

# Recovering Pre-Deletion Measurement Quality from Published Statistics: An Inverse Meta-Analysis of the Technology Readiness Index 2.0

Nishant Joshi

nishant.j9@gmail.com

Sankalchand Patel University <https://orcid.org/0009-0004-3788-9066>

---

## Research Article

**Keywords:** technology readiness, TRI 2.0, measurement validity, item deletion, meta-analysis, censored model, questionable measurement practices

**Posted Date:** October 20th, 2026

**DOI:** <https://doi.org/10.21203/rs.3.rs-11405055/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

**Additional Declarations:** The authors declare no competing interests.

---

# Abstract

When a measurement model fails conventional thresholds, authors often delete the weakest items and report statistics only for the remaining items. The Technology Readiness Index 2.0 (TRI 2.0) is the most widely used measure of customers' readiness to adopt technology-based services. In its original validation, the two inhibitor dimensions, discomfort and insecurity, had an average variance extracted (AVE) below .50, yet most published applications report values that meet the usual thresholds. In Study 1, we develop an inverse meta-analysis that recovers pre-deletion AVE and composite reliability from published statistics by treating deleted items as censored observations. Across 94 published studies, inhibitor items were deleted more often than motivator items. Once the deleted items were restored, the inhibitor dimensions had lower AVE and composite reliability than the motivators. It had about one in five insecurity dimensions failing the AVE criterion. A simulation (Study 2) shows that the procedure is accurate and errs on the conservative side. Two raw datasets (Study 3) show the same weakness when nothing is deleted. When a typical author's deletion rule was applied to these data, the reported AVE rose by up to 0.07 and the procedure moved it back toward the full-scale value. The article offers a procedure that can be applied to any multi-item scale for which authors report AVE, CR and alpha, so that the effect of deletion on a body of research can be estimated without raw data. Our advice to researchers is they should report every administered item.

## Introduction

Multi-item scales are rarely reported as they were administered. When a measurement model falls short of conventional thresholds, authors drop the weakest items and report reliability and validity statistics for the items that remain. Each such decision may be defensible but a literature built from such reports describes scales shaped by the data rather than the scales respondents answered. How large this gap is cannot be read from the published values themselves. We take the Technology Readiness Index 2.0 (TRI 2.0) as a test case because it is widely used and its validation reported weak dimensions. Further, its later applications report much stronger ones.

Since Parasuraman (2000) introduced the Technology Readiness Index (TRI), readiness has been measured as a balance of motivators and inhibitors with optimism and innovativeness on one side and discomfort and insecurity on the other. Parasuraman and Colby (2015) later reduced the original 36 items to 16, four per dimension and this shorter TRI 2.0 has become the standard instrument.

A search conducted for this article found 15,387 works published since 2014 that cite either version of the scale or name technology readiness in their title or abstract and 1,427 of them cite Parasuraman and Colby (2015). The construct has also been the subject of a meta-analysis (Blut & Wang, 2020).

From the beginning, the inhibitor dimensions have been the weak part of the scale. In the validation of TRI 2.0, discomfort and insecurity had an average variance extracted (AVE) of .38 and .40 (Parasuraman & Colby, 2015), well below the .50 commonly required for convergent validity (Fornell & Larcker, 1981). The authors attributed this to the complexity of the inhibitor constructs. Published applications seldom show the same problem. Most report discomfort and insecurity with AVE and composite reliability (CR) values that pass the usual thresholds. There are two ways to read this. Either the weakness disappeared once the scale was applied to other

technologies and samples or the published values no longer describe the scale that respondents actually answered.

We argue that the second reading deserves attention because of a routine practice in structural equation modeling. Authors estimate a measurement model, remove items whose loadings fall below a cut-off such as .70 (Hair et al., 2019), re-estimate and report the values for the items that are left. Few report the loadings of the removed items and some do not mention the deletion at all, so the reader sees a scale shaped after the data were in. Flake and Fried (2020) count such undisclosed changes among questionable measurement practices and Hussey and Hughes (2020) show how they hide the weaknesses of established measures. If inhibitor items are removed more often than motivator items, the published record would make the two halves of TRI 2.0 look equally sound even if they are not.

This article estimates how large that difference is. We develop a procedure that we call an inverse meta-analysis. It starts from the AVE, CR and alpha that authors publish, works back to the items they administered and treats every deleted item as a censored observation. This allows the strength of a dimension before deletion to be estimated from published reports (Study 1). Study 2 tests the procedure in simulated literatures where the true values are known and Study 3 turns to two raw datasets in which no item was deleted. What the inhibitor items measure and what they predict is examined in a companion paper (Author, under review).

The article contributes to measurement research in two ways. It shows that the inhibitor dimensions of TRI 2.0 have weak convergent validity and that selective deletion has made the published record look better than the scale is. It offers a procedure that can be applied to any multi-item scale for which authors report AVE, CR and alpha, so that the effect of deletion on a body of research can be estimated without raw data. In fact, we are not aware of an earlier procedure that recovers pre-deletion measurement quality in this way.

## **Conceptual Background**

### **Technology Readiness and Its Measurement**

Parasuraman (2000) proposed technology readiness as a fairly stable individual disposition that combines positive and negative feelings about technology. The idea drew on evidence that consumers hold contradictory views of technology at the same time and find it both liberating and enslaving, efficient and wasteful (Mick & Fournier, 1998). Optimism is a positive view of technology and the belief that it offers people more control, flexibility and efficiency in their lives and innovativeness is a tendency to be a technology pioneer and thought leader. Discomfort is a perceived lack of control over technology and a feeling of being overwhelmed by it. Insecurity is distrust of technology that arises from safety concerns and from worry about its other negative consequences (Parasuraman, 2000; Parasuraman & Colby, 2015).

TRI 2.0 kept some of the original items and added new ones (Parasuraman & Colby, 2015). Users of the scale may analyze the four dimensions separately or combine them into an overall readiness score that is the mean of the four dimensions after discomfort and insecurity have been reverse-coded. Item content and the logic of the overall score are examined in a companion paper (Author, under review).

### **Item Deletion as a Reporting Practice**

Authors who estimate reflective measurement models usually expect standardized loadings of .70 and an AVE of at least .50 (Fornell & Larcker, 1981; Hair et al., 2019). When a dimension falls short, the usual remedy is to drop the weakest item and estimate the model again. Because AVE is the mean of the squared loadings of the items that remain, a deletion raises it almost mechanically. Readers can judge whether this is a defensible refinement only if the deletion, the item and its loading are reported. Reviews of published research show that validity evidence for established scales is often reported selectively or not at all (Flake et al., 2017; Hussey & Hughes, 2020).

Published statistics contain more information than they appear to. Since AVE, CR and alpha are all functions of the loadings of the retained items, values reported together constrain one another and checks based on such constraints have already exposed errors in published means and test statistics (Brown & Heathers, 2017; Nuijten et al., 2016). We apply the same reasoning to measurement statistics. When AVE and CR fix the sums of the retained loadings and when it is known that a deletion removed an item whose loading lay below a cut-off, the values before deletion can be estimated. Our expectations were that inhibitor items would be deleted more often than motivator items, that the inhibitor dimensions would have lower AVE and CR than the motivator dimensions once the deleted items were restored. We further anticipated that the inhibitor dimensions would fail the AVE criterion in raw data in which all items were kept.

## Study 1: Inverse Meta-Analysis of Published TRI 2.0 Reports

Study 1 asks whether the published record understates the weakness of the inhibitor dimensions. A conventional meta-analysis cannot answer this because the AVE, CR and alpha it would pool describe only the items that survived deletion and therefore the scale as authors chose to present it. To estimate the scale as respondents answered it, we developed a procedure that works backward from the reported statistics to the items that were administered and treats each deleted item as a censored observation whose loading is known only to lie below the author's cut-off. We call this an inverse meta-analysis.

## Literature Search and Eligibility

On October 5, 2026, we searched OpenAlex (Priem et al., 2022), an open bibliographic index, for works published in 2014 or later that cite Parasuraman and Colby (2015) or Parasuraman (2000) or that contain "technology readiness," "technological readiness" or "TRI 2.0" in the title or abstract (which would make this procedure fully taxonomy free). The search combined a local copy of the OpenAlex snapshot with queries to the OpenAlex interface because the snapshot held only part of the works citing Parasuraman and Colby (2015) and the same matching rules were applied to both sources. It returned 15,387 unique records, of which 1,427 cite Parasuraman and Colby (2015). Since only studies that cite TRI 2.0 can be expected to use its items, we restricted the retrieval of full texts to these 1,427 records. The 1,403 records with a DOI were imported into a reference manager (Zotero) which located full texts for 317 of them; 308 remained after duplicates, such as the same article published in two languages, were removed. Two further searches for legal open-access copies of the remaining records using OpenAlex, Unpaywall, Semantic Scholar, CORE and Europe PMC, added 198 full texts, for a total of 506, of which 11 duplicated other full texts. Studies that use TRI 2.0 without citing it, or that have no DOI, could not be found this way and records without an accessible full text could not be analyzed. Both are limitations of the search.

Eligibility was decided on the full texts. For the main analysis, a study had to administer the TRI 2.0 items and model the four dimensions as separate first-order constructs. Further it had to report AVE or CR for each construct. Studies that modeled fewer than four dimensions or that combined the dimensions into composite or second-order constructs, were kept for a sensitivity analysis. Of the 495 unique full texts, 161 were excluded because they did not model the TRI dimensions or reported no measurement statistics for them. This left 334 studies, 192 for the main analysis and 142 for the sensitivity analysis (120 with fewer than four dimensions and 22 with composite or second-order constructs).

## Coding

For every study and dimension, we recorded the number of items administered and retained, along with the codes and loadings of any deleted items. We also noted the authors' loading cut-off, the reported AVE, CR and alpha and the loadings of the retained items. A parser that reads table layouts and column headers extracted these values from the full texts. It assigned a printed value to AVE, CR or alpha only when a column header identified it, recomputed AVE and CR from the printed loadings and flagged every dimension for which the two disagreed. As this whole procedure was done by a machine learning algorithm, we personally cross checked 25 studies out of 192. The extracted AVE, CR, alpha and TRI 2.0 four facet values matched the machine learning extracted values.

Usable values were obtained for 297 dimensions from 94 of the 192 main studies and CR for 272 dimensions from 86 studies. Contributing and non-contributing studies had similar sample sizes and the contributing studies were slightly more recent. Where they differed was in reporting. Contributing studies stated a loading cut-off more often (76% against 36%), mentioned a deletion more often (52% against 27%) and more often used covariance-based estimation and consumer samples (all predictors jointly:  $\chi^2(6) = 45.24$ ,  $p < .001$ ).

The results below therefore describe the better-reported part of the literature. Since better-reported studies are also expected to disclose their deletions, this selection understates the effect of deletion than exaggerate it. It means that whatever results we report here are conservative. Where a study did not state its cut-off, we set it to .70, the conventional threshold for standardized loadings and repeated the analysis with .50.

## Analysis

### Recovering the Retained Items

AVE and CR are both functions of two sums of the  $k$  retained standardized loadings,  $S_1 = \sum \lambda_i$  and  $S_2 = \sum \lambda_i^2$ . Following Fornell and Larcker (1981),

$$AVE = S_2 / k \quad CR = S_1^2 / (S_1^2 + k - S_2).$$

Solving the two equations for the sums gives, for a reported AVE, CR and number of retained items,

$$S_2 = k \times AVE \quad S_1^2 = CR(k - S_2) / (1 - CR).$$

If a study provided all retained loadings, we computed the sums from the loadings themselves and used the inverted values only otherwise. A pair of AVE and CR values is feasible only if  $S_2 \leq S_1^2 \leq k \cdot S_2$  and pairs outside this range, after allowing for rounding, were treated as inconsistent and not inverted.

## A Censored Model for Deleted Items

The loadings of deleted items are rarely available in most of the published studies. Each deletion nevertheless carries information, namely that the loading was below the author's cut-off. We used this information in a censored (Tobit) model (Tobin, 1958) of the item loadings, estimated on the Fisher z scale so that loadings stay within their bounds. The z-transformed loading of item j in study s is

$$z_{js} = \mu_j + u_s + e_{js}, \quad u_s \sim N(0, \tau^2), \quad e_{js} \sim N(0, \sigma_f^2),$$

where  $\mu_j$  is the mean of item j,  $u_s$  a study effect and  $\sigma_f$  the residual standard deviation of dimension f. Retained items with a printed loading enter the likelihood through their density. A deleted item without a printed loading enters as left-censored, contributing the probability that its loading lies below the study's cut-off. Retained items whose loadings were not printed enter as right-censored at the cut-off or at .40 when no cut-off was stated which is a conservative lower bound leaving them out would bias the item means downward. The study effect was integrated with 25-point Gauss-Hermite quadrature.

## Pre-Deletion AVE and CR

For each deleted item, the unknown loading was replaced with its expected value below the cut-off,  $E[\lambda \mid \lambda < c]$  and its expected square, both taken from the fitted distribution of that item. Where the loading of a deleted item was printed, we used it as it stood. The pre-deletion values are then

$$AVE_{pre} = (S_2 + \sum E[\lambda^2]) / k_{adm} \quad CR_{pre} = (S_1 + \sum E[\lambda])^2 / [(S_1 + \sum E[\lambda])^2 + k_{adm} - S_2 - \sum E[\lambda^2]],$$

with  $k_{adm}$  the number of items administered. Setting the deleted loadings to zero and to the cut-off gives lower and upper bounds.

## Unreported Deletions

Some authors drop items without saying so and the joint reporting of alpha, AVE and CR makes it possible to detect this. Alpha implies a mean inter-item correlation of  $r_a = \alpha / [k - (k - 1)\alpha]$  while AVE and CR imply another,  $\bar{r} = (S_1^2 - S_2) / [k(k - 1)]$  and both depend on k. For each dimension reported with four items, we looked for the number of items between two and six that brought the two correlations closest together. When that number was below four, we flagged the dimension as a suspected unreported deletion and recomputed it with the smaller number of items.

## Tests

The unit of analysis is a dimension within a study. Because the inhibitor and motivator dimensions of a study share its sample and estimation method, we compared them with linear mixed models that include a random intercept for each study (Pineiro & Bates, 2000). Outcomes were the reported and recovered AVE and CR whether any item had been deleted and how many. We also tested, within each dimension whether deletions

were spread evenly across items ( $\chi^2$  test against administrations). We further tested whether reported AVE values cluster just above .50 by fitting a quadratic Poisson trend to the distribution of AVE values between .30 and .85 leaving out the window from .45 to .55. Then we compared the number of values observed just above .50 with the number the trend predicts (Gerber & Malhotra, 2008).

## Results

The censored model was estimated on 708 item observations from 67 studies, 92 of which were deleted items without a printed loading. Authors reported a deletion for 84 dimensions and the consistency check flagged 30 more as suspected unreported deletions. Table 1 summarizes the reported and recovered values for each dimension and Table 2 reports the test.

**Table 1. Reported and Recovered AVE and CR by Dimension (Main Studies)**

| Dimension      | $k_f$ | $k_{del}$ | $AVE_{rep}$ | $AVE_{rec}$ | $F_{rep}$ | $F_{rec}$ | $F_{unr}$ | $CR_{rep}$ | $CR_{rec}$ |
|----------------|-------|-----------|-------------|-------------|-----------|-----------|-----------|------------|------------|
| Optimism       | 80    | 16        | .669        | .643        | 5.0%      | 8.8%      | 15.0%     | .875       | .867       |
| Innovativeness | 83    | 13        | .634        | .613        | 7.2%      | 10.8%     | 18.1%     | .861       | .856       |
| Discomfort     | 65    | 12        | .629        | .612        | 4.6%      | 7.7%      | 16.9%     | .848       | .846       |
| Insecurity     | 69    | 25        | .627        | .570        | 10.1%     | 21.7%     | 27.5%     | .833       | .811       |

Note.  $k_f$  = dimensions with usable values;  $k_{del}$  = dimensions with at least one reported deleted item; rep = reported (after deletion); rec = recovered (before deletion, censored model); F = share of dimensions with AVE < .50; unr = recovered after recoding suspected unreported deletions. AVE = average variance extracted; CR = composite reliability. Means are unweighted.

In the published values, the inhibitor dimensions looked only slightly weaker than the motivators. Their mean reported AVE was 0.023 lower, a difference that fell just short of significance ( $p = .055$ ). Further, their reported CR was 0.028 lower ( $p = .002$ ). Restoring the deleted items almost doubled both gaps. Before deletion, the inhibitor dimensions had an AVE 0.042 lower and a CR 0.036 lower than the motivator dimensions (both  $p < .001$ ). The AVE difference grew to 0.045 ( $p < .001$ ) when the suspected unreported deletions were included. The choice of default cut-off made little difference; with unknown cut-offs set to .50 instead of .70, the recovered AVE difference was -0.041 ( $p < .001$ ).

**Table 2. Inhibitor Minus Motivator Differences (Linear Mixed Models)**

| Outcome                                    | Facets | Studies | Difference | SE    | p      |
|--------------------------------------------|--------|---------|------------|-------|--------|
| AVE, reported                              | 297    | 94      | -0.023     | 0.012 | .055   |
| AVE, recovered                             | 297    | 94      | -0.042     | 0.012 | < .001 |
| AVE, recovered, unknown cut-offs at .50    | 297    | 94      | -0.041     | 0.012 | < .001 |
| AVE, recovered, incl. unreported deletions | 297    | 94      | -0.045     | 0.012 | < .001 |
| CR, reported                               | 272    | 86      | -0.028     | 0.009 | .002   |
| CR, recovered                              | 272    | 86      | -0.036     | 0.008 | < .001 |
| CR, recovered, incl. unreported deletions  | 272    | 86      | -0.038     | 0.008 | < .001 |
| Any item deleted (probability)             | 297    | 94      | 0.121      | 0.037 | .001   |
| Number of items deleted                    | 297    | 94      | 0.164      | 0.054 | .002   |

Note. Difference = inhibitor (discomfort, insecurity) minus motivator (optimism, innovativeness) dimensions from linear mixed models with a random intercept for each study. Recovered = pre-deletion values from the censored model.

Deletion explains why the published gap is smaller than the recovered one. Inhibitor dimensions were 12.1 percentage points more likely than motivator dimensions to lose at least one item ( $p = .001$ ) and lost 0.16 more items on average ( $p = .002$ ). Insecurity accounts for most of this difference. Authors reported a deletion in 25 of the 69 insecurity dimensions, compared with 12 to 16 in each of the other dimensions. With the deleted items restored, the share of insecurity dimensions with an AVE below .50 rose from 10.1% to 21.7%, and to 27.5% when suspected unreported deletions were counted as well. For the other three dimensions the share stayed between 7.7% and 10.8% (15.0% to 18.1% with unreported deletions).

Results were much the same under the wider eligibility rule. With the sensitivity studies added, 128 studies contributed 353 dimensions. The reported AVE difference became smaller ( $-0.018$ ,  $p = .096$ ) while the recovered AVE difference ( $-0.032$ ,  $p = .006$ ), the recovered CR difference ( $-0.030$ ,  $p < .001$ ) and the difference in the probability of deletion (8.6 percentage points,  $p = .015$ ) kept their direction and significance. Further, 21.5% of insecurity dimensions failed the AVE criterion after recovery. If we consider this, the 34 sensitivity studies with usable values (56 dimensions) showed no difference in AVE or CR. In these studies, motivator dimensions were deleted more often than inhibitor dimensions. Most of them modeled only one or two dimensions, often in adapted form.

Table 3 lists the recovered loadings of the 16 items. The two items with the lowest pre-deletion loadings were also the two deleted most often and both belong to insecurity: INS4, confidence in a business reachable only online (0.628, deleted in 31.4% of the studies that administered it) and INS1, dependence on technology (0.671, 22.0%). INS3 came next (0.718, 15.8%). Every other item had a recovered loading of 0.725 or more and was deleted in no more than 13% of studies. Insecurity also showed the largest spread of loadings across studies ( $\sigma = 0.41$  on the z scale, against 0.26 to 0.34 for the other dimensions).

**Table 3. Recovered Pre-Deletion Item Loadings**

| Item | Loading | SE   | Administered | Deleted | Item | Loading | SE   | Administered | Deleted |
|------|---------|------|--------------|---------|------|---------|------|--------------|---------|
| OPT1 | .773    | .018 | 48           | 6.2%    | DIS1 | .725    | .026 | 31           | 12.9%   |
| OPT2 | .780    | .018 | 48           | 6.2%    | DIS2 | .747    | .024 | 32           | 12.5%   |
| OPT3 | .790    | .017 | 48           | 4.2%    | DIS3 | .793    | .019 | 33           | 6.1%    |
| OPT4 | .776    | .019 | 43           | 7.0%    | DIS4 | .763    | .022 | 33           | 12.1%   |
| INN1 | .735    | .020 | 42           | 9.5%    | INS1 | .671    | .033 | 41           | 22.0%   |
| INN2 | .769    | .017 | 42           | 9.5%    | INS2 | .746    | .026 | 39           | 7.7%    |
| INN3 | .752    | .019 | 42           | 7.1%    | INS3 | .718    | .029 | 38           | 15.8%   |
| INN4 | .778    | .018 | 37           | 8.1%    | INS4 | .628    | .041 | 35           | 31.4%   |

Note. Loading = mean pre-deletion standardized loading from the censored model; Administered = studies in the censored model that administered the item; Deleted = share of these studies that deleted it. OPT = optimism; INN = innovativeness; DIS = discomfort; INS = insecurity. Item content is paraphrased in below Table 4.

**Table 4. Objects Appraised by the TRI 2.0 Items**

| Item          | Content (paraphrased)                                                                                                                           | What agreement indicates                  |
|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------|
| OPT1,<br>OPT3 | New technologies improve the quality of life; technology gives people more control over their daily lives                                       | Societal techno-optimism                  |
| OPT2,<br>OPT4 | Technology gives us more freedom of mobility; it makes us more productive in my personal life                                                   | Personal instrumental benefit             |
| INN1,<br>INN2 | Others ask us for advice on new technologies; We am among the first of my friends to acquire new technology                                     | Early-adopter identity and status         |
| INN3          | I can usually work out new high-tech products and services without help                                                                         | Technology self-efficacy                  |
| INN4          | I keep up with the latest technological developments in my areas of interest                                                                    | Technology involvement                    |
| DIS1,<br>DIS2 | In technical support We sometimes feel taken advantage of by someone who knows more; support lines do not explain things in terms We understand | Power and clarity in technical support    |
| DIS3,<br>DIS4 | Technology systems are not designed for ordinary people; manuals are not written in plain language                                              | Inclusiveness and accessibility of design |
| INS1-<br>INS3 | People depend too much on technology; technology distracts people to a harmful degree; it lowers the quality of relationships                   | Societal and relational critique          |
| INS4          | I am not confident doing business with a place that can only be reached online                                                                  | Channel-specific trust                    |

Note. OPT = optimism; INN = innovativeness; DIS = discomfort; INS = insecurity. Items paraphrased from Parasuraman and Colby (2015).

The two tests, aimed directly at selective deletion, gave weaker evidence. Deletions within insecurity concentrated on INS4 but the  $\chi^2$  test of deletions against administrations was not significant ( $p = .112$ ). For the other dimensions the  $p$  values lay between .84 and 1.00. With 11 to 29 deletions per dimension these tests have little power, so the lack of a pattern should not be taken to mean that items were deleted at random. Reported inhibitor AVE values were more frequent just above .50 than the local trend predicts, with 24 values between .50 and .55 where 10.3 were expected ( $p = .074$ ), compared with 19 against 12.3 for the motivator dimensions ( $p = .152$ ). This excess fits the idea that some deletions were made to reach .50, though it does not reach conventional significance. Conclusions about single items and any distinction between a method factor and random error, should therefore depend on raw data instead of on recovered values.

## Discussion of Study 1

In the published record, the inhibitor dimensions of TRI 2.0 appear almost as sound as the motivator dimensions. Much of that appearance is due to deletion. Authors remove inhibitor items, especially insecurity items, more often than motivator items. The values, they then report, conceal a lower AVE and CR as deletion cuts the AVE gap roughly in half, from 0.042 to 0.023. After the deleted items are restored, more than one in five insecurity dimensions fails the AVE criterion that the published tables appear to meet. It is worth noting which items drive this result. INS4 and INS1 are the two insecurity items whose content is furthest from a personal feeling of insecurity, since one concerns a type of business and the other society's dependence on technology.

A recovered AVE gap of about 0.04 rests on the treatment of the deleted items which is not only the purpose of the method but also its main assumption. Only 94 of the 192 main studies provided usable values. These studies report their measurement more completely than the rest. The censored model draws on item-level information from 67 studies.

Mean recovered AVE values remain above .50 for every dimension which suggests that published studies are not a random sample of TRI 2.0 administrations. Whether the recovery procedure gives accurate values when the truth is known is the subject of Study 2; Study 3 then examines the items in raw data in which nothing was deleted.

The contrast between the two tests is the central result: the published values do not show a reliable difference between inhibitor and motivator dimensions whereas the values before deletion do.

## Study 2: Simulation of the Recovery Procedure

The accuracy of the procedure used in Study 1 cannot be checked against the published record since nobody knows the true pre-deletion values of published studies. For this reason, we created simulated literatures in which those values are known. Simulated TRI 2.0 studies were generated under different structures of the inhibitor items, simulated authors deleted and reported items in the way published authors do. Then the procedure of Study 1 was applied to what they reported so that its results could be compared with the truth. The design follows the recommendations of Morris et al. (2019) for simulation studies.

# Design

Each simulated study administered the 16 TRI 2.0 items to N respondents. Motivator items loaded between .76 and .82 on their dimension while the inhibitor items followed one of four structures. In the random error structure, all eight inhibitor items were equally weak. The broad domain structure let the strength of each inhibitor item vary at random from study to study. In the referent structure, three inhibitor items (DIS1, DIS2 and INS4) shared an additional factor representing a common object of these items rather than the respondent's readiness. INS4 loaded only weakly on its own dimension. The method factor structure gave all eight inhibitor items a shared factor of the kind that negative wording can produce. Across structures, the dimensions correlated at .60 (optimism and innovativeness), .50 (discomfort and insecurity) and .10 (any motivator with any inhibitor).

We used two calibrations of the inhibitor items. In the first, the inhibitor dimensions had a pre-deletion AVE of about .38 to .42, as in Parasuraman and Colby (2015). The second raised the inhibitor loadings by 0.15 which brings the inhibitor AVE close to the values reported in the published studies of Study 1. Samples of 200, 400 and 800 respondents were drawn. Each study was estimated either as a confirmatory factor model by maximum likelihood in lavaan (Rosseel, 2012) or as a partial least squares composite model (mode A), the widely used two approaches found in the TRI literature.

The simulated authors then behaved as published authors do. Each one drew a loading cut-off of .70, .50 or .40 (with probabilities .6, .3 and .1), deleted the item with the lowest loading below the cut-off, re-estimated the model and continued until no loading fell below the cut-off keeping at least two items per dimension. AVE, CR and alpha were reported to three decimals or to two in 30% of studies. Reporting was incomplete in the ways observed in Study 1 (see the Online Appendix for the rates). A study that neither named its deleted items nor printed its loadings reported each dimension as if all four items had been kept which produces the unreported deletions the procedure tries to detect.

One replication was a literature of 40 such studies analyzed with the procedure of Study 1 and each of the 48 cells (four structures, three sample sizes, two estimators and two calibrations) was replicated 100 times. We measured how much the recovered and reported AVE differed from the original AVE before any items were deleted. We also compared the differences between the inhibitor and motivator dimensions and checked how often the test correctly or incorrectly detected deletions that had not been reported.

# Results

Table 5 shows the cells with N = 400 for both calibrations; the pattern at N = 200 and N = 800 was the same. In the second calibration, the inhibitor items were strong enough that few of them were deleted under the random error and method factor structure. Reported and recovered values hardly differed there.

**Table 5. Accuracy of Recovered and Reported AVE in Both Calibrations (N = 400)**

| Structure                                                        | Estimator | RMSE recovered | RMSE reported | Gap true | Gap recovered | Gap reported | Detection rate |
|------------------------------------------------------------------|-----------|----------------|---------------|----------|---------------|--------------|----------------|
| <b>Panel A: inhibitor AVE as in Parasuraman and Colby (2015)</b> |           |                |               |          |               |              |                |
| Random error                                                     | CB-SEM    | .019           | .031          | -.226    | -.212         | -.197        | 1.00           |
|                                                                  | PLS       | .014           | .022          | -.171    | -.162         | -.157        | .00            |
| Broad domain                                                     | CB-SEM    | .033           | .061          | -.224    | -.203         | -.158        | 1.00           |
|                                                                  | PLS       | .031           | .053          | -.173    | -.148         | -.126        | .04            |
| Referent                                                         | CB-SEM    | .037           | .069          | -.196    | -.169         | -.113        | 1.00           |
|                                                                  | PLS       | .031           | .058          | -.155    | -.134         | -.104        | .01            |
| Method factor                                                    | CB-SEM    | .016           | .025          | -.192    | -.181         | -.169        | 1.00           |
|                                                                  | PLS       | .007           | .011          | -.145    | -.142         | -.141        | .00            |
| <b>Panel B: inhibitor loadings 0.15 higher</b>                   |           |                |               |          |               |              |                |
| Random error                                                     | CB-SEM    | .002           | .003          | -.017    | -.017         | -.017        | 1.00           |
|                                                                  | PLS       | .001           | .002          | -.013    | -.013         | -.013        | n/a            |
| Broad domain                                                     | CB-SEM    | .021           | .043          | -.022    | -.019         | +.012        | 1.00           |
|                                                                  | PLS       | .008           | .018          | -.020    | -.018         | -.014        | .15            |
| Referent                                                         | CB-SEM    | .027           | .055          | +.013    | +.024         | +.059        | 1.00           |
|                                                                  | PLS       | .024           | .050          | +.002    | +.014         | +.041        | .91            |
| Method factor                                                    | CB-SEM    | .002           | .003          | -.006    | -.006         | -.006        | 1.00           |
|                                                                  | PLS       | .001           | .002          | -.004    | -.004         | -.004        | n/a            |

Note. RMSE = root mean squared error of the dimension-level AVE against the true pre-deletion AVE; recovered = procedure of Study 1; reported = AVE as reported after deletion. Gap = mean AVE of inhibitor minus motivator dimensions. Detection rate = share of unreported deletions detected. n/a = no unreported deletions occurred. The false positive rate was 0 in every cell. 100 replications of 40 studies per cell. CB-SEM = confirmatory factor model, maximum likelihood; PLS = partial least squares, mode A.

Recovery reduced the error everywhere. The root mean squared error of the recovered AVE was smaller than that of the reported AVE in every cell, by roughly 30% to 60%. The recovered difference between inhibitor and motivator dimensions was always at least as close to the true difference as the reported one. In no cell did recovery make the inhibitor dimensions look weaker, relative to the motivators, than they really were. In that sense the procedure is conservative.

The second calibration illustrates why recovery matters even when the true difference is small. With stronger inhibitor items, deletion concentrated on the inhibitor dimensions and in the broad domain structure estimated by CB-SEM the reported difference even changed sign. At  $N = 400$ , the true inhibitor AVE was 0.022 below the motivator AVE whereas the reported AVE was 0.012 above it and the recovered AVE 0.019 below. A published record of this kind would suggest that the inhibitor dimensions measure better than the motivators when the opposite is true.

The test for unreported deletions never flagged a dimension that had kept all four items, so its false positive rate was 0 in all 48 cells. With CB-SEM it found 99% to 100% of unreported deletions. With PLS it found few of them (0% to 15%) in every cell except the referent structure in the second calibration (87% to 91%) because PLS loadings change little after a deletion and AVE, CR and alpha then stay consistent with four items. A flag can therefore be trusted whereas the absence of a flag in a PLS study says little.

We see, two limitations of the procedure also became clear. Recovery is less accurate for single items than for dimensions. The recovered loading of INS4 in the referent structure was too low (CB-SEM,  $N = 200$ : 0.23 against a true 0.37), although the error shrank as samples grew. The procedure could also tell the referent structure apart from the others but not the method factor structure. Repeated deletion of the same items identified the referent structure in all 1,200 replications in which it was true (false identification: 0.4% to 2.4%). The method factor structure, by contrast, was classified as random error or as a broad domain in nearly every replication, since a factor shared by all inhibitor items leaves no trace in AVE, CR and alpha.

## Discussion of Study 2

Under the conditions simulated, the procedure of Study 1 recovers pre-deletion AVE more accurately than the reported values do. It does not exaggerate the difference between inhibitor and motivator dimensions and does not flag deletions that never happened. Its estimates can thus be read as conservative. Two kinds of conclusion from Study 1 deserve more caution: item-level loadings of frequently deleted items and anything concerning wording effects which aggregate statistics cannot reveal. Study 3 examines how deletion and recovery work in raw data where every loading is known; wording effects are examined in the companion paper (Author, under review).

## Study 3: Deletion and Recovery in Raw Data

Studies 1 and 2 only used the results that other authors chose to publish.

In Study 3, we looked at the actual survey items in a full dataset where nothing had been removed. We checked two things:

- Whether the “inhibitor” scales still show good convergent validity when all 16 items are kept.
- How much a typical author’s decision to drop items would change the numbers they report.

## Samples

Sample A consists of 649 people in India who were aware of mobile health (mHealth) applications and answered an online and paper questionnaire. The TRI 2.0 items were adapted to mHealth applications by changing only the reference to technology. Of the 817 questionnaires received, 136 came from respondents who were not aware of mHealth applications and a further 32 were removed as duplicate submissions (3), random responding (22) or straight-lining (7).

Sample B consists of 273 students who answered the TRI 2.0 items in a study of artificial intelligence in education that was published as an open dataset (Lim et al., 2024).

In both samples the items used five-point agreement scales and the inhibitor items were not reverse-coded, so that agreement means more discomfort or insecurity. The screening of both samples for uniform responding, and the exclusion of a third open dataset that failed it, are described in the companion paper (Author, under review).

## Analysis

Confirmatory factor models were estimated in lavaan (Rosseel, 2012) in two ways. Treating the items as ordinal (weighted least squares with mean and variance adjustment, WLSMV) suits five-point items whereas treating them as continuous (robust maximum likelihood) reproduces how most published TRI studies estimate them. AVE and CR were computed from the standardized loadings (Fornell & Larcker, 1981).

To connect Study 3 with the first two studies, we imitated a typical author on each dataset. The four-dimension model was estimated by maximum likelihood, the item with the lowest loading below .70 was deleted and the process was repeated until no loading fell below .70 keeping at least three items per dimension. We then recorded the AVE that would have been reported. The same exercise was repeated in 200 subsamples at each of several sample sizes and the censored model of Study 1 was applied to the reported values.

## Results

Convergent validity of the inhibitor dimensions was not acceptable. With all items retained, both discomfort and insecurity had an AVE below .50 in both samples and under both estimators where values were between .35 and .49. The full measurement results for both samples are reported in the companion paper (Author, under review).

The weakest items were largely those identified in Study 1. INS4 had the lowest loading of all 16 items in Sample B (.52), followed by INS3 and in Sample A INS3, DIS4, INS1 and INS4 were among the five weakest items, the weakest being OPT1 (.54).

A typical author would have removed one item from each dimension in both samples, namely OPT1, INN3, DIS4 and INS3 in Sample A and OPT3, INN4, DIS1 and INS4 in Sample B. These deletions raised every reported AVE, by 0.02 to 0.07 under continuous estimation. In Sample B, insecurity would have risen from .42 to .49 and innovativeness from .45 to .52, enough for innovativeness to pass the .50 criterion. In the subsamples, the recovery procedure moved the reported values back toward the full-scale values for both inhibitor dimensions in both samples. For insecurity in Sample A, the full-scale AVE was .348, the reported value .378 and the

recovered value .363; in Sample B the corresponding values were .420, .494 and .437. As Study 2 predicts, the correction was partial. At one subsample size, optimism in Sample A could only be bounded because OPT1 was deleted in almost every subsample.

## Discussion of Study 3

When nothing is deleted, the inhibitor dimensions of TRI 2.0 fail to reach acceptable convergent validity in either sample. Deletion of the kind found in the literature raises their reported AVE by amounts that can carry a dimension across the .50 criterion. The recovery procedure moves the values back in the right direction that agrees with Studies 1 and 2.

## General Discussion

Our three studies point to the same main finding.

When every item is kept, the inhibitor parts of TRI 2.0 do not show good convergent validity. This is true in the original validation study (Parasuraman & Colby, 2015), in two full datasets and in a larger number of published studies once the missing items are put back in.

The published results make the problem look smaller than it is. Authors drop inhibitor items (especially the insecurity ones) more often, so the numbers they report only describe the items that were left in.

## Theoretical Implications

Beyond TRI 2.0, the findings say something about a literature as evidence. When authors delete items in response to their data, a body of studies can accumulate apparent support for a measure that each study, taken before deletion, would have questioned. Measurement validity in a literature therefore depends on reporting as well as on items (Flake & Fried, 2020). The inverse meta-analysis offers a way to estimate the gap between reported and administered measurement quality. It can be applied to other scales wherever AVE, CR and alpha are reported.

The procedure adds to existing checks of published statistics. Tests such as GRIM (Brown & Heathers, 2017) and statcheck (Nuijten et al., 2016) detect values that cannot be correct. The inverse meta-analysis works with values that are correct for the items that were kept and asks what they would have been for the items that were administered. Study 2 shows why this matters even when the true difference between dimensions is small. When deletion falls more often on one set of dimensions, the reported difference can shrink or even change sign, so that a literature points in the wrong direction while every single study reports correct numbers.

## Implications for Research Practice

For researchers who use TRI 2.0 or similar scales, several recommendations follow. Every administered item should be reported with its loading including items that were later deleted and the deletion rule should be stated before estimation. AVE and CR for the full set of items belong alongside any values computed after deletion. Reviewers and editors can easily check whether the reported AVE, CR and alpha of a dimension are

consistent with one another and with the stated number of items. In Study 2, this check never flagged a dimension that had kept all its items.

## Limitations and Future Research

Study 1 relies on values extracted from full texts by a parser and checked by hand. Only 94 of the 192 main studies yielded usable values. Because these studies stated their cut-offs and deletions more often than the rest, the results describe the better-reported part of the literature. Adding the sensitivity studies did not change the main results but only studies with an accessible full text, 506 of the 1,403 records with a DOI, could be assessed. The censored model assumes that deleted items fell below a known or default cut-off and that loadings are normal on the Fisher  $z$  scale. The simulations of Study 2 cannot cover every way in which authors select and report items.

Study 3 tests two samples, mHealth users in India and students in China. The inverse meta-analysis could also be applied to other service measures whose dimensions are routinely trimmed.

## Conclusion

Published measurement statistics can be read backward to estimate how a scale performed before items were deleted. TRI 2.0 measures four distinct dimensions but its two inhibitor dimensions measure their constructs weakly. They have been made to look stronger in published research through the selective deletion of items. Researchers who use the scale should report all items. Readiness research would benefit from a reporting standard that makes the effects of item deletion visible.

## Declarations

We used AI-assisted tools (Claude) for language refinement and for support with code generation and debugging. All analytical decisions and theoretical arguments were developed and validated by us only. R was used for statistical calculations.

## References

1. Author (under review). What does the Technology Readiness Index measure? Item content and the inhibitor dimensions of TRI 2.0. Manuscript under review.
2. Blut, M., & Wang, C. (2020). Technology readiness: A meta-analysis of conceptualizations of the construct and its impact on technology usage. *Journal of the Academy of Marketing Science*.  
<https://doi.org/10.1007/s11747-019-00680-8>
3. Brown, N. J. L., & Heathers, J. A. J. (2017). The GRIM test: A simple technique detects numerous anomalies in the reporting of results in psychology. *Social Psychological and Personality Science*.  
<https://doi.org/10.1177/1948550616673876>

4. Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science*.  
<https://doi.org/10.1177/1948550617693063>
5. Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*.  
<https://doi.org/10.1177/2515245920952393>
6. Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*. <https://doi.org/10.1177/002224378101800104>
7. Gerber, A. S., & Malhotra, N. (2008). Publication bias in empirical sociological research: Do arbitrary significance levels distort published results?. *Sociological Methods & Research*.  
<https://doi.org/10.1177/0049124108318973>
8. Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*. <https://doi.org/10.1108/EBR-11-2018-0203>
9. Hussey, I., & Hughes, S. (2020). Hidden invalidity among 15 commonly used measures in social and personality psychology. *Advances in Methods and Practices in Psychological Science*.  
<https://doi.org/10.1177/2515245919882903>
10. Lim, W. J., Tan, Y. H., See, S., Lim, L., Lee, J., Subramanian, S., & Tan, W. P. (2024). Unravelling students' perspectives, personalities and practices of using AI in education dataset. *Mendeley Data*.  
<https://doi.org/10.17632/cp3wpfmdt.1>
11. Mick, D. G., & Fournier, S. (1998). Paradoxes of technology: Consumer cognizance, emotions, and coping strategies. *Journal of Consumer Research*. <https://doi.org/10.1086/209531>
12. Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*. <https://doi.org/10.1002/sim.8086>
13. Nuijten, M. B., Hartgerink, C. H. J., van Assen, M. A. L. M., Epskamp, S., & Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985-2013). *Behavior Research Methods*.  
<https://doi.org/10.3758/s13428-015-0664-2>
14. Parasuraman, A. (2000). Technology Readiness Index (TRI): A multiple-item scale to measure readiness to embrace new technologies. *Journal of Service Research*. <https://doi.org/10.1177/109467050024001>
15. Parasuraman, A., & Colby, C. L. (2015). An updated and streamlined technology readiness index: TRI 2.0. *Journal of Service Research*. <https://doi.org/10.1177/1094670514539730>
16. Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. Springer.  
<https://doi.org/10.1007/b98882>
17. Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv*. <https://doi.org/10.48550/arXiv.2205.01833>
18. Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*.  
<https://doi.org/10.18637/jss.v048.i02>
19. Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*.  
<https://doi.org/10.2307/1907382>