

KR-20 and KR-21 for some non-dichotomous data (It's not just Cronbach's alpha)

Robert C. Foster

Bettis Atomic Power Laboratory

Abstract

This paper presents some equivalent forms of the common Kuder-Richardson Formula 21 and 20 estimators for non-dichotomous data belonging to certain other exponential families, such as Poisson count data, exponential data, or geometric counts of trials until failure. Using the generalized framework of Foster (2020), an equation for the reliability for a subset of the natural exponential family have quadratic variance function is derived for known population parameters, and both formulas are shown to be different plug-in estimators of this quantity. The equivalent Kuder-Richards formulas 20 and 21 are given for six different natural exponential families, and these match earlier derivations in the case of binomial and Poisson data. Simulations show performance exceeding that of Cronbach's alpha in terms of root mean squared error when the formula matching the correct exponential family is used, and a discussion of Jensen's inequality suggests explanations for peculiarities of the bias and standard error of the simulations across the different exponential families.

Keywords— Reliability, KR-20, KR-21, Cronbach's Alpha, Exponential Families

1 Introduction

Formulas 20 and 21 of Kuder and Richardson (1937), abbreviated throughout this paper as KR20 and KR21, are some of the earliest and most well-known formulas in assessing reliability of a test. For dichotomous data, the formulas are given by

$$KR20 = \frac{k}{k-1} \left(1 - \frac{\sum_{j=1}^k p_j(1-p_j)}{\sigma_X^2} \right)$$
$$KR21 = \frac{k}{k-1} \left(1 - \frac{k\bar{p}(1-\bar{p})}{\sigma_X^2} \right)$$

where k is the test length, σ_X^2 is the variance of sum test scores, p_j is the proportion of correct responses to test item j , and \bar{p} is the average correct response over all items. The most common estimator of reliability, Cronbach's alpha, is often seen as a general version of KR20 (Cronbach, 1951). Beyond Cronbach's alpha, there does not seem to have been many attempts to determine equivalent variants of KR20 and KR21 to specific types non-dichotomous data, such as count data. Allison (1978) derived a KR21 equivalent for Poisson distributed data, but most have stuck to Cronbach's alpha as a general estimator of reliability.

The topic of reliability estimation, and Cronbach's alpha in particular, has been a subject of much recent debate within the psychometric literature. Alpha has been criticized as having assumptions which are not realistic in practice (McNeish, 2018; Schmitt, 1996; Sijtsma, 2009). Criticisms have focused on alpha's assumptions of tau equivalence and uncorrelated errors. McNeish (2018) claims that normality is an assumption of Cronbach's alpha, but Raykov and Marcoulides (2019) rebut that no assumptions of normality are made in the derivation of alpha and consistency as an estimator does not depend on an assumption of normality. However, Zumbo (1999) notes that though the classical test theory derivation of alpha makes no assumptions of normality, estimators of alpha often do, in agreement with Bay (1973) who noted in his time that early derivations of statistical properties of estimators such as Cronbach's alpha and KR20 as in L. S. Feldt (1965) were developed using an ANOVA model which includes an assumption of normality. More recent research in Zyl, Neudecker, and Nel (2000) on the sampling distribution of cronbach's alpha using maximum likelihood techniques also assumes normality, though nonparametric methods have been developed for psychometric calculations using techniques like the bootstrap as in Raykov (1998). The statistical properties of alpha and other estimators have been little explored outside of these assumptions of normality, with simulation studies in Sheng and Sheng (2012) and Zimmerman, Zumbo, and Lalonde (1993). Beyond classical test theory, Geldhof, Preacher, and Zyphur (2014) explicitly state that non-normal data is a limitation for reliability estimation in a multilevel confirmatory factor analysis framework. As Zinbarg, Yovel, Revelle, and McDonald (2006) note, properties of alternatives to Cronbach's alpha based on a factor analysis are not well known under non-normality, despite most psychological data being non-normal. It is clear that there is a need for analysis of reliability for non-normal data.

The purpose of this paper is to make a contribution towards the analysis of reliability under non-normality by reversing direction from Cronbach's alpha once again to derive

formulas to estimate test reliability in special cases – to find equivalent forms of KR20 and KR21 which can be used to assess reliability of tests for specific types of non-dichotomous data. For example, chapter 21 of Lord, Novick, and Birnbaum (1968), discussing the work of Rasch (1960), notes that the number of an examinee’s misreadings in an oral reading test may be modeled as a Poisson random variable. If the response of interest were instead the time between misreadings, an exponential distribution would then be appropriate. Meredith (1971) also showed that a Poisson process is appropriate for tests of speed, and noted that under certain assumptions the distributions of observed scores ought to be negatively binomial distributed. The key idea is this: KR20 and KR21 can be seen as a specific version of an estimator for reliability for Bernoulli distributed item responses where the mean-variance relationship of the Bernoulli distribution is exploited for variance calculations rather than using standard sample variances. Other exponential family distributions also have mean-variance relationships. By working within the generalized framework of Foster (2020) and deriving a formula for reliability in exponential families which uses the mean-variance relationship and which matches traditional KR20 and KR21 in the binomial case, equivalent versions of KR20 and KR21 are obtained for data from other exponential family distributions. The KR20 and KR21 equivalent formulas are shown for when sum test scores can be said to follow one of the six natural exponential family distributions with quadratic variance function (NEF-QVF): the normal, the binomial, the Poisson, the gamma, the negative binomial, or the natural exponential family generated by a convolution of generalized hyperbolic secant functions (NEF-GHS).

Section 2 describes the properties of the generalized framework for reliability described in Foster (2020) which form the basis for derivations of the formulas and definition of reliability as parallel-test correlation, and states which assumptions are slightly modified for the purposes of this paper. Section 3 shows how the mean-variance relationship can be exploited to obtain test reliability and derives the general KR20 and KR21 formulas as estimators of this relationship, giving the formulas for each NEF-QVF distribution. These are shown to match the traditional KR20 and KR21 of Kuder and Richardson (1937) in the binomial case, and the Poisson KR21 of Allison (1978) in the Poisson case. The conditions for algebraic equivalence between KR20 and Cronbach’s alpha are also discussed. Section 4 performs a simulation study showing that these formulas do appear to converge to the population reliability as the number of subjects and test length increases and comparing the equivalent KR20 and KR21 formulas to Cronbach’s alpha and each other in terms of root mean squared error, bias, and standard deviation. Results indicate that when the formulas are used for data following the appropriate exponential family, performance is improved over Cronbach’s alpha in terms of RMSE, though whether KR20 or KR21 is superior depends on the variance function of the exponential distribution. A brief discussion of Jensen’s inequality indicates a possible explanation for why one formula is superior to another for a given distribution.

2 Framework

This paper follows the generalized framework of Foster (2020), where “generalized” is used in the same sense of a generalized linear model which may deal with non-normal exponential

family data. A complete theoretical description of the framework is given in Foster (2020), but this paper will only state without proof elements which are necessary. One major difference is that while the framework of Foster (2020) applies to the entire natural exponential family, this paper focuses only on the members of the exponential family having quadratic variance function so that the variance is a polynomial function of the mean up to degree two. For example, a Bernoulli distribution with success probability p has mean p and variance $p(1-p) = p - p^2$, so the variance is a quadratic function of the mean. Such distributions and their properties are extensively described in Morris (1982) and Morris (1983), which serve as a general reference. Furthermore, while the framework of Foster (2020) allows the test length to vary between subjects, this paper assumes a common test length for all subjects for the sake of simplicity. The framework of Foster (2020) is most closely related to the strong true-score theory found in chapters 21 through 24 of Lord et al. (1968) and associated papers such as Lord (1965) and Keats and Lord (1962) in deriving properties of reliability when specific distributional forms can be assumed, but rather than dealing individually with the binomial and Poisson distributions, the framework of Foster (2020) derives properties for the exponential family in general. The framework makes assumptions of true score distributions in such a way as to ensure that the regression of true score on observed score is linear, though the observed scores are non-normal.

Let $i = 1, 2, \dots, n$ index test subject and let $j = 1, 2, \dots, k$ index test item. In this framework, each test item for each subject Y_{ij} is assumed to identically follow a natural exponential family distribution with quadratic variance function (NEF-QVF), with independence conditional on subject ability θ_i . The six NEF-QVF distributions which may be used as generator distributions in this fashion are the normal, Bernoulli, Poisson, exponential, geometric, and generalized hyperbolic secant densities. Each of these distributions is closed under convolution. For example, the normal is a sum of normals, the binomial is the sum of Bernoullis, and the gamma is a sum of exponentials. Let $X_i = \sum_{j=1}^k Y_{ij}$ be the sum score for subject i , summing over all k test items. The six possible distributions for X_i are then the normal, binomial, Poisson, gamma, negative binomial, and NEF-GHS.

Furthermore, let abilities θ_i follow the corresponding conjugate prior $g(\theta_i|\mu, M)$ for the natural exponential family distribution, where $\mu = E[\theta_i]$ and $M = E[V(\theta_i)]/Var(\theta_i)$. For example, if the test items Y_{ij} are Bernoulli distributed with mean given by subject ability θ_i , then subject sum scores X_i are binomial distributed, and abilities θ_i themselves follow a beta distribution with parameters $\mu = \alpha/(\alpha + \beta)$ and $M = \alpha + \beta$. A complete description of several common NEF-QVF distributions, their conjugate priors, and the appropriate parameterizations in terms of μ and M is given in the appendix of Foster (2020). The model can be thought of as a hierarchy, with observed scores at the top level and abilities at the bottom level.

For natural exponential family distributions with quadratic variance function, the variance is a polynomial function of the mean of up to degree two.

$$Var(Y_{ij}|\theta_i) = V(\theta_i) = v_0 + v_1\theta_i + v_2\theta_i^2 \tag{1}$$

For example, the Bernoulli distribution has $V(\theta_i) = \theta_i(1 - \theta_i) = \theta_i - \theta_i^2$, so $v_0 = 0, v_1 = 1$, and $v_2 = -1$. The normal distribution, which assumes the variance around each item response is known to be σ^2 , has $V(\theta_i) = \sigma^2$ constant.

In this framework, the conditional and unconditional expectations of test item Y_{ij} are given by

$$\begin{aligned} E[Y_{ij}|\theta_i] &= \theta_i \\ E[Y_{ij}] &= \mu \end{aligned} \tag{2}$$

where μ is the population mean ability. The conditional expectation being equal to ability θ_i in Equation (2) is an implicit assumption that each item is of equal difficulty conditional on subject with ability θ_i , the necessity of which is discussed in Section 6. The conditional and unconditional variances are given by

$$\begin{aligned} Var(Y_{ij}|\theta_i) &= V(\theta_i) \\ Var(Y_{ij}) &= E[V(\theta_i)] + Var(\theta_i) \end{aligned} \tag{3}$$

where $V(\theta_i)$ is the variance function in Equation (1) applied to abilities θ_i . Then sum scores X_i have conditional and unconditional expectations

$$\begin{aligned} E[X_i|\theta_i] &= k\theta_i \\ E[X_i] &= k\mu \end{aligned} \tag{4}$$

Correspondingly, the conditional and unconditional variances of X_i are given by

$$\begin{aligned} Var(X_i|\theta_i) &= kV(\theta_i) \\ Var(X_i) &= kE[V(\theta_i)] + k^2Var(\theta_i) \end{aligned} \tag{5}$$

Within this framework, the test reliability is defined as the correlation between parallel tests, where parallel means that each test consists of identical, conditionally independent items Y_{ij} following the same exponential family model with means and variances given by Equations (2), (3), (4), and (5). As shown in Foster (2020), when this condition is met the test reliability is equal to

$$\rho = \frac{k}{M + k}$$

where k is the test length and $M = E[V(\theta_i)]/Var(\theta_i)$ is the parameter of the underlying distribution of abilities $g(\theta_i|\mu, M)$. This is also one minus the shrinkage parameter in the Bayesian posterior distribution. Cronbach's alpha reduces to this quantity in the framework.

Because the response to each test item Y_{ij} is conditionally independent and identical on ability θ_i , this framework is unidimensional. As shown in Equation (3) and in Foster (2020), the unconditional variances of each test item and unconditional covariances between test items are equal. The variance-covariance matrix between test items implied in this framework is thus tau-equivalent with equal variances, though the variance around each test item Y_{ij} is different for each subject i because of the mean-variance relation given in Equation (1).

3 KR20 and KR21 as Estimators of Reliability

Conjugate priors for natural exponential families have a close relationship between their mean μ and their variance $Var(\theta_i)$, connecting to the variance function of the original exponential family distribution they are conjugate to. From Morris (1983), the variance of the conjugate prior for natural exponential families with a quadratic variance function is

$$V(\mu) = Var(\theta_i)(M - v_2) \quad (6)$$

where $V(\mu)$ is the variance function of Equation (1) applied to the mean μ , M is the parameter of the conjugate distribution of abilities $g(\theta_i|\mu, M)$, and v_2 is the coefficient of the quadratic term of the variance function in Equation (1). For example, a beta distribution for θ_i with parameters $\mu = \alpha/(\alpha + \beta)$ and $M = \alpha + \beta$ has variance $Var(\theta_i) = \mu(1 - \mu)/(M + 1)$. With $V(\theta_i) = \theta_i(1 - \theta_i)$ for the Bernoulli distribution, this gives $V(\mu) = (M + 1)Var(\theta_i)$.

Define the the quantity R as

$$R = \frac{k}{k + v_2} \left(1 - \frac{kV(\mu)}{Var(X_i)} \right) \quad (7)$$

where $Var(X_i)$ is the unconditional variance. Using the mean-variance relationship given in Equation (6) and the variances given in Equation (5), this quantity becomes

$$\begin{aligned} R &= \frac{k}{k + v_2} \left(1 - \frac{kV(\mu)}{Var(X_i)} \right) \\ &= \frac{k}{k + v_2} \left(1 - \frac{kVar(\theta_i)(M - v_2)}{kE[V(\theta_i)] + k^2Var(\theta_i)} \right) \\ &= \frac{k}{k + v_2} \left(\frac{kE[V(\theta_i)] + k^2Var(\theta_i) - kVar(\theta_i)(M - v_2)}{kE[V(\theta_i)] + k^2Var(\theta_i)} \right) \\ &= \frac{k}{k + v_2} \left(\frac{M + k - (M - v_2)}{M + k} \right) \\ &= \frac{k}{k + v_2} \left(\frac{k + v_2}{M + k} \right) \\ &= \frac{k}{M + k} \\ &= \rho \end{aligned} \quad (8)$$

The last line in Equation (7) is the parallel-test reliability, as previously stated. For dichotomous, Bernoulli distributed item responses, which have $V(\theta_i) = \theta_i - \theta_i^2$ and $v_2 = -1$, substitution into the quantity R in Equation (7) gives

$$R = \frac{k}{k - 1} \left(1 - \frac{k\mu(1 - \mu)}{Var(X_i)} \right)$$

which is strikingly similar to the original KR20 and KR21 estimators of Kuder and Richardson (1937), but depends on known population parameters rather than sample quantities.

The parameters in Equation (7) are unknown, however, or else there would be no need to administer any sort of test. The question is then, is it possible to construct a

consistent estimator for Equation (7) from sample quantities? The answer is yes. Given a consistent estimator for $V(\mu)$ and a consistent estimator for $Var(X_i)$, then by Slutsky's theorem, using these as plug-ins for the numerator and denominator of Equation (7) will produce a consistent estimator for the population test reliability.

For $Var(X_i)$, the standard unbiased variance moment-based estimator of variance is used for raw sum scores x_i .

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

A note on the equivalence of KR20 and Cronbach's alpha for dichotomous data: these are often stated to be exactly equal; however, algebraic equivalence only occurs when the biased estimator of the sample variance which divides by the number of subjects n is used for $Var(X_i)$ rather than the unbiased estimator which divides by $n-1$ (for Cronbach's alpha itself, either estimator of variance yields identical estimates so long as it is used consistently). For all calculations in this paper, the unbiased estimator of sample variance is used. The historical reason for introducing bias into the denominator of the KR20 estimator of reliability appears to be obtaining algebraic equivalence Cronbach's alpha. As the results of the simulation study in Section 4 show, however, using the unbiased estimator of variance produces improved performance over alpha.

For the quantity $V(\mu)$, more than one estimator is possible. Because $V(\theta_i)$ is a continuous polynomial for all six distributions considered, any consistent estimator for μ is consistent for $V(\mu)$ by the continuous mapping theorem. The simplest estimator is to note that from Equation (4), $E[X_i] = k\mu$, and so $\frac{1}{k}$ times the sample mean of sum scores has expectation $E[\frac{1}{k}\bar{x}] = \frac{1}{k}(k\mu) = \mu$. Hence, $\frac{1}{k}\bar{x}$ is a consistent estimator for μ by the law of large numbers, and so $V(\frac{1}{k}\bar{x})$ is a consistent estimator for $V(\mu)$.

An alternative estimator is constructed by observing that the terms in the mean of sum scores \bar{x} can be rearranged to equal the sum of test item means.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k y_{ij} = \frac{1}{n} \sum_{j=1}^k \sum_{i=1}^n y_{ij} = \sum_{j=1}^k \bar{y}_j \quad (9)$$

From Equation (2), a single test item Y_{ij} has $E[Y_{ij}] = \mu$. By the law of large numbers, the item sample mean \bar{y}_j converges to μ for each item j . Then $V(\bar{y}_j)$ is also a consistent estimator for $V(\mu)$.

In Equation (7), the desired quantity is not $V(\mu)$, but $kV(\mu)$. Because the variance function applied each item mean $V(\bar{y}_j)$ is independently consistent for $V(\mu)$, a sum of $V(\bar{y}_j)$ over all k items gives the desired result. Plugging this into Equation (7) with unbiased sample variance s_x^2 for the denominator yields the generalized form of KR20:

$$KR20 = \frac{k}{k + v_2} \left(1 - \frac{\sum_{j=1}^k V(\bar{y}_j)}{s_x^2} \right) \quad (10)$$

For dichotomous, Bernoulli distributed item responses with $V(\theta_i) = \theta_i - \theta_i^2$ so that $v_2 = -1$, this gives:

$$\frac{k}{k-1} \left(1 - \frac{\sum_{j=1}^k \bar{y}_j (1 - \bar{y}_j)}{s_x^2} \right)$$

which is the traditional KR20 formula, with $\bar{y}_j = p_j$ as the proportion of correct responses on item j .

The alternative is to use $kV(\frac{1}{k}\bar{x})$ as a consistent estimator for $kV(\mu)$. Plugging this and s_x^2 into Equation (7) yields the generalized form of KR21:

$$KR21 = \frac{k}{k+v_2} \left(1 - \frac{kV(\frac{1}{k}\bar{x})}{s_x^2} \right) \quad (11)$$

For dichotomous, Bernoulli distributed item responses, this is

$$\frac{k}{k-1} \left(1 - \frac{k(\frac{1}{k}\bar{x})(1 - \frac{1}{k}\bar{x})}{s_x^2} \right)$$

which is the traditional KR21 formula, as \bar{x} is the average sum score and so $\frac{1}{k}\bar{x}$ is the average proportion correct \bar{p} .

The difference between KR20 and KR21, as far as this framework is concerned, is whether the variance function $V(\cdot)$ is applied “inside” the sum of item means on the right-hand side of Equation (9), yielding KR20, or “outside” the sum, yielding KR21. A table showing the six different NEF-QVF distributions for item responses, their variance functions, and their corresponding KR20 and KR21 estimators is given in Table 1. As shown, the binomial formulas match the formulas originally derived in Kuder and Richardson (1937). The Poisson KR21 formula here exactly matches equation thirteen of Allison (1978), which originally derived KR21 for Poisson count data. This paper shows that it is also the Poisson KR20 equivalent because the variance function for the Poisson distribution is simply the identity $V(\theta_i) = \theta_i$, and so applying Equation (9) gives equality. The rest of the formulas are new, so far as can be determined. For the normal distribution, the error variance σ^2 around the response to each item is assumed known. As this is extremely unlikely to be the case in practice, Cronbach’s alpha is recommended instead.

4 Simulation Study

4.1 Simulation Method

A reasonable question is, why use these generalized KR20 or KR21 estimators when Cronbach’s alpha is available? What advantage do these estimators have? This question is answered with a simulation study.

The purpose of this simulation study is to show that when data is simulated from the correct NEF-QVF model, the generalized KR20 and KR21 estimators corresponding to that family both converge to the population reliability as the test length and number of subjects increase, and that they do so as or more efficiently than Cronbach’s alpha. It is not a simulation study to determine all properties of the estimators or to investigate all potential sources and magnitudes of bias, though several are discussed. For data generation,

the method described in the appendix of Foster (2020) is used, which is identical to the simulation method of Huynh (1979) in the case of the binomial-beta model. Test data is simulated using the following algorithm:

1. Choose the appropriate NEF-QVF distribution for test item responses y_{ij} , the mean ability μ , the number of subjects n , the number of test items k , and desired population reliability ρ .
2. Calculate the M required to obtain the desired population reliability ρ as

$$M = \left(\frac{1 - \rho}{\rho} \right) k$$

3. Simulate n abilities θ_i from the conjugate prior $g(\theta_i|\mu, M)$, one for each subject i . The appropriate parameterization of the prior in terms of μ and M is given in the appendix of Foster (2020)
4. For each θ_i , simulate k responses from the NEF-QVF distribution $p(y_{ij}|\theta_i)$. The sum score x_i for each subject i is calculated as the sum of these responses over the k items.

This algorithm implies the underlying distribution of talent levels $g(\theta_i|\mu, M)$ is different for each (k, ρ) pair, as the parameter M of the distribution is calculated anew for each. Hence, it is difficult to directly compare simulation results across two identical k values if ρ differs, or across identical ρ values if k differs. Again, the primary goal is to show convergence. The alternative is to keep the distribution of talent levels constant by keeping M constant and varying the the reliability through the test length k , but this was not chosen because it limits the potential values of the desired population reliability ρ and preliminary simulations indicated that the shape of the distribution of talent levels $g(\theta_i|\mu, M)$ was less influential than the value of ρ , with the exception of extremely skewed distributions which may be of interest in other simulations. The balance of choices in producing a desired population reliability ρ is delicate and it is not possible within this framework to have complete freedom in all choices. A more complete simulation study might achieve a desired ρ through different methods in order to determine the effect of each. Though the mean ability μ does not affect the population reliability directly, it does have the potential to affect the sampling distribution of reliability through manipulating aspects of the distribution of abilities $g(\theta_i|\mu, M)$ such as skew and kurtosis. In particular, preliminary simulations indicated that μ values near the edges of the distribution support may lead to very skewed $g(\theta_i|\mu, M)$ distributions which could potentially have a large influence, but these were not considered for further study by simulation in this paper. The mean μ is instead kept constant throughout the simulation studies to avoid yet another quantity which could potentially affect results, though it could be of interest in a further simulation study. This algorithm also produces items which are of equal difficulty conditional on subject with ability θ_i , though items may have different algebraic means.

For the choice of exponential family model, the Poisson-gamma, gamma-inverse gamma, negative binomial-F, and binomial-beta distributions are used, where the first distribution named is the distribution of sum scores x_i and the second is the conjugate prior

$g(\theta_i|\mu, M)$. The binomial-beta distribution corresponds to the Kuder and Richardson (1937) coefficients and has been extensively studied, but is also shown here for completeness and comparison to results for other estimators. The normal-normal assumes the variance σ^2 is known, and as such is not of practical use. The distribution in the last row of Table 1, the natural exponential family generated by a convolution of generalized hyperbolic secant distributions, is neither common nor easy to simulate from.

For population ability ρ , the values 0.3, 0.6, and 0.8 are used, representing low, moderate, and high reliability. For test length, the values $k = 5, 10$, and 30 are used, representing a short, medium, and long test. For number of subjects, the values $n = 30, 75$, and 500 are used, representing a small, moderate, and large number of subjects. The value of the mean ability μ is constant within each set of simulations at $\mu = 1$ for the Poisson-gamma model and the gamma-inverse gamma model, $\mu = 1.01$ for the negative binomial-F model, and $\mu = 0.5$ for the binomial-beta model. These were simply chosen as reasonable values which worked well in simulation.

This gives a total of 27 sets of (ρ, k, n) values for each exponential family model and four models. Details of the implied variance-covariance matrix are given in the Appendix. For each set of values, a total of $N_{sim} = 1,000,000$ data sets were simulated and Cronbach's alpha and the KR20 and KR21 equivalents given by Equations (10) and (11) were calculated for each. To judge the estimator for a given set of estimated reliabilities $\hat{\rho}$, the sample root mean squared error is used.

$$RMSE(\hat{\rho}) = \sqrt{\frac{1}{N_{sim}} \sum_{s=1}^{N_{sim}} (\rho - \hat{\rho}_s)^2}$$

The root mean squared error is further decomposed into the traditional bias and variance (squared standard deviation) of the estimator.

$$RMSE(\hat{\rho}) = \sqrt{Bias^2(\hat{\rho}) + Var(\hat{\rho})} = \sqrt{\left(\frac{1}{N_{sim}} \sum_{s=1}^{N_{sim}} (\rho - \hat{\rho}_s)\right)^2 + \frac{1}{N_{sim}} \sum_{s=1}^{N_{sim}} (\hat{\rho}_s - \bar{\rho}_s)^2}$$

where $\bar{\rho}_s$ is the sample mean of the $\hat{\rho}_s$. Note that decomposition requires the biased variance formula for exact algebraic equivalence, but for $N_{sim} = 10^6$ total simulations using either the biased or unbiased variance estimate gives identical results within the first three decimal places. The decomposition allows for further inspection of results in order to determine exact properties of the estimator and answer questions about why a particular estimator might be performing better or worse than another.

The results of the simulation study are shown in Table 2 and Figure 1 for the Poisson-gamma model, Table 3 and Figure 2 for the gamma-inverse gamma model, Table 4 and Figure 3 for the negative binomial-F model, and Table 5 and Figure 4 for the binomial-beta model. In the results the sample standard deviation is given rather than the sample variance because the sample variance was often extremely small. Results are reported to three decimal places. The largest sample variance reported for any estimator is 0.251² for Cronbach's alpha in the first row of Table 3; thus, the approximate Monte Carlo standard error is less than or equal to $0.251/\sqrt{10^6} = 0.000251$, and quite often much smaller. For

this reason, the reporting of three decimal places in results is appropriate. Again, note that the unbiased estimator of sample variance s_x^2 which divides by $n - 1$ is used in all cases. Furthermore, intervals around derived quantities are small. A 99% bootstrap interval for the RMSE of Cronbach’s alpha in the first row of Table 3, based on 10,000 bootstrap samples of the one million simulated alpha values, is (0.251, 0.252). The standard error of the point estimate of RMSE is thus smaller than the rounding amount 0.001, so confidence intervals surrounding the point estimates are not included.

4.2 Analysis of Results

The results of Tables 2, 3, 4, and 5 show that the primary aims of the simulation study are met. Both Cronbach’s alpha and the estimators in Table 1 decrease in RMSE and appear to be converging towards zero as the test length k and the number of subjects n increase. This can be seen for $\rho = 0.3$ in Figures 1, 2, 3, and 4. Furthermore, the estimators in Table 1 are nearly universally superior to Cronbach’s alpha, answering the initial question posed of why they might be preferred over alpha. It is only in high-data situations with a relatively large number of test items k and large number of subjects n that all estimators are nearly equal and the root mean squared error is approximately the same no matter which estimator you choose. In general, it appears that increasing the number of subjects n can greatly lead to a reduction in RMSE through reduction of both bias and variance. For shorter tests with a smaller number of subjects, the generalized estimators provided by Equations (10) and (11) often strongly outperform alpha in terms of RMSE. However, there are noticeable differences in the exact manner the increase occurs. For example, the generalized KR21 estimator outperforms the generalized KR20 estimator for the gamma-inverse gamma and negative-binomial F results in Tables 3 and 4, but the reverse occurs for the binomial-beta model results given in Table 5. The bias of all three estimators is also generally negative, except for KR20 and KR21 in the binomial-beta model results given in Table 5. There are also differences in variances of the estimators across models. Though formal proof is not offered, some probability theory may help to explain the idiosyncrasies within each table and between tables.

First, it is noticeable that in all three sets of simulations the bias is negative for Cronbach’s alpha in all cases, though converging to zero as the number of subjects n increases. Why? From Foster (2020), Cronbach’s alpha is estimating the quantity

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{j=1}^k \text{Var}(Y_j)}{\text{Var}(X)} \right)$$

using plug-in estimates of the sample variances $s_{y_j}^2$ and s_x^2 , though as in Foster (2020) this paper assumes $\text{Var}(Y_j)$ is identical for all test items j . Essentially, it is a ratio of sample variances, and ratios of sample variances tend to be biased upwards. For independent variances, Jensen’s inequality shows that the bias is always high, as $1/s_x^2$ is a convex function of s_x^2 and so $E[1/s_x^2] \geq 1/E[s_x^2] = 1/\text{Var}(X)$; however, the lack of independence between s_x^2 and $\sum s_{y_j}^2$ in the calculation of Cronbach’s alpha makes precise determination of the bias for alpha difficult. When the bias of the ratio is positive, then the act of subtracting the ratio from one in the calculation of Cronbach’s alpha flips the bias of the estimator as a

whole to underestimation.

Also noticeable is that the bias of the KR21 and KR20 estimators is often larger than the bias of Cronbach's alpha for the gamma-inverse gamma and negative-binomial F models, identical for the Poisson-gamma model, and smaller for the binomial-beta model. Why? Though it is difficult to state exactly the reasons, as Equations (10) and (11) are slightly different than alpha due to the presence of the $k/(k + v_2)$ in front and because the two estimators are estimating the reliability using different quantities, a very strong clue is offered by Jensen's inequality. The numerators of KR20 and KR21 in Equations (10) and (11) are still estimating variances similarly to alpha, only in an indirect manner. From Equation (7), the mean-variance relationship for exponential family random variables is being exploited to estimate the variance through estimation of the variance function $V(\mu)$. Because $V(\mu)$ is not available, however, the plug-in estimate $V(\frac{1}{k}\bar{x})$ is used for the KR21 estimator. By Jensen's inequality, $E[V(\frac{1}{k}\bar{x})] \geq V(E[\frac{1}{k}\bar{x}]) = V(\mu)$ for convex functions $V(\cdot)$. In Table 1, the variance function $V(\mu)$ for both the exponential (gamma-inverse gamma model) and geometric (negative binomial-F model) distributions is convex. Hence, additional bias is introduced to the numerator of the ratio estimate, as the expected value of the variance function of the sample mean is larger than the variance function of the population mean. This is seen in the increased bias over Cronbach's alpha in Tables 3 and 4. A similar argument applies for the KR20 estimator. The bias for the Poisson-gamma model in Table 2, however, is equal for both Cronbach's alpha and the KR20 and KR21 equivalents. This is because the variance function for the Poisson distribution $V(\mu) = \mu$ is linear, and thus $E[V(\frac{1}{k}\bar{x})] = V(E[\frac{1}{k}\bar{x}]) = V(\mu)$. The bias for KR20 and KR21 in the binomial-beta model is smaller than the bias of Cronbach's alpha, as seen in Table 5, because the variance function $V(\mu) = \mu - \mu^2$ is concave, and so $E[V(\frac{1}{k}\bar{x})] \leq V(E[\frac{1}{k}\bar{x}]) = V(\mu)$ by Jensen's inequality. For most estimators, a downward bias would be a disadvantage. However, the bias of the standard binomial KR20 and KR21 appear in the numerator of a ratio estimator that is already biased upwards. By all accounts, it appears that the downward bias for the traditional KR20 and KR21 from the use of the variance function applied to the sample mean acts as a counterweight to the upward bias coming from taking the ratio of sample variances, which is a rather remarkable idea.

It has long been noted that in the standard binomial case, KR21 serves as a lower bound for KR20 (Kuder & Richardson, 1937). This is also easily seen as an application of Jensen's inequality in its algebraic form. Because $V(\theta) = \theta - \theta^2$ is concave, Jensen's inequality gives

$$kV\left(\frac{1}{k}\bar{x}\right) = kV\left(\frac{1}{k}\sum_{j=1}^k \bar{y}_j\right) \geq k\left(\frac{1}{k}\right)\sum_{j=1}^k V(\bar{y}_j) = \sum_{j=1}^k V(\bar{y}_j)$$

and once again, the act of subtracting from one in Equations (10) and (11) flips the inequality to obtain the standard result for the inequalities as a whole. Conversely, for convex variance functions, as is the case with the exponential (gamma-inverse gamma model) and geometric (negative binomial-F model) data, the inequality is reversed and KR20 serves as a lower bound to KR21. For Poisson data with identity variance function, the two estimators are equal, as previously discussed. Because KR20 and KR21 are both generally underestimating the reliability, this has the effect that KR20 is less biased than KR21 for

the binomial-beta model, with a smaller variance, mirroring the traditional wisdom that KR20 is superior to KR21 for dichotomous data. KR20 is more biased than KR21 for the gamma-inverse gamma and negative binomial-F models, with a larger variance.

Lastly, even though the bias of the KR20 and KR21 estimators is sometimes larger, it compensates for this by a greatly decreased variance. What is the reason for this decrease in variance? A possible explanation is that it comes from using sample means in combination with the mean-variance relationship of exponential families to estimate a variance rather than using sample variances themselves. Sample means converge to population means at a rate of \sqrt{n} , where n is the number of test subjects, while sample variances converge to population variances at a rate of n (Casella & Berger, 2002). Simply put, sample means converge faster than sample variances. Hence, estimating item variances using sample means in combination with a mean-variance relationship should tend to be more efficient than using sample variances alone.

4.3 Alternative Reliability Estimators

Cronbach’s alpha, though popular is far from the only estimator available, and its use has been criticized (McNeish, 2018; Sijtsma, 2009) and alternatively defended (Raykov & Marcoulides, 2019). One alternative based on analysis of a factorial structure is McDonald’s omega (McDonald, 1999), which has been proposed as preferable to Cronbach’s alpha (Hayes & Coutts, 2020; McNeish, 2018). Different estimators called omega exist, but two in particular, omega hierarchical (ω_H) which calculates the general factor saturation of a test and omega total (ω_T) which also includes specific factors, have been a focus of interest (Revelle & Zinbarg, 2009; Savalei & Reise, 2019). Another commonly proposed alternative is the greatest lower bound (Sijtsma, 2009; Trizano-Hermosilla & Alvarado, 2016), but Revelle and Zinbarg (2009) note that it is potentially smaller than ω_T , and its performance has been examined in Trizano-Hermosilla and Alvarado (2016). It is, however, worth comparing the generalized estimators of Table 1 to ω_H and ω_T to evaluate performance. Because the data is tau-equivalent, omega will equal reliability and the simulation study will evaluate performance as an estimator.

To do so, a smaller simulation study similar to that of Section 4.1 was performed. The Poisson-gamma (P-G), gamma - inverse gamma (G-IG), and binomial-beta (B-B) models were used to simulate data with a desired population reliability of $\rho = 0.8$, chosen to avoid possible numerical issues with low reliabilities. Test lengths of $k = 5, 10$ and number of subjects $n = 30, 75$ were chosen to present results with large root mean squared errors so that differences between the estimators are more easily observed. A total of $N_{sim} = 1,000,000$ data sets were generated for each model. The generalized KR21 estimator was chosen for comparison because it performed the best on the Poisson-gamma and gamma-inverse gamma models, as seen in Tables 2 and 3. Calculation of ω_H and ω_T was performed using the *psych* package in R (Revelle, 2020). Root mean squared error and both bias and variance of $KR21, \omega_H$, and ω_T are shown in Table 6.

The results of the study generally show that ω_H has the worst performance in terms of root mean squared error, including both elements of bias and variance, while the generalized KR21 formulas of Table 1 are competitive with ω_T for superior RMSE. In some scenarios one is larger, while in the others the second is larger. Noticeably, it occasionally

occurs that one estimator has a larger RMSE but a lower standard deviation, suggesting that bias correction could be a worthwhile strategy to improve the estimator. Also noticeable is that the bias of ω_T is generally small, but positive. One possible explanation for the small bias is that ω_T may be overfitting the model, which causes an upward bias in estimation of reliability that counters the general downward bias from ratios of sample statistics seen in other estimators. However, as Savalei and Reise (2019) note, further study regarding the properties of reliability coefficients such as ω_H and ω_T is needed to determine their statistical properties.

5 Example

A small example shows how and why these formulas can be applied in practice. The data in Table 7 are taken from Table 3 of Moore (1970) and represent a count of the number of times a specific letter was seen in a clerical speed test, for which the time has been subdivided into blocks of 10 units. Only the first 9 blocks of time are used in Table 7 in order to keep test length equal for all subjects. Chi-squared goodness of fit tests performed by Moore (1970) show that a Poisson process is a good fit to the data, and so the number of responses in each block of time may be treated as an independent and identical Poisson random variable with mean θ_i for subject i . The data may therefore also be treated as unidimensional and tau-equivalent. If the data were larger, either in terms of number of subjects or number of items, a regression of sample variance on sample means could indicate a linear relationship which would indicate that the Poisson model is appropriate. For small data, recognizing that the observation is a count should draw attention to the Poisson distribution for responses.

The reliability of this test represents the theoretical correlation of the sum scores x_i with sum scores on a parallel test. Ideally, this correlation will be high to indicate that the sum scores are accurately capturing each subject’s counting ability under time constraints. However, the data contains only $k = 9$ responses for each of $n = 10$ subjects, and the data is discrete and non-negative. Cronbach’s alpha for this data set is $\alpha = 0.079$, which is low. A problem is that the small number of responses per subject indicates that item variances are not being efficiently estimated. McDonald’s omega for this data set gives $\omega_H = 0.726$ and $\omega_T = 0.802$, but factor analysis on a small, exceedingly non-normal data set should call for caution with respect to the statistical properties, and the results of Table 6 indicate that this could be an overestimate.

The generalized KR20 and KR21 estimators of Table 1, originally derived in Allison (1978) and which are equal in this case, resolve this issue by utilizing the mean-variance relationship for Poisson data to obtain reliability. For Poisson data, the variance of a subject’s response for each item is equal to their mean ability. The generalized KR21 estimator therefore uses sample means to obtain an estimate of $\hat{\rho} = 0.626$. The simulation results of Table 2 indicate that this estimate is likely still biased downwards, but is a more efficient estimator than alternatives.

6 Conclusion, Suggestions, and Future Research Directions

This paper has introduced formulas which may be seen as the equivalent of the classical KR20 and KR21 formulas of Kuder and Richardson (1937) for some non-dichotomous data, extending to data which may be modeled using an NEF-QVF distribution. Simulations show that when the model is satisfied, the KR20 and KR21 equivalent formulas appear to converge to the population reliability and are generally a more efficient estimator of reliability than Cronbach's alpha. Exponential families are defined by mean-variance relations, so these formulas may be considered when a quadratic relationship between subject mean and subject variance appears to be present, as in Equation (1), which can be checked with analysis of sample variances as a function of sample means, exactly in the way a generalized linear model is chosen. Certain types of data such as dichotomous data (binomial), count data (Poisson), or data measuring time between events (exponential) also naturally fit this framework, especially when each response may be assumed to have roughly the same difficulty.

It is important to state the assumptions made in the derivation of these estimators: that the response to each test item identically follows a natural exponential family conditionally independent given ability θ_i for subject i , implying that all item difficulties are equal for a subject, and that abilities themselves follow the corresponding conjugate prior distribution. Not all of these assumptions are made in the derivations of the traditional binomial KR20 and KR21 in Kuder and Richardson (1937) and the Poisson KR21 in Allison (1978), suggesting that some assumptions may be superfluous, particularly the assumption that abilities follow the conjugate prior distribution. L. J. Feldt (1984) also derives the traditional binomial KR20 and KR21 formulas simply from the binomial error model. It is likely that these generalized KR20 and KR21 formulas may similarly be derived using a general error model from the mean-variance relationship of natural exponential families, and possibly for exponential families without quadratic variance functions. Furthermore, the difference in use between KR20 and KR21 in Kuder and Richardson (1937) is taken to be whether or not all items have equal difficulties. Zimmerman (1972) showed that the traditional binomial KR21 does not equal reliability when item difficulties are unequal, and it seems incredibly likely that some similar property may hold for the generalized KR20 and KR21 estimators of Table 1. How these estimators perform with items of varying difficulty, and whether there exist theoretical guarantees of performance, remains unknown excepting the case of the traditional binomial KR20 and KR21 formulas.

As stated in Section 2, the data in this framework are implied to be unidimensional and tau equivalent. If either of these assumptions are incorrect, none of KR20, KR21, or alpha are appropriate, and a reliability estimator based on analysis of a factorial structure, such as ω_T of McDonald (1999), is more appropriate. Unidimensionality should be tested before reliability is calculated using these formulas. When assumptions are appropriate, Section 4.3 shows that the KR20 and KR21 values are competitive with ω_T in terms of root mean squared error.

The advantage of these formulas is that when the data can be assumed to come from the correct model, use of distribution-specific formulas will generally provide a su-

perior estimate of the test reliability as measured by mean-squared error as compared to Cronbach’s alpha. This reduction occurs primarily through a reduction of the variance of the estimator, though for some estimators with an increase in bias. In general, for negative binomial and gamma data, the KR21 estimator outperforms the KR20 estimator and has less bias, though both generally outperform Cronbach’s alpha. For Poisson data, the two estimators are equal, and both outperform Cronbach’s alpha. For binomial data, simulations agree with conventional wisdom that the traditional KR20 estimator is superior to KR21. This is verified through simulations in Section 4. Notably, it is only in the case of dichotomous data, assumed to come from a binomial distribution, that KR20 outperforms KR21. Though it is not fully clear the reasons for this, Jensen’s inequality provides a strong clue.

Cronbach’s alpha can be used when no assumptions are made regarding the distribution of responses. Simulations show that while it is not the most efficient estimator of reliability, it is a consistent estimator that performs well over all the exponential families tested. The use of distribution-specific formulas such as KR20 and KR21 can then be seen as a riskier method of estimating reliability – when the parametric assumption is correct, the distribution-specific formula provides improved performance; however, if an incorrect distribution is assumed, performance may be disastrously worse. This raises the obvious question of how “wrong” a parametric assumption must be in order for an estimate based on it to be worse than Cronbach’s alpha, which could potentially be the subject of a future simulation study.

Bias is present in both Cronbach’s alpha and the equivalent KR20 and KR21 estimators. Noticeably, the bias is universally negative for alpha and almost universally negative for both KR20 and KR21, excepting only a few instances of positive bias for KR20 in the binomial-beta model. As discussed in Section 4, Jensen’s inequality may help to explain the direction of the bias, though further study is needed. As noted in Section 4.3, bias correction of the estimator could lead to improved performance. In general, a strategy to fix bias in an estimate of reliability could be to perform a bootstrap, parametric or otherwise, to obtain an estimate of the bias and then bias correct the original estimate of reliability. The accuracy of this procedure would depend on whether the bias the same at both the true and estimated values of reliability, which is not yet known; however, even an inaccurate bootstrap bias corrected estimate may prove to be more useful than an uncorrected estimate. As the RMSE of KR20 and KR21 for the gamma-inverse gamma and negative-binomial F models is smaller than the RMSE of alpha despite the increase in bias, a bootstrap bias corrected version of the generalized KR20 and KR21 estimators for these models might prove incredibly useful.

Lastly, Section 4.3 showed that ω_T remains a strong estimator of reliability in the presence of non-normality, sometimes superior to the KR21 estimates, but other times inferior. The exact mechanism for this remains unknown, and the framework of Section 2 could potentially be used to analyze statistical properties of estimators of factorial structures in the presence of non-normality. It could also potentially be possible to derive estimators of the factorial structure which utilize the mean-variance relationship of exponential families with improved performance when the distribution of responses is known to take a certain form.

These formulas have been shown to be strong estimators of reliability in the scenario

of perfect match to the assumed exponential model, and represent a small step towards the derivation of a more complete theory for properties of reliability estimators in the presence of non-normality. However, questions still remain about their use in the presence of items of unequal difficulty, lack of unidimensionality, and what their properties are both when the data does and does not perfectly fit the assumed exponential model. Further research is needed, and these formulas offer many fruitful avenues for such research.

Y_{ij} Distribution	$V(\theta)$	KR20	KR21
Normal	σ^2	$\left(1 - \frac{\sum_{j=1}^k \sigma^2}{s_x^2}\right)$	$\left(1 - \frac{k\sigma^2}{s_x^2}\right)$
Bernoulli	$\theta - \theta^2$	$\frac{k}{k-1} \left(1 - \frac{\sum_{j=1}^k \bar{y}_j(1 - \bar{y}_j)}{s_x^2}\right)$	$\frac{k}{k-1} \left(1 - \frac{k(\frac{1}{k}\bar{x})(1 - \frac{1}{k}\bar{x})}{s_x^2}\right)$
Poisson	θ	$\left(1 - \frac{\sum_{j=1}^k \bar{y}_j}{s_x^2}\right)$	$\left(1 - \frac{\bar{x}}{s_x^2}\right)$
Exponential	θ^2	$\frac{k}{k+1} \left(1 - \frac{\sum_{j=1}^k \bar{y}_j^2}{s_x^2}\right)$	$\frac{k}{k+1} \left(1 - \frac{\frac{1}{k}\bar{x}^2}{s_x^2}\right)$
Geometric	$\theta + \theta^2$	$\frac{k}{k+1} \left(1 - \frac{\sum_{j=1}^k (\bar{y}_j + \bar{y}_j^2)}{s_x^2}\right)$	$\frac{k}{k+1} \left(1 - \frac{\bar{x} + \frac{1}{k}\bar{x}^2}{s_x^2}\right)$
GHS	$1 + \theta^2$	$\frac{k}{k+1} \left(1 - \frac{\sum_{j=1}^k (1 + \bar{y}_j^2)}{s_x^2}\right)$	$\frac{k}{k+1} \left(1 - \frac{k(1 + \frac{1}{k^2}\bar{x}^2)}{s_x^2}\right)$

Table 1: Natural exponential family distributions with quadratic variance functions and their corresponding KR20 and KR21 estimators, given by Equations (10) and (11), respectively. The sum scores x_i for subject i have sample mean \bar{x} and sample variance s_x^2 , where sample mean and variance are calculated over subjects. Each test item has response y_{ij} , with i indexing subject and j indexing item, and where \bar{y}_j is the mean response for item j , again averaging over subjects. The test length is k items and the number of subjects is n . Note that the use of the normal density assumes that the noise variance σ^2 around each test item is known, which is unlikely to be the case in practice. Cronbach's alpha is recommended when the data are believed to be normal. Where possible, terms are cancelled to simplify formulas, except in the case of the binomial so as to preserve KR20 and KR21 in their original forms. Also note that because $\sum_{j=1}^k \bar{y}_j = \bar{x}$ as shown in Equation (9), the Poisson estimator is identical for both KR20 and KR21. The binomial estimators were first derived in Kuder and Richardson (1937) and the Poisson estimator was first derived in Allison (1978).

Poisson-Gamma

Parameters					Cronbach's Alpha			KR20			KR21		
μ	ρ	k	n	M	RMSE	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD
1	0.30	5	30	11.67	0.247	-0.057	0.241	0.224	-0.057	0.216	0.224	-0.057	0.216
1	0.30	5	75	11.67	0.142	-0.022	0.140	0.128	-0.022	0.126	0.128	-0.022	0.126
1	0.30	5	500	11.67	0.052	-0.003	0.052	0.047	-0.003	0.047	0.047	-0.003	0.047
1	0.30	10	30	23.33	0.232	-0.054	0.225	0.221	-0.054	0.214	0.221	-0.054	0.214
1	0.30	10	75	23.33	0.132	-0.021	0.130	0.126	-0.021	0.124	0.126	-0.021	0.124
1	0.30	10	500	23.33	0.048	-0.003	0.048	0.046	-0.003	0.046	0.046	-0.003	0.046
1	0.30	30	30	70.00	0.223	-0.053	0.217	0.220	-0.053	0.213	0.220	-0.053	0.213
1	0.30	30	75	70.00	0.126	-0.020	0.124	0.124	-0.020	0.122	0.124	-0.020	0.122
1	0.30	30	500	70.00	0.046	-0.003	0.046	0.045	-0.003	0.045	0.045	-0.003	0.045
1	0.60	5	30	3.33	0.156	-0.041	0.151	0.141	-0.041	0.135	0.141	-0.041	0.135
1	0.60	5	75	3.33	0.090	-0.017	0.089	0.081	-0.016	0.080	0.081	-0.016	0.080
1	0.60	5	500	3.33	0.033	-0.003	0.033	0.030	-0.003	0.030	0.030	-0.003	0.030
1	0.60	10	30	6.67	0.139	-0.036	0.135	0.133	-0.036	0.128	0.133	-0.036	0.128
1	0.60	10	75	6.67	0.080	-0.014	0.079	0.076	-0.014	0.075	0.076	-0.014	0.075
1	0.60	10	500	6.67	0.029	-0.002	0.029	0.028	-0.002	0.028	0.028	-0.002	0.028
1	0.60	30	30	20.00	0.130	-0.032	0.126	0.128	-0.032	0.124	0.128	-0.032	0.124
1	0.60	30	75	20.00	0.074	-0.012	0.073	0.072	-0.012	0.071	0.072	-0.012	0.071
1	0.60	30	500	20.00	0.027	-0.002	0.027	0.026	-0.002	0.026	0.026	-0.002	0.026
1	0.80	5	30	1.25	0.096	-0.031	0.091	0.087	-0.031	0.081	0.087	-0.031	0.081
1	0.80	5	75	1.25	0.055	-0.013	0.054	0.050	-0.013	0.048	0.050	-0.013	0.048
1	0.80	5	500	1.25	0.020	-0.002	0.020	0.018	-0.002	0.018	0.018	-0.002	0.018
1	0.80	10	30	2.50	0.079	-0.023	0.075	0.075	-0.023	0.071	0.075	-0.023	0.071
1	0.80	10	75	2.50	0.045	-0.009	0.044	0.043	-0.009	0.042	0.043	-0.009	0.042
1	0.80	10	500	2.50	0.017	-0.001	0.017	0.016	-0.001	0.016	0.016	-0.001	0.016
1	0.80	30	30	7.50	0.068	-0.018	0.065	0.067	-0.018	0.064	0.067	-0.018	0.064
1	0.80	30	75	7.50	0.039	-0.007	0.038	0.038	-0.007	0.037	0.038	-0.007	0.037
1	0.80	30	500	7.50	0.014	-0.001	0.014	0.014	-0.001	0.014	0.014	-0.001	0.014

Table 2: Simulation results for Cronbach's alpha and the KR20 and KR21 equivalents for the Poisson-gamma model, where abilities θ_i follow a gamma distribution and sum test scores x_i follow a Poisson distribution. The population reliability is ρ , the test length is k , the number of subjects is n , and the parameters of the gamma distribution for abilities are μ and M . The exact parameterizations of μ and M are given in the appendix of Foster (2020). The number in the first column beneath the estimator is the sample root mean squared error, RMSE. The numbers in the second and third columns are the sample bias and sample standard deviation (SD) of the estimator. The relationship between the three numbers is given by the formula $RMSE = \sqrt{Bias^2 + SD^2}$ (the formula may not be exact for results in the table due to rounding). For example, for $\hat{\alpha}$ in the first row of the table, $0.247 = \sqrt{(-0.057)^2 + 0.241^2}$. One million simulated data sets are used for each set of parameters. The results are identical for KR20 and KR21 for the Poisson-gamma model because the two estimators are equal. The generalized KR21 for Poisson count data was first derived in Allison (1978).

Gamma-Inverse Gamma

Parameters					Cronbach's Alpha			KR20			KR21		
μ	ρ	k	n	M	RMSE	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD
1	0.30	5	30	11.67	0.262	-0.075	0.251	0.233	-0.098	0.211	0.221	-0.081	0.206
1	0.30	5	75	11.67	0.157	-0.033	0.154	0.134	-0.041	0.127	0.131	-0.035	0.126
1	0.30	5	500	11.67	0.061	-0.005	0.061	0.051	-0.007	0.050	0.051	-0.006	0.050
1	0.30	10	30	23.33	0.239	-0.063	0.230	0.233	-0.087	0.216	0.220	-0.066	0.210
1	0.30	10	75	23.33	0.140	-0.026	0.137	0.131	-0.035	0.126	0.127	-0.027	0.124
1	0.30	10	500	23.33	0.052	-0.004	0.052	0.048	-0.005	0.047	0.047	-0.004	0.047
1	0.30	30	30	70.00	0.226	-0.055	0.219	0.233	-0.080	0.219	0.220	-0.057	0.212
1	0.30	30	75	70.00	0.128	-0.021	0.127	0.128	-0.031	0.124	0.124	-0.022	0.123
1	0.30	30	500	70.00	0.047	-0.003	0.047	0.046	-0.004	0.046	0.046	-0.003	0.046
1	0.60	5	30	3.33	0.212	-0.099	0.187	0.191	-0.115	0.152	0.181	-0.104	0.148
1	0.60	5	75	3.33	0.135	-0.054	0.124	0.115	-0.060	0.098	0.112	-0.056	0.097
1	0.60	5	500	3.33	0.062	-0.015	0.060	0.051	-0.016	0.048	0.051	-0.015	0.048
1	0.60	10	30	6.67	0.168	-0.063	0.156	0.164	-0.078	0.144	0.155	-0.066	0.140
1	0.60	10	75	6.67	0.101	-0.029	0.097	0.095	-0.035	0.089	0.093	-0.030	0.088
1	0.60	10	500	6.67	0.041	-0.005	0.041	0.038	-0.006	0.037	0.038	-0.006	0.037
1	0.60	30	30	20.00	0.139	-0.040	0.134	0.144	-0.055	0.133	0.136	-0.041	0.129
1	0.60	30	75	20.00	0.080	-0.016	0.079	0.080	-0.022	0.077	0.078	-0.016	0.076
1	0.60	30	500	20.00	0.030	-0.002	0.030	0.029	-0.003	0.029	0.029	-0.003	0.029
1	0.80	5	30	1.25	0.199	-0.137	0.144	0.183	-0.150	0.104	0.175	-0.142	0.102
1	0.80	5	75	1.25	0.139	-0.095	0.102	0.120	-0.100	0.067	0.118	-0.097	0.067
1	0.80	5	500	1.25	0.081	-0.051	0.062	0.063	-0.052	0.035	0.062	-0.052	0.035
1	0.80	10	30	2.50	0.126	-0.072	0.103	0.124	-0.083	0.092	0.117	-0.075	0.090
1	0.80	10	75	2.50	0.080	-0.042	0.068	0.075	-0.046	0.060	0.073	-0.043	0.059
1	0.80	10	500	2.50	0.038	-0.014	0.035	0.034	-0.015	0.031	0.034	-0.014	0.031
1	0.80	30	30	7.50	0.083	-0.032	0.077	0.086	-0.040	0.077	0.081	-0.033	0.074
1	0.80	30	75	7.50	0.049	-0.014	0.047	0.049	-0.017	0.046	0.048	-0.014	0.046
1	0.80	30	500	7.50	0.020	-0.002	0.020	0.019	-0.003	0.019	0.019	-0.003	0.019

Table 3: Simulation results for Cronbach's alpha and the KR20 and KR21 equivalents for the gamma-inverse gamma model, where abilities θ_i follow an inverse gamma distribution and sum test scores x_i follow a gamma distribution. The population reliability is ρ , the test length is k , the number of subjects is n , and the parameters of the inverse gamma distribution for abilities are μ and M . The exact parameterizations of μ and M are given in the appendix of Foster (2020). The number in the first column beneath the estimator is the sample root mean squared error, RMSE. The numbers in the second and third columns are the sample bias and sample standard deviation (SD) of the estimator. The relationship between the three numbers is given by the formula $RMSE = \sqrt{Bias^2 + SD^2}$ (the formula may not be exact for results in the table due to rounding). For example, for $\hat{\alpha}$ in the first row of the table, $0.262 = \sqrt{(-0.075)^2 + 0.251^2}$. One million simulated data sets are used for each set of parameters.

Negative Binomial - F

Parameters					Cronbach's Alpha			KR20			KR21		
μ	ρ	k	n	M	RMSE	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD
1.01	0.30	5	30	11.67	0.251	-0.067	0.242	0.220	-0.090	0.201	0.210	-0.073	0.197
1.01	0.30	5	75	11.67	0.151	-0.028	0.149	0.126	-0.037	0.120	0.123	-0.030	0.119
1.01	0.30	5	500	11.67	0.058	-0.005	0.058	0.047	-0.006	0.047	0.047	-0.005	0.047
1.01	0.30	10	30	23.33	0.235	-0.061	0.227	0.228	-0.084	0.212	0.215	-0.064	0.206
1.01	0.30	10	75	23.33	0.137	-0.024	0.135	0.127	-0.033	0.123	0.124	-0.025	0.122
1.01	0.30	10	500	23.33	0.051	-0.004	0.051	0.046	-0.005	0.046	0.046	-0.004	0.046
1.01	0.30	30	30	70.00	0.225	-0.055	0.218	0.232	-0.080	0.218	0.219	-0.057	0.212
1.01	0.30	30	75	70.00	0.128	-0.021	0.126	0.128	-0.031	0.124	0.124	-0.022	0.122
1.01	0.30	30	500	70.00	0.047	-0.003	0.047	0.046	-0.005	0.046	0.046	-0.003	0.045
1.01	0.60	5	30	3.33	0.164	-0.053	0.155	0.132	-0.068	0.113	0.125	-0.058	0.111
1.01	0.60	5	75	3.33	0.103	-0.024	0.100	0.076	-0.030	0.070	0.074	-0.026	0.069
1.01	0.60	5	500	3.33	0.042	-0.004	0.042	0.029	-0.005	0.029	0.029	-0.005	0.029
1.01	0.60	10	30	6.67	0.142	-0.043	0.136	0.135	-0.058	0.122	0.128	-0.046	0.119
1.01	0.60	10	75	6.67	0.085	-0.018	0.083	0.077	-0.024	0.073	0.075	-0.019	0.072
1.01	0.60	10	500	6.67	0.032	-0.003	0.032	0.028	-0.004	0.028	0.028	-0.003	0.028
1.01	0.60	30	30	20.00	0.132	-0.036	0.127	0.136	-0.050	0.126	0.128	-0.037	0.122
1.01	0.60	30	75	20.00	0.076	-0.014	0.074	0.075	-0.019	0.073	0.073	-0.014	0.072
1.01	0.60	30	500	20.00	0.028	-0.002	0.028	0.027	-0.003	0.027	0.027	-0.002	0.027
1.01	0.80	5	30	1.25	0.114	-0.032	0.109	0.064	-0.043	0.048	0.060	-0.038	0.046
1.01	0.80	5	75	1.25	0.080	-0.014	0.079	0.029	-0.018	0.022	0.027	-0.016	0.022
1.01	0.80	5	500	1.25	0.042	-0.003	0.042	0.009	-0.004	0.009	0.009	-0.003	0.009
1.01	0.80	10	30	2.50	0.075	-0.026	0.070	0.065	-0.035	0.055	0.061	-0.029	0.053
1.01	0.80	10	75	2.50	0.047	-0.012	0.045	0.037	-0.015	0.033	0.035	-0.013	0.033
1.01	0.80	10	500	2.50	0.019	-0.002	0.019	0.014	-0.003	0.014	0.014	-0.002	0.014
1.01	0.80	30	30	7.50	0.066	-0.020	0.063	0.068	-0.027	0.062	0.063	-0.021	0.060
1.01	0.80	30	75	7.50	0.038	-0.008	0.038	0.038	-0.011	0.036	0.037	-0.008	0.036
1.01	0.80	30	500	7.50	0.014	-0.001	0.014	0.014	-0.002	0.014	0.014	-0.001	0.014

Table 4: Simulation results for Cronbach's alpha and the KR20 and KR21 equivalents for the negative binomial - F model, where abilities θ_i follow an F distribution and sum test scores x_i follow a negative binomial distribution. The population reliability is equal to ρ , the test length is k , the number of subjects is n , and the parameters of the F distribution for abilities are μ and M . The exact parameterizations of μ and M are given in the appendix of Foster (2020). The number in the first column beneath the estimator is the sample root mean squared error, RMSE. The numbers in the second and third columns are the sample bias and sample standard deviation (SD) of the estimator. The relationship between the three numbers is given by the formula $RMSE = \sqrt{Bias^2 + SD^2}$ (the formula may not be exact for results in the table due to rounding). For example, for $\hat{\alpha}$ in the first row of the table, $0.251 = \sqrt{(-0.067)^2 + 0.242^2}$. One million simulated data sets are used for each set of parameters.

Binomial-Beta

Parameters					Cronbach's Alpha			KR20			KR21		
μ	ρ	k	n	M	RMSE	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD
0.50	0.30	5	30	11.67	0.240	-0.048	0.235	0.228	-0.015	0.228	0.238	-0.039	0.235
0.50	0.30	5	75	11.67	0.136	-0.018	0.135	0.133	-0.005	0.133	0.136	-0.015	0.135
0.50	0.30	5	500	11.67	0.050	-0.003	0.050	0.050	-0.001	0.050	0.050	-0.002	0.050
0.50	0.30	10	30	23.33	0.229	-0.050	0.223	0.217	-0.021	0.216	0.228	-0.046	0.223
0.50	0.30	10	75	23.33	0.129	-0.019	0.128	0.126	-0.008	0.126	0.129	-0.017	0.127
0.50	0.30	10	500	23.33	0.047	-0.003	0.047	0.047	-0.001	0.047	0.047	-0.002	0.047
0.50	0.30	30	30	70.00	0.221	-0.051	0.216	0.210	-0.025	0.208	0.221	-0.050	0.215
0.50	0.30	30	75	70.00	0.125	-0.019	0.124	0.122	-0.009	0.122	0.125	-0.019	0.124
0.50	0.30	30	500	70.00	0.045	-0.003	0.045	0.045	-0.001	0.045	0.045	-0.003	0.045
0.50	0.60	5	30	3.33	0.134	-0.023	0.132	0.127	-0.001	0.127	0.132	-0.015	0.131
0.50	0.60	5	75	3.33	0.077	-0.009	0.077	0.076	0.000	0.076	0.077	-0.005	0.077
0.50	0.60	5	500	3.33	0.028	-0.001	0.028	0.028	0.000	0.028	0.028	-0.001	0.028
0.50	0.60	10	30	6.67	0.127	-0.025	0.124	0.121	-0.007	0.120	0.126	-0.022	0.124
0.50	0.60	10	75	6.67	0.072	-0.009	0.072	0.071	-0.003	0.071	0.072	-0.008	0.072
0.50	0.60	10	500	6.67	0.026	-0.001	0.026	0.026	0.000	0.026	0.026	-0.001	0.026
0.50	0.60	30	30	20.00	0.125	-0.028	0.122	0.118	-0.012	0.118	0.125	-0.026	0.122
0.50	0.60	30	75	20.00	0.071	-0.010	0.070	0.069	-0.004	0.069	0.071	-0.010	0.070
0.50	0.60	30	500	20.00	0.026	-0.001	0.026	0.026	-0.001	0.026	0.026	-0.001	0.026
0.50	0.80	5	30	1.25	0.068	-0.009	0.067	0.065	0.006	0.065	0.067	-0.001	0.067
0.50	0.80	5	75	1.25	0.040	-0.003	0.040	0.039	0.003	0.039	0.040	0.000	0.040
0.50	0.80	5	500	1.25	0.015	0.000	0.015	0.015	0.000	0.015	0.015	0.000	0.015
0.50	0.80	10	30	2.50	0.061	-0.010	0.061	0.059	0.000	0.059	0.061	-0.007	0.060
0.50	0.80	10	75	2.50	0.036	-0.004	0.035	0.035	0.000	0.035	0.035	-0.002	0.035
0.50	0.80	10	500	2.50	0.013	-0.001	0.013	0.013	0.000	0.013	0.013	0.000	0.013
0.50	0.80	30	30	7.50	0.061	-0.012	0.059	0.057	-0.004	0.057	0.060	-0.011	0.059
0.50	0.80	30	75	7.50	0.034	-0.005	0.034	0.034	-0.001	0.034	0.034	-0.004	0.034
0.50	0.80	30	500	7.50	0.013	-0.001	0.013	0.013	0.000	0.013	0.013	-0.001	0.013

Table 5: Simulation results for Cronbach's alpha and KR20 and KR21 for the binomial - beta model, where abilities θ_i follow a beta distribution and sum test scores x_i follow a binomial distribution. The population reliability is equal to ρ , the test length is k , the number of subjects is n , and the parameters of the beta distribution for abilities are μ and M . The exact parameterizations of μ and M are given in the appendix of Foster (2020). The number in the first column beneath the estimator is the sample root mean squared error, RMSE. The numbers in the second and third columns are the sample bias and sample standard deviation (SD) of the estimator. The relationship between the three numbers is given by the formula $RMSE = \sqrt{Bias^2 + SD^2}$ (the formula may not be exact for results in the table due to rounding). For example, for $\hat{\alpha}$ in the first row of the table, $0.240 = \sqrt{(-0.048)^2 + 0.235^2}$. One million simulated data sets are used for each set of parameters. The KR20 and KR21 estimators for this model are the traditional ones derived in Kuder and Richardson (1937).

Parameters						ω_H			ω_T			KR21		
Model	μ	ρ	k	n	M	RMSE	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD
P-G	1	0.80	5	30	1.25	0.201	-0.153	0.130	0.090	0.061	0.066	0.087	-0.031	0.081
P-G	1	0.80	10	30	2.50	0.310	-0.281	0.131	0.077	0.050	0.059	0.075	-0.023	0.071
P-G	1	0.80	5	75	1.25	0.145	-0.114	0.088	0.060	0.041	0.045	0.050	-0.013	0.048
P-G	1	0.80	10	75	2.50	0.250	-0.230	0.096	0.053	0.039	0.036	0.043	-0.009	0.042
G-IG	1	0.80	5	30	1.25	0.309	-0.248	0.183	0.116	0.034	0.110	0.175	-0.142	0.102
G-IG	1	0.80	10	30	2.50	0.355	-0.317	0.159	0.090	0.035	0.083	0.118	-0.076	0.090
G-IG	1	0.80	5	75	1.25	0.270	-0.221	0.156	0.091	0.030	0.085	0.118	-0.097	0.066
G-IG	1	0.80	10	75	2.50	0.331	-0.306	0.124	0.068	0.035	0.059	0.073	-0.043	0.059
B-B	0.5	0.80	5	30	1.25	0.173	-0.135	0.108	0.084	0.066	0.052	0.067	-0.001	0.067
B-B	0.5	0.80	10	30	2.50	0.296	-0.272	0.118	0.073	0.055	0.048	0.061	-0.007	0.061
B-B	0.5	0.80	5	75	1.25	0.124	-0.101	0.071	0.054	0.041	0.035	0.040	0.000	0.040
B-B	0.5	0.80	10	75	2.50	0.231	-0.215	0.084	0.049	0.040	0.029	0.035	-0.002	0.035

Table 6: Simulation results for the KR21 equivalent for the Poisson-gamma (P-G), gamma - inverse gamma (I-G), and binomial-beta (B-B) models compared to omega hierarchical and omega total, ω_H and ω_T , as measured by root mean squared error. The omega reliability calculations are performed by the *psych* package in R (Revelle, 2020). The number in the first column beneath the estimator is the sample root mean squared error, RMSE. The numbers in the second and third columns are the sample bias and sample standard deviation (SD) of the estimator. The relationship between the three numbers is given by the formula $RMSE = \sqrt{Bias^2 + SD^2}$ (the formula may not be exact for results in the table due to rounding). For example, for ω_H in the first row of the table, $0.201 = \sqrt{(-0.153)^2 + 0.130^2}$. One million simulated data sets are used for each set of parameters.

Subject	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	x_i	$\frac{1}{k}x_i$	s^2
1	13	10	12	8	12	10	9	16	8	98	10.89	6.86
2	14	9	9	13	10	10	14	9	11	99	11.00	4.50
3	11	10	5	12	12	8	12	11	5	86	9.56	4.50
4	12	13	4	8	7	10	14	1	14	83	9.22	21.19
5	16	8	9	7	11	12	16	8	11	98	10.89	11.11
6	11	13	7	8	9	7	8	7	4	74	8.22	6.69
7	13	11	10	9	15	9	18	10	5	100	11.11	14.36
8	11	5	9	9	9	13	11	15	6	88	9.78	9.94
9	13	8	10	11	11	12	8	7	4	84	9.33	8.00
10	12	9	12	9	14	8	12	10	7	93	10.33	5.25

Table 7: Data from Moore (1970) showing the count of the number of responses for each of ten subjects in consecutive intervals of transformed time. This data may be treated as a Poisson process, where each response is a realization of a Poisson random variable with mean given by a subject's ability θ_i .

Difference in RMSE (Alpha - KR21) for the Poisson-Gamma Model for Varying Test Length, $\rho = 0.3$

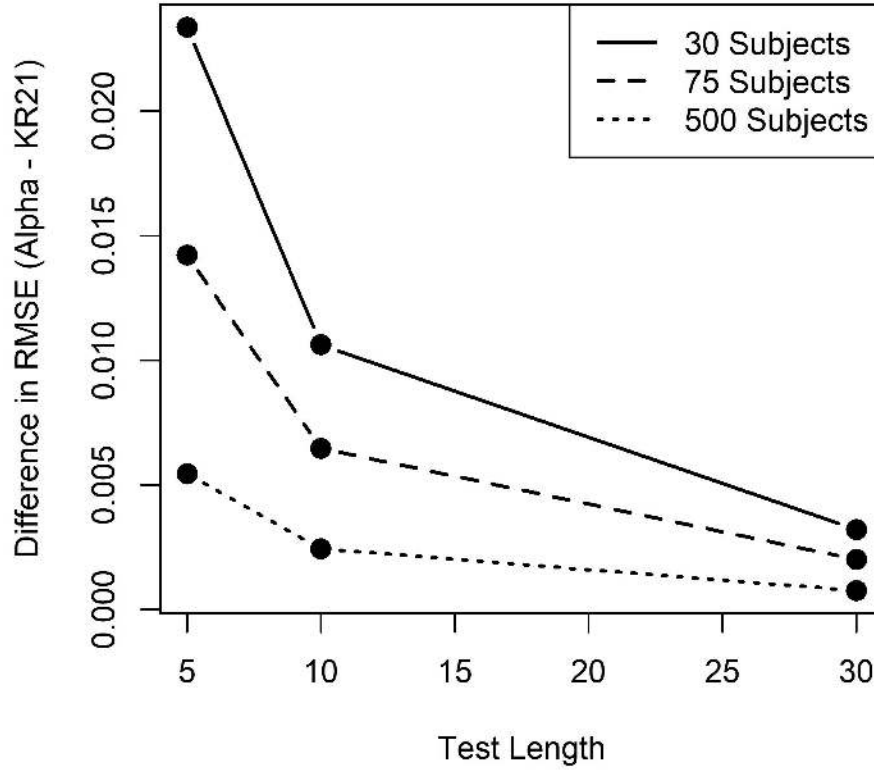


Figure 1: Difference in RMSE (Alpha - KR21) for tests of different lengths for the poisson-gamma model with $\rho = 0.3$, grouped by number of subjects. The plot clearly shows that $\hat{\alpha}$ has a larger RMSE than KR21, but this effect decreases towards zero as both the test length k and number of subjects n increases. The data in this plot are from Table 2.

Difference in RMSE (Alpha - KR21) for the Gamma-Inverse Gamma Model for Varying Test Length, $\rho = 0.3$

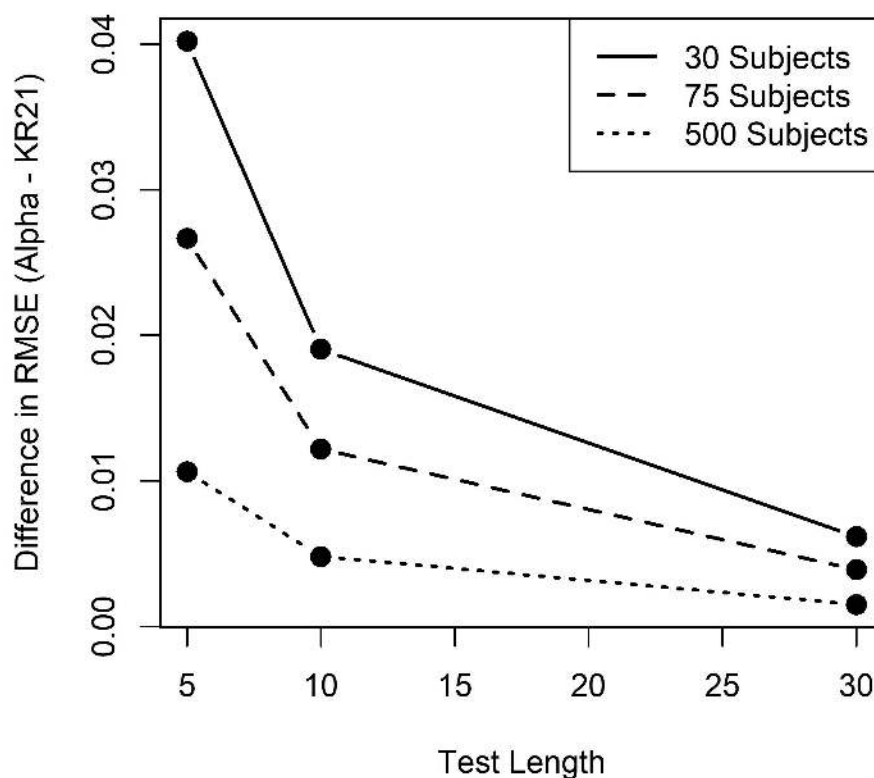


Figure 2: Difference in RMSE (Alpha - KR21) for tests of different lengths for the gamma-inverse gamma model with $\rho = 0.3$, grouped by number of subjects. The plot clearly shows that $\hat{\alpha}$ has a larger RMSE than KR21, but this effect decreases towards zero as both the test length k and number of subjects n increases. The data in this plot are from Table 3.

Difference in RMSE (Alpha - KR21) for the Neg. Binomial - F Model for Varying Test Length, $\rho = 0.3$

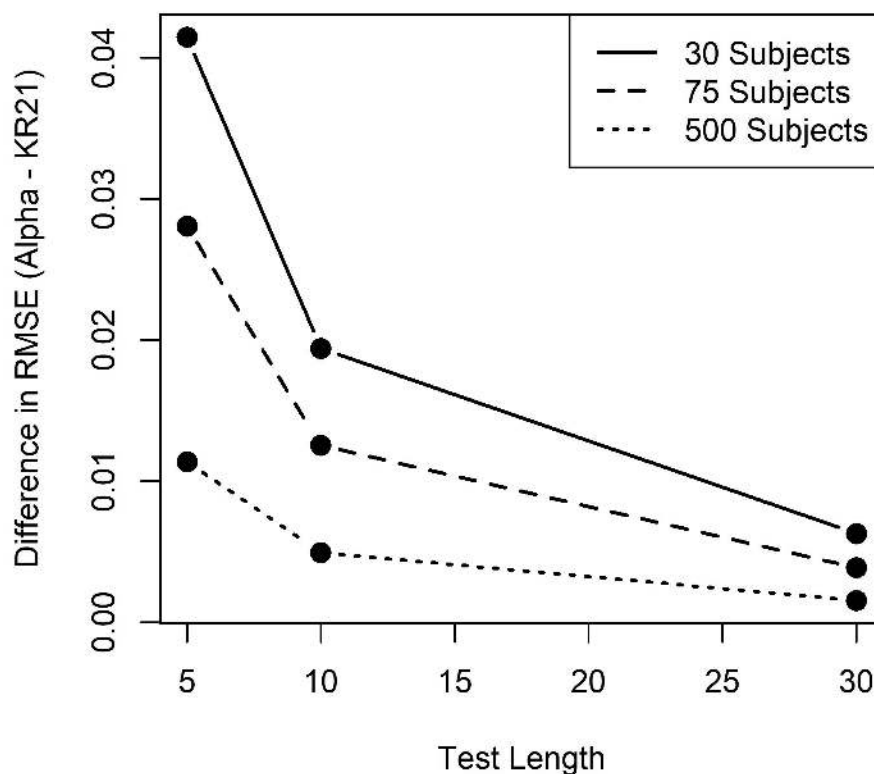


Figure 3: Difference in RMSE (Alpha - KR21) for tests of different lengths for the negative binomial-F model with $\rho = 0.3$, grouped by number of subjects. The plot clearly shows that $\hat{\alpha}$ has a larger RMSE than KR21, but this effect decreases towards zero as both the test length k and number of subjects n increases. The data in this plot are from Table 4.

Difference in RMSE (Alpha - KR21) for the Binomial-Beta Model for Varying Test Length, $\rho = 0.3$

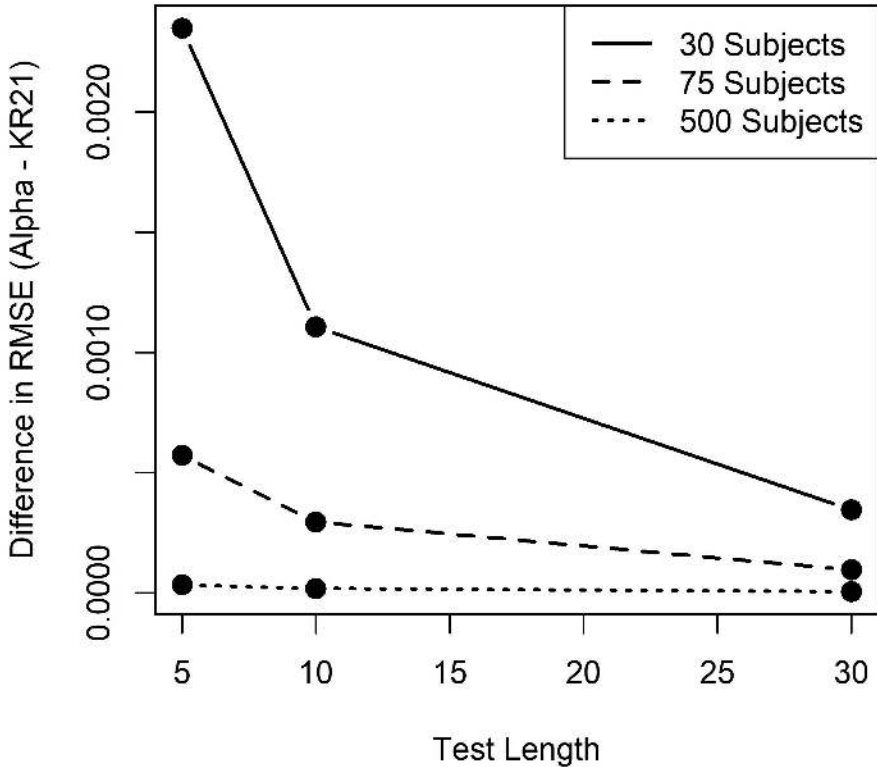


Figure 4: Difference in RMSE (Alpha - KR21) for tests of different lengths for the binomial-beta model with $\rho = 0.3$, grouped by number of subjects. The plot clearly shows that $\hat{\alpha}$ has a larger RMSE than KR21, but this effect decreases towards zero as both the test length k and number of subjects n increases. The data in this plot are from Table 5.

References

- Allison, P. D. (1978). The reliability of variables measured as the number of events in an interval of time. *Sociological Methodology*, 9, 238 –253.
- Bay, K. S. (1973). The effect of non-normality on the sampling distribution of standard error of reliability coefficient estimates under an analysis of variance model. *British Journal of Mathematical and Statistical Psychology*, 26(1), 45-57.
- Casella, G., & Berger, R. (2002). *Statistical inference, second edition*. Duxbury Press.
- Cronbach, L. J. (1951, September). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297 – 334.
- Feldt, L. J. (1984). Some relationships between the binomial error model and classical test theory. *Educational and Psychological Measurement*, 44(4), 883 – 891.
- Feldt, L. S. (1965, September). The approximate sampling distribution of kuder-richardson reliability coefficient twenty. *Psychometrika*, 30(3), 357 – 370.
- Foster, R. C. (2020, June). A generalized framework for classical test theory. *The Journal of Mathematical Psychology*, 96.
- Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, 19(1), 72-91.
- Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than cronbach’s alpha for estimating reliability. but... *Communication Methods and Measures*, 14(1), 1-24.
- Huynh, H. (1979). Statistical inference for two reliability indices in mastery testing based on the beta-binomial model. *Journal of Educational Statistics*, 4(3), 231-246.
- Keats, J. A., & Lord, F. M. (1962, March). A theoretical distribution for mental test scores. *Psychometrika*, 27(1), 59 – 72.
- Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2(3), 151 – 160.
- Lord, F. M. (1965, September). A strong true-score theory, with applications. *Psychometrika*, 30(3), 239 – 270.
- Lord, F. M., Novick, M. R., & Birnbaum, A. (1968). *Statistical theories of mental test scores*. Oxford, England: Addison-Wesley.
- McDonald, R. (1999). *Test theory: A unified treatment*. Taylor & Francis.
- McNeish, D. (2018). Thanks coefficient alpha, we’ll take it from here. *Psychological Methods*, 23(3), 412-433.
- Meredith, W. (1971, 08). Poisson distributions of error in mental test theory. *British Journal of Mathematical and Statistical Psychology*, 24, 49 - 82.
- Moore, W. E. (1970, November). *Stochastic processes as true-score models for highly speeded mental tests* (Tech. Rep.). Princeton, New Jersey: Educational Testing

- Service.
- Morris, C. N. (1982). Natural exponential families with quadratic variance functions. *The Annals of Statistics*, *10*(1), 65 – 80.
- Morris, C. N. (1983). Natural exponential families with quadratic variance functions: Statistical theory. *The Annals of Statistics*, *11*(2), 515 – 529.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danmarks Paedagogiske Institut.
- Raykov, T. (1998). A method for obtaining standard errors and confidence intervals of composite reliability for congeneric items. *Applied Psychological Measurement*, *22*(4), 369-374.
- Raykov, T., & Marcoulides, G. A. (2019). Thanks coefficient alpha, we still need you! *Educational and Psychological Measurement*, *79*(1), 200-210.
- Revelle, W. (2020). psych: Procedures for psychological, psychometric, and personality research [Computer software manual]. Evanston, Illinois. Retrieved from <https://CRAN.R-project.org/package=psych> (R package version 2.0.12)
- Revelle, W., & Zinbarg, R. (2009, 03). Coefficients alpha, beta, omega, and the glb: Comments on sijtsma. *Psychometrika*, *74*, 145-154.
- Savalei, V., & Reise, S. P. (2019). Don't forget the model in your model-based reliability coefficients: A reply to mcneish (2018). *Collabra:Psychology*, *5*(1).
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*, 350-353.
- Sheng, Y., & Sheng, Z. (2012, February). Is coefficient alpha robust to non-normal data? *Frontiers in Psychology*.
- Sijtsma, K. (2009, 2009). On the use, the misuse, and the very limited usefulness of cronbach's alpha. *Psychometrika*, *74*(1), 107 – 120.
- Trizano-Hermosilla, I., & Alvarado, J. (2016, 05). Best alternatives to cronbach's alpha reliability in realistic conditions: Congeneric and asymmetrical measurements. *Frontiers in Psychology*, *7*.
- Zimmerman, D. W. (1972). Test reliability and the kuder-richardson formulas: Derivation from probability theory. *Educational and Psychological Measurement*, *32*(4), 939 – 954.
- Zimmerman, D. W., Zumbo, B. D., & Lalonde, C. (1993). Coefficient alpha as an estimate of test reliability under violation of two assumptions. *Educational and Psychological Measurement*, *53*(1), 33-49.
- Zinbarg, R., Yovel, I., Revelle, W., & McDonald, R. (2006, 03). Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for ω_h . *Applied Psychological Measurement*, *30*, 121-144.
- Zumbo, B. (1999). *A glance at coefficient alpha with an eye towards robustness studies : Some mathematical notes and a simulation model*. (Paper No. ESQBS-99-1). Prince George, B.C.: University of Northern British Columbia. Edgeworth Laboratory for Quantitative Behavioural Science.
- Zyl, J., Neudecker, H., & Nel, D. (2000, 09). On the distribution of the maximum

likelihood estimator of cronbach's alpha. *Psychometrika*, 65, 271-280.

Appendix

The elements of the variance-covariance matrices used in the simulations of Section 4 are easily obtained through applying the formulas of Foster (2020). For an individual test item y_j or pair of test items $y_{j,1}, y_{j,2}$, the unconditional variance and covariance in this framework are

$$\begin{aligned} \text{Var}(y_j) &= E[V(\theta_i)] + \text{Var}(\theta_i) = \text{Var}(\theta_i)(M + 1) = \left(\frac{V(\mu)}{M - v_2} \right) (M + 1) \\ \text{Cov}(y_{j,1}, y_{j,2}) &= \text{Var}(\theta_i) = \frac{V(\mu)}{M - v_2} \end{aligned}$$

These equalities hold for all natural exponential family distributions except for the last line, which depends on the mean-variance relationship given in Equation (6) and holds only for those with quadratic variance function, as discussed in this paper. The function $V(\mu)$ is the variance function of the natural exponential family applied to the underlying population mean, several of which are shown in Table 1, and v_2 is the coefficient of the quadratic term of the quadratic variance function, as shown in Equation (1).

If the parameter M is defined by the desired population value of Cronbach's alpha ρ and test length k by the relation $M = [(1 - \rho)/\rho]k$ as in Section 4, these formulas reduce to

$$\begin{aligned} \text{Var}(y_j) &= V(\mu) \left[\frac{(1 - \rho)k + \rho}{(1 - \rho)k - \rho v_2} \right] \\ \text{Cov}(y_{j,1}, y_{j,2}) &= V(\mu) \left[\frac{\rho}{(1 - \rho)k - \rho v_2} \right] \end{aligned}$$

Within the variance-covariance matrix, all test items have identical variances and all covariances between them are equal. The variances and covariance for the simulations in Section 4 are as follows.

Poisson-Gamma

For the Poisson-gamma model, the variance function is $V(\theta) = \theta$ with $v_2 = 0$. This yields variance and covariance elements as

$$\begin{aligned} \text{Var}(y_j) &= \frac{\mu}{M}(M + 1) \\ \text{Cov}(y_{j,1}, y_{j,2}) &= \frac{\mu}{M} \end{aligned}$$

ρ	k	μ	M	$V(y_j)$	$Cov(y_{j,1}, y_{j,2})$
0.3	5	1	11.67	1.0857	0.0857
0.3	10	1	23.33	1.0429	0.0429
0.3	30	1	70	1.0143	0.0143
0.6	5	1	3.33	1.3	0.3
0.6	10	1	6.67	1.15	0.15
0.6	30	1	20	1.05	0.05
0.8	5	1	1.25	1.8	0.8
0.8	10	1	2.5	1.4	0.4
0.8	30	1	7.5	1.1333	0.1333

Gamma-Inverse Gamma

For the gamma-inverse gamma model, the variance function is $V(\theta) = \theta^2$ with $v_2 = 1$. This yields variance and covariance elements as

$$Var(y_j) = \left(\frac{\mu^2}{M-1} \right) (M+1)$$

$$Cov(y_{j,1}, y_{j,2}) = \frac{\mu^2}{M-1}$$

ρ	k	μ	M	$V(y_j)$	$Cov(y_{j,1}, y_{j,2})$
0.3	5	1	11.67	1.0857	0.0857
0.3	10	1	23.33	1.0429	0.0429
0.3	30	1	70	1.0143	0.0143
0.6	5	1	3.33	1.3	0.3
0.6	10	1	6.67	1.15	0.15
0.6	30	1	20	1.05	0.05
0.8	5	1	1.25	1.8	0.8
0.8	10	1	2.5	1.4	0.4
0.8	30	1	7.5	1.1333	0.1333

Note that for small values of M the resulting distributions of y_j become heavy tailed, and so sample variances and covariances may not match theoretical values due to outliers.

Negative Binomial-F

For the negative binomial-F model, the variance function is $V(\theta) = \theta + \theta^2$ with $v_2 = 1$. This yields variance and covariance elements as

$$Var(y_j) = \left(\frac{\mu + \mu^2}{M - 1} \right) (M + 1)$$

$$Cov(y_{j,1}, y_{j,2}) = \frac{\mu + \mu^2}{M - 1}$$

ρ	k	μ	M	$V(y_j)$	$Cov(y_{j,1}, y_{j,2})$
0.3	5	1.01	11.67	2.4107	0.1903
0.3	10	1.01	23.33	2.2119	0.0909
0.3	30	1.01	70	2.0889	0.0294
0.6	5	1.01	3.33	3.7702	0.87
0.6	10	1.01	6.67	2.7466	0.3583
0.6	30	1.01	20	2.2438	0.1068
0.8	5	1.01	1.25	18.2709	8.1204
0.8	10	1.01	2.5	4.7369	1.3534
0.8	30	1.01	7.5	2.6547	0.3123

Binomial-Beta

For the binomial-beta model, the variance function is $V(\theta) = \theta(1 - \theta)$ with $v_2 = -1$. This yields variance and covariance elements as

$$Var(y_j) = \mu(1 - \mu)$$

$$Cov(y_{j,1}, y_{j,2}) = \frac{\mu(1 - \mu)}{M + 1}$$

ρ	k	μ	M	$V(y_j)$	$Cov(y_{j,1}, y_{j,2})$
0.3	5	0.5	11.67	0.25	0.0197
0.3	10	0.5	23.33	0.25	0.0103
0.3	30	0.5	70	0.25	0.0035
0.6	5	0.5	3.33	0.25	0.0577
0.6	10	0.5	6.67	0.25	0.0326
0.6	30	0.5	20	0.25	0.0119
0.8	5	0.5	1.25	0.25	0.1111
0.8	10	0.5	2.5	0.25	0.0714
0.8	30	0.5	7.5	0.25	0.0294