

RASCH MEASUREMENT

Transactions of the Rasch Measurement SIG
American Educational Research Association

Vol. 28 No. 4

Spring 2015

ISSN 1051-0796

Testing Unidimensionality Using the PCA/t-test Protocol with the Rasch Model: A Cautionary Note

One approach that has gained popularity for testing unidimensionality within the Rasch measurement framework is the Principal Component Analysis (PCA) and t-test based method first proposed by Smith (Smith, 2002). This procedure first identifies two item sets potentially representing different dimensions from a PCA of residuals that are used to estimate two separate sets of person measures. A series of t-tests is then conducted to compare the two estimates on a person-by-person basis to determine the proportion of instances where the two item sets yield different person measures. It has been suggested that unidimensionality can be inferred if $\leq 5\%$ of the t-tests are significant or if the lower bound of a binomial 95% confidence interval (CI) of the observed proportion overlaps 5% (Horton & Tennant, 2010; Smith, 2002; Tennant & Conaghan, 2007; Tennant & Pallant, 2006). Simulation studies have suggested that this protocol performs well as a unidimensionality test in comparison to traditional fit analysis, as well as raw score or residual based PCA (Horton & Tennant, 2010; Tennant & Pallant, 2006). The implementation of the procedure in popular Rasch analysis software (Andrich, Sheridan, & Luo, 1997-2012) and its suggested function as a test of strict unidimensionality (Tennant & Conaghan, 2007), has rendered the procedure increasingly popular and it is often interpreted as “definite” evidence for or against unidimensionality (Forjaz et al., 2013; Ramp, Khan, Misajon, & Pallant, 2009; Riazi, Aspden, & Jones, 2014; Young, Mills, Woolmore, Hawkins, & Tennant, 2012).

A central aspect of the PCA/t-test protocol is the binomial 95% CI, which is the basis for deciding whether scales are unidimensional or not. However, there is a number of procedures available for estimating the 95% binomial CI (Brown, Cai, & DasGupta, 2001; Newcombe, 1998), and sample size impacts the CI width and hence interpretation of results (Feinstein, 1998; McCormack, Vandermeer, & Allan, 2013). These aspects were explored in a recent paper (Hagell, 2014) addressing the impact of sample size and 95% binomial CI estimation method on the resulting

conclusions according to published heuristics (Horton & Tennant, 2010; Tennant & Conaghan, 2007; Tennant & Pallant, 2006).

Binomial 95% CIs were calculated according to the normal approximation 95% CI (“Wald” method), the “exact” binomial CI, and the Wilson, Agresti-Coull, and Jeffreys methods for hypothesized observed proportions of 6%, 8% and 10% and sample sizes ranging from $n=100$ to $n=2500$. Results for the normal approximation, and the Wilson and Agresti-Coull 95% CI estimations are shown in Figure 1 (for complete results, see (Hagell, 2014)). It can be seen that normal approximation 95% CIs included 5% with sample sizes of $n=100-2000$ and a 6% observed proportion, $n=100-300$ with an 8% observed proportion, and $n=100$ with a 10% observed proportion. The Wilson and Agresti-Coull CIs all included 5% with sample sizes of $n=100-1500$ and a 6% observed proportion as well as with sample sizes of $n=100-200$ with an 8% observed proportion, but not for any sample size with a 10% observed proportion.

These results are fully expected (Brown et al., 2001; Feinstein, 1998; McCormack et al., 2013; Newcombe, 1998), although aspects do not appear to be commonly acknowledged when applying the procedure. For example, Ramp et al. (Ramp et al., 2009) used the PCA/t-test protocol to test the unidimensionality of the 20-item physical impact scale of the Multiple Sclerosis Impact Scale with a sample of 92 people, and found that 9.2% of

Table of Contents

Testing Unidimensionality Using the PCA/t-test Protocol with the Rasch Model (Hagell).....	1487
Unfolding Rater Accuracy in Performance Assessments (Engelhard & Wang).....	1489
Investigation and Application of the Person Aberrant Detection Indices (Chien & Djaja).....	1491
Rasch as a Basis for Metrologically Traceable Standards (Fisher).....	1492
Message from Rasch SIG Chair (Tognolini).....	1494

the person measures from two item subsets differed and the lower 95% binomial CI bound was 4%, leading the authors to infer unidimensionality. Young et al. (Young et al., 2012) used the protocol with a 17-item self-efficacy scale among 309 people with multiple sclerosis and found that 12.2% of the person measures differed (lower 95% binomial CI bound, 9.8%), interpreted as “considerable multidimensionality” (p 1329). Despite similar observed proportions the two conclusions contrast as an effect of different CI widths

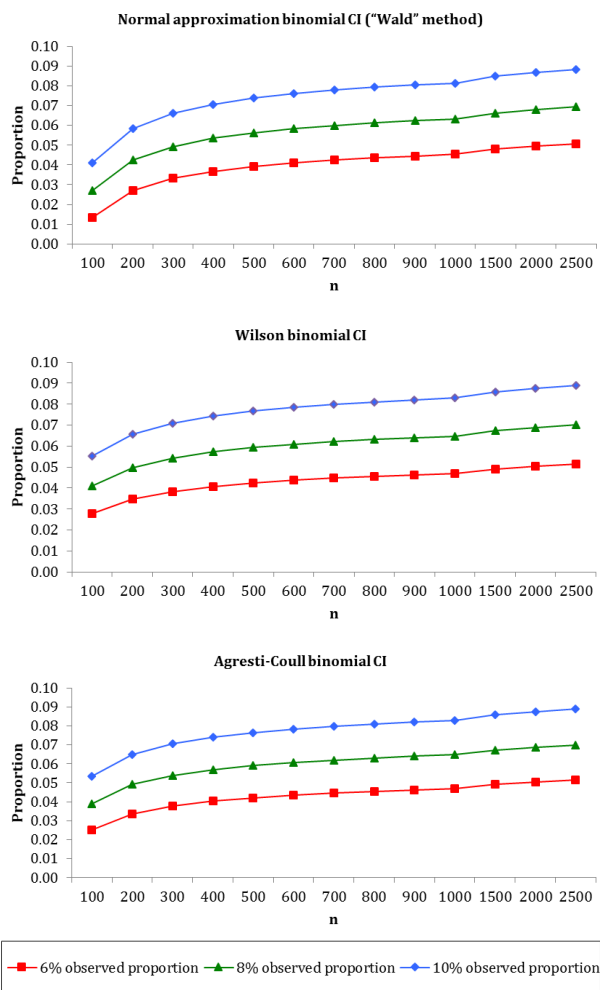


Figure 1. Lower 95% CIs according to (a) the normal approximation (“Wald”), (b) Wilson, and (c) Agresti-Coull estimation methods.

The normal approximation 95% CI seems to be the most common but also the most problematic binomial CI estimation. For example, it has been found to be highly erratic in terms of the actual interval covered, which rarely approximated 95% (Brown et al., 2001). In contrast, the Wilson and Agresti-Coull 95% CIs behaved much more reliably, particularly for small and large sample sizes, respectively (Brown et al., 2001). This aspect is rarely considered in published studies. However, authors using the PCA/t-test protocol (or any other procedure involving the binomial CI) are recommended to

report the estimation method used and there are good reasons to avoid the normal approximation estimation.

Unidimensionality is a relative matter and the decision whether a scale is sufficiently unidimensional should ultimately come from outside the data and be driven by the purpose of measurement and clinical/theoretical considerations (Andrich, 1988; Cano, Barrett, Zajicek, & Hobart, 2011; Hobart & Cano, 2009; Rasch, 1960). Use and interpretation of results from the PCA/t-test protocol must be made with the same considerations as with any hypothesis testing procedure and is dependent on sample size as well as choice of estimation method for the 95% binomial CI. The PCA/t-test procedure should not be viewed as a “definite” test for unidimensionality and does not replace an integrated quantitative/qualitative interpretation based on an explicit variable definition and in view of the perspective, context and purpose of measurement. Statistical procedures and reliance on P-values and CIs cannot compensate for conceptual and theoretical considerations.

Peter Hagell, RN PhD

The PRO-CARE Group, School of Health and Society, Kristianstad University, Kristianstad, Sweden

References

- Andrich, D. (1988). Rasch models for measurement. Beverly Hills: Sage Publications, Inc.
- Andrich, D, Sheridan, B., & Luo, G. (1997-2012). RUMM2030: Rasch Unidimensional Models for Measurement. Perth, Australia: RUMM Laboratory.
- Brown, L. D., Cai, T. T., & DasGupta, A. (2001). Interval estimation for a binomial proportion. *Stat Sci*, 16(2), 101–133.
- Cano, S. J., Barrett, L. E., Zajicek, J. P., & Hobart, J. C. (2011). Dimensionality is a relative concept. *Multiple Sclerosis*, 17(7), 893–894.
- Feinstein, A. R. (1998). P-values and confidence intervals: two sides of the same unsatisfactory coin. *J Clin Epidemiol*, 51(4), 355-360.
- Forjaz, M. J., Martinez-Martin, P., Dujardin, K., Marsh, L., Richard, I. H., Starkstein, S. E., & Leentjens, A. F. (2013). Rasch analysis of anxiety scales in Parkinson's disease. *J Psychosom Res*, 74(5), 414-419. doi: 10.1016/j.jpsychores.2013.02.009
- Hagell, P. (2014). Testing Rating Scale Unidimensionality Using the Principal Component Analysis (PCA)/t-Test Protocol with the Rasch Model: The Primacy of Theory over Statistics. *Open Journal of Statistics*, 4(6), 456-465. doi: 10.4236/ojs.2014.46044
- Hobart, J., & Cano, S. (2009). Improving the evaluation of therapeutic interventions in multiple sclerosis: the role

of new psychometric methods. *Health Technology Assessment*, 13(12), iii, ix-x, 1-177.

Horton, M., & Tennant, A. (2010). Assessing unidimensionality using Smith's (2002) approach in RUMM 2030. Paper presented at the Probabilistic models for measurement in education, psychology, social science and health, Copenhagen, Denmark.

McCormack, J., Vandermeer, B., & Allan, G. M. (2013). How confidence intervals become confusion intervals. *BMC Med Res Methodol*, 13, 134. doi: 10.1186/1471-2288-13-134

Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: comparison of seven methods. *Statistics in Medicine*, 17(8), 857-872.

Ramp, M., Khan, F., Misajon, R. A., & Pallant, J. F. (2009). Rasch analysis of the Multiple Sclerosis Impact Scale MSIS-29. *Health Qual Life Outcomes*, 7, 58. doi: 10.1186/1477-7525-7-58

Rasch, Georg. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen: Danmarks Paedagogiske Institut.

Riazi, A., Aspden, T., & Jones, F. (2014). Stroke Self-efficacy Questionnaire: a Rasch-refined measure of confidence post stroke. *J Rehabil Med*, 46(5), 406-412. doi: 10.2340/16501977-1789

Smith, E. V., Jr. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas*, 3(2), 205-231.

Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis and Rheumatism*, 57(8), 1358-1362.

Tennant, A., & Pallant, J. (2006). Unidimensionality matters. *Rasch Measurement Transactions*, 20, 1048-1051.

Young, C. A., Mills, R. J., Woolmore, J., Hawkins, C. P., & Tennant, A. (2012). The unidimensional self-efficacy scale for MS (USE-MS): developing a patient based and patient reported outcome. *Multiple Sclerosis*, 18(9), 1326-1333. doi: 10.1177/1352458512436592