purpose, certainty of the evidence). The results clearly support our assumption that fit indicators are differentially effective. Beyond the hit rate for the best performing indicator (in bold), the gray bars in Table 4 indicate the deviation of the individual indicator from the best performing indicator in the respective condition. For example, for novel CFA models tested with larger samples (N > 200), RMSEA with flexible cutoffs shows a hit rate of 95.3% and differs by 2.5 from the best-performing indicator in this condition SRMR with a hit rate of 97.8%. Overall, the diverging patterns across the conditions support our notion that the fit indicator selection needs to be tailored. We next elaborate on each of the three decisions that scholars have to make.

4.4. Decision 1: Estimating an established model vs. testing a novel model

The first fundamental decision scholars must make is regarding the purpose of research, which may affect selecting the appropriate index and cutoff (fixed or flexible). Our assumption is rooted in the fact that fit indicators differ in the extent to which they are capable of balancing Type I and Type II errors. Note that both error types are differently important to test a novel model as compared to estimating a well-established model. When studying novel models, researchers typically adopt a critical rationalist perspective and employ a more conservative approach by testing the model with a stronger emphasis on Type II errors (erroneously accepting a false model). Conversely, research that estimates an established model also accentuates avoiding Type I errors (erroneously rejecting a correct model), which requires a more nuanced balance between both types of errors.

To analyze the impact of the research purpose, we systematically varied the weighing of Type I and Type II errors from very conservative (1:4) to equally balanced (1:1). To facilitate the understanding of both errors' implications, we analyzed the calculated hit rates in the different conditions of model evaluation. In the following, we report and spotlight only the results for the best-performing indicators (Table 2) and their most relevant combinations, namely, CFI, SRMR, RMSEA, and TSB (TLI performed comparably to CFI in all conditions). Our results reveal notable differences among the widely used fit indicators and their combinations. As visualized in Fig. 2, the use of flexible cutoffs largely shows improved performance than when applying fixed cutoffs to fit indices. Notably, this difference in the hit rates between both cutoff paradigms increases with more conservative testing (towards a 1:4 ratio of Type I and Type II errors). For this reason, flexible cutoffs appear to be particularly efficient when estimating novel and so-far untested models.

In the analysis, we also address the question of whether researchers should rely on combinations of indicators or whether an indicator exists with superior performance. The flexible form of SRMR seems to perform reasonably well across all manipulated models. As illustrated in Fig. 3, $SRMR_{flex}$ may be considered a "silver bullet" for research when both types of errors are equally relevant (i.e., estimating well-established models). The index performs here even better than in combination with another index (e.g., CFI & SRMR, RMSEA & SRMR), which is due to a lower accuracy of SRMR's counterpart in detecting a correctly specified model. Evidently, index combinations can introduce additive errors from both indices so that the likelihood of falsely rejecting a correct model is artificially inflated if two fit indices are applied in combination. Researchers are hence well-advised to refrain from cherry-picking several fit indicators in their reporting but, instead, rely on specific previously tested fit indicators or their combinations.

In sum, when estimating established models (equal weight to Type I and Type II errors), the following indicators perform best with the following highest hit rates: $SRMR_{flex\ .05} = 90.55\%$, CFI & $SRMR_{flex\ .05} = 89.64\%$, TSB & $SRMR_{flex\ .05} = 89.58\%$. Note that all differences are highly significant ($ps < .001$). By contrast, when evaluating novel models (more strongly weighing Type II), this set of indicators is slightly different: CFI & $SRMR_{flex\ .05} = 88.77\%$, TSB & $SRMR_{flex\ .05} = 88.71\%$, $SRMR_{flex\ .05} = 88.59\%$. Again, all differences are significant (CFI & $SRMR_{flex\ .05}$ vs. TSB & $SRMR_{flex\ .05}$: $p < .001$; TSB & $SRMR_{flex\ .05}$ vs. $SRMR_{flex\ .05}$: $p = .029$). Consequently, while for established models, a single index strategy seems advisable (SRMR), combinations of indices (CFI & SRMR) are recommended in situations when a more conservative approach is required, such as when testing a novel model.

4.5. Decision 2: CFA vs. structural models

We next distinguish our recommendations concerning the purpose of

Table 4
Hit rates contingent on the research purpose and data certainty

	Established models (1:1 weighting)				**Novel models** (1:4 weighting)			
Focus	***CFA*** (measurement model)		***Structural model*** (structural paths)		***CFA*** (measurement model)		***Structural model*** (structural paths)	
Sample size	***small***	***large***	***small***	***large***	***small***	***large***	***small***	***large***
Fixed cutoffs								
CFI	-7.4	**97.1**	-15.9	-35.5	**84.2**	-2.1	-12.2	-59.7
SRMR	-17.2	-39.8	-15.3	-43.2	-33.2	-66.1	-28.9	-72.3
TSB	-6.3	-1.5	-17.4	-19.2	-6.4	-.9	-23.0	-29.9
RMSEA	-13.7	-29.0	-19.9	-40.1	-27.9	-48.7	-36.5	-67.1
CFI & SRMR	-7.4	**97.1**	-15.2	-35.4	**84.2**	-2.1	-11.1	-59.5
TSB & SRMR	-4.4	-2.1	-4.1	-1.6	-1.7	-.6	**68.5**	-.6
RMSEA & SRMR	-12.7	-29.0	-17.6	-39.9	-26.2	-48.7	-32.7	-66.8
Flexible cutoffs (p = .05)								
CFI	-5.0	-3.6	-19.2	-19.2	-14.2	-1.5	-35.6	-27.3
SRMR	**84.1**	-.5	**75.1**	**94.5**	-5.3	**97.8**	-4.1	**94.4**
TSB	-5.0	-.8	-17.6	-19.4	-13.6	-.7	-32.5	-31.2
RMSEA	-6.5	-6.4	-19.3	-19.9	-12.2	-2.5	-31.2	-24.8
CFI & SRMR	-.1	-1.4	-.9	-.9	-3.5	-.3	-3.4	-.3
TSB & SRMR	-.3	-1.5	-.9	-1.0	-3.9	-.3	-3.4	-.3
RMSEA & SRMR	-2.6	-6.7	-3.3	-6.2	-3.8	-2.4	-3.7	-2.2

Notes. Best-performing indicator in the column in **bold**. All other values indicate the difference of the indicator from this best performing indicator under the given condition. Sample size small: $n \leq 200$, large: $n > 200$.

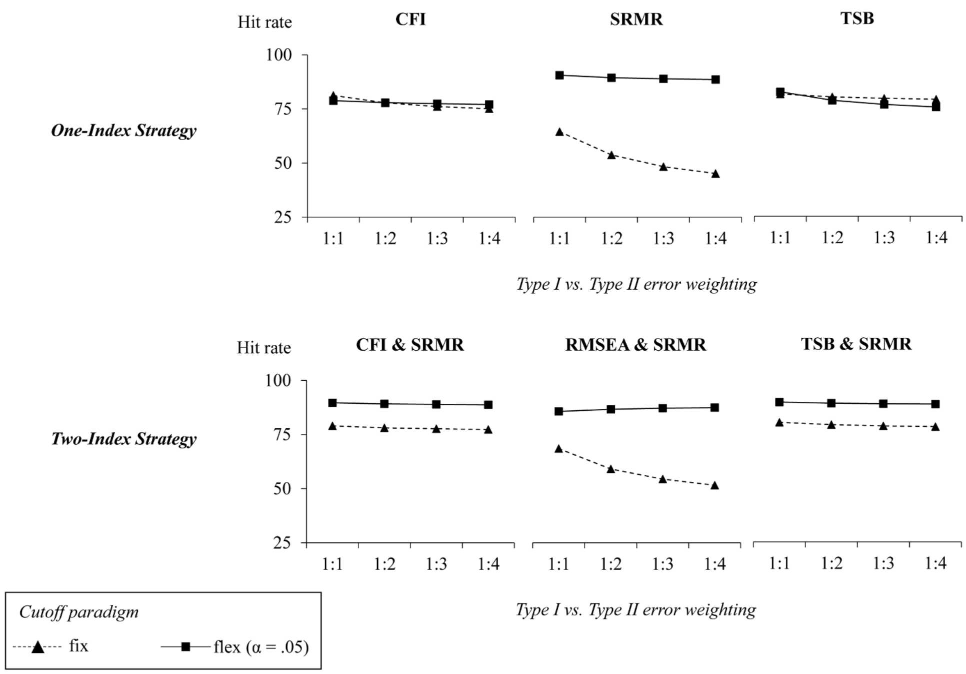

Fig. 2. Comparison of the fixed vs. flexible cutoff paradigm
Notes. Type I to Type II error weighing indicates the relative weight of the respective error type for the overall hit rate (e.g., 1:4 implies that the imprint of Type II error is four times greater than that of the Type I error).

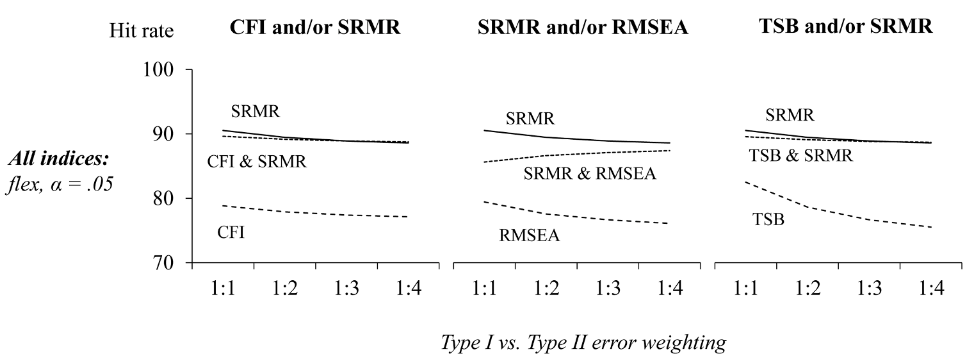

Fig. 3. Comparison of top-performing single indicators vs. combinations
Notes. Type I to Type II error weighting indicates the relative weight of the respective error type for the overall hit rate (e.g., 1:4 implies that the imprint of Type II error is four times greater than that of the Type I error).

model specification. Our results support previous findings that fit indicators are differentially effective in detecting misspecifications in the measurement model and the structural model (e.g., Hu and Bentler, 1999; Niemand and Mai, 2018), such that scholars may need to refer to different fit indicators when running CFAs or to test a theoretically derived structural model. For established models, $SRMR_{flex}$ is best when testing structural models (84.81%) and significantly better than the second-best two-index strategy (CFI_{fix} and $SRMR_{flex}$: 83.89%, $t_{(1152)} = 10.849$, $p < .001$). For novel models, CFI_{fix} (89.98%) as well as the combination of CFI_{fix} & $SRMR_{fix}$ (89.98%) and TSB_{fix} & $SRMR_{fix}$ (89.85%, $t_{(1152)} = .455$, $p > .05$) perform well when conducting CFA. However, for novel structural models, this is the case for fixed (81.17%) and flexible combinations of TSB & $SRMR_{fix}$ (79.59%, $t_{(1152)} = 3.864$, $p < .001$). This is a noteworthy finding because this indicator combination has not been considered or tested thus far.

4.6. Decision 3: Low vs. high certainty

As a third and final major decision, the model evaluation strategy may need to be adjusted to the certainty about the empirical data. All fit indicators are biased by the volume of data. As visualized in Fig. 4, the hit rates of fit indicators substantially depend on sample size. Especially for $SRMR_{fix}$, the indicator's accuracy weakens dramatically with larger samples, irrespective of whether misspecification is present in the model. As was shown in Table 3, the average SRMR values for the very same model specification range from .077 for a sample of 100 subjects to .024 for 1,000 subjects. It is consequently much more likely that the fit score will exceed the fixed cutoff for a misspecified model with larger samples. Flexible cutoffs can partly compensate for this issue; hit rates even increase with larger samples. Nonetheless, flexible cutoffs cannot compensate for fit indicators that are less precise per se. Here, the proposed flexible model evaluation strategy appears to help compensate for individual fit indicator shortcomings by optimizing the selection of the best performing indicators under the given situation (Table 4).

The consequences of the sample size-induced changes in fit indicators are further illustrated in Fig. 4 using the example of SRMR and CFI. The application of flexible (vs. fixed) cutoffs substantially improved the quality of the decision about a model (i.e., hit rates are improved;

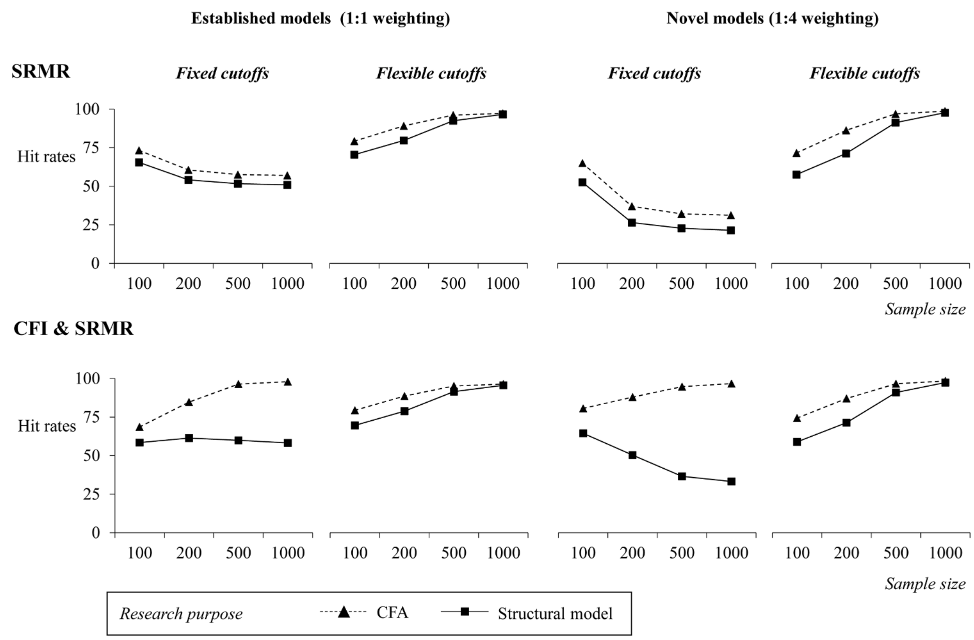

Fig. 4. Hit rates of the fit indicators conditional on the research purpose
Notes. Type I to Type II error weighting indicates the relative weight of the respective error type for the overall hit rate (e.g., 1:4 implies that the imprint of Type II error is four times greater than that of the Type I error).

overall: $t_{(2304)} = 16.638$, $p < .001$). The need for more precise decisions is particularly pressing for novel models that need to be judged more conservatively (Fig. 4). Furthermore, the fit indicators seem to have slightly greater difficulty in detecting misspecification in the structural model than in the measurement model. If both types of misspecification occur in a model (not shown), the hit rates reach the upper bond (i.e., a hit rate of 100%) because most fit indicators successfully detect the presence of several types of misspecification. Evidently, this is not a challenging evaluation of a model.

4.7. Recommendations for model evaluation

Through this simulation study, we found that the precision with which a model is evaluated is contingent on the research purpose,

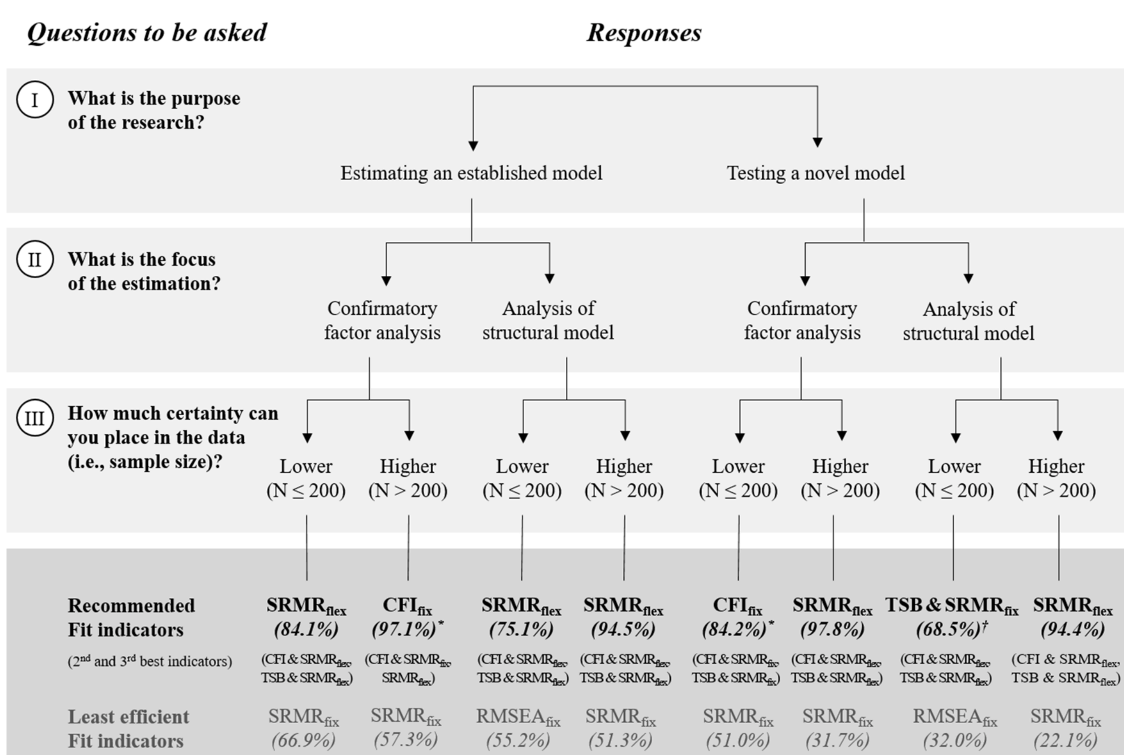

Fig. 5. Decision tree for identifying the optimal fit indicator
Notes. Hit rate of the best-fit indicator are shown in parentheses. *CFI & $SRMR_{fix}$ performed equally well, but is omitted for reasons of parsimoniousness. †: Recommended but needs to be considered with caution due to low overall hit rates in this condition.

purpose of the specification, and certainty about the available data. Researchers are advised to answer three key questions to determine the most effective fit indicator or their combinations. The first question should be, is this a novel model that has not been explored before, or is it already well established in the literature with sufficient evidence? The second question should be, is the purpose of the specified model to verify a measurement model in CFA, or does the research focus on confirming a theoretically derived structural model? The last question should be, is the sample size large enough to offer sufficient statistical power, or is the sample size rather small (e.g., $\leq$ 200) considering the number of parameters to be estimated?

By answering these three simple questions, Fig. 5 provides a simple decision tree to determine the best performing fit indicator(s) and cutoffs (fixed or flexible). The disparity of recommended fit indicators across the different research conditions supports our premise that guided this research. Evidently, fit indicators to evaluate model fit need to be carefully selected, and—although SRMR with flexible cutoffs ($p =$ .05) proved most accurate in many conditions—no one-fits-all-purposes indicator emerged. When testing established models (balancing Type I and Type II errors), a single index performed reasonably well and, in terms of SRMR, even often outperformed combinations of indicators. This is an important observation because scholars may intuitively apply a more-is-better approach with the intent to be more rigorous. Yet, consulting multiple indicators boosts the danger of erroneously rejecting a correct model, inflating Type I errors. By contrast, carefully selected combinations of indicators (esp. TSB & SRMR) are beneficial for novel models, where the ramifications of falsely accepting a misspecified model are exceptionally harmful (i.e., Type II errors).

Our recommendations even suggest a fit indicator that has not been addressed in previous research: TSB is particularly useful when testing structural paths (i.e., detecting a misspecification in the structural model) with small samples (up to 200 subjects) that include substantial uncertainty. The finding is notable because this—according to our review—is a frequent and problematic condition in research on technology management and social change. Bear in mind that for the models estimated in this research, the fit indicators had greater difficulty in detecting misspecification in the structural model compared to misspecification in the measurement model. The simple decision tree for a flexible model evaluation strategy (Fig. 5) therefore helps researchers rely on the most precise fit indicator(s) to make better decisions about their model, measurement, or theory.

5. General discussion

5.1. Discussion of our results in the light of previous recommendations

This research shifted the focus to optimize the model evaluation strategy. The various corrections of T and an extensive set of fit indices with flexible and fixed cutoffs were included in our model evaluation strategy. Rather than searching for a "one-fits-all-purposes" index, our findings suggest exploiting their specific benefits for flexibly adjusting the evaluation strategy to the situation at hand. Accordingly, different fit indicators need to be evaluated depending on (a) the research purpose (testing a novel model vs. estimating a well-established model), (b) the purpose of the model specification (CFA vs. estimating a theoretically derived model), and (c) the certainty that can be placed in the data (small vs. large sample). For these important questions, we developed specific recommendations about which indicators to consult (Fig. 5).

Overall, SRMR with a flexible cutoff detected correct and misspecified models better than all other indices examined in this study. Still, the indicator substantially responded to sample size, such that other indicators showed better performance in specific conditions. For example, CFI with a fixed cutoff of .95 seemed optimal for running CFA on a well-established measurement model using larger samples. This is in line with prior recommendations because—under these rather forgiving conditions—fixed cutoffs work well (Bagozzi and Yi, 2012; Hu and Bentler, 1999; Marsh et al., 2004). Also, fit indicator combinations that have not been proposed so far emerged as optimal in specific conditions. When testing a novel structural model with a small sample, the T statistic TSB combined with SRMR outperforms established fit index combinations, such as CFI & SRMR or TLI & SRMR. Yet, this is limited by the generally low hit rates in these very demanding conditions (i.e., a novel model is evaluated with a rather small sample).

Although hit rates tend to be lower in demanding conditions (e.g., Marsh et al., 2004), previous guidelines fail to address the issue of inflated Type I errors in these conditions. Across all conditions, the pairing of TSB & SRMR with flexible cutoffs seems to be a solid choice because the largest difference compared to the top-performing indicator in each condition does not exceed 3.9% (or 4.4% for fixed cutoffs). The combination of CFI & $SRMR_{flex}$ is even slightly more effective as the hit rate differences do not exceed 3.5% across all model conditions. Although not the intention of this research, TSB & $SRMR_{flex}$ or CFI & $SRMR_{flex}$ would serve the one-fits-all indicator philosophy in a decent number of cases. This is illustrated in Fig. 6, in which a smaller area under the curve indicates better unconditional performance across all conditions.

Assistance by a second indicator is only needed for novel models and smaller samples. Our findings also accord with previous observations that SRMR is better in detecting structural misspecification, whereas CFI and TSB are better in detecting measurement misspecification (Fan and Sivo, 2005; Niemand and Mai, 2018). In this way, our recommendations (Fig. 5) help overcome the misuse (e.g., cherry-picking of fit indicators) or lack of practical simplicity that is often addressed in methodological research (Marsh et al., 2004).

The proposed flexible model evaluation strategy also helps overcome some shortcomings of the recent advances towards more contingency-based approaches in model evaluation. As argued by Niemand and Mai (2018) and consistent with our findings, flexible cutoffs partially remedy the poor performance of widely used cutoff rules in the critical conditions where substantial uncertainty exists in the model and data. However, as they are limited to flexibility in the cutoffs, flexible cutoffs are no safeguard against ineffective indicators and do not ensure researchers use the most precise toolbox they have at hand for their specific problem. Instead of applying the contingency-based notion to a classical "one-size-fits-all-purposes" approach, we widened the flexibility focus towards a "best-to-fit-a-specific purpose" thinking in model testing.

5.2. Implications for future research

This paper and the extensive simulation study provided evidence that a flexible model evaluation strategy improves hit rates and, in turn, the chance to detect correct models or sort out misspecified models. Neither the established combinations of fit indicators with fixed cutoffs nor single or paired fit indicators with flexible cutoffs clearly emerged as being consistently superior in all conditions. Still, in five of our eight recommendations, $SRMR_{flex}$ is the endorsed index. In line with Bagozzi and Yi (2012), this is due to more conservative cutoff values of $\leq$ .06 / .07 (Table 4) compared to the .08 / .09 recommendations by Hu and Bentler (1999). Apart from the challenging case of testing novel structural models with small samples, $SRMR_{flex}$ alone or combined with CFI or TSB performs relatively consistent across the board and better than any other index or index pair, partially confirming Niemand and Mai (2018). The reasons why the combination of TSB with SRMR (fixed or flexible cutoffs) is such a promising index pair needs to be explained through future research. To the best of our knowledge, no study has yet investigated this combination.

Furthermore, the inferior performance of the corrections of the T statistic may stem from the factors manipulated in the simulation study. The corrections are recommended with minimum sample sizes, such as $n \geq \max(50, 2p)$ (Yuan et al., 2015) or $n \geq 4p$ (Shi et al., 2018), indicating that only small models can be analyzed with small samples. It is possible

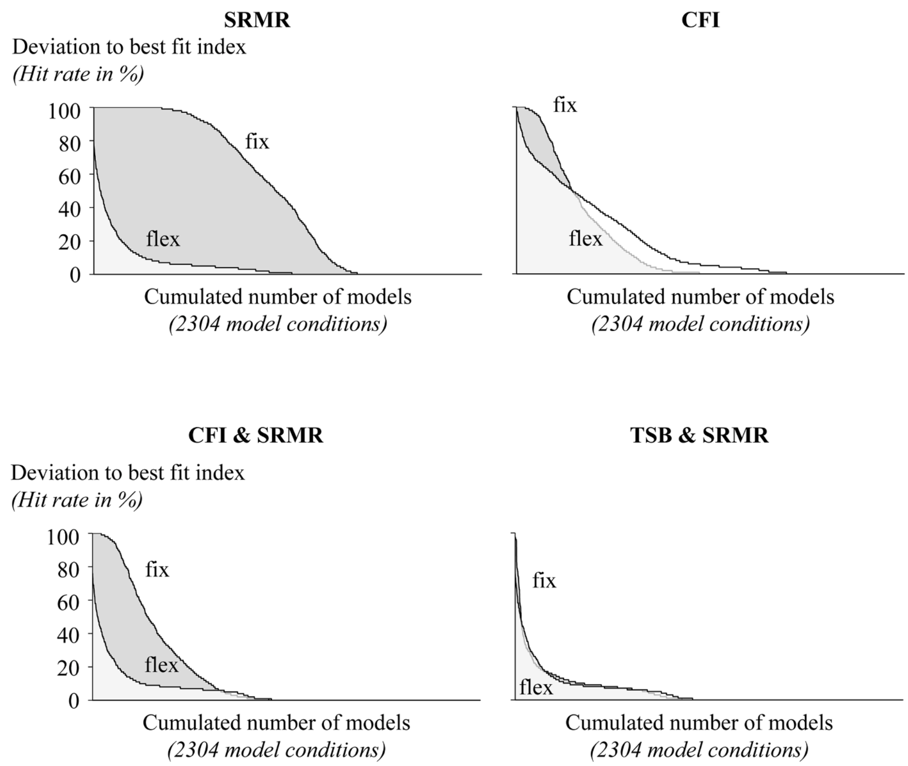

Fig. 6. Cumulated differences of selected fit indicators to the best performing indicator (across all model conditions)
Note. The smaller area under the curve indicates better unconditional performance.

that the corrections of T perform better at the limits of those rules (e.g., a six-factor CFA with two items each, v = 50, against a sample size of n = 100 or 200). Supplementing our decision tree in the future with model size (small vs. large models) may further improve our flexible model evaluation strategy.

Another avenue for further research lies in a more detailed manipulation of the tradeoff between sample and model size (e.g., Gagne and Hancock, 2006) when comparing T-based fit indicators and fit indices. Future simulation studies may start, for example, with a ratio of sample size–to–number of items of 1 and relax this tradeoff with higher ratios (2 to 5) to gain knowledge about the limits at which fit indicator performance increases.

A strength of contingency-based approaches lie in their independence from misspecification. Recommendations for fixed cutoffs (e.g., .95) require the manipulation of rather subjective types of misspecification (Marsh et al., 2004), whereas flexible cutoffs require correct models to be determined (Niemand and Mai, 2018). Our focus on binary levels of SME and MME may be too simplistic here. Future research may investigate less obvious levels (e.g., a weaker correlation restricted to zero) or other forms of misspecification (e.g., equality constraints).

In light of the ongoing discussion on the overreliance on *p*-values (Camerer et al., 2018), we warn against a "dogmatic insistence" in SEM research (Baumgartner and Homburg, 1996, p. 157). Very much as a *p*-value of .051 should not decide alone about rejecting a hypothesis, SRMR values of .081 should not be interpreted as the absolute rejection of a model when the cutoff is .08. Cutoffs can only serve as a benchmark to which fit indicators should come "close to" (Hu and Bentler, 1999, p. 27). The flexible paradigm seconds this understanding because the ideal fit value of 0 (or 1) is often unrealistic given the indicator's dependency on sample and model size. In line with Baumgartner and Homburg (1996), we recommend a subjective interpretation of model fit. Does my fit indicator come close to the benchmark? Although this research included an extensive set of global fit indicators, future research may integrate local fit criteria (such as composite reliability, average variance extracted, or discriminant validity). For example, reliability estimates may detect issues in the measurement model more precisely compared to fit indicators, such as T or CFI.

5.3. Limitations

Some limitations of the simulation design and estimation may weaken the generalizability of our findings. First, despite their independence from misspecification, flexible cutoffs may depend on other design characteristics (e.g., equal factor loadings and correlations among factors). Although being eminent for fixed cutoffs and corrections (e.g., TSB), this issue requires further simulation studies involving more factors and factor levels. For example, studies with less than 100 or more than 1,000 subjects are harmed by the fact that flexible cutoffs were not derived for such levels. Unequal factor loadings, a larger variety of non-normality patterns, or an unequal number of items per factor are also not considered. Since flexible cutoffs require extensive simulations of CFA models (which are non-trivial to compute), the current limiting factor is processing power. Simulating replicable flexible cutoffs for specific models under investigation is a future venture as processing power increases over time.

Second, our review of SEM in TFSC only provides a snapshot of the past five years, demonstrating some tentative results that motivated our research. Evidently, a larger sample (e.g., 20 years of research in TFSC) and a more holistic view of other outlets in the field will generate more precise and generalizable results.

Third, as every simulation study rests on certain assumptions, beliefs, and tradeoffs (Boomsma, 2013), generalizability cannot be achieved in a single study. Despite our efforts to provide feasible recommendations by investigating a variety of factors and distortions relevant to TFSC research, generalizing the results from this research may require some caution.

By proposing the idea of a tailored-fit evaluation strategy, this research aimed to place a fresh take on the ongoing debate about assessing structural equation models. The suggested flexible model evaluation strategy, coupled with the empirical findings, urges us to also account for conceptual considerations as well as the research purpose. This shift in the prevailing paradigm allows the selection of the most powerful fit indicators and, in turn, can lead to better decisions in the domain of technology knowledge and social change. Adopting this flexible principle will support a cleaner measurement of technology knowledge and social change. It will ideally enable managers to advance technology development and social change strategies.

CRediT authorship contribution statement

Robert Mai: Conceptualization, Validation, Writing – original draft, Writing – review & editing. **Thomas Niemand:** Conceptualization, Methodology, Validation, Writing – original draft, Writing – review & editing. **Sascha Kraus:** Writing – original draft, Writing – review & editing, Project administration.

Appendix

Corrections of T

The difference between both matrices follows a (non-)central chi-square distribution and can be written as (Curran et al. 2002, p. 8):

$$\mathrm{TML} = \hat{\mathrm{F}}(\mathrm{n}-1) \tag{Eq. A.1}$$

F is the Maximum Likelihood-estimator of minimizing the discrepancy function of the estimated covariance matrix for a given sample size *n* (Yuan et al., 2015, p. 380). Since this value is not necessarily chi-square distributed in most cases, we hereafter use the term T instead of χ^2 to acknowledge this issue. For Eq. (4), we also use TML to indicate that it is derived from Maximum Likelihood.

Sample size corrections account for the typically small sample sizes in research and can be traced back to factor analysis techniques (Bartlett, 1951) [TB]. Subsequent approaches are proposed by Swain (1975) [TS] and Yuan (2005) [TY]. All three aim to replace n in Eq. (4) by a smaller n estimated through the number of items *v*, the degrees-of-freedom *df*, the, the number of free parameters *q* or the number of moments *m* (in that order: Eqs. (5)-(7)). Since v, df, q, and m estimate the model size, they also help overcome adverse ratios of sample and model size.

$$n^* = n - \frac{v}{3} - \frac{2m}{3} - \frac{11}{6} \tag{Eq. A.2}$$

$$N^* = n - 1 - \frac{\left[v(2v^2+3v-1) - h(2h^2+3h-1)\right]}{12df} \text{ with } h = \frac{\left[\sqrt{(1+8q)}-1\right]}{2} \tag{Eq. A.3}$$

$$n^* = n - \frac{v}{3} - \frac{m}{3} - \frac{13}{6} \tag{Eq. A.4}$$

Non-normality corrections instead use a scaling factor *c* on T to account for non-normal data conditions that are based on an asymptotic robust estimator of a mixture of chi-square distributions with 1 df each (Satorra and Bentler, 2001) [TSB]:

$$\mathrm{TSB} = \frac{\mathrm{TML}}{c} \text{ with } c = \frac{\mathrm{tr(M)}}{\mathrm{df}} \tag{Eq. A.5}$$

Since the full rank of the sample fourth-order moment matrix (M) is not always available, Jiang and Yuan (2017) proposed a version based on the rank of M [TJY]:

$$\mathrm{TJY} = \frac{\mathrm{TML}}{c} \text{ with } c = \frac{\mathrm{tr(M)}}{\mathrm{rank(M)}} \tag{Eq. A.6}$$

As a consequence of inclusive results on the performance of the proposed indicators (TML, TB, TS, TY, TSB, TJY), Yuan et al. (2015) proposed three empirical corrections [TY1, TY2, TY3] based on 342 simulated conditions of n, v, q, and m:

$$n_1^* = n - (2.381 + .361v + .003q) \tag{Eq. A.7}$$

$$n_2^* = n - (2.229 + .365v + .038m) \tag{Eq. A.8}$$

$$n_3^* = n - (2.262 + .369v + .052m - .002q) \tag{Eq. A.9}$$

Results showed that Eqs. (10) to (12) performed predominantly compared to all previous corrections if $n \geq \max(50, 2v)$. This finding is confirmed by Shi et al. (2018), but a ratio of $n \geq 4v$ is recommended to achieve reasonable Type I error rates. Finally, McNeish (2020) recently introduced a variant of testing fit with T by replacing T with *F* by substitution with df [FN]:

$$\mathrm{FN} = \frac{\mathrm{TML}}{\mathrm{df}} \tag{Eq. A.10}$$

This correction is especially suited as an alternative to T and corrected Ts when $n < 200$ and $n/df < 3$ since the overall rejection rates for TML (Type I error) of correct models exceeded 5%. However, other sources of variation, such as model size effects (v, q, m), have not been investigated.

Fit indices

Root Mean Square of Error Approximation (RMSEA) relies on the estimated model (T_T) and its degrees of freedom (df) and sample size (n) (Hu and Bentler, 1998, p. 428):

$$\mathrm{RMSEA} = \sqrt{\frac{\max(T_T - df_T\ 0)}{df_T(n-1)}} \tag{Eq. A. 11}$$

An alternative to RMSEA as an absolute badness-of-fit index is the Standardized Root Mean Square Residual (SRMR, Hu and Bentler, 1998, p. 428):

$$\mathrm{SRMR} = \sqrt{\frac{(\mathrm{Cov_o} - \mathrm{Cov_e})^2 / \mathrm{SD_o}}{\mathrm{v}(\mathrm{v}+1)/2}} \qquad \text{(Eq. A.12)}$$

SRMR is sometimes also termed SRMSR (Nye and Drasgow, 2011). Remarkably, SRMR is the lone badness-of-fit index that is based on the observed ($\mathrm{Cov_o}$) and estimated ($\mathrm{Cov_e}$) covariance matrices instead of T, standardized by the observed standard deviations ($\mathrm{SD_o}$). An example of an absolute fit index is the Goodness of Fit Index (GFI, Jöreskog and Sörbom, 1982, p. 407):

$$\mathrm{GFI} = 1 - \frac{\mathrm{tr}(\mathrm{Cov_e}^{-1}\mathrm{Cov_o} - \mathrm{I})^2}{\mathrm{tr}(\mathrm{Cov_e}^{-1}\mathrm{Cov_o})^2} \qquad \text{(Eq. A.13)}$$

I is a unit matrix of the same dimensions as the covariance matrices. That is, GFI is the lone goodness-of-fit index not based on T. Additionally present in the main script are the Tucker Lewis Index (TLI) and the Comparative Fit Index (CFI, Hu and Bentler, 1998, p. 428):

$$\mathrm{TLI} = \frac{\frac{\mathrm{T_B}}{\mathrm{df_B}} - \frac{\mathrm{T_T}}{\mathrm{df_T}}}{\left(\frac{\mathrm{T_B}}{\mathrm{df_B}} - 1\right)} \qquad \text{(Eq. A.14)}$$

$$\mathrm{CFI} = 1 - \frac{\max(\mathrm{T_T} - \mathrm{df_T}, 0)}{\max(\mathrm{T_T} - \mathrm{df_T},\ \mathrm{T_B} - \mathrm{df_B},\ 0)} \qquad \text{(Eq. A.15)}$$

The same notation as before applies ($\mathrm{T_T}$: T target model, $\mathrm{T_B}$: T base model, $\mathrm{df_T}$: df target model, $\mathrm{df_B}$: df base model).

Finally, the Normed Fit Index (NFI) is simply (Hu and Bentler, 1998, p. 428):

$$\mathrm{NFI} = \frac{\mathrm{T_B} - \mathrm{T_T}}{\mathrm{T_B}} \qquad \text{(Eq. A.16)}$$

TFSC studies included in the review

ID	Authors	Title	Year
1	Hatak, I., Kautonen, T., Fink, M., Kansikas, J.	Innovativeness and family firm performance: The moderating effect of family commitment	2016
2	Chen, S.-C., Hung, C.-W.	Elucidating the factors influencing the acceptance of green products: An extension of theory of planned behavior	2016
3	Yoon, S. N., Lee, D., Schiederjans, M.	Effects of innovation leadership and supply chain innovation on supply chain efficiency: Focusing on hospital size	2016
4	Staub, S., Kaynak, R., Gok, T.	What affects sustainability and innovation — Hard or soft corporate identity?	2016
5	Cleven, A., Mettler, T., Rohner, P., Winter, R.	Healthcare quality innovation and performance through process orientation: Evidence from general hospitals in Switzerland	2016
6	Cui, Y., Sun, C., Xiao, H., Zhao, C.	How to become an excellent entrepreneur: The moderating effect of risk propensity on alertness to business ideas and entrepreneurial capabilities	2016
7	Khasar, S. M. S., Khosla, R., Chu, M. T., Shahmehr, F. S.	Service innovation using social robot to reduce social vulnerability among older people in residential care facilities	2016
8	Blohmke, J., Kemp, R., Türkeli, S.	Disentangling the causal structure behind environmental regulation	2016
9	Staphorst, L., Pretorius, L., Pretorius, M. W.	Technology forecasting in the national research and education network technology domain using context sensitive data fusion	2016
10	Huang, C.-Y., Lin, C.-P.	Enhancing performance of contract workers in the technology industry: Mediation of proactive commitment and moderation of need for social approval and work experience	2016
11	Popa, S., Soto-Acosta, P., Martinez-Conesa, I.	Antecedents, moderators, and outcomes of innovation climate and open innovation: An empirical study in SMEs	2017
12	Barnes, S. J., Mattson, J.	Understanding collaborative consumption: Test of a theoretical model	2017
13	Parguel, B., Lunardo, R., Benoit-Moreau, F.	Sustainability of the sharing economy in question: When second-hand peer-to-peer platforms stimulate indulgent consumption	2017
14	Huang, M.-H., Chen, D.-Z.	How can academic innovation performance in university-industry collaboration be improved?	2017
15	Garcia Martinez, M.	Inspiring crowdsourcing communities to create novel solutions: Competition design and the mediating role of trust	2017
16	Bogicevic, V., Bujisic, M., Bilgihan, A., Yang, W., Cobanoglu, C.	The impact of traveler-focused airport technology on traveler satisfaction	2017
17	Gkypali, A., Filiou, D., Tsekouras, K.	R&D collaborations: Is diversity enhancing innovation performance?	2017
18	Rahman, S. A., Taghizadeh, S. K., Ramayah, T., Alam, M. M. D.	Technology acceptance among micro-entrepreneurs in marginalized social strata: The case of social innovation in Bangladesh	2017
19	Fotiadis, A., Stylos, N.	The effects of online social networking on retail consumer dynamics in the attractions industry: The case of 'E-da' theme park, Taiwan	2017
20	Karikari, S., Osei-Frimpong, K., Owusu-Frimpong, N.	Evaluating individual level antecedents and consequences of social media use in Ghana	2017
21	Radeaelli, G., Lettieri, E., Luzzini, D., Boaretto, A.	Users' search mechanisms and risks of inappropriateness in healthcare innovations: The role of literacy and trust in professional contexts	2017
22	Santoro, G., Vrontis, D., Thrassou, A., Dezi, L.	The internet of things: Building a knowledge management system for open innovation and knowledge management capacity	2017
23	Mital, M., Choudhary, P., Chang, V., Papa, A., Pani, A. K.	Adoption of internet of things in India: A test of competing models using a structured equation modeling approach	2017
24	Osei-Frimpong, K., McLean, G.	Examining online social brand engagement: A social presence theory perspective	2018
25	Osei-Frimpong, K., Wilson, A., Lemke, F.	Patient co-creation activities in healthcare service delivery at the micro level: the influence of online access to healthcare information	2018

(continued on next page)

(*continued*)

ID	Authors	Title	Year
26	Naqshbandi, M. M., Tabche, I.	The interplay of leadership, absorptive capacity, and organizational learning culture in open innovation: Testing a moderated mediation model	2018
27	Popa, S., Soto-Acosta, P., Perez-Gonzales, D.	An investigation of the effect of electronic business on financial performance of Spanish manufacturing SMEs	2018
28	Akhtar, P., Khan, Z., Tarba, S. Y., Jayawickrama, U.	The internet of things, dynamic data and information processing capabilities and operational agility	2018
29	Eva, M.-C., Cegarra-Navarro, J.-G., Garcia-Perez, A., Fait, M.	Healthcare service evolution towards the internet of things: An end-user perspective	2018
30	Leal-Rodriguez, A. L., Ariza-Ontes, A. J., Morales-Fernandez, E., Albort-Morant, G.	Green innovation, indeed a cornerstone in linking market requests and business performance.	2018
31	Gkypali, A., Arvanitis, S., Tsekouras, K.	Absorptive capacity, exporting activities, innovation openness and innovation performance: A SEM approach towards a unifying framework	2018
32	Ahmadi, H., Nilashi, M., Shahmoradi, L., Ibrahim, O., Sadoughi, F., Alizadeh, M., Alizadeh, A.	The moderating effect of hospital size on inter and intra-organizational factors of hospital information system adoption	2018
33	Mettler, T., Pinto, R.	Evolutionary paths and influencing factors towards digital maturity: An analysis of the status quo in Swiss hospitals	2018
34	Andrade-Valbuena, N., Torres, J. P.	Technological reflectiveness from a managerial capability perspective	2018
35	El-Kassar, A.-N., Singh, S. K.	Green innovation and organizational performance: The influence of big data and the moderating role of management commitment and HR practices	2019
36	Dubey, R., Gunasekaran, A., Childe, S. J., Papadopoulos, T., Luo, Z., Wamba, S. F., Roubaud, D.	Can big data and predictive analytics improve social and environmental sustainability?	2019
37	Sepasgozar, S. M. E., Hawken, S., Sargolzaei, S., Foroozanfa, M.	Implementing citizen centric technology in developing smart cities: A model for predicting the acceptance of urban technologies	2019
38	Chen, S.-C., Lin, C.-P.	Understanding the effect of social media marketing activities: The mediation of social identification, perceived value, and satisfaction	2019
39	De Luna, Liebana-Cabanillas, F., Sanchez-Fernandez, J., Munoz-Leiva, F.	Mobile payment is not all the same: The adoption of mobile payment systems depending on the technology applied	2019
40	Singh, S. K., Chen, J., Del Giudice, M., El-Kassar, A.-N.	Environmental ethics, environmental performance, and competitive advantage: Role of environmental training	2019
41	Sanchez-Barrioluengo, M., Benneworth, P.	Is the entrepreneurial university also regionally engaged? Analysing the influence of structural configuration on third mission performance	2019
42	Foroudi, P, Yu, Q., Gupta, S., Foroudi, M. M.	Enhancing university brand image and reputation through customer value co-creation behaviour	2019
43	de Zubielqui, G., Fryges, H., Jones, J.	Social media, open innovation & HRM: Implications for performance	2019
44	Del Guidice, M., Scuotto, V., Garcia-Perez, A., Petruzzelli, A. M.	Shifting Wealth II in Chinese economy. The effect of the horizontal technology spillover for SMEs for international growth	2019
45	Rialti, R., Zollo, L., Ferraris, A., Alon, I.	Big data analytics capabilities and performance: Evidence from a moderated multi-mediation model	2019
46	Nestle, V., Täube, F. A., Heidenreich, S., Bogers, M.	Establishing open innovation culture in cluster initiatives: The role of trust and information asymmetry	2019
47	Herz, M., Rauschnabel, P. A.	Understanding the diffusion of virtual reality glasses: The role of media, fashion and technology	2019
48	Kizgin, H., Jamal, A., Rana, N. P., Dwivedi, Y. K., Weerakkody, V. J. P.	The impact of social networking sites on socialization and political engagement: Role of acculturation	2019
49	Martin-Rojas, R., Garcia-Morales, V. J., Gonzalez-Alvarez, N.	Technological antecedents of entrepreneurship and its consequences for organizational performance	2019
50	Vlacic, E., Dabic, M., Daim, T., Vlacic, D.	Exploring the impact of the level of absorptive capacity in technology development firms	2019
51	Yu, Q., Foroudi, P., Gupta, S.	Far apart yet close by: Social media and acculturation among international students in the UK	2019
52	Ghouri, A. M., Akhtar, P., Shahbaz, M., Shabbir, H.	Affective organizational commitment in global strategic partnerships: The role of individual-level microfoundations and social change	2019
53	Mazzucchelli, A., Chierici, R., Abbate, T., Fontana, S.	Exploring the microfoundations of innovation capabilities. Evidence from a cross-border R&D partnership	2019
54	Li, S., Modi, P., Wu, M.-S. S., Chen, C.-H. S. Nguyen, B.	Conceptualizing and validating the social capital construct in consumer-initiated online brand communities (COBCs)	2019
55	McLean, G., Osei-Frimpong, K.	Examining satisfaction with the experience during a live chat service encounter- implications for website providers	2019
56	Steininger, D. M., Gatzemeier, S.	Digitally forecasting new music product success via active crowdsourcing	2019
57	Lalicic, L., Dickinger, A.	An assessment of user-driven innovativeness in a mobile computing travel platform	2019
58	Singh, S. K., Del Giudice, M., Chierici, R., Graziano, D.	Green innovation and environmental performance: The role of green T transformational leadership and green human resource management	2020
59	Papa, A., Mital, M., Del Giudice, M.	E-health and wellbeing monitoring using smart healthcare devices: An empirical investigation	2020
60	Nadeem, W., Juntunen, M., Shirazi, F., Hajli, N.	Value co-creation in sharing economy: The role of social support, consumers' ethical perceptions and relationship quality	2020
61	Falahat, M., Ramayah, T., Soto-Acosta, P., Lee, Y.-Y.	SMEs internationalization: The role of product innovation, market intelligence, pricing and marketing communication capabilities as drivers of SMEs' international performance	2020
62	Islam, A. K. M. N., Laato, S., Talukder, M. S., Sutinen, E.	Misinformation sharing and social media fatigue during COVID-19: An affordance and cognitive load perspective	2020
63	Bugshan, H., Attar, R. W.	Social commerce information sharing and their impact on consumers	2020
64	Kraus, S., Rehman, S. U., Garcia, F. J. S.	Corporate social responsibility and environmental performance: The mediating role of environmental strategy and green innovation	2020
65	Youssef, A. B., Boubaker, S., Dedaj, B., Carabregu-Vokshi, M.	Digitalization of the economy and entrepreneurship intention	2020
66	Califf, C. B., Brooks, S., Longstreet, P.	Human-like and system-like trust in the sharing economy: The role of context and humanness	2020
67	Gali, N., Niemand, T., Shaw, E., Hughes, M., Kraus, S., Brem, A.	Social entrepreneurship orientation and company success: The mediating role of social performance	2020
68	Rehman, S. U., Kraus, S., Shah, S. A., Khanin, D., Mahto, R.	Analyzing the relationship between green innovation and environmental performance in large manufacturing firms	2020
69	Luo, N., Wang, Y., Zhang, M., Niu, T., Tu, J.	Integrating community and e-commerce to build a trusted online second-hand platform: Based on the perspective of social capital	2020
70	Donbesuur, F., Ampong, G. O. A., Owusu-Yirenkyi, D., Chu, I.	Technological innovation, organizational innovation and international T performance of SMEs: The moderating role of domestic institutional environment	2020

(*continued on next page*)

(*continued*)

ID	Authors	Title	Year
71	Wang, Y., Zhu, Q., Krikke, H., Hazen, B.	How product and process knowledge enable consumer switching to remanufactured laptop computers in circular economy	2020
72	Nastjuk, I., Herrenkind, B., Marrone, M., Brendel, A. B., Kolbe, L. M.	What drives the acceptance of autonomous driving? An investigation of acceptance factors from an end-user's perspective	2020

Ranking of the fit indices under the different conditions

Established models (1:1 weighting)				Novel models (1:4 weighting)			
CFA (measurement model)	SEM (structural model)	CFA (measurement model)	SEM (structural model)	CFA (measurement model)	SEM (structural model)	CFA (measurement model)	SEM (structural model)
Small sample (≤200)	Large sample (>200)	Small sample (≤200)	Large sample (>200)	Small sample (≤200)	Large sample (>200)	Small sample (≤200)	Large sample (>200)
Top 3 fit indices							
SRMR$_{flex}$	CFI$_{fix}$	SRMR$_{flex}$	SRMR$_{flex}$	CFI$_{fix}$	SRMR$_{flex}$	TSB & SRMR$_{fix}$	SRMR$_{flex}$
CFI & SRMR$_{flex}$	CFI & SRMR$_{fix}$	CFI & SRMR$_{flex}$	CFI & SRMR$_{flex}$	CFI & SRMR$_{fix}$	CFI & SRMR$_{flex}$	CFI & SRMR$_{flex}$	CFI & SRMR$_{flex}$
TSB & SRMR$_{flex}$	SRMR$_{flex}$	TSB & SRMR$_{flex}$	TSB & SRMR$_{flex}$	TSB & SRMR$_{fix}$	TSB & SRMR$_{flex}$	TSB & SRMR$_{flex}$	TSB & SRMR$_{flex}$
RMSEA & SRMR$_{flex}$	TSB$_{flex}$	RMSEA & SRMR$_{flex}$	TSB & SRMR$_{fix}$	CFI & SRMR$_{flex}$	TSB & SRMR$_{fix}$	RMSEA & SRMR$_{flex}$	TSB & SRMR$_{fix}$
TSB & SRMR$_{fix}$	CFI & SRMR$_{flex}$	TSB & SRMR$_{fix}$	RMSEA & SRMR$_{flex}$	RMSEA & SRMR$_{flex}$	TSB$_{flex}$	SRMR$_{flex}$	RMSEA & SRMR$_{flex}$
CFI$_{flex}$	TSB$_{fix}$	CFI & SRMR$_{fix}$	TSB$_{fix}$	TSB & SRMR$_{flex}$	TSB$_{fix}$	CFI & SRMR$_{fix}$	RMSEA$_{flex}$
TSB$_{flex}$	TSB & SRMR$_{flex}$	SRMR$_{fix}$	CFI$_{flex}$	SRMR$_{flex}$	CFI$_{flex}$	CFI$_{fix}$	CFI$_{flex}$
TSB$_{fix}$	TSB & SRMR$_{fix}$	CFI$_{fix}$	TSB$_{flex}$	TSB$_{fix}$	CFI$_{fix}$	TSB$_{fix}$	TSB$_{fix}$
RMSEA$_{flex}$	CFI$_{flex}$	TSB$_{fix}$	RMSEA$_{flex}$	RMSEA$_{flex}$	CFI & SRMR$_{fix}$	SRMR$_{fix}$	TSB$_{flex}$
CFI$_{fix}$	RMSEA$_{flex}$	RMSEA & SRMR$_{fix}$	CFI & SRMR$_{fix}$	TSB$_{flex}$	RMSEA & SRMR$_{flex}$	RMSEA$_{flex}$	CFI & SRMR$_{fix}$
CFI & SRMR$_{fix}$	RMSEA & SRMR$_{flex}$	TSB$_{flex}$	CFI$_{fix}$	CFI$_{flex}$	RMSEA$_{flex}$	TSB$_{flex}$	CFI$_{fix}$
Bottom 3 fit indices							
RMSEA & SRMR$_{fix}$	RMSEA$_{fix}$	CFI$_{flex}$	RMSEA & SRMR$_{fix}$	RMSEA & SRMR$_{fix}$	RMSEA$_{fix}$	RMSEA & SRMR$_{fix}$	RMSEA & SRMR$_{fix}$
RMSEA$_{fix}$	RMSEA & SRMR$_{fix}$	RMSEA$_{flex}$	RMSEA$_{fix}$	RMSEA$_{fix}$	RMSEA & SRMR$_{fix}$	CFI$_{flex}$	RMSEA$_{fix}$
SRMR$_{fix}$	SRMR$_{fix}$	RMSEA$_{fix}$	SRMR$_{fix}$	SRMR$_{fix}$	SRMR$_{fix}$	RMSEA$_{fix}$	SRMR$_{fix}$

References

Ajzen, I., 1991. The theory of planned behavior. Organ. Behav. Hum. Decis. Process. 50, 179–211.

Bagozzi, R.P., Yi, Y., 2012. Specification, evaluation, and interpretation of structural equation models. J. Acad. Mark. Sci. 40, 8–34.

Bagozzi, R.P., Yi, Y., 1988. On the evaluation of structural equation models. J. Acad. Mark. Sci. 16, 74–94.

Barnes, S.J., Mattsson, J., 2017. Understanding collaborative consumption: test of a theoretical model. Technol. Forecast. Soc. Change 118, 281–292.

Bartlett, M.S., 1951. The effect of standardization on a χ2 approximation in factor analysis. Biometrika 38, 337–344.

Baumgartner, H., Homburg, C., 1996. Applications of structural equation modeling in marketing and consumer research: a review. Int. J. Res. Mark. 13, 139–161.

Bentler, P.M., 1990. Comparative fit indexes in structural models. Psychol. Bull. 107, 238–246.

Bentler, P.M., Bonett, D.G., 1980. Significance tests and goodness of fit in the analysis of covariance structures. Psychol. Bull. 88, 588–606.

Blohmke, J., Kemp, R., Türkeli, S., 2016. Disentangling the causal structure behind environmental regulation. Technol. Forecast. Soc. Change 103, 174–190.

Boomsma, A., 2013. Reporting Monte Carlo studies in structural equation modeling. Struct. Equ. Model. 20, 518–540.

Boomsma, A., 1985. Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. Psychometrika 50, 229–242.

Camerer, C.F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B.A., Pfeiffer, T., 2018. Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. Nat. Hum. Behav. 2, 637–644.

Chen, F., Curran, P.J., Bollen, K.A., Kirby, J., Paxton, P., 2008. An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. Sociol. Methods Res. 36, 462–494.

Cheung, G.W., Rensvold, R.B., 2001. The effects of model parsimony and sampling error on the fit of structural equation models. Organ. Res. Methods 4, 236–264.

Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences. Lawrence Erlbaum, Hillsdale.

Curran, P.J., Bollen, K.A., Paxton, P., Kirby, J., Chen, F., 2002. The noncentral chi-square distribution in misspecified structural equation models: finite sample results from a Monte Carlo simulation. Multivar. Behav. Res. 37, 1–36.

de Zubielqui, G.C., Fryges, H., Jones, J., 2019. Social media, open innovation & HRM: implications for performance. Technol. Forecast. Soc. Change 144, 334–347.

Fan, X., Sivo, S.A., 2005. Sensitivity of fit indexes to misspecified structural or measurement model components: rationale of two-index strategy revisited. Struct. Equ. Model. 12, 343–367.

Foldnes, N., Grønneberg, S., 2015. How general is the Vale–Maurelli simulation approach? Psychometrika 80, 1066–1083.

Fotiadis, A.K., Stylos, N., 2017. The effects of online social networking on retail consumer dynamics in the attractions industry: the case of 'E-da'theme park, Taiwan. Technol. Forecast. Soc. Change 124, 283–294.

Fouladi, R.T., 2000. Performance of modified test statistics in covariance and correlation structure analysis under conditions of multivariate nonnormality. Struct. Equ. Model. 7, 356–410.

Gagne, P., Hancock, G.R., 2006. Measurement model quality, sample size, and solution propriety in confirmatory factor models. Multivar. Behav. Res. 41, 65–83.

Gali, N., Niemand, T., Shaw, E., Hughes, M., Kraus, S., Brem, A., 2020. Social entrepreneurship orientation and company success: the mediating role of social performance. Technol. Forecast. Soc. Change 160, 120230.

Herzog, W., Boomsma, A., Reinecke, S., 2007. The model-size effect on traditional and modified tests of covariance structures. Struct. Equ. Model. 14, 361–390.

Hu, L., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct. Equ. Model. 6, 1–55.

Hu, L., Bentler, Peter M., 1998. Fit indices in covariance structure modeling: sensitivity to underparameterized model misspecification. Psychol. Methods 3, 424–453.

Hyndman, R.J., Fan, Y., 1996. Sample quantiles in statistical packages. Am. Stat. 50, 361–365.

Jackson, D.L., Voth, J., Frey, M.P., 2013. A note on sample size and solution propriety for confirmatory factor analytic models. Struct. Equ. Model. 20, 86–97.

Jalal, S., Bentler, P.M., 2018. Using Monte Carlo normal distributions to evaluate structural models with nonnormal data. Struct. Equ. Model. 25, 541–557.

Jiang, G., Yuan, K.-H., 2017. Four new corrected statistics for SEM with small samples and nonnormally distributed data. Struct. Equ. Model. 24, 479–494.

Jöreskog, K.G., Sörbom, D., 1982. Recent developments in structural equation modeling. J. Mark. Res. 19, 404–416.

Kenny, D.A., McCoach, D.B., 2003. Effect of the number of variables on measures of fit in structural equation modeling. Struct. Equ. Model. 10, 333–351.
Kraus, S., Rehman, S.U., García, F.J.S., 2020. Corporate social responsibility and environmental performance: the mediating role of environmental strategy and green innovation. Technol. Forecast. Soc. Change 160, 120262.
Marsh, H.W., Hau, K.-T., Wen, Z., 2004. In search of golden rules: comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. Struct. Equ. Model. 11, 320–341.
McNeish, D., 2020. Should we use F-tests for model fit instead of Chi-Square in overidentified structural equation models? Organ. Res. Methods 23, 487–510.
Moshagen, M., 2012. The model size effect in SEM: inflated goodness-of-fit statistics are due to the size of the covariance matrix. Struct. Equ. Model. 19, 86–98.
Nadeem, W., Juntunen, M., Shirazi, F., Hajli, N., 2020. Consumers' value co-creation in sharing economy: the role of social support, consumers' ethical perceptions and relationship quality. Technol. Forecast. Soc. Change 151, 119786.
Niemand, T., Mai, R., 2018. Flexible cutoff values for fit indices in the evaluation of structural equation models. J. Acad. Mark. Sci. 46, 1148–1172.
Nye, C.D., Drasgow, F., 2011. Assessing goodness of fit: simple rules of thumb simply do not work. Organ. Res. Methods 14, 548–570.
Ringle, C.M., Sarstedt, M., Mitchell, R., Gudergan, S.P., 2020. Partial least squares structural equation modeling in HRM research. Int. J. Hum. Resour. Manag. 31, 1617–1643.
Satorra, A., Bentler, P.M., 2001. A scaled difference chi-square test statistic for moment structure analysis. Psychometrika 66, 507–514.
Savalei, V., 2010. Small sample statistics for incomplete nonnormal data: extensions of complete data formulae and a Monte Carlo comparison. Struct. Equ. Model. 17, 241–264.
Schuberth, F., Henseler, J., Dijkstra, T.K., 2018. Confirmatory composite analysis. Front. Psychol. 9, 2541.
Sharma, S., Mukherjee, S., Kumar, A., Dillon, W.R., 2005. A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models. J. Bus. Res. 58, 935–943.
Shi, D., Lee, T., Terry, R.A., 2018. Revisiting the model size effect in structural equation modeling. Struct. Equ. Model. 25, 21–40.
Steiger, J.H., 1990. Structural model evaluation and modification: an interval estimation approach. Multivar. Behav. Res. 25, 173–180.
Swain, A.J., 1975. Analysis of parametric structures for variance matrices/by Anthony. J. Swain. Adelaide.
West, S.G., Taylor, A.B., Wu, W., 2012. Model fit and model selection in structural equation modeling. In: Hoyle, R.H. (Ed.), Handbook of Structural Equation Modeling. Guilford Press, New York, pp. 209–231.
Williams, L.J., O'Boyle Jr., E., 2011. The myth of global fit indices and alternatives for assessing latent variable relations. Organ. Res. Methods 14 (2), 350–369.
Williams, L.J., Vandenberg, R.J., Edwards, J.R., 2009. 12 structural equation modeling in management research: a guide for improved analysis. Acad. Manag. Ann. 3, 543–604.
Yang, M., Jiang, G., Yuan, K.-H., 2018. The performance of ten modified rescaled statistics as the number of variables increases. Struct. Equ. Model. 25, 414–438.
Youssef, A.B., Boubaker, S., Dedaj, B., Carabregu-Vokshi, M., 2021. Digitalization of the economy and entrepreneurship intention. Technol. Forecast. Soc. Change 164, 120043.
Yuan, K.-H., 2005. Fit indices versus test statistics. Multivar. Behav. Res. 40, 115–148.
Yuan, K.-H., Tian, Y., Yanagihara, H., 2015. Empirical correction to the likelihood ratio statistic for structural equation modeling with many variables. Psychometrika 80, 379–405.

Robert Mai is Full Professor of Marketing at Grenoble Ecole de Management. His research focuses on consumer behavior and, more specifically, on sustainability innovations, health care, and food decision making as well as international management and cross-cultural marketing. He publishes his research in academic journals, such as Journal of the Academy of Marketing Science, Journal of Consumer Research, Journal of Consumer Psychology, and Journal of Retailing.

Thomas Niemand is Associate Professor (JProf) for Management and Market Research at Clausthal University of Technology. He holds a doctorate in Business Administration from Dresden University of Technology. His research focuses on advances in market research and the interplay of entrepreneurship, marketing and innovation management. He publishes his research in academic journals, such as Journal of the Academy of Marketing Science, Journal of Product Innovation Management, Industrial Marketing Management and Journal of Business Research.

Sascha Kraus is Full Professor of Management at the Free University of Bozen-Bolzano in South Tyrol, Italy. He holds a doctorate in Social and Economic Sciences from Klagenfurt University, Austria, a Ph.D. in Industrial Engineering and Management from Helsinki University of Technology and a Habilitation (Venia Docendi) from Lappeenranta University of Technology, both in Finland. Before, he held Full Professor positions at Utrecht University, The Netherlands, the University of Liechtenstein, École Supérieure du Commerce Extérieur Paris, France, and at Durham University, United Kingdom.