Psychometrika (2026), **91**, 1348–1360
doi:10.1017/psy.2026.10124

PSYCHOMETRIC SOCIETY

THEORY AND METHODS

Revelle's Beta: The Wait Is Over—Computation Becomes Possible

Jan O. Bauer[1,2]

[1]Vrije Universiteit Amsterdam, Netherlands; [2]Tinbergen Institute, Netherlands

Corresponding author: Jan O. Bauer; Email: j.bauer@vu.nl

(Received 20 October 2025 UTC; revised 13 May 2026 UTC; accepted 18 May 2026 UTC)

Abstract

Over 45 years ago, William Revelle proposed a reliability measure based on the worst split-half of a test or scale, commonly known as *Revelle's beta*, to assess the general factor saturation. However, to this day, there is no reliable method for computing this measure, as existing approaches are either computationally infeasible or insufficiently accurate in identifying the worst split-half. This difficulty arises because the number of candidate splits increases exponentially with the number of items. In this article, we show that computing Revelle's beta is conceptually equivalent to divisive ("top-down") hierarchical clustering. This insight allows us to reduce the number of candidate splits to a quadratic problem, making the computation feasible. We specify theoretical conditions under which this approach is guaranteed to recover the worst split-half. To validate the efficiency of our approach, we conduct simulation studies and analyze real-world data. Code implementations accompanying this work are available online, together with Supplementary Material.

Keywords: hierarchical clustering; internal consistency; reliability; Revelle's beta; split-half

1. Introduction

The questions of what is being measured and how consistent it is measured are central concerns in psychometrics. In this context, validity refers to the first question, while reliability refers to the second question and quantifies the consistency of measurement.

Internal consistency coefficients summarize the degree of shared covariance among items within a test scale and are commonly used as indices of scale reliability. Although the term internal consistency is often interpreted as implying unidimensionality, meaning that items measure a single latent construct, such an interpretation is not justified on the basis of coefficient values alone (Sijtsma, 2009). Coefficients of internal consistency quantify statistical properties of the covariance structure but do not, by themselves, establish the factorial composition of an item set.

Among the most widely used coefficients is Cronbach's alpha (Cronbach, 1951; Guttman, 1945). Alpha is derived under the classical test model, which decomposes an observed score into a true score and a random measurement error, assuming that errors are uncorrelated with true scores and with each other. Under these assumptions, alpha constitutes a lower bound to reliability. Equality between alpha and reliability holds under restrictive assumptions such as essential tau-equivalence, which, if misinterpreted, may lead to incorrect conclusions about scale reliability (Dunn et al., 2014; Revelle & Zinbarg, 2009; Sijtsma, 2009).

© The Author(s), 2026. Published by Cambridge University Press on behalf of Psychometric Society.
This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/psy.2026.10124 Published online by Cambridge University Press

These limitations and interpretational ambiguities have motivated the development of alternative reliability coefficients. One such alternative is Revelle's beta (Revelle, 1979), which is defined as the worst split-half reliability across all possible item partitions. Because beta is based on the most unfavorable split, it is sensitive to multidimensional structure: when items form relatively distinct clusters or subscales, beta tends to decrease, reflecting reduced homogeneity across the full item set. In contrast to alpha, which may attain high values even in the presence of multiple correlated factors, beta more strongly penalizes heterogeneous item groupings. Despite these conceptual advantages, beta remains underutilized in applied research (Reise & Haviland, 2025; Revelle & Zinbarg, 2009; Rossiter, 2002; Tang et al., 2014).

This is largely due to the computational challenges involved in its calculation as beta is based on the worst split-half reliability of a test scale. For example, given a scale (data set) containing p variables (items), there are $2^{p-1}-1$ candidate splits that must be evaluated to identify the worst split-half. For $p = 30$, this already results in 536,870,911 candidate splits, and checking all of them exceeds hours of computation time, as will be discussed later in this work. Such computational demands are impractical for most applications.

In the context of clustering, however, a recent approach reduces this exponential problem to a quadratic one, making it computational feasible (Bauer, 2025b). The idea is to find the best block diagonal matrix approximation of the correlation matrix among a set of block diagonal matrices comprising two blocks. We demonstrate in this work that computing Revelle's beta is conceptually identical to finding the best approximation with respect to an average linkage loss function. Consequently, this allows to adapt the approach introduced by Bauer (2025b) to the computation of beta.

The use of hierarchical clustering to explore the internal structure of tests was pioneered by Revelle (1979), who illustrated on small item pools how clusters of items can reveal subscales and patterns of internal consistency. Our approach builds on this tradition by extending the methodology to larger item set.

The remainder of this article is organized as follows. In Section 2, after introducing basic notation and definitions, we present the conceptual framework for computing Revelle's beta. We also discuss how the full item hierarchy based on the worst split-half reliability can be visualized. Section 3 presents extensive simulation studies to demonstrate the robustness of our approach compared to existing methods. In Section 4, we compute Revelle's beta on real data, and we conclude with a general discussion in Section 5.

2. Computation of Revelle's beta

2.1. Setup

The following notations are used throughout the article. We use lowercase letters x, boldface lowercase letters $\boldsymbol{v}$, and boldface capital letters $\boldsymbol{X}$ to represent scalars, vectors, and matrices, respectively. Additionally, for any vector $\boldsymbol{v}^T = (v_1, \ldots, v_p)$, $\|\boldsymbol{v}\|_1 = \sum_{k=1}^{p} |v_k|$ denotes the ℓ_1 vector norm and $\|\boldsymbol{v}\|_2 = \left(\sum_{k=1}^{p} v_k^2\right)^{1/2}$ denotes the ℓ_2 vector norm. For any matrix $\boldsymbol{A} = (a_{kl})$, $\|\boldsymbol{A}\|_{1,1} = \|\text{vec}(\boldsymbol{A})\|_1 = \sum_k \sum_l |a_{kl}|$ denotes the element-wise ℓ_1 vector norm of the matrix, $\|\boldsymbol{A}\|_{\infty,\infty} = \max_{k,l} |a_{kl}|$ denotes the element-wise ℓ_∞ vector norm of the matrix, and $\|\boldsymbol{A}\|_F = \|\text{vec}(\boldsymbol{A})\|_2 = \left(\sum_k \sum_l a_{kl}^2\right)^{1/2}$ denotes the Frobenius norm, which corresponds to the ℓ_2 norm of the vectorized matrix.

Let $\boldsymbol{X}$ be an $n \times p$ matrix of data that contains n realizations of p variables. We denote the corresponding sample correlation matrix as $\boldsymbol{R}$. The singular value decomposition of the correlation matrix, respectively, called eigendecomposition, can be written as follows:

$$\boldsymbol{R} = \boldsymbol{V}\boldsymbol{L}\boldsymbol{V}^T,\ \boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}_p,\ l_1 \geq \cdots \geq l_p \geq 0,$$

where $\boldsymbol{L}$ contains the eigenvalues (singular values) of $\boldsymbol{R}$ on the diagonal, and $\boldsymbol{V} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p)$ are the eigenvectors (singular vectors).

https://doi.org/10.1017/psy.2026.10124 Published online by Cambridge University Press

As will be discussed below, our goal is to identify two distinct submatrices $\boldsymbol{X}_j$ $(n \times p_j)$ and $\boldsymbol{X}_{j'}$ $(n \times p_{j'})$ of $\boldsymbol{X}$, with $p = p_j + p_{j'}$, which exhibit the smallest average correlation between them for the purpose of computing Revelle's beta. $\boldsymbol{X}_j$ and $\boldsymbol{X}_{j'}$ can be considered as clusters that identify p_j and $p_{j'}$ many variables.

Although the exact submatrices are currently unknown, we can still assume the convenient order $\boldsymbol{X} = (\boldsymbol{X}_j, \boldsymbol{X}_{j'})$ without loss of generality, since any such ordering can be obtained using column permutation. The sample correlation matrix can thus be written as

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{R}_j & \boldsymbol{R}_{jj'} \\ \boldsymbol{R}_{jj'}^T & \boldsymbol{R}_{j'} \end{pmatrix}, \tag{1}$$

where $\boldsymbol{R}_j$ $(p_j \times p_j)$ is the correlation matrix of $\boldsymbol{X}_j$, $\boldsymbol{R}_{j'}$ $(p_{j'} \times p_{j'})$ is the correlation matrix of $\boldsymbol{X}_{j'}$, and $\boldsymbol{R}_{jj'}$ $(p_j \times p_{j'})$ contains the correlations between these two submatrices, that is, the between-test correlations.

Revelle (1979) proposed a reliability measure based on the worst split-half of a test, commonly known as *Revelle's beta*. For a split of $\boldsymbol{X}$ into the two halves $\boldsymbol{X}_j$ and $\boldsymbol{X}_{j'}$, this split-half reliability estimate is defined as the average between-half correlations $\|\boldsymbol{R}_{jj'}\|_{1,1}/(p_j p_{j'})$ times the total number of items squared p^2, and scaled by the total correlation $\|\boldsymbol{R}\|_{1,1}$. Thus, the coefficient can be written as

$$\frac{p^2 \|\boldsymbol{R}_{jj'}\|_{1,1}/(p_j p_{j'})}{\|\boldsymbol{R}\|_{1,1}} = \frac{\|\boldsymbol{R}_{jj'}\|_{1,1}/(p_j p_{j'})}{\|\boldsymbol{R}\|_{1,1}/p^2}. \tag{2}$$

Consequently, this estimate scales the average of absolute components of $\boldsymbol{R}_{jj'}$ by the average of absolute components of $\boldsymbol{R}$, yielding a value between zero and one.

Although split-half reliability can also be formulated in terms of the covariance matrix, the present manuscript focuses on correlation matrices and therefore on standardized tests. This focus stems from an assumption in Bauer (2025b), whose methodology underlies the proposed computation of Revelle's beta. In particular, the assumption ensures that individual items remain distinguishable. A brief overview of the methodology is provided later.

This restriction should be kept in mind when interpreting the coefficient. Since the use of correlations removes differences in item variances, the resulting value is Revelle's beta for the standardized item scores, rather than necessarily for a raw-score total. This distinction is expected to be minor when items are measured on the same response scale and have similar variances, but it may matter when items have different response formats, substantially different variances, or are intentionally weighted differently. In such cases, standardization may affect both the value of beta and the split identified as the worst split-half.

2.2. Computation of the worst split-half reliability

In practice, when $\boldsymbol{X}_j$ and $\boldsymbol{X}_{j'}$ are unknown, for example, when they represent underlying latent factors, we aim to find these by minimizing Equation (2). This minimum is the worst split-half reliability, known as *Revelle's beta*. Consequently, to identify the disjoint submatrices $\boldsymbol{X}_j$ and $\boldsymbol{X}_{j'}$, we solve the following optimization problem:

$$\beta = \min_{\boldsymbol{X}_j, \boldsymbol{X}_{j'}} \frac{\|\boldsymbol{R}_{jj'}\|_{1,1}/(p_j p_{j'})}{\|\boldsymbol{R}\|_{1,1}/p^2}. \tag{3}$$

We note that identifying the submatrices with the worst split-half reliability is conceptually identical to identifying the submatrices with the lowest average absolute between-test correlations.

Remark 2.1. The arguments $\boldsymbol{X}_j$ and $\boldsymbol{X}_{j'}$ that minimize the optimization problem in (3) are identical to the ones of the optimization problem

$$\min_{\boldsymbol{X}_j, \boldsymbol{X}_{j'}} \frac{\|\boldsymbol{R}_{jj'}\|_{1,1}}{p_j p_{j'}}.$$

Proof. The proof is provided in Section A of the Supplementary Material. □

https://doi.org/10.1017/psy.2026.10124 Published online by Cambridge University Press

This optimization problem is known as average linkage clustering in clustering contexts, and computation of Revelle's beta is therefore conceptually identical to a divisive ("top-down") clustering task: we iteratively split $\boldsymbol{X}$ into subsequently smaller subtests using average linkage, until each variable is contained in a single subtest. Consequently, we start by splitting $\boldsymbol{X}$ into $\boldsymbol{X}_j$ and $\boldsymbol{X}_{j'}$, then splitting $\boldsymbol{X}_j$ as well as $\boldsymbol{X}_{j'}$, and so on.

However, a drawback of divisive clustering is its computational cost: when splitting p variables, there are $2^{p-1}-1$ candidate combinations which cannot all be checked for large p. In the context of clustering, however, a recent approach reduces this exponential problem to a quadratic one. The idea is to find the best block-diagonal approximation of the correlation matrix from a set of block-diagonal matrices comprising two blocks, with respect to some function $f : \mathbb{R}^{p\times p} \to \mathbb{R}$ via

$$\min_{\boldsymbol{X}_j,\, \boldsymbol{X}_{j'}} f\left(\begin{pmatrix} \boldsymbol{R}_j & \boldsymbol{R}_{jj'} \\ \boldsymbol{R}_{jj'}^T & \boldsymbol{R}_{j'} \end{pmatrix} - \begin{pmatrix} \boldsymbol{R}_j & 0 \\ 0 & \boldsymbol{R}_{j'} \end{pmatrix}\right). \tag{4}$$

For clustering using average linkage, this corresponds to the optimization problem in Equation (4) with $f(\boldsymbol{A}) \mapsto \|\boldsymbol{A}\|_{1,1}/(2p_jp_{j'})$ given as

$$\min_{\boldsymbol{X}_j,\, \boldsymbol{X}_{j'}} \frac{1}{2p_jp_{j'}} \left\| \begin{pmatrix} \boldsymbol{R}_j & \boldsymbol{R}_{jj'} \\ \boldsymbol{R}_{jj'}^T & \boldsymbol{R}_{j'} \end{pmatrix} - \begin{pmatrix} \boldsymbol{R}_j & 0 \\ 0 & \boldsymbol{R}_{j'} \end{pmatrix} \right\|_{1,1},$$

or, equivalently,

$$\min_{\boldsymbol{X}_j,\, \boldsymbol{X}_{j'}} \frac{1}{p_jp_{j'}} \|\boldsymbol{R}_{jj'}\|_{1,1},$$

where the approximation is with respect to the average absolute matrix components of $\boldsymbol{R}_{jj'}$. Bauer (2025b, 2025c) demonstrated that *sparse eigenvectors* of the correlation matrix, which are sparse approximations of the eigenvectors $\boldsymbol{V}$ that contain many zeros, identify the correlation matrix $\boldsymbol{R}_{jj'}$ that minimizes the split-half in (3). Specifically, the sparse eigenvectors mirror the block diagonal matrix approximation $\text{diag}(\boldsymbol{R}_j, \boldsymbol{R}_{j'})$ of the correlation matrix $\boldsymbol{R}$ in (1), thereby identifying $\boldsymbol{X}_j$ and $\boldsymbol{X}_{j'}$.

We now give a recap of this method, named hierarchical clustering using singular vectors (HC-SVD), and refer to Bauer (2025b) for an elaborate introduction to HC-SVD. First, we note that if the correlation matrix $\boldsymbol{R}$ in (1) was a block diagonal matrix, that is, if $\boldsymbol{R}_{jj'} = 0$, then its eigenvectors $\boldsymbol{V}$ exhibit also a block diagonal structure that mirrors this block diagonal shape (Bauer, 2025b; Bauer & Drabant, 2021). If however $\boldsymbol{R}_{jj'} \neq 0$, which is commonly the case in real-world application, the eigenvectors are perturbed and do not mirror the underlying block diagonal structure $\text{diag}(\boldsymbol{R}_j, \boldsymbol{R}_{j'})$. In HC-SVD, sparse eigenvectors are computed by shrinking these perturbations back toward zero, thereby revealing again the block diagonal structure of the eigenvectors which aligns with the block diagonal structure of the correlation matrix. According to the Eckart–Young theorem (Eckart & Young, 1936), the best rank-one approximation of a symmetric matrix in the Frobenius norm is determined by its first eigenvector and eigenvalue. This can be formulated as

$$\min_{l,\, \boldsymbol{v}} \|\boldsymbol{R} - l\boldsymbol{v}\boldsymbol{v}^T\|_F^2 \text{ subject to } \|\boldsymbol{v}\|_2^2 = 1.$$

Building on this, sparse eigenvectors are computed using the following optimization problem (Shen & Huang, 2008):

$$\min_{\boldsymbol{v},\, \tilde{\boldsymbol{v}}} \|\boldsymbol{R} - \tilde{\boldsymbol{v}}\boldsymbol{v}^T\|_F^2 \text{ subject to } \|\boldsymbol{v}\|_2^2 = 1,\ \|\boldsymbol{v}\|_1 \leq c. \tag{5}$$

This formulation yields the first sparse eigenvector $\check{\boldsymbol{v}}_1$ with $\|\check{\boldsymbol{v}}_1\|_2^2 = 1$ subject to the ℓ_1 sparsity constraint c. $\tilde{\boldsymbol{v}}$ is allowed to vary and absorbs the corresponding "eigenvalue" $\check{l}_1$, which means that the optimal $\tilde{\boldsymbol{v}}$ corresponds to the scaled vector $\check{l}_1\check{\boldsymbol{v}}_1$. The remaining sparse eigenvectors $\check{\boldsymbol{v}}_k$ for $k > 1$ can be calculated iteratively where the correlation matrix $\boldsymbol{R}$ must be replaced by the residual matrices of the sequential matrix approximations.

https://doi.org/10.1017/psy.2026.10124 Published online by Cambridge University Press

The approach underlying HC-SVD is then computing p sparse eigenvectors, each with $s \in \{1, \ldots, p-1\}$ components shrunk toward zero, to identify $\text{diag}(\boldsymbol{R}_j, \boldsymbol{R}_{j'})$. This reduces the number of candidate splits to check from $2^{p-1}-1$ to $p(p-1)$, resulting in a quadratic problem that is computationally feasible even for large p. Checking a reduced subset of candidate splits carries the theoretical risk that the split (argument) minimizing (3) is missed. However, we can establish a regime in which the procedure is guaranteed to identify the minimum. Bauer (2025b) derived a bound on the off-diagonal block $\boldsymbol{R}_{jj'}$ that ensures the optimal split is perfectly recovered by HC-SVD. For completeness, we restate this recovery guarantee below. However, when only $p(p-1)$ instead of all $2^{p-1}-1$ candidate splits are checked, there is a risk that the split (argument) minimizing Equation (3) is not considered. However, Bauer (2025b) provides a bound on $\boldsymbol{R}_{jj'}$ that guarantees that the optimal split is recovered by HC-SVD. For completeness, we restate the corresponding result below.

Theorem 2.2 (From Theorem 1 of Bauer, 2025b). *Consider a correlation matrix as in (1). The sparse eigenvector $\check{\boldsymbol{v}}$ with eigenvalue $\check{l}$ computed using the optimization problem in (5) mirrors the block diagonal matrix* $\text{diag}(\boldsymbol{R}_j, \boldsymbol{R}_{j'})$ *if*

$$\frac{1}{p}\|\boldsymbol{R}_{jj'}\|_{\infty,\infty} \leq \max\{|\check{\boldsymbol{v}}|_{(p_j)}, |\check{\boldsymbol{v}}|_{(p_{j'})}\}. \tag{6}$$

Here, $|\check{\boldsymbol{v}}|_{(p)}$ denotes the p-th largest absolute component of the vector $\check{\boldsymbol{v}}$.

Proof. The proof is available in Section A of the Supplementary Material and in the Supplementary Material of Bauer (2025b) (Section A of the Supplementary Material). □

When this condition holds, the method is guaranteed to identify the true worst split-half partition. Bauer (2025b) further argues that detection of the block diagonal matrix can be improved by computing the sparse eigenvectors of $\boldsymbol{R}^{\circ h}$ instead of $\boldsymbol{R}$. Here, $\odot$ denotes the Hadamard (element-wise) product, and $\boldsymbol{R}^{\circ h} = \underbrace{\boldsymbol{R} \odot \cdots \odot \boldsymbol{R}}_{h \text{ times}} = (r_{kl}^h)$ defines the Hadamard power of order h. Since $|r_{kl}| \leq 1$, the intuition is that taking the Hadamard powers shrinks most correlation coefficients toward zero. Importantly, the off-block coefficients in $\boldsymbol{R}_{jj'}$ shrink faster than the within block coefficients in $\boldsymbol{R}_j$ and $\boldsymbol{R}_{j'}$, because $\|\boldsymbol{R}_{jj'}\|_{1,1}$ minimizes the between-test correlation. This allows for better detection according of the optimal split, as the bound in (6) from Theorem 2.2 is more likely to be satisfied.

2.3. Visualization of Revelle's beta

Let us now denote β_j as Revelle's beta of any matrix $\boldsymbol{X}_j$ with correlation matrix $\boldsymbol{R}_j$, and β be the corresponding measure of $\boldsymbol{X}$ as in (3). The result of hierarchical clustering methods is interpretable without inversions via a dendrogram if the resulting distance matrix is ultrametric (Hartigan, 1967; Johnson, 1967). However, Revelle's beta satisfies the ultrametric property only under restrictive assumptions.

Theorem 2.3. *Consider a correlation matrix as in (1), and let $\boldsymbol{X}_j$ and $\boldsymbol{X}_{j'}$ be the arguments that yield the worst split-half reliability in (3). Revelle's beta is ultrametric, meaning that $\beta \leq \min\{\beta_j, \beta_{j'}\}$, if*

$$\frac{\|\boldsymbol{R}_j\|_{1,1}}{p_j^2} \leq \frac{\|\boldsymbol{R}\|_{1,1}}{p^2} \quad \textit{and} \quad \frac{\|\boldsymbol{R}_{j'}\|_{1,1}}{p_{j'}^2} \leq \frac{\|\boldsymbol{R}\|_{1,1}}{p^2}.$$

Proof. The proof is available in Section A of the Supplementary Material. □

Consequently, if the average absolute correlation coefficients of $\boldsymbol{X}$ are larger than the average absolute correlation coefficients within its two subtests, Revelle's beta is ultrametric. However, since $\boldsymbol{R}$ includes the submatrix $\boldsymbol{R}_{jj'}$, which contains the correlations between the two subtests $\boldsymbol{X}j$ and $\boldsymbol{X}j'$, these conditions are not necessarily satisfied in practice.

In contrast, HC-SVD using average linkage produces an ultrametric distance matrix (Bauer, 2025b). As discussed above, the worst-split half aligns with the splits identified by HC-SVD using average

https://doi.org/10.1017/psy.2026.10124 Published online by Cambridge University Press

linkage. Consequently, all splits according to Revelle's beta can be hierarchically visualized using a dendrogram, where the heights of this dendrogram are computed via average linkage. While the heights of this dendrogram do not correspond to the actual values of Revelle's beta, its hierarchical structure is identical to that implied by Revelle's beta and thus accurately represents all splits.

2.4. Comments on ICLUST

Revelle (1978) introduced the item cluster analysis (ICLUST) procedure to identify subtests exhibiting the worst split-half reliabilities. This "bottom-up" approach iteratively merges subtests using agglomerative clustering with complete linkage to form candidate clusters. A candidate cluster $\boldsymbol{X} = (\boldsymbol{X}_j, \boldsymbol{X}_{j'})$ is constructed by merging $\boldsymbol{X}_j$ and $\boldsymbol{X}_{j'}$ if two internal consistency criteria are satisfied. Specifically, the merged cluster $\boldsymbol{X}$ must yield a higher Cronbach's alpha than either subtest ($\alpha > \min\{\alpha_j, \alpha_{j'}\}$), and a higher Revelle's beta than the average of the corresponding betas ($\beta > (\beta_j + \beta_{j'})/2$).

Nowadays, alternative combinations of these internal consistency tests are also considered. For Cronbach's alpha, the conditions may take one of the forms

$$\alpha > \min\{\alpha_j, \alpha_{j'}\}, \quad \alpha > (\alpha_j + \alpha_{j'})/2, \quad \text{or } \alpha > \max\{\alpha_j, \alpha_{j'}\},$$

and analogously for Revelle's beta,

$$\beta > \min\{\beta_j, \beta_{j'}\}, \quad \beta > (\beta_j + \beta_{j'})/2, \quad \text{or } \beta > \max\{\beta_j, \beta_{j'}\}.$$

Implementations of these consistency tests are discussed in Section 3 and in the corresponding part of the Supplementary Material.

Among these, the conditions $\alpha > \min\{\alpha_j, \alpha_{j'}\}$ and $\beta > \min\{\beta_j, \beta_{j'}\}$ are the most relaxed, as they are satisfied whenever any of the other conditions hold. Consequently, if these relaxed conditions are not met, none of the stricter ones will hold either. However, these conditions align with Revelle's beta only under restrictive assumptions, as can be concluded from the following theorem.

Theorem 2.4. *Consider a correlation matrix as in (1), and let $\boldsymbol{X}_j$ and $\boldsymbol{X}_{j'}$ be the arguments that yield the worst split-half reliability in (3). The condition $\alpha > \min\{\alpha_j, \alpha_{j'}\}$ is satisfied if*

$$\frac{\|\boldsymbol{R}_j\|_{1,1}}{p_j} \geq \frac{\|\boldsymbol{R}\|_{1,1}}{p} \quad \textit{and} \quad \frac{\|\boldsymbol{R}_{j'}\|_{1,1}}{p_{j'}} \geq \frac{\|\boldsymbol{R}\|_{1,1}}{p}. \tag{7}$$

Proof. The proof is provided in Section A of the Supplementary Material. □

Using the identity $\|\cdot\|_{1,1} = \|\cdot\|_{1,1}^{1/2} \|\cdot\|_{1,1}^{1/2}$, the left condition in Equation (7) can be rewritten as

$$\left(\frac{\|\boldsymbol{R}_j\|_{1,1}}{p_j^2}\right)^{1/2} \|\boldsymbol{R}_j\|_{1,1}^{1/2} \geq \left(\frac{\|\boldsymbol{R}\|_{1,1}}{p^2}\right)^{1/2} \|\boldsymbol{R}\|_{1,1}^{1/2}. \tag{8}$$

The right condition can be rewritten analogously and is omitted here for brevity.

In the context of Theorem 2.3, we argued that $\|\boldsymbol{R}_j\|_{1,1}/p_j^2 \geq \|\boldsymbol{R}\|_{1,1}/p^2$, and therefore also $(\|\boldsymbol{R}_j\|_{1,1}/p_j^2)^{1/2} \geq (\|\boldsymbol{R}\|_{1,1}/p^2)^{1/2}$, might generally be valid. However, regarding the remaining terms $\|\boldsymbol{R}_j\|_{1,1}^{1/2}$ and $\|\boldsymbol{R}\|_{1,1}^{1/2}$ on the left- and right-hand sides of Equation (8), we note that $\|\boldsymbol{R}\|_{1,1} = \|\boldsymbol{R}_j\|_{1,1} + \|\boldsymbol{R}_{j'}\|_{1,1} + 2\|\boldsymbol{R}_{jj'}\|_{1,1} > \|\boldsymbol{R}_j\|_{1,1}$, which imposes restrictions on the terms in Equation (8) and thus also in (7). Below, we discuss an example in which the ICLUST conditions are not satisfied, meaning that the procedure stops identifying the worst split-half reliability.

To the best of our knowledge, the conditions about Revelle's beta cannot be simplified further. Nevertheless, the following statement follows directly from Theorem 2.3.

https://doi.org/10.1017/psy.2026.10124 Published online by Cambridge University Press

Corollary 2.5. *Consider a correlation matrix as in (1), and let $\boldsymbol{X}_j$ and $\boldsymbol{X}_{j'}$ be the arguments that yield the worst split-half reliability in (3). The condition $\beta > \min\{\beta_j, \beta_{j'}\}$ is not satisfied if*

$$\frac{\|\boldsymbol{R}_j\|_{1,1}}{p_j^2} \leq \frac{\|\boldsymbol{R}\|_{1,1}}{p^2} \quad \textit{and} \quad \frac{\|\boldsymbol{R}_{j'}\|_{1,1}}{p_{j'}^2} \leq \frac{\|\boldsymbol{R}\|_{1,1}}{p^2}.$$

We now consider an example where none of the ICLUST conditions are satisfied, implying that the algorithm fails to identify the worst split-half reliability by construction.

Example 2.6. Consider the correlation matrices $\boldsymbol{R}$, $\boldsymbol{R}_j$, and $\boldsymbol{R}_{j'}$ as in (1) with

$$\boldsymbol{R} = \begin{pmatrix} 1 & 0.9 & 0.8 & 0.4 & 0.4 & 0.4 \\ 0.9 & 1 & 0.8 & 0.4 & 0.4 & 0.4 \\ 0.8 & 0.8 & 1 & 0.4 & 0.4 & 0.4 \\ 0.4 & 0.4 & 0.4 & 1 & 0.9 & 0.8 \\ 0.4 & 0.4 & 0.4 & 0.9 & 1 & 0.8 \\ 0.4 & 0.4 & 0.4 & 0.8 & 0.8 & 1 \end{pmatrix}, \; \boldsymbol{R}_j = \boldsymbol{R}_{j'} = \begin{pmatrix} 1 & 0.9 & 0.8 \\ 0.9 & 1 & 0.8 \\ 0.8 & 0.8 & 1 \end{pmatrix}.$$

Then, $\alpha \approx 0.89 \not> \min\{\alpha_j, \alpha_{j'}\} \approx 0.94$ and $\beta \approx 0.62 \not> \min\{\beta_j, \beta_{j'}\} = 0.9$, and consequently none of the ICLUST conditions are satisfied and the procedure stops.

Code to replicate this example is available in the Supplementary Material.

2.5. Comments on the computation of Guttman's lambda four

Guttman (1945) introduced a series of lower bounds for reliability. Similar to Revelle's beta, the fourth of these bounds, commonly referred to as Guttman's lambda four, considers a split of the test into two parts. Although originally defined for an arbitrary split-half, it is nowadays typically interpreted as the greatest split-half reliability (Jackson & Agunwamba, 1977; Revelle & Zinbarg, 2009). Thus, as with Revelle's beta, for a data matrix comprising p items, there are $2^{p-1} - 1$ candidate splits.

However, the greatest split-half reliability cannot be computed using the same approach as that used for identifying the worst split-half. To elaborate, Guttman's lambda four is defined as

$$\lambda_4 = \max_{\boldsymbol{X}_j, \boldsymbol{X}_{j'}} \frac{4\|\boldsymbol{R}_{jj'}\|_{1,1}}{\|\boldsymbol{R}\|_{1,1}}. \tag{9}$$

Similar as for Revelle's beta, we note that identifying the submatrices with the greatest split-half reliability is conceptually identical to identifying the submatrices that exhibit the largest sum of between-test correlation in absolute terms.

Remark 2.7. The arguments $\boldsymbol{X}_j$ and $\boldsymbol{X}_{j'}$ that minimize the optimization problem in (9) are identical to the ones of the optimization problem

$$\max_{\boldsymbol{X}_j, \boldsymbol{X}_{j'}} \|\boldsymbol{R}_{jj'}\|_{1,1}. \tag{10}$$

Proof. The proof is provided in Section A of the Supplementary Material. □

However, this optimization problem does not align with Equation (4) as it corresponds to finding the worst instead of best block diagonal matrix approximation with respect to the function $f(\boldsymbol{A}) \mapsto 2\|\boldsymbol{A}\|_{1,1}$.

Moreover, to the best of our knowledge, it is not possible to transform the optimization problem in Equation (10) into the form of Equation (4). Such a transformation would require mapping large (absolute) values of $\boldsymbol{R}$ to small (absolute) ones, and vice versa, effectively turning a maximization problem into a minimization problem. This transformation would not necessarily preserve positive definiteness, as shown, for example, in Theorem 2.1 of FitzGerald et al. (1995), which is however a requirement for the approach proposed by Bauer (2025b).

https://doi.org/10.1017/psy.2026.10124 Published online by Cambridge University Press

3. Simulation study

3.1. Preliminaries

All computational results in Section 3 and 4 were conducted in `R` 4.1.3 (R Core Team, 2022) on a PC running macOS Sequoia 15.6.1. with 8 GB of RAM. Code to replicate the simulations and examples of these sections alongside details about computational implementations are provided in the Supplementary Material. Furthermore, computing Revelle's beta using HC-SVD has been implemented specifically for this research project in the `R` package `blox` (Bauer, 2025a).

3.2. Methods to compare

We compare our approach to ICLUST, which is implemented in the `R` package `psych` (Revelle, 2023). Revelle (1979) suggested that agglomerative ("bottom-up") hierarchical clustering using average linkage could be employed to identify the worst split-halves. However, he noted that agglomerative clustering does not always find the true worst split. Bauer (2025b) elaborates on this argument and showcases the differences between agglomerative and divisive clustering. Nonetheless, given that agglomerative hierarchical clustering with average linkage is a well-established method that exhibits fast computation times, we include it in our simulation study for comparison in the context of identifying Revelle's beta. Furthermore, an ad-hoc procedure is to consider only a randomly chosen subset of candidate splits. However, such a procedure cannot guarantee to find the worst split-half by construction.

There are also other non-agglomerative clustering methodologies like divisive analysis clustering (DIANA) introduced by Kaufman and Rousseeuw (1990) or an approach by Bottazzi Schenone et al. (2025) and Cavicchia et al. (2020) that aims to identify hierarchical clusters by finding a so-called ultrametric correlation matrix that fits the underlying data. However, their performances are less robust compared to HC-SVD (see Bauer, 2025b for a comparison) and are therefore not considered in this work.

3.3. Simulation designs

We consider two simulation designs to evaluate the performance of the different methods. In both designs, we construct a correlation matrix $\boldsymbol{R}$ from a data matrix $\boldsymbol{X}$ $(n \times p)$ of rank k. Specifically, the data matrix is generated as $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^T + \boldsymbol{E}$, where $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}_k$, $\boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}_k$, $\boldsymbol{D} = \text{diag}(d, \ldots, d)$ $(k \times k)$, and $\boldsymbol{E}$ is a noise matrix with components drawn independently from an $N(0, 0.5^2)$ distribution. To ensure orthonormality of $\boldsymbol{U}$, we generate it as the left orthonormal matrix of a QR decomposition of a matrix with independently $N(0, 1)$ components. Across both simulation designs, we set $n = 5p$. Thus, the matrix $\boldsymbol{E}$ introduces noise to the factor structure, while general sample noise is added by observing n realizations. The two simulation designs A and B differ in how the matrix $\boldsymbol{V}$ is constructed.

A. In design A, the data matrix $\boldsymbol{X}$ is generated by two underlying factors ($k = 2$) as

$$\boldsymbol{V} = \begin{pmatrix} \tilde{\boldsymbol{v}}_1 & 0 \\ 0 & \tilde{\boldsymbol{v}}_2 \end{pmatrix} (p \times 2),$$

with column vectors $\tilde{\boldsymbol{v}}_1$ (5×1) and $\tilde{\boldsymbol{v}}_2$ $((p-5) \times 1)$ whose components are drawn independently from a uniform $U(0.2, 0.8)$ distribution. Afterward, $\boldsymbol{V}$ is scaled to have unit-length columns. Thus, $\boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}_2$ holds by construction. We consider $d \in \{0.5, 1, 1.5\}$, where smaller values of d reduce the latent factor structure. A visualization is provided in Figure A.1 in the Supplementary Material.

B. In design B, the factor structure is randomly generated. First, we construct a symmetric and positive semi-definite matrix

$$\boldsymbol{S} = \boldsymbol{F}\boldsymbol{F}^T (p \times p),$$

https://doi.org/10.1017/psy.2026.10124 Published online by Cambridge University Press

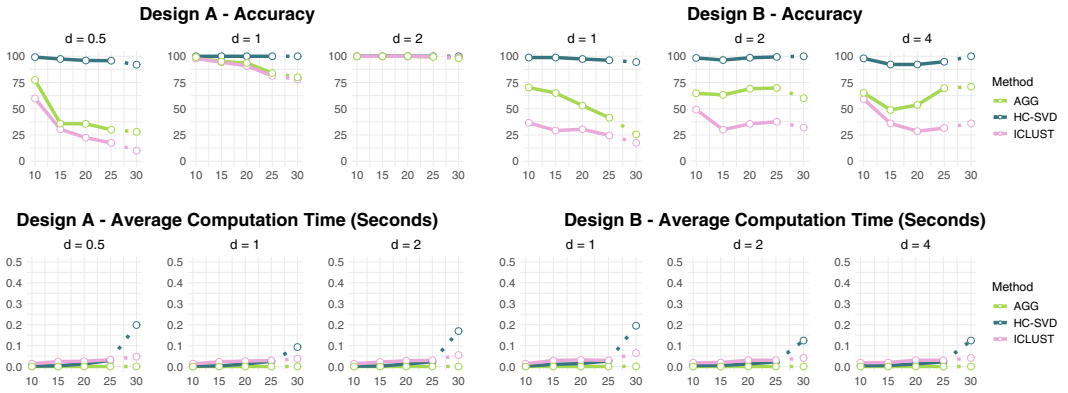


Figure 1. Simulation results for finding the worst split-half by agglomerative clustering using average linkage (AGG, *green*), HC-SVD (*blue*), and ICLUST (*pink*) when $p \in \{10, 15, 20, 25, 30\}$. For both simulation designs A (*left*) and B (*right*), we give the accuracy, defined as the percentage of correctly identified worst split-halves, (*top*) and average computation times in seconds (*bottom*). .

where $\boldsymbol{F}$ $(p \times 2)$ is a matrix with all components drawn independently from an $N(0, 1)$ distribution. The matrix $\boldsymbol{V}$ is then constructed by first $k = 2$ eigenvectors of $\boldsymbol{S}$. Thus, the factor structure is randomly generated. We consider $d \in \{1, 2, 4\}$. Note that the correlation matrix converges toward the identity matrix for smaller values of d. A visualization is provided in Figure A.1 in the Supplementary Material.

For both designs, the true worst split is unknown a priori and must therefore be identified by exhaustively evaluating all possible split-halves. This restricts the number of variables due to the computational constraints, as the number of candidate splits grows exponentially with the number of variables. For example, evaluating all splits for $p = 30$ required approximately 4 hours computation time per simulation iteration. As a result, the simulation study consists of two parts: of evaluating the total performance and of evaluating the relative performance. For the total performance, the simulation study was conducted for $p \in \{10, 15, 20, 25\}$ with 500 simulation iterations performed for each design. For $p = 30$, we ran 100 simulation iterations due to the computational cost. For the relative performance, we compare ICLUST and agglomerative clustering for identifying Revelle's beta to the HC-SVD-based approach. Since we compare relative performances, we do not evaluate all $2^{p-1} - 1$ candidate splits. This avoids the exponentially increasing computational costs by exhaustive search and we run the simulation study for $p \in \{50, 60, 70, 80, 90, 100\}$. The complete code used to run the simulation study is available in the Supplementary Material.

3.4. Simulation results

First, we report the accuracy of the three methods, defined as the percentage of correctly identified worst split-halves. In addition, we provide the average computation times in seconds.

The simulation study highlights the limitations of using agglomerative clustering or ICLUST for identifying the worst split-half. For both designs A and B, the accuracy of these two methods declines as the number of variables increases (*top* in Figure 1). Intuitively, as these methods are "bottom-up" approaches, they rely on identifying the correct hierarchical merging "path" from the bottom to the top-level structure. As the number of variables grows, the space of possible merge paths becomes increasingly complex and nested, reducing the likelihood of identifying this path and thus the correct worst split.

Design A supports this intuition. Both agglomerative clustering and ICLUST perform well when the factor structure is strong ($d = 2$), but their performance drops substantially under weaker signal

https://doi.org/10.1017/psy.2026.10124 Published online by Cambridge University Press

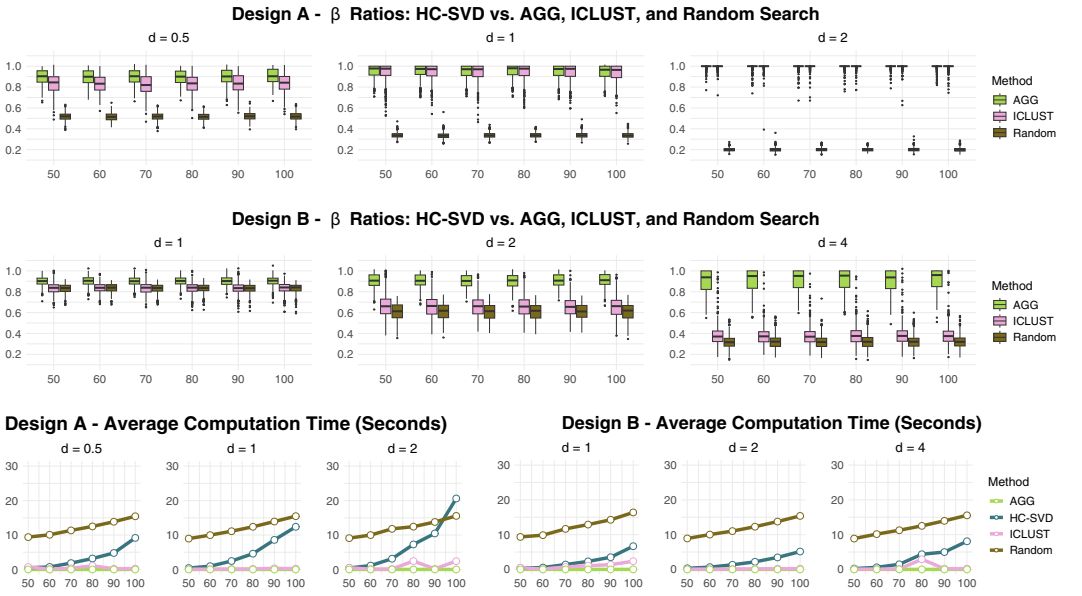


Figure 2. Simulation results for the relative performance by agglomerative clustering using average linkage (AGG, *green*), ICLUST (*pink*), and random search (*yellow*) compared to HC-SVD using the ratios $\beta_{\text{HC-SVD}}/\beta_{\text{AGG}}$, $\beta_{\text{HC-SVD}}/\beta_{\text{ICLUST}}$, and $\beta_{\text{HC-SVD}}/\beta_{\text{Random}}$ (*first two rows*) when $p \in \{50, 60, 70, 80, 90, 100\}$. For both simulation designs A and B, we also give the average computation times in seconds (*bottom*).

conditions ($d = 0.5$). This difference in signal strength also explains the performance drop observed from Design A to Design B. In Design A, the data matrix is driven by two underlying factors, which facilitates the hierarchical clustering process. Conversely, Design B exhibits a randomly generated correlation structure, providing less bottom-up guidance for agglomerative approaches.

In contrast to the agglomerative clustering and ICLUST baseline methods, HC-SVD consistently achieves substantially higher accuracy across all simulation settings. In most conditions, its accuracy relative to the exhaustive search is at or very close to 100%, indicating that the proposed method almost perfectly recovers the optimal worst split-half solution. Although HC-SVD exhibits a slight decrease in accuracy in the more challenging Design B scenarios, its performance remains high and clearly superior to the competing approaches. These small deviations from 100% accuracy arise in settings where the condition stated in Equation (6) of Theorem 2.2 is not satisfied. When this condition holds, the procedure is theoretically guaranteed to reproduce the exhaustive solution, that is, to recover the exact worst split-half partition.

Second, we report the relative performance of identifying the worst split-half reliability using agglomerative clustering β_{AGG}, ICLUST β_{ICLUST}, and random search β_{Random} compared to HC-SVD via the ratios $\beta_{\text{HC-SVD}}/\beta_{\text{AGG}}$, $\beta_{\text{HC-SVD}}/\beta_{\text{ICLUST}}$, and $\beta_{\text{HC-SVD}}/\beta_{\text{Random}}$. Consequently, a ratio of one indicates that both approaches identified the same split. A ratio greater than one means the respective approach found a smaller Revelle's beta, while a ratio less than one indicates it found a larger beta. Thus, values below one favor HC-SVD in identifying the worst split-half. For the random search, $2^{20-1} - 1$ random candidate splits were evaluated. For larger number of candidates, computation times strongly increased without showing substantially better performance. In addition, we provide the average computation times in seconds.

Figure 2 (*top*) presents these results. Excluding minor outliers, ratios remain mostly below one across all designs and values of p, demonstrating that HC-SVD consistently outperforms both agglomerative clustering and ICLUST. The performance margin is smallest for design A ($d = 2$), aligning with the generally high accuracy of all methods observed in Figure 1. Overall, random search (*yellow*) yields the weakest performance.

https://doi.org/10.1017/psy.2026.10124 Published online by Cambridge University Press

Table 1. Revelle's beta coefficients for firstborns and laterborns in the personality item survey

β	ALL	EXT	EST	AGR	CSN	OPN
Firstborns	0.43	0.76	0.62	0.54	0.60	0.55
Laterborns	0.46	0.77	0.58	0.53	0.60	0.55

Note: β is computed using all 50 items (ALL) and separately for the items corresponding to each of the factors extraversion (EXT), emotional stability (EST), agreeableness (AGR), conscientiousness (CSN), and openness to experience (OPN). Results are rounded to two decimal places.

Computation times for identifying the worst split-half reliability using HC-SVD remain within a range we consider feasible for application (see *bottom* in Figure 1 and *bottom* in Figure 2). Specifically, for $p \leq 30$ and for $p = 100$, the methods complete in under 0.3 seconds and under 24 seconds on average, respectively. Figure 2 indicates the quadratic grow in computational complexity of HC-SVD with increasing p. As shown in Bauer (2025b), computational time of HC-SVD could further be reduced by evaluating fewer candidate splits. The increase of computation time by the random search approach is caused by our implementation and is discussed in the Supplementary Material.

4. Real data example

In this section, we illustrate computation of Revelle's beta on a personality item survey. We analyze a data set comprising 50 personality items, categorized into five groups according the IPIP five factor markers (Goldberg, 1992): extraversion (EXT), emotional stability (EST), agreeableness (AGR), conscientiousness (CSN), and openness to experience (OPN), using the IPIP five factor markers. Data were collected with a research survey on the internet and can be accessed online at https://openpsychometrics.org/_rawdata/. Each personality trait has 10 Likert-rated items and we provide all item questions in the Supplementary Material for completeness.

The survey also contains information on participants' birth order. Since birth order has been studied in relation to personality traits (see, e.g., Black et al., 2018; Rohrer et al., 2015), we compute Revelle's beta separately for firstborn participants with siblings ($n_{FB} = 15,340$) and laterborn participants ($n_{LB} = 20,667$). This comparison is intended as an empirical illustration of the proposed computation. The results are reported in Table 1.

The estimates are broadly similar across the two birth-order groups. Moreover, the value of beta based on all 50 items is lower than the factor-specific values, which is consistent with the fact that the full item set combines items from multiple personality domains.

Similar to the simulation study, we compared our identified worst split-half to the one identified by an exhaustive search for all five factors among both groups, firstborns and laterborns, and our proposed methodology does find the worst split-half. For all items, an exhaustive search is computationally too demanding. As in the simulation study, we therefore compared our identified split to the one identified by ICLUST and agglomerative hierarchical clustering using average linkage. For both groups, our method identified a split that separates the items stronger, that is, our method identified a worse split-half compared to ICLUST and agglomerative hierarchical clustering using average linkage.

As in the simulation study, we compared the worst split-half identified by our method with the split-half obtained by exhaustive search. This comparison was feasible for each of the five factors and for both birth-order groups. In all cases, our proposed method identified the worst split-half. For the full set of 50 items, exhaustive search is computationally infeasible. We therefore compared the split identified by our method with those obtained by ICLUST and agglomerative hierarchical clustering using average linkage. For both birth-order groups, our method identified a split with lower average absolute between-half correlation, that is, a worse split-half than the splits identified by ICLUST and average-linkage

https://doi.org/10.1017/psy.2026.10124 Published online by Cambridge University Press

hierarchical clustering. In particular, this means that neither ICLUST nor agglomerative hierarchical clustering found the true worst split-half.

5. Discussion

In this work, we introduced an approach for computing Revelle's beta. Essentially, we elaborated that identifying the worst split-half reliability is conceptually identical to divisive ("top-down") hierarchical clustering using average linkage. Recently, a method was introduced to reduce the exponentially increasing computation problem for finding these splits in clustering contexts into a quadratic one. Thus, adaption of this approach allows computation of Revelle's beta.

Additionally, we implemented the approach in an `R` package publicly available on CRAN to facilitate its application.

The proposed methodology also suggests several directions for further research. One possible extension is to use Revelle's beta as an exploratory diagnostic tool for investigating potential substructures within a factor. Since beta is based on the worst split-half reliability and may reflect departures from internal homogeneity, repeated or recursive splits into internally coherent item subsets may provide indications of secondary structure or subfacets. Such use should, however, be regarded as exploratory: beta does not constitute a formal test of dimensionality and should therefore be interpreted in conjunction with factor-analytic evidence and substantive theory.

A related possibility is to use Revelle's beta as an exploratory tool for identifying items that reduce the internal homogeneity of a factor. For example, increases in beta following item deletion may indicate that particular items introduce variation that is not fully aligned with the dominant common factor. This approach could provide additional insight into the internal homogeneity structure of an item set and may inform the construction or refinement of scales. Nevertheless, such extensions should not be interpreted as definitive evidence for item removal, since beta reflects statistical homogeneity rather than theoretical relevance or construct coverage. Further theoretical and empirical work is needed to clarify the usefulness and limitations of such exploratory procedures, similar to item-deleted reliability approaches based on Cronbach's alpha or McDonald's omega (see, e.g., Dunn et al., 2014; Raykov, 2007).

Finally, we discussed that Revelle's beta does not satisfy the ultrametric property. Consequently, visualizing all worst split-halves of an item set in the form of a dendrogram may lead to inversions. However, the corresponding splits can be visualized using the average absolute correlation between split-halves, since this quantity satisfies the ultrametric property. A direction for future research is therefore to investigate whether such dendrogram-based visualizations of worst split-halves can contribute to the interpretation of the factors or substructures underlying a set of items.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/psy.2026.10124.

Data availability statement. Replication data and code can be found in the Supplementary Material.

Acknowledgements. I am grateful to Mauricio Garnier Villarreal for helpful discussions and valuable insights on issues related to reliability. I am also grateful to the Editor and the anonymous reviewers for their thoughtful and constructive feedback, which substantially improved the manuscript.

Funding statement. This research received no specific grant funding from any funding agency, commercial, or not-for-profit sectors.

Competing interests. The author declares none.

Software. HC-SVD and extensions of it to compute Revelle's beta as outlined in Section 2 of this work is available in the CRAN contributed `R` package `blox` (Bauer, 2025a).

References

Bauer, J. O. (2025a). Blox: Block diagonal matrix approximation. [R package version 0.0.1].

Bauer, J. O. (2025b). *Divisive hierarchical clustering using block diagonal matrix approximations*. Preprint. arXiv:2308.06820v4.

Bauer, J. O. (2025c). High-dimensional block diagonal covariance structure detection using singular vectors. *Journal of Computational and Graphical Statistics*, *34*(3), 1005–1016.

Bauer, J. O., & Drabant, B. (2021). Principal loading analysis. *Journal of Multivariate Analysis*, *184*, 104754.
Black, S. E., Grönqvist, E., & Öckert, B. (2018). Born to lead? The effect of birth order on noncognitive abilities. *The Review of Economics and Statistics*, *100*(2), 274–286.
Bottazzi Schenone, M., Cavicchia, C., Vichi, M., & Zaccaria, G. (2025). Ultrametric factor analysis for building hierarchies of reliable and unidimensional latent concepts. *Psychometrika*, *90*(2), 1–20.
Cavicchia, C., Vichi, M., & Zaccaria, G. (2020). The ultrametric correlation matrix for modelling hierarchical latent concepts. *AStA Advances in Statistical Analysis*, *14*, 837–853.
Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.
Dunn, T. J., Baguley, T. S., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399–412.
Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, *1*, 211–218.
FitzGerald, C. H., Micchelli, C. A., & Pinkus, A. (1995). Functions that preserve families of positive semidefinite matrices. *Linear Algebra and its Applications*, *221*, 83–102.
Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment*, *4*, 26–42.
Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, *10*(4), 255–282.
Hartigan, J. A. (1967). Representation of similarity matrices by trees. *Journal of the American Statistical Association*, *62*(320), 1140–1158.
Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. (2nd ed.) Springer.
Horn, R. A., & Johnson, C. R. (2013). *Matrix analysis*. (2nd ed.) Cambridge University Press.
Jackson, P. H., & Agunwamba, C. C. (1977). Lower bounds for the reliability of the total score on a test composed of non-homogeneous items: I: Algebraic lower bounds. *Psychometrika*, *42*(4), 567–578.
Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, *32*, 241–254.
Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis*. John Wiley.
Lee, M., Shen, H., Huang, J. Z., & Marron, J. S. (2010). Biclustering via sparse singular value decomposition. *Biometrics*, *66*(4), 1087–1095.
R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
Raykov, T. (2007). Reliability if deleted, not 'alpha if deleted': Evaluation of scale reliability following component deletion. *The British Journal of Mathematical and Statistical Psychology*, *60*(2), 201–216.
Reise, S. P., & Haviland, M. G. (2025). Understanding alpha and beta and sources of common variance: Theoretical underpinnings and a practical example. *Journal of Personality Assessment*, *107*(3), 267–282.
Revelle, W. (1978). ICLUST: A cluster analytic approach to exploratory and confirmatory scale construction. *Behavior Research Methods*, *10*(5), 739–742.
Revelle, W. (1979). Hierarchical cluster analysis and the internal structure of tests. *Multivariate Behavioral Research*, *14*(1), 57–74.
Revelle, W. (2023). psych: Procedures for psychological, psychometric, and personality research. [R package version 2.3.3].
Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, *74*(1), 145–154.
Rohrer, J. M., Egloff, B., & Schmukle, S. C. (2015). Examining the effects of birth order on personality. *Proceedings of the National Academy of Sciences*, *112*(46), 14224–14229.
Rossiter, J. R. (2002). The C-OAR-SE procedure for scale development in marketing. *International Journal of Research in Marketing*, *19*(4), 305–335.
Shen, H., & Huang, J. Z. (2008). Sparse principal component analysis via regularized low rank matrix approximation. *Journal of Multivariate Analysis*, *99*(6), 1015–1034.
Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107–120.
Tang, W., Cui, Y., & Babenko, O. (2014). Internal consistency: Do we really know what it is and how to assess it? *Journal of Psychology and Behavioral Science*, *2*(2), 205–220.
Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, *58*(1), 267–288.

Cite this article: Bauer, J. O. (2026). Revelle's Beta: The Wait Is Over—Computation Becomes Possible. *Psychometrika*, 1348–1360. https://doi.org/10.1017/psy.2026.10124

https://doi.org/10.1017/psy.2026.10124 Published online by Cambridge University Press