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 ≤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

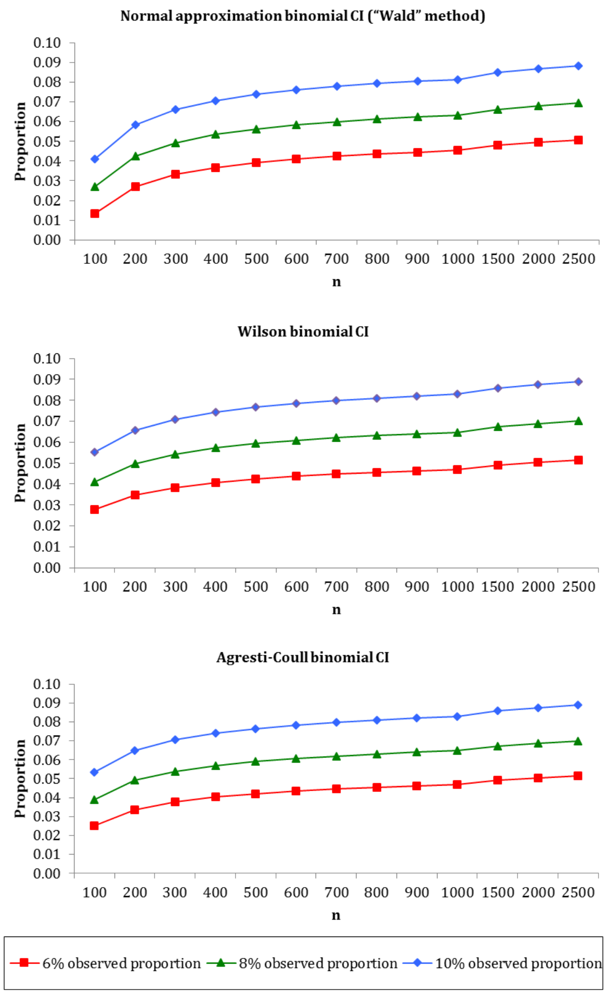


Figure 1. Lower 95% CIs according to (a) the normal approximation ("Wald"), (b) Wilson, and (c) Agresti-Coull estimation methods.

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

Unfolding Rater Accuracy in Performance Assessments

One of the persistent problems in the scoring of performance assessments is how to monitor the ratings that are obtained from raters. The most common methods for evaluating rating quality are based on rater agreement indices (Johnson, Penny, & Gordon, 2009). An emergent approach for evaluating the quality of ratings obtained in rater-mediated performance assessments is based on an examination of the accuracy of the ratings relative to a set of benchmark performances that have been assigned *true ratings* by a panel of experts. For example, Engelhard (2013) suggested dichotomously recoding the differences between operational and expert ratings, and then using these scores to define rater accuracy using a dichotomous scoring procedure. Although this method provides useful information about interesting aspects of rater accuracy, it ignores the directionality underlying the dichotomous accuracy ratings. In other words, the simple Rasch model for accuracy ratings does not parameterize directionality and *zone of accuracy* for each rater. In other words, dichotomous ratings as currently modeled do not indicate whether or not a rater tends to be lenient or severe relative to the true ratings assigned by an expert panel.

The purpose of this note is to suggest the use of unfolding models for examining rater accuracy. Unfolding models offer the potential to provide information about directionality, as well as zone of accuracy. We believe that unfolding models, such as the Hyperbolic Cosine Model (HCM; Andrich, 1996), offer a potentially useful way to unfold dichotomous accuracy ratings (0=inaccurate, 1=accurate) into three latent categories: inaccurate (below expert rating), accurate, and inaccurate (above the expert rating).

The HCM can be written as follows:

$$P\{x_{ni}=0\}=\frac{\cosh(\rho_i)}{\cosh(\rho_i)+\cosh(\beta_n-\delta_i)} \qquad (1)$$

$$P\{x_{ni}=1\}=\frac{\cosh(\beta_n-\delta_i)}{\cosh(\rho_i)+\cosh(\beta_n-\delta_i)} \qquad (2)$$

where x_{ni} represents accuracy (0=inaccurate, 1= accurate), β_n represents the location of benchmark performance *n*, δ_i represents the location of rater *i*, and ρ_i represents the unit for rater *i*. In the context of attitude measurement, the unit parameter has been interpreted as a latitude of acceptance parameter, and we suggest interpreting the unit parameter as zone of accuracy for each rater. The hyperbolic cosine function is $\cosh(x) = [\exp(-x)+\exp(x)]/2$. The parameters of this model can be estimated with *RateFOLD* software (Andrich and Luo, 2002).

Table 1 illustrates the structure of data that may be suitable for analyses with HCM. There are seven raters (Raters 1 to 6) who are evaluated on their accuracy using seven benchmark performances with known ratings. The entries in the table represent accuracy (0=inaccurate, 1=accurate) in terms of agreement with the known ratings on the benchmark performances (Engelhard, 2013). Rater accuracy rates vary from 14.3% to 57.1%. For example, Raters 2 and 5 have the same accuracy rates (42.9%) with

Rater 2 tending to be accurate on Performances A, B and C, and Rater 5 accurate on Performances E, F and G.

Table 1. Six raters evaluated on seven benchmark performances (0=inaccurate, 1=accurate)

Raters	Benchmark Performances							Accuracy Rate	Rater Location
	A	B	C	D	E	F	G		
1	1	1	0	0	0	0	0	28.6	5.98
2	1	1	1	0	0	0	0	42.9	4.48
3	0	1	1	1	1	0	0	57.1	1.44
4	0	0	0	1	1	1	1	57.1	-2.20
5	0	0	0	0	1	1	1	42.9	-3.38
6	0	0	0	0	0	0	1	14.3	-6.32

These data were analyzed with the *RateFOLD* computer program (Andrich & Luo, 2002). The locations for each rater are shown in the last column of Table 1. Figure 1 shows the variable map for rater accuracy with the locations of the benchmarks and the six raters. Figure 2 gives the probability curves for unfolded latent categories (inaccurate below, accurate, and accurate above) for the location of Rater 3. Figure 3 shows the nonlinear relationship between accuracy rates and rater locations on the unfolding scale.

Figure 1. Variable map for unfolding rater accuracy

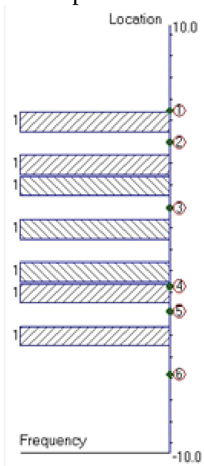


Figure 2. Probability curves for three latent ordered categories (Rater 3)

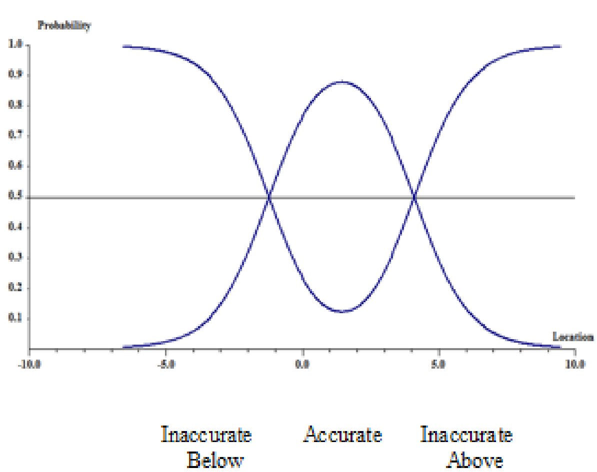
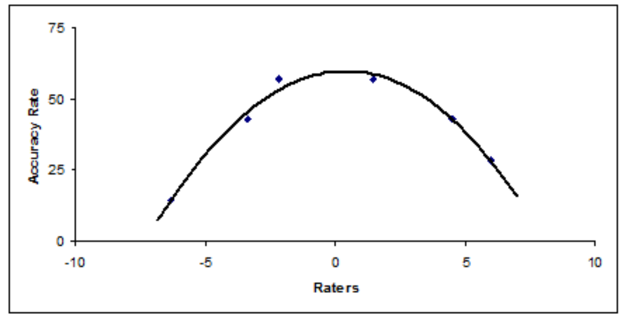


Figure 3. Plot of accuracy rate on rater locations on unfolding scale



Previous research on accuracy (Engelhard, 2013) with a dichotomous Rasch model does not provide evidence on the direction of the inaccurate ratings. The HCM holds promise for unfolding dichotomous accuracy data into three categories:

- Raters who give ratings that tend to be higher than the true ratings,
- Raters who give ratings that tend to be accurate, and
- Raters who give ratings that tend to be lower than the true ratings

We are currently conducting several studies on applying unfolding models for rater accuracy. We believe that unfolding models may be a promising approach for identifying direction of inaccuracy in a single index for each rater in conjunction with a zone of accuracy on the benchmark performances. Future research should also examine why some performances receive inaccurate ratings.

Acknowledgment: We would like to thank Professor David Andrich for providing us with the RateFOLD program.

George Engelhard, Jr. & Jue Wang
The University of Georgia

References

Andrich, D. (1996). A hyperbolic cosine latent model for unfolding polytomous responses: Reconciling Thurstone and Likert methodologies. British Journal of Mathematical and Statistical Psychology, 49, 347–365.

Andrich, D., & Luo, G. (2002). RateFOLD Computer Program. Social Measurement Laboratory: School of Education, Murdoch University. Western Australia.

Andrich, D. & Luo, G. (1993). A hyperbolic cosine latent trait model for unfolding dichotomous single-stimulus responses. Applied Psychological Measurement, 17, 253-276.

Engelhard, G. (2013). Invariant measurement: Using Rasch models in the social, behavioral, and health sciences. New York: Routledge.

Johnson, R. L., Penny, J. A., & Gordon, B. (2009). Assessing performance: Designing, scoring, and validating performance tasks. New York: Guilford Press.

Investigation and Application of the Person Aberrant Detection Indices

Many studies investigated and compared the aberrant response detection indices (Karabatsos, 2003; Linacre, 1997/2012; Li, Olejnik, 1997). Although a large number of statistics is available for detecting the person aberrant response pattern of a test, there is little consensus as to which ones are most useful.

Referring to the data and indices from Linacre (1997), we added the Point-biserial correlation coefficient (r_{pbis}) and replaced items with persons using the values of eigenvector, unrotated and rotated factor loading in current study to verify the usefulness and application in practice. The unrotated loading (denoted by Loading_1 in Table 1) represents the person loading yielded by the formula = values of eigenvector * sqrt(respective Eigenvalue). Loading_2 represents the rotated loading. Eigen denotes eigenvector values. The PTME is the r_{pbis} in terms of Rasch measures shown in Winstesp. The r_{pbis} for an examinee n can be shown with the formula (Linacre. 2014) as below:

$$r_{pbis} = \frac{\sum_{i=1}^{L}(X_{ni} - \sum_{m=1}^{L} X_{nm}/L)(\sum_{j=1}^{N} X_{ni} - \sum_{m=1}^{L}\sum_{j=1}^{N} X_{mj}/L)}{\sqrt{\sum_{i=1}^{L}(X_{ni} - \sum_{m=1}^{L} X_{nm}/L)^2 \sum_{i=1}^{L}(\sum_{j=1}^{N} X_{ni} - \sum_{m=1}^{L}\sum_{j=1}^{N} X_{mj}/L)^2}}$$

Whereas, L denotes the item length, N represents the sample size, X is the observation responses of examinees against the respective item. With which, the r_{pbis} can be calculated, and the higher the value, the greater the association.

We found that some different values (marked with an asterisk in Table 1) in the article (Linacre, 1997) might be attributable to typos for the l_z index (Li,Olejnik, 1997). Correlation coefficients between the abovementioned indices are shown in Table 2, indicating the PTME earns the highest average values compared the other four counterparts (i.e., Eigen, Loading_1, Loading_2, and r_{pbis}).

Diagnosis-related group (DRG) is a system to classify hospital inpatient cases into one of originally 467 groups (Fetter. Et al., 1980), with the last group (coded as 470 through v24, 999 thereafter) being "Ungroupable". The system is also referred to as "the DRGs", and its intent was to identify the "products" that a hospital provides. The r_{pbis} (named as coherence coefficient) can be referred to the patient item-score response pattern of medical fees for a DRG code (like in a class). We illustrated the coherence coefficients and 95% confidence interval for 38 DRGs in diseases and disorders of the circulatory system in a hospital (Figure 1). From which, it can be seen that the DRG 11003 displays a negative coherence coefficient, indicating the structure of medical fees is significantly different from its peers. From the hospital point of view, it might miss either medical fees on a specific ITEM and/or important diagnosis codes (e.g., complication or comorbidity) resulting in a low fee DRG code. From the issuance institute point of view, it can use the coherence coefficients to detect any DRG payment with a cheating up-code to claim more medical fees due to different ITEM response pattern from its peers within the same DRG.

Table 1. Investigation of the Person Aberrant Detection Indices

IN.MSQ	OUT.MSQ	lz Index	Eigen	Loading_1	Loading_2	r_{pbis}	PTME
0.98	0.99	0.49	0.22	0.86	0.77	0.77	0.78
0.98	1.04	0.75	0.21	0.85	0.65	0.81	0.81
1.06	0.97	1.15*	0.21	0.85	0.55	0.91	0.87
1.03	1.00	0.72	0.21	0.84	0.68	0.81	0.81
0.18	0.22	1.95	0.23	0.92	0.88	0.84	0.92
0.31	0.36	2.67	0.24	0.96	0.79	0.94	0.97
0.80	0.77	1.95	0.25	0.97	0.84	0.95	0.93
0.21	0.26	1.39	0.23	0.93	0.87	0.81	0.89
0.52	0.54	1.27	0.21	0.83	0.56	0.84	0.82
0.24	0.24	1.27	0.24	0.96	0.95	0.87	0.87
0.25	0.34	0.67	0.24	0.96	0.95	0.87	0.87
0.16	0.20	1.93*	0.22	0.87	0.61	0.90	0.93
0.94	1.22	-0.85	0.14	0.57	0.42	0.48	0.55
1.25	1.09	0.21	0.17	0.69	0.39	0.79	0.77
1.37	1.20	0.67	0.24	0.96	0.95	0.87	0.87
1.49	1.40	-0.56	0.19	0.74	0.51	0.76	0.72
0.85	1.21	-2.06	0.00	0.00	0.00	0.00	0.00
1.50	1.96	-3.73	-0.02	-0.09	0.07	0.10	-0.09
2.99	3.59	-6.73	-0.03	-0.13	-0.11	-0.02	-0.01
3.62	4.61	-9.34	-0.06	-0.23	-0.31	0.00	-0.19
5.14	6.07	-12.54	-0.02	-0.09	0.07	0.10	-0.09
1.75	2.02	-4.13	0.00	-0.02	-0.05	0.03	0.00
2.11	4.14	-5.64*	-0.24	-0.96	-0.95	-0.87	-0.87
2.56	3.20	-7.34	-0.24	-0.96	-0.95	-0.87	-0.87
4.00	5.58	-11.82	-0.23	-0.92	-0.88	-0.84	-0.92
8.30	9.79	-23.35	-0.24	-0.96	-0.95	-0.87	-0.87

Table 2. Correlation coefficients between the study indices

Indices	IN.MSQ	OUT.MSQ	lz Index	Eigen	Loading_1	Loading_2	r_{pbis}	PTME
IN.MSQ	1	0.985	-0.976	-0.749	-0.75	-0.735	-0.728	-0.749
OUT.MSQ	0.985	1	-0.987	-0.828	-0.828	-0.812	-0.811	-0.828
lz Index	-0.976	-0.987	1	0.847	0.847	0.823	0.832	0.848
Eigen	-0.749	-0.828	0.847	1	1	0.987	0.994	**0.997**
Loading_1	-0.75	-0.828	0.847	1	1	0.986	0.994	**0.998**
Loading_2	-0.735	-0.812	0.823	0.987	0.986	1	0.977	**0.977**
r_{pbis}	-0.728	-0.811	0.832	0.994	0.994	0.977	1	**0.995**
PTME	-0.749	-0.828	0.848	**0.997**	**0.998**	**0.977**	**0.995**	1
Rank				2	1	4	3	

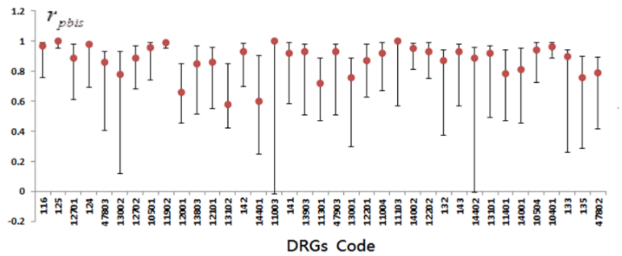


Figure 1. Inpatient medical fees r_{pbis} and 95% CI of the patient DRGs codes

The first three indices in Table 2 are suitable for category response test. In contrast, the last five are appropriate when variables are continuous like the medical fees in the illustrative example of this article.

Tsair-Wei Chien – Chi Mei Medical Center, Taiwan
Ngadiman Djaja - School of Public Health and Social Work, Queensland University of Technology, Brisbane, Australia.

References:

Donlan, T. F., Fischer, F.E. (1968) An index of an individual's agreement with group determined item difficulties. Educational and Psychological Measurement, 28, 105-113.

Fetter, R.B., Shin, Y., Freeman, J.L., Averill, R.F., Thompson, J.D. (1980) Case mix definition by diagnosis related groups. Medical Care 18(2),1–53.

Karabatsos, G.(2003) Comparing the aberrant response detection performance of thirty-six person-fit statistics. Applied Measurement in Education, 16(4), 277-298.

Li, M.F, Olejnik, S. (1997) The power of Rasch person-fit statistics in detecting unusual response patterns. Applied Psychological Measurement 21(3), 215-231.

Linacre, J.M.(1997) An All-Purpose Person Fit Statistic? Rasch Measurement Transactions, 11(3), 582-3.

Linacre, J.M.(2012) A Comment on the HT Person Fit Statistic. Rasch Measurement Transactions, 26(1), 1358.

Linacre, J.M.(2014) Correlations: point-biserial, point-measure, residual. Winsteps user's guide.2014/12/14 retrieved at
http://www.winsteps.com/winman/correlations.htm

Rasch Measurement as a Basis for Metrologically Traceable Standards

Leslie Pendrill, a past president of the European Association of National Metrology Institutes, recently published an article entitled "Man as a Measurement Instrument" in *NCSLI Measure: The Journal of Measurement Science*, published by the National Conference of Standards Laboratories International (Pendrill, 2014). Some key passages in this text include:

"The Rasch approach...is not simply a mathematical or statistical approach, but instead [is] a specifically metrological approach to human-based measurement" (p. 26).

"...the Rasch approach, with its explicit separation of person and item attribute estimation, is well suited for introducing metrological traceability to human-based measurement" (p. 28).

"If the Rasch attributes for persons, tasks or products are 'quantities' as opposed to mere numerical values, then there should be metrological 'references' for them if we are to be consistent with the definition of 'quantity' in the international metrology vocabulary VIM [1.1, 58]. In a note to that definition, a 'reference' in this context can be a 'measurement unit, a measurement procedure, a reference material, or a combination of such.' Access to metrological references for psychometric quantities would--in addition to the mathematical logit units--enable the scales of different 'rulers' for a given quantity, e.g. person ability or task challenge, to be objectively compared with each other" (p. 28)

"Thus, the measurement units (discussed for instance by Humphry [20]) associated with the Rasch attribute parameters [theta] and [beta] should be intimately related to metrological traceability and measurement standards. Perhaps the closest analogies to references in psychometrics can be found with reference materials that are utilized as references for metrological traceability in chemistry. In psychometrics, we could imagine a certified reference for knowledge challenge, for example, a particular concept in understanding physics or for product quality of a certain health care service" (p. 28).

Pendrill rightly indicates that the idea of unit definitions based on Rasch models, and so also the potential of Rasch measurement to support metrological unit traceability, are controversial. It must also be emphasized that fit to a Rasch model does not automatically confer properties of invariance, parameter separation, unidimensionality, etc. on scores or measures. As Rasch (1960, pp. 37-38; 1973/2011) was at pains to convey, data never fit models; the point is not truth but usefulness, and this must be evaluated in terms informed by not just the measures, but also by their qualitative expression relative to the learning progression or developmental sequence, by the uncertainty (error) obtained, and by any available anomalous departures from the modeled expectations. The latter may be particularly valuable as a source of practical ideas for better understanding the construct and for improving outcomes (Linacre, 1993; Fisher, 2013).

Further, though an item or performance level may well be taken as a reference standard in the way Pendrill indicates, that reference might also be provided in a more general form by a construct map (Wilson, 2005, 2009), an LLTM (Fischer, 1973; Embretson & Daniel, 2008), or a specification equation (Stenner, Smith, & Burdick, 1983; Stenner, Fisher, Stone, & Burdick, 2013) capable of explaining and predicting variation in the item difficulties by means of theory.

William P. Fisher
University of California-Berkeley

References

Embretson, S. E., & Daniel, R. C. (2008). Understanding and quantifying cognitive complexity level in mathematical problem solving items. Psychology Science Quarterly, 50, 328-344.

Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. Acta Psychologica, 37, 359-374.

Fisher, W. P., Jr. (2013). Imagining education tailored to assessment as, for, and of learning: Theory, standards, and quality improvement. Assessment and Learning, 2, 6-22.

Linacre, J. M. (1993). Quality by design: Taguchi and Rasch. Rasch Measurement Transactions, 7(2), 292.

Pendrill, L. (2014). Man as a measurement instrument [Special Feature]. NCSLI Measure: The Journal of Measurement Science, 9(4), 22-33.

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests (Reprint, with Foreword and Afterword by B. D. Wright, Chicago: University of Chicago Press, 1980). Copenhagen, Denmark: Danmarks Paedogogiske Institut.

Rasch, G. (1973/2011). All statistical models are wrong! Comments on a paper presented by Per Martin-Löf, at the Conference on Foundational Questions in Statistical Inference, Aarhus, Denmark, May 7-12, 1973. Rasch Measurement Transactions, 24(4), 1309.

Stenner, A. J., Fisher, W. P., Jr., Stone, M. H., & Burdick, D. S. (2013). Causal Rasch models. Frontiers in Psychology: Quantitative Psychology and Measurement, 4(536), 1-14.

Stenner, A. J., Smith, M., III, & Burdick, D. S. (1983). Toward a theory of construct definition. Journal of Educational Measurement, 20(4), 305-316.

Wilson, M. (2005). Constructing measures: An item response modeling approach. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Wilson, M. R. (2009). Measuring progressions: Assessment structures underlying a learning progression. Journal of Research in Science Teaching, 46, 716-730.

Rasch-related Coming Events

Mar. 11-13, 2015, Wed.-Fri. In-person workshop: Introductory Rasch (A. Tennant, RUMM), Leeds, UK, www.leeds.ac.uk/medicine/rehabmed/psychometric

Mar. 11-13, 2015, Wed.-Fri. In-person workshop: Einfuhrung in die Raschanalyse mit RUMM2030 (Bocker, Salzberger, RUMM), Aachen, Germany, www.psychometrikon.de

Mar. 20, 2015, Fri. UK Rasch User Group Meeting, London, United Kingdom, www.rasch.org.uk

Mar. 26-27, 2015, Thur.-Fri. In-person workshop: Introduction to Rasch Measurement with Winsteps (W. Boone), Cincinnati, www.raschmeasurementanalysis.com

April 16-20, 2015, Thurs.-Mon. AERA Annual Meeting, Chicago, IL, www.aera.net

April 21-22, 2015, Tues.-Wed. IOMC 2015: International Outcomes Measurement Conference, Chicago, IL, www.jampress.org

May 13-15, 2015, Wed.-Fri. In-person workshop: Introductory Rasch (A. Tennant, RUMM), Leeds, UK,

May 18-20, 2015, Mon.-Wed. In-person workshop: Intermediate Rasch (A. Tennant, RUMM), Leeds, UK,

May 29-June 26, 2015, Fri.-Fri. Online workshop: Practical Rasch Measurement – Core Topics (E. Smith, Winsteps), www.statistics.com

July 3-31, 2015, Fri.-Fri. Online workshop: Practical Rasch Measurement – Further Topics (E. Smith, Winsteps), www.statistics.com

Aug. 14-Sept. 11, 2015, Fri-Fri. Online workshop: Many-Facet Rasch Measurement (E. Smith, Facets), www.statistics.com

Sept. 4-Oct. 16, 2015, Fri.-Fri. Online workshop: Rasch Applications, Part 1: How to Construct a Rasch Scale (W. Fisher), www.statistics.com

Sept. 9-11, 2015, Wed.-Fri. In-person workshop: Introductory Rasch (A. Tennant, RUMM), Leeds, UK,

Sept. 14-16, 2015, Mon.-Wed. In-person workshop: Intermediate Rasch (A. Tennant, RUMM), Leeds, UK,

Sept. 17-18, 2015, Thur.-Fri. In-person workshop: Advanced Rasch (A. Tennant, RUMM), Leeds, UK,

Oct. 16-Nov. 13, 2015, Thur.-Fri. Online workshop: Practical Rasch Measurement – Core Topics (E. Smith, Winsteps), www.statistics.com

Message from Rasch Measurement SIG Chair

Greetings Rasch SIG colleagues,

As you all know, AERA is just around the corner (so to speak) and I thought it time to focus your minds on the coming event.

I would like to start by thanking our Program Co-Chairs Jessica Cunningham and Sara Hennings for preparing this year's Rasch Measurement SIG sessions. As with every year, the upcoming program is full of exciting and relevant papers dealing with aspects of Rasch Theory. I hope that you will all make an extra effort to support the Rasch offerings at this year's conference.

I am also very pleased to announce that George Engelhard is our Keynote Speaker at this year's business meeting. His topic and the accompanying abstract are as follows.

Invariant Measurement with Raters and Rating Scales

The purpose of this keynote address is to briefly describe the concept of invariant measurement within the context of rater-mediated assessments. As the number of performance assessments continues to increase around the world, it is of critical importance to develop indices for evaluating the psychometric quality of ratings obtained from raters using rating scales. Three categories of indices of rating quality will be described: Agreement indices, rater-error indices, and accuracy indices. A small illustration will show how these indices of rating quality can be used within the context of a large-scale writing assessment. The implication of invariant measurement for rater-mediated assessments offers a promising framework for examining rating quality in the 21st century.

I am sure that you can see the relevance of this work in the current climate. In addition to George's address there will be a brief summary of how our SIG is progressing. The meeting is scheduled for Thursday, April 16 from 6:15pm to 7:45pm. Hors d'oeuvres and a cash bar will be provided. I will send out more detailed information on all presentations and logistics prior to the AERA conference.

We will also hopefully be presenting the plaque to the winner of The Georg William Rasch Early Career Publication Award.

I look forward to catching up with you all in Chicago.

Jim Tognolini
Rasch Measurement SIG Chair

Journal of Applied Measurement
Vol. 16, No. 1, 2015

A Mathematical Theory of Ability Measure Based on Partial Credit Item Responses, *Nan L. Kong*

Differential Item Functioning Analysis by Applying Multiple Comparison Procedures, *Paolo Eusebi and Svend Kreiner*

Visually Discriminating Upper Case Letters, Lower Case Letters and Numbers, *Janet Richmond, Russell F. Waugh, and Deslea Konza*

Testing the Multidimensionality of the Inventory of School Motivation in a Dutch Student Sample, *Hanke Korpershoek, Kun Xu, Magdalena Mo Ching Mok, Dennis M. McInerney, and Greetje van der Werf*

Measuring Teaching Assistants' Efficacy using the Rasch Model, *Zi Yan, Chun Wai Lum, Rick Tze Leung Lui, Steven Sing Wa Chu, and Ming Lui*

Detecting Measurement Disturbance Effects: The Graphical Display of Item Characteristics, *Randall E. Schumacker*

Criteria Weighting with Respect to Institution's Goals for Faculty Selection, *Sheu Hua Chen, Yen Ting Chen, and Hong Tau Lee*

Gendered Language Attitudes: Exploring Language as a Gendered Construct using Rasch Measurement Theory, *Kris A. Knisely and Stefanie A. Wind*

Richard Smith, Editor, www.jampress.org

Rasch News

The California Community Colleges Common Assessment Initiative involves an exciting implementation of advanced measurement ideas. The goal of the Common Assessment Initiative (CAI) "is to develop a comprehensive, common assessment system (CAS) that will reduce unnecessary remediation, align to state legislation, and provide statewide efficiencies for the academic placement process within and between California colleges all of which will ultimately benefit student success."

Additional information about the project can be found at: http://cccassess.org/

AERA 2015 Rasch-related Papers Chicago, Illinois Thur., April 16 – Mon., April 20, 2014

Examining the Validity of the Test of Preschool Early Literacy Print Knowledge for English- and Spanish-Speaking Children Using Rasch Modeling, **Mihaiela Ristei Gugiu, Ohio State University – Columbus; *Sabrina Francesca Sembiante, Ohio State University – Columbus*

A Rasch-Based Borderline Method for Standard Setting in an Objective Structured Clinical Examination Station, **Jean-Sebastien Renaud, Université Laval; *Gilles Raiche, L'Université du Québec à Montréal; *Eric Dionne, University of Ottawa; *François Ratté, Université Laval; *Julie F. Thériault, Université Laval*

Comparing Classical Psychometric and Rasch Modeling Results Using the International Consultation on Incontinence Questionnaire–Bowel, **T. Mark Beasley, University of Alabama – Birmingham; *Shannon Lyn David, North Dakota State University*

Validating Student Feedback Surveys for Educator Evaluation Using the Rasch Construct Validity Framework, **Shelagh M. Peoples, Massachusetts Department of Elementary and Secondary Education; *Claire Abbott, Massachusetts Department of Elementary and Secondary Education; * Kathleen Marie Flanagan, Massachusetts Department of Elementary and Secondary Education*

Examining the Validity of the Test of Preschool Early Literacy Print Knowledge for English- and Spanish-Speaking Children Using Rasch Modeling; **Mihaiela Ristei Gugiu, Ohio State University – Columbus; *Sabrina Francesca Sembiante, Ohio State University - Columbus*

Measuring Mathematics Testing Confidence and Anxiety: A Scale Analysis Using Rasch Modeling; **Caroline Vuilleumier, Boston College; *Kelsey Klein, Boston College*

Rasch Calibration: Reform-Oriented Teaching Practices; **Hye Sun You, University of Texas - Austin*

The Effect of Research on Changes in Interest in Science and Technology in College; **William Lee Romine, Wright State University; *Troy D. Sadler, University of Missouri - Columbia*

A Cross-Classified Explanatory Partial-Credit Model; **Luke Stanke, University of Minnesota; *Okan Bulut, University of Alberta*

Adaptive Measure of Change: Impacts of Item Selection, Test Length, and Hypothesis Testing; **Wei He, Northwest Evaluation Association*

Investigating Theta Equating for the Rasch Testlet Model Under Nonequivalent Groups Anchor Test Design; **Hirotaka Fukuhara, Pearson; *Insu Paek, Florida State University*

Unidimensionality or Multidimensionality: The Application of a Rasch Testlet Model in a Mixed-Format Reading Proficiency Test; **Lihong Yang, Michigan State University*

The Detection of Severity and Centrality in Raters Under Various Levels of Double Scoring; **Rose Stafford, University of Texas – Austin; *Edward W. Wolfe, Pearson; *Jodi M. Casabianca, University of Texas – Austin; *Tian Song, Pearson Assessment & Information*

Estimating Interrater Reliability Using Latent Variable Modeling and Incomplete Data; **Grant B. Morgan, Baylor University; *Robert L. Johnson, University of South Carolina; *Kari Hodge, Baylor University*

Trifactor Model for the Multiple Ratings Data; **Hyo Jeong Shin, University of California – Berkeley; *Mark R. Wilson, University of California – Berkeley*

Evaluating Rater Accuracy With a Hyperbolic Cosine Unfolding Model; **Jue Wang, University of Georgia – Athens; *Edward W. Wolfe, Pearson; *George Engelhard, University of Georgia*

Trade-Offs in the Implementation of Observational Ratings Systems; **Stephen Ponisciak, University of Wisconsin – Madison; *Nandita Gawade, Wisconsin Center for Education Research; *Yang Wang, Education Analytics; *Robert H. Meyer, University of Wisconsin - Madison*

Applying the Many-Faceted Rasch Measurement Model to Explore Reviewer Ratings of AERA Annual Conference Proposals; **Kelly D. Bradley, University of Kentucky; *Michael Peabody, American Board of Family Medicine; *Richard Kweku Mensah, University of Kentucky*

Using the Rasch Model to Examine the 2006 PISA (Programme for International Student Assessment) Measure of Middle School Resources; **Ruixue Liu, University of Kentucky; *Kelly D. Bradley, University of Kentucky*

Examination of a Parent School Climate Survey Using Rasch Methodology; **Elizabeth Leighton, University of South Carolina; *Mihaela Ene, University of South Carolina; *Christine DiStefano, University of South Carolina; *Diane M. Monrad, University of South Carolina*

Validity of the Adapted Online Self-Regulated Learning Questionnaire; **Beyza Aksu Dunya, University of Illinois at Chicago; *Kubra Karakaya Ozyer, University of Illinois at Chicago*

Building a Learning Progression for Scientific Imagination: A Measurement Perspective; *Chia-Chi Wang, National Sun Yat-Sen University; *Hsiao-Chi Ho, National Sun Yat-Sen University; *Ying-Yao Cheng, National Sun Yat-Sen University*

Detecting Rater Effects in Writing Assessment: A Multilevel Modeling Approach; **Mihaela Ene, University of South Carolina; *Robert L. Johnson, University of South Carolina; *Edward W. Wolfe, Pearson*

Exploring the Effects of Rater Linking Designs and Rater Fit on Achievement Estimates Within the Context of Music Performance Assessments; **Stefanie Anne Wind, Georgia Institute of Technology, *Brian Wesolowski, University of Georgia; *George Engelhard, University of Georgia*

Statistical and Social Validation of Leadership Scales; **Stephan Gerhard Huber, University of Teacher Education Zug; *Rolf Olsen; *Alexandra Petridou, University of Manchester; *Marius Schwander, University of Teacher Education Zug; *Christian Brandmo, University of Oslo; *Jonas Melker Hoog, Umea University*

Evaluating the Quality of Analytic Ratings With Mokken Scaling; **Stefanie Anne Wind, Georgia Institute of Technology*

Secondary Analysis of PISA (Programme for International Student Assessment) 2012 Data Using a Mixture Rasch Model With a Covariate; **Tugba Karadavut, University of Georgia; *Seock-Ho Kim, University of Georgia*

Investigating Adjudicator Bias in Concert Band Evaluations: An Application of the Many-Facets Rasch Model; **D. Gregory Springer, Boise State University; *Kelly D. Bradley, University of Kentucky*

Measuring Preservice Teachers' Competencies Regarding Linguistically Diverse Classrooms in Germany: Test Development and Its Validation; **Timo Ehmke, Leuphana University – Lueneburg; *Svenja Hammer, Luephana University – Leuenburg*

The Validation of Computer Game Engagement Instrument Using Rasch Model; **Sunha Kim, University at Buffalo – SUNY; *Mido Change, Florida International University*

Analysis of Students' College Experiences: Many-Facet Rasch Rating Scale Analysis; **Zongmin Kang, DePaul University; *Gregory E. Stone, University of Toledo*

Developing an Engineering Design Process Assessment Using Mixed Methods: An Illustration With Rasch Measurement Theory and Cognitive Interviews; **Stefanie Anne Wind, Georgia Institute of Technology; *Meltem Alemdar, Georgia Institute of Technology; *Jessica Gale, Georgia Institute of Technology; *Jeremy Lingle, Georgia Institute of Technology; *Roxanne Moore, Georgia Institute of Technology*

Conceptualizing and Measuring Opportunities to Learn and the Contexts of Teaching; **Michael C. Rodriguez, University of Minnesota; *Maria Teresa Tatto, Michigan State University*

Developing Student Feedback Surveys for Educator Evaluation: Combining Stakeholder Engagement and Psychometric Analyses; **Shelagh M. Peoples, Massachusetts Department of Elementary and Secondary Education; *Claire Abbott, Massachusetts Department of Elementary and Secondary Education; *Kathleen Marie Flanagan, Massachusetts Department of Elementary and Secondary Education*

Investigating Effects of the Complexity of Leveled Texts on Student Comprehension; **Yukie Toyama, University of California – Berkeley; *Alexandra N. Spichtig, Reading Plus/Taylor Associates*

Exploring the Synergies Between Distal Future and Proximal Achievement Goals and Metacognition in Academic Achievement; **Dennis M. McInerney, The Hong Kong Institute of Education; *Fraide A. Ganotice, The Hong Kong Institute of Education; *Ronnel Bornasal King, The Hong Kong Institute of Education*

Interactive Knowledge Building Through Research on Teacher Evaluation in Chicago; **Susan E. Sporte, University of Chicago; Jennie Jiang, University of Chicago*

Differences in Beliefs and Knowledge for Teaching Mathematics: An International Study of Future Teachers; **Traci Shizu Kutaka, University of Nebraska – Lincoln; *Wendy M. Smith, University of Nebraska – Lincoln; *Anthony Albano, University of Nebraska – Lincoln; *Chansuk Kang, University of Nebraska – Lincoln*

Addressing Measurement Challenges in the Use of Rubrics to Evaluate General Education Outcomes; **Tracy Bartholomew, Concordia University – Chicago; *Beth Venzke, Concordia University – Chicago; *Elizabeth Owolabi, Concordia University – Chicago*

Using Scenarios to Assess Teachers' Justification of Actions in Mathematics Teaching; **Ander Willard Erickson, University of Michigan; *Justin Kelly Dimmel, University of Michigan; *Kristi Hanby, University of Michigan; *Inah Ko, University of Michigan*

A Novel Method for Evaluating Item Quality in Medical and Professional School Exams; **Kenneth D. Royal, North Carolina State University; *Mari-Wells Hedgpeth, North Carolina State University; *Ryan Madanick, University of North Carolina at Chapel Hill*

A Mixture Response Time Model for Test Speededness; **Aijun Wang, Federation of State Boards of Physical Therapy; *Yu Zhang, Federation of State Boards of*

*Physical Therapy; *Lorin Mueller, Federation of State Boards of Physical Therapy*

Student Variables and the Engagement Balance in Western Australian Secondary School Classroom Learning Environments; **Robert Frederick Cavanagh, Curtin University; *David J. Shernoff, Rutgers University*

IOMC 2015 Conference Program Chicago, Illinois Tues., April 21 - Wed., April 22, 2015

*Program subject to change

William P. Fisher, Jr. - Building the field of dreams: The unrealized scientific and economic power of health care outcome metrology

Jeremy Hobart - Alternate paths to successful clinical outcome measurement

Stefan Cano - Individual vs group level measurement: Implications for health care outcome economics

Laurie Burke - On systematically measuring the right things well enough vs locally measuring the wrong things really well

Nikolaus Bezruczko, Teresa Stanley, Maureen Battle, Cynthia Latty, and Shu-Pi Chen - Pathological consequences of evaluating simulation caregiver training with nonlinear, ordinal ratings: Measuring caregiver response to tracheostomy emergencies

Michael R. Peabody, Kelly D. Bradley, and Melba Custer - Assessing the validity of a continuum-of-care survey: A Rasch measurement approach

Alexandra Rouquette, Jean-Benoit Hardouin, and Joël Coste - Differential Item Functioning (DIF) and subsequent bias in group comparisons using a multi-item scale: A simulation study

James J. Thompson - What are you measuring? Dimensionality and reliability analysis of ability and speed in medical school didactic examinations

Chih-Ying Li and Craig A. Velozo - Using Rasch analysis to generate Medicare G-Code Modifiers and develop a treatment framework in the attention domain

Craig Velozo, Leigh Lehman, Ickpyo Hong, and Chih-Ying Li - Use of Rasch analysis to generate G-Code Modifiers for CMS outpatient reimbursement

Matthew W. Grady, Haiqin Chen, Chien-Lin Yang, and David Waldschmidt - Comparing the Rasch and Rasch testlet model for a health care licensureexamination

Haiqin Chen, Matthew W. Grady, Chien-Lin Yang, and David Waldschmidt - Application of the multilevel Rasch testlet model for dual local dependence to empirical data in the health care field

Ngadiman Djaja, Monika Janda, Catherine Olsen, and David Whiteman - Diagnostic discrimination of the Skin Cancer Risk (SCR) scale: Application of item response theory

Sherri L. LaVela, Sara Locatelli, Carol Kostovich, and Megan Gosch - Developing the Respirator Comfort, Wearing Experience, and Function Instrument using Rasch partial credit model analysis

Benjamin Fox - The application of Rasch measurement theory to dementia research

Nick Marosszeky - Not a fan of Fan (1998)! Item response theory and classical test theory: An empirical comparison of their item / person statistics

Peter Hagell and Albert Westergren - Sample size and statistical conclusions from tests of fit to the Rasch measurement model according to the RUMM2030 program

Robert Furter - Test speededness: Collecting evidence to support form length decisions

Melissa Hofmann - Assessment of acute trauma exposure response for FIRE-EMS personnel

Jane Summer and William P. Fisher, Jr. - Ontological midwifery of caring in nursing: Practical measures for management

Brett Berg, Karen Atler, and Anne G. Fisher - Constructing a health outcome measure of occupational experience: An application of Rasch measurement methods

Chris Wera - Development of a brief screening measure for depression and problem drinking

Ickpyo Hong, Annie N. Simpson, Chih-Ying Li, and Craig A. Velozo - Development of an upper extremity function measurement model

Jack Stenner - Reading measurement in education as a model metrology network for health care

Rob Cavanagh - An unmodern perspective on the role of educational measurement in globalization

Robert Massof - Lions Low Vision Rehabilitation Network (LOVRNET): A system that uses outcome measures for quality improvement through continuous professional

Maureen K. Powers, William P. Fisher, Jr., Robert Massof, and Mark Wilson - Integrating visual symptoms and visual skills to model and measure functional binocular vision

Maria do Céu Ferreira and Ana Sofia Matos - A comparison study to evaluate the role of metrological traceability in health care

Myriam Blanchin, Elodie De Bock, Gildas Kubis, Tanguy Le Néel, Véronique Sébille, and Jean-Benoit Hardouin - Rasch and CTT-based approaches for joint analysis of group and time effects of longitudinal Patient Reported Outcomes: impact of informative and non-informative missing data

Véronique Sébille, Myriam Blanchin, Alice Guilleux, Mohand-Larbi Feddag, and Jean-Benoit Hardouin - Methods of power and sample size determination of clinical studies based on Rasch measurement

Sarah Thomas, Karen M. Schmidt, Monica Erbacher, and Cindy Bergeman - Sliding scales and changing rulers: Anchoring the longitudinal measurement of positive affect

Carol T. Kostovich, Beyza Aksu Dünya, Lee A. Schmidt, and Eileen G. Collins - A Rasch rating scale analysis of the Presence of Nursing Scale-RN

Christophe Chénier, Gilles Raîche, Céline Gélinas, Nadine Talbot, and Bianca Carignan - Rasch analysis of the Critical-Care Pain Observation Tool (CPOT)

Trudy Mallinson - Addressing response dependence in repeated measures of rehabilitation outcomes in the unidimensionsl Rasch model

Robert Massof and Judith Goldstein - The role of item filtering in measures of low vision rehabilitation outcomes

Deborah M. Rooney, Bruce L. Tai, Oren Sagher, Albert J. Shih, and Luis Savastano - Validation of performance measures from a Novel Ventriculostomy Simulator using *Standards* framework

Thomas Salzberger and Stefan Cano - Investigating a lack of discrimination between two adjacent response categories in the Rasch model for ordered categories in health measurement

Curt Hagquist - Do 7 items provide as good measurement as 13 items? A comparison of a short and long version of Kidscreen

Paula Petry and William Fisher - Applying published instrument development research in a workshop evaluation: Practical use of Rasch-calibrated instruments with small samples

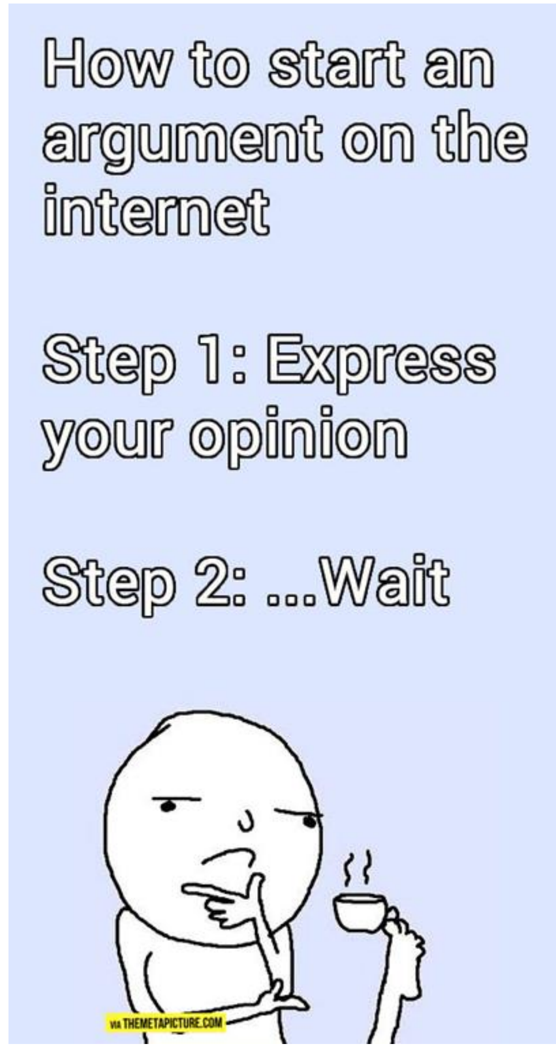
Shu-Ren Chang, Gene A. Kramer, and Shu-Mei Lien - A comparison of objective and Book mark standard setting methods on pass/fail decisions

Ying Du - Investigating knowledge growth during pediatric residency training using Rasch and linear fixed models

Gunnar Grimby - On the treatment of ordinal scale data in rehabilitation medicine research

Nikolaus Bezruczko - Standards and practices to guide health outcomes measurement: A strategy to avoid measurement malpractice

Richard Smith, Lee McKenna, and Christie Plackner - A comparison of the information available in various measurement models



Call for Submissions

Research notes, news, commentaries, tutorials and other submissions in line with *RMT*'s mission are welcome for publication consideration. All submissions need to be short and concise (approximately 400 words with a table, or 500 words without a table or graphic). The next issue of *RMT* is targeted for June 1, 2015, so please make your submission by May 1, 2015 for full consideration. Please email Editor\at/Rasch.org with your submissions and/or ideas for future content.