Summer 2017, VOL. 31, NO. 1; ISSN 1051-0796

RMT

RASCH MEASUREMENT TRANSACTIONS

Research Notes and Editorials Include:

- **Measurement Decision Theory from a Rasch Perspective – John M. Linacre**
- **A New Fit Statistic for the Dichotomous Rasch Model: Modified Cut-off Values – Purya Baghaei, Takuya Yanagida, & Moritz Heene**
- **An Online Multidimensional Computerized Adaptive Testing (MCAT) Module Using APP – Tsair-Wei Chien, Wen-Chung Wang**
- **Note from SIG Chair – Leigh Harrell Williams**
- **Profiles in Rasch Measurement – Sarah Thomas**

Transactions of the Rasch Measurement SIG
American Educational Research Association

Measurement Decision Theory from a Rasch Perspective

"Measurement Decision Theory (MDT), developed by A. Wald (1947) and now widely used in engineering, agriculture, and computing, provides a simple model for the analysis of categorical data. Measurement decision theory requires only one key assumption -- that the items are independent" (Rudner, 2001). Independent items are also a Rasch assumption. MDT and Rasch agree thus far.

Let's look at a worked example of MDT (ABA, 2016). ABA provide the responses and calculations for fictitious examinee, Dr. Able. These are shown in Table 1 with additional computations:

Dr. Able is a new examinee. Previous examinees have fallen into two groups, Group S of successful examinees and Group F of failing examinees. For each item, there is the success rate of the Group, the item p-value (probability value). The p-values are higher for Group S than for Group F. In Table 1, the odds-ratio and log-odds (logit) of each item for each group is shown. We see that, on average, the items are 0.8 logits easier for Group S than for Group F. This indicates that the average difference in Group abilities is 0.8 logits.

The MDT Latent Variable from a Rasch Perspective

97% of the previous examinees are in Group S, only 3% in Group F, so we can use the Group S item difficulties to draw a Wright item-person map of the latent variable. This is shown in Figure 1. The item difficulties are slightly different for Group F indicating that Group F slightly misfits with the variable defined by Group S. This is not unexpected, because the failing Group is more likely to have idiosyncratic knowledge, guessing, skipping and other off-dimensional behaviors.

In Figure 2, the Rasch measure for each raw score is added to the map. Rasch produces the same measure for every way of getting a raw score. Dr. Able scored 3 and so is higher on the latent variable than the average examinee in Group S. A score of 2 is probably a failure, depending on the decision of the Examination Board about the exact location of the pass-fail cut-point on the latent variable. A score of 1 is a definite failure.

MDT Likelihoods and Probabilities

MDT takes a different approach. First off, it computes the likelihood of the examinee's response string for each of the two Groups, and then estimates the probability that the string was produced by an average member of each Group. The likelihood of Dr. Able's responses for an average member of Group S is the product of their probability of success on the first three items and failure on the fourth item = 0.80 * 0.73 * 0.58 * (1 - 0.55) = 0.15. Similarly, for an average member of Group F, the likelihood is 0.07. So the probability that this is scored by an average member of Group S and not an average member of Group F is 0.15 / (0.15 + 0.07) = 0.68. So, Dr. Able probably belongs to Group S.

We can do this for all the 16 possible different response strings. The computations are shown in Table 2. We can transform the probability of success for an average member of Group S into its logit value. These are plotted in Figure 3 with .5 probability positioned halfway between Group S and Group F. Dr. Able's probability is the most probable score of 3 shown in red. Most scores of 2 are probably Group F. Most scores of 3 are probably Group S. However, one score of 3 (failure on the easiest item) is a fail and one score of 2 (success on the two easiest items) is a success. MDT is highly influenced by performance on very easy and very hard items. In contrast, a rationale for treating all items equally is given at "The Eternal Question about Raw Scores" www.rasch.org/rmt/rmt154k.htm

Rasch Measurement Transactions
www.rasch.org/rmt
Editor: Kenneth Royal
Email submissions to: Editor \at/ Rasch.org

RMT Editor Emeritus: John M. Linacre
Rasch SIG Chair: Leigh Harrell-Williams
Secretary: Mikaela Raddatz
Treasurer: Matt Schulz
Program Chairs: Liru Zhang & Eli Jones
Rasch SIG website: www.raschsig.org

Table 1. Four items administered to Dr. Able

Item	Dr. Able's response	Group S: successful examinees			Group F: failing examinees			Group difference logits
		Item p-value	Odds-ratio	Log-odds = logits	Item p-value	Odds-ratio	Log-odds = logits	
1	✓ = 1	0.80	0.80/0.20	1.39	0.60	0.60/0.40	0.41	0.98
2	✓ = 1	0.73	0.73/0.37	0.68	0.53	0.53/0.47	0.12	0.56
3	✓ = 1	0.58	0.58/0.42	0.32	0.38	0.38/0.62	-0.49	0.81
4	× = 0	0.55	0.55/0.45	0.20	0.35	0.35/0.65	-0.62	0.82
Score:	3 of 4		Average:	0.65		Average:	-0.15	0.80

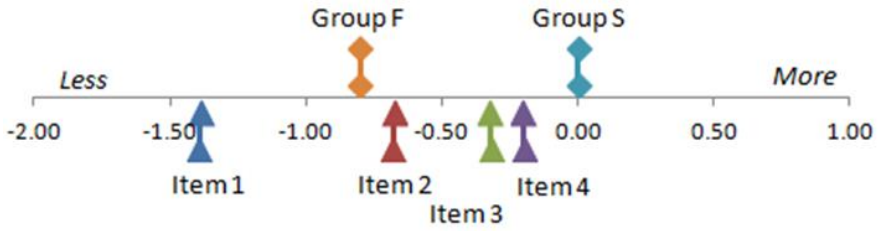

Figure 1. Wright latent-variable map of MDT

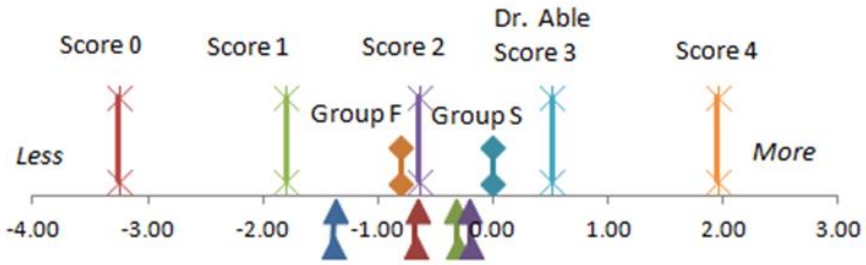

Figure 2. Wright map with person measures for possible scores.

MDT Weighted Probabilities

MDT goes one step further. It weights the likelihoods of the response strings based on the proportion of all examinee believed to belong to each group. In the ABA example, 97% of previous examinees belong to Group S. 3% belong to Group F. These percentages multiply the likelihoods in Table 2 and the resulting probabilities are shown in the rightmost column of Table 2.The probability for Dr. Able agrees with the ABA number. The weighted probabilities of membership in Group S rather than Group F vary from 0.85 to 0.99. Even a person with a score of 0 would succeed! Only a score of 0 on more than 8 items like these would produce a probability less than 0.5 and so a fail. Clearly, this Bayesian weighting based on previous samples is problematic.

Conclusion

MDT is a "simple model" (Rudner, 2001), but this is only computationally. The weighting by previous success rates produces nonsensical pass-fail decisions, at least for this example. Even ignoring the weighting, it would be challenging to explain to your audience that a failure on a very easy item, perhaps a careless mistake or misunderstanding ("the item couldn't be this

Table 2. Probability of response string for average member of Group S relative to Group F.

Response String				Score	Group S Likelihood	Group F Likelihood	Group S Probability	Group S logit -0.4	Weighted Probability
0	0	0	0	0	0.01	0.06	0.15	-2.12	0.85
0	0	1	0	1	0.01	0.05	0.21	-1.72	0.90
0	0	1	1	2	0.02	0.05	0.26	-1.46	0.92
0	1	0	0	1	0.03	0.06	0.31	-1.18	0.94
1	0	0	0	1	0.04	0.07	0.36	-0.98	0.95
0	1	0	1	2	0.03	0.06	0.37	-0.92	0.95
0	1	1	0	2	0.04	0.06	0.41	-0.78	0.96
1	0	0	1	2	0.05	0.07	0.42	-0.72	0.96
1	0	1	0	2	0.06	0.07	0.46	-0.58	0.96
0	1	1	1	3	0.05	0.05	0.47	-0.52	0.97
1	0	1	1	3	0.07	0.06	0.52	-0.32	0.97
1	1	0	0	2	0.11	0.08	0.59	-0.04	0.98
1	1	0	1	3	0.13	0.07	0.65	0.22	0.98
1	1	1	0	3	0.15	0.07	0.68	0.36	0.99
1	1	1	1	4	0.19	0.07	0.74	0.62	0.99

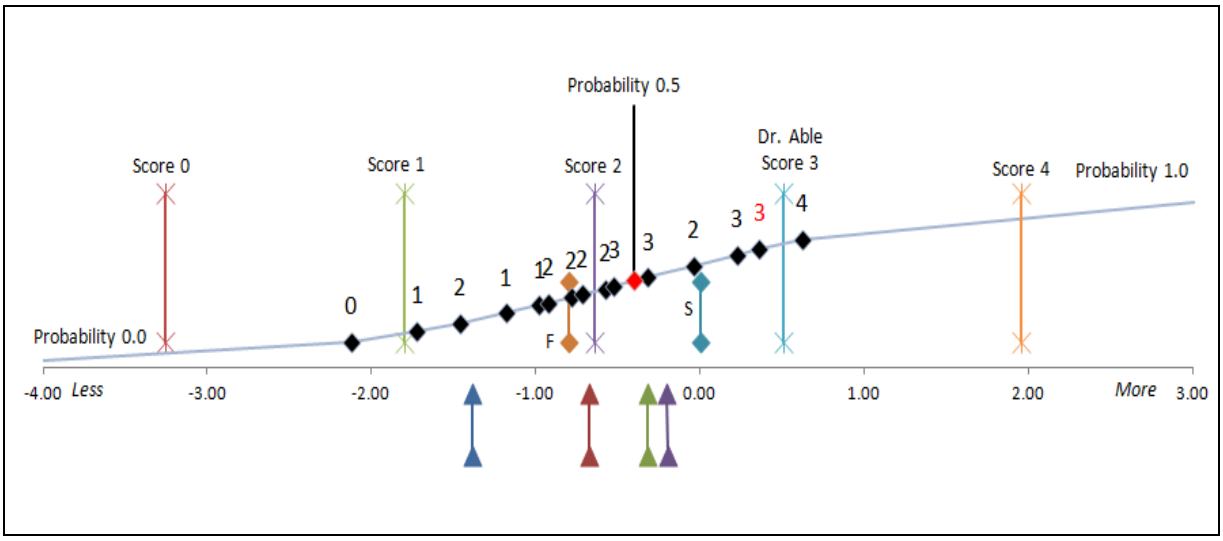

Figure 3. The Group S (relative to Group F) probability of each response string is plotted with its score.

easy!") can lead to failure when there has been success on much more difficult items.

Rasch produces a simple picture in which the two Groups, every item and every raw score can be positioned unambiguously. An intelligent, informed decision can be made about the precise location of the pass-fail cut-point. The sizes of previous pass and fail groups are irrelevant to decisions about the current examinee. They merely "put a thumb on the scales".

John Michael Linacre
mike@winsteps.com

ABA, American Board of Anesthesiology (2016) How Measurement Decision Works. http://www.theaba.org/PDFs/MOCA/Sample-MDT-Calculation

Rudner L.M. (2001) Measurement Decision Theory. National Inst. on Student Achievement, Curriculum, and Assessment (ED/OERI),

Washington, DC. ERIC ED 457 164 http://files.eric.ed.gov/fulltext/ED457164.pdf

Wald, A. (1947). Sequential analysis. New York: Wiley.

A New Fit Statistic for the Dichotomous Rasch Model: Modified Cut-off Values

Most of the fit tests available for the Rasch model rely on the principles of statistical hypothesis testing. However, statistical tests for the Rasch model have certain disadvantages that affect their utility and reliability as model checks. The most serious problem of statistical significance testing is their dependency of statistical power on sample size and the number of items. Moreover, they do not provide direct information about the degree of misfit between the data and the model. Baghaei, Yanagida and Heene (2017) suggested a descriptive fit value for the Rasch model based on Andersen's likelihood ratio test. Later examinations of the proposed fit value revealed that it is rather stringent. In this note, modified values for the test are suggested.

Simulation study

To investigate the properties of the proposed measures, simulations based on two general conditions were carried out: (1) without differential item functioning, that is under null hypothesis conditions and (2) with differential item functioning or alternative hypothesis conditions. In the alternative hypothesis conditions, data were simulated with 8 DIF items. The magnitude of DIF was |DIF| = 0.6 or 1/10 range of the simulated item parameters.

The item parameters were set as equally spaced within the interval [-3, 3], which corresponds to the whole spectrum of item difficulties that usually arise in practice. Meanwhile, the person parameters of examinees were randomly drawn from $N(0, 1.5)$, again corresponding to the values of person parameters that are likely to occur in practice. Simulations were conducted in R (R Core Team, 2015) using the eRm package (Mair, 2015).

To compute the proposed fit statistics, data sets were split into high and low scorers, based on the mean of the raw scores. Next, the item parameters were estimated separately in the two subsamples. Lastly, the item parameters where brought on to a

common scale. In each condition, the fit statistic in question was computed for 10,000 replications. In addition, for each fit statistic, we computed mean, standard deviation as well as minimum and maximum over all replications. The following statistics were computed:

Root-mean-square deviation (RMSD)

RMSD is the square root of the mean square difference between item parameters estimated in two subgroups after bringing them onto a common scale:

$$RMSD = \sqrt{\frac{\sum_{i=1}^{k}(\hat{\beta}_{i1} - \hat{\beta}_{i2})^2}{k}}$$

where $\hat{\beta}_{i1}$ is the estimated item parameter in the first subgroup (e.g., examinees with low scores), $\hat{\beta}_{i2}$ is the estimated item parameter in the second subgroup (e.g., examinees with high scores), and k is the number of items. Following the rationale of the Andersen's (1973) likelihood ratio (LR) test, if the Rasch model holds in the population, equivalent item parameter estimates should be obtained, apart from sampling error, which means the RMSD should be zero.

Standardized root-mean-square deviation (SRMSD)

SRMSD is the RMSD divided by the pooled standard deviation (SD_{pooled}) of item parameters for both subgroups:

$$SRMSD = \frac{RMSD}{SD_{pooled}}$$

The pooled standard deviation is given by:

$$SD_{pooled} = \frac{SD(\beta_{i1}) + SD(\beta_{i2})}{2}$$

If the Rasch model holds, the RMSD should be zero.

Normalized root-mean-square deviation (NRMSD)

The NRMSD is the RMSD divided by the range of estimated item parameters in both subgroups:

$$SRMSD = \frac{RMSD}{\max(\hat{\beta}_{i1}, \hat{\beta}_{i2}) - \min(\hat{\beta}_{i1}, \hat{\beta}_{i2})}$$

where $\max(\hat{\beta}_{i1}, \hat{\beta}_{i2})$ is the maximum of the item parameters in both subgroups and $\min(\hat{\beta}_{i1}, \hat{\beta}_{i2})$ is the minimum of the item parameters in both subgroups. Again, if the Rasch model holds, the SRMSD should be near zero.

Chi square to degree of freedom ($\frac{\chi^2}{df}$)

The chi square to degree of freedom ($\frac{\chi^2}{df}$) is commonly applied in the framework of structural equation modeling (SEM) to assess model fit (see West, 2012). The rationale is that the expected value of the χ^2 for a correct model equals its degrees of freedom. Thus, if the Rasch model holds, $\frac{\chi^2}{df}$ should be close to one. The current study investigated $\frac{\chi^2}{df}$for both the Andersen's LR test and the Fischer and Scheiblechner'sS statistic.

Root mean square error of approximation (RMSEA)

The RMSEA (Steiger, 1980) is a widely used fit measure in structural equation modeling:

$$RMSEA = \sqrt{max\left\{\frac{(\chi^2 - df, 0)}{df(n-1)}\right\}}$$

The RMSEA is the square root of the normalized mean non-centrality parameter, given through $(\chi^2 - df)$/(*n*-1), per degree of freedom. The noncentrality parameter is an estimate of the squared distance between a model and data. Because the degrees of freedom are the number of dimensions in which data can differ from a model after its parameters were estimated, the RMSEA serves as an average measure of lack of fit per dimension of potential lack of fit.

When the chi-square is less than the degree of freedom, the RMSEA is set to zero. In the current study, the RMSEA based on both the Andersen's LR test and the Fischer and Scheiblechner's S statistic is investigated. If the Rasch model holds, the RMESA should be near zero.

Results

Table 1 show the simulation results in the null hypothesis condition where there is no DIF, i.e., when the data fit the Rasch model perfectly. Table 2 shows the results where there are 8 DIF items.

The tables show that most of the statistics depend on sample size. Only $\frac{\chi^2}{df}$ for the Andersen's LR test is robust against sample size but it is sensitive to test length. In Baghaei, Yanagida, and Heene (2017), to be on the safe side, we set 68% confidence intervals for the χ^2/df values. Further examinations revealed that those criteria were more stringent than Andersen's (1973) LR test. That is, when the Rasch model fits according to the LR test it does not fit according to χ^2/df values set on the basis of 68% confidence intervals. Therefore, modified criteria based on 95% confidence intervals are suggested in this study (Table 3). To compute the suggested fit value you do not need a new Rasch model computer program. If your software performs Andersen's LR test, then just divide the chi square value by the relevant degrees of freedom.

Purya Baghaei, Takuya Yanagida, Moritz Heene,

Table 1. Simulation Results of the Null Hypothesis Conditions

		RMSD				SRMSD				SRMSD				Andersen χ^2/df				S Statistic χ^2/df				Andersen RMSEA				S Statistic RMSEA			
k	N	Jd	SD	Min	Max	Jd	SD	Min	Max	Jd	SD	Min	Max	Jd	SD	Min	Max	Jd	SD	Min	Max	Jd	SD	Min	Max	Jd	SD	Min	Max
10	100	0.01	0.17	0.16	1.47	0.28	0.07	0.07	0.58	0.09	0.02	0.02	0.17	1.01	0.47	0.06	3.94	1.08	0.46	0.07	3.81	0.03	0.03	0.00	0.17	0.03	0.04	0.00	0.17
	200	0.42	0.12	0.11	1.04	0.20	0.05	0.05	0.48	0.06	0.02	0.02	0.13	1.02	0.48	0.08	3.88	1.12	0.52	0.08	3.95	0.02	0.02	0.00	0.12	0.02	0.03	0.00	0.12
	300	0.34	0.09	0.04	0.87	0.16	0.04	0.02	0.41	0.05	0.01	0.01	0.11	1.01	0.47	0.02	3.84	1.11	0.52	0.02	3.12	0.02	0.02	0.00	0.10	0.02	0.02	0.00	0.12
	400	0.29	0.08	0.08	0.70	0.14	0.04	0.04	0.32	0.05	0.01	0.01	0.09	1.02	0.49	0.07	3.91	1.13	0.54	0.08	4.59	0.01	0.02	0.00	0.09	0.02	0.02	0.00	0.09
	500	0.28	0.07	0.06	0.59	0.13	0.03	0.03	0.27	0.04	0.01	0.01	0.08	1.00	0.47	0.08	3.80	1.11	0.53	0.07	3.97	0.01	0.02	0.00	0.07	0.01	0.02	0.00	0.08
	600	0.24	0.06	0.07	0.52	0.12	0.03	0.03	0.23	0.04	0.01	0.01	0.07	1.01	0.47	0.08	4.04	1.11	0.52	0.07	4.33	0.01	0.01	0.00	0.07	0.01	0.02	0.00	0.07
	700	0.22	0.06	0.04	0.47	0.11	0.03	0.02	0.23	0.03	0.01	0.01	0.07	1.01	0.47	0.04	3.91	1.11	0.53	0.04	4.22	0.01	0.01	0.00	0.06	0.01	0.01	0.00	0.07
	800	0.20	0.05	0.06	0.44	0.10	0.03	0.03	0.21	0.03	0.01	0.01	0.07	1.00	0.48	0.07	3.79	1.11	0.53	0.08	4.34	0.01	0.01	0.00	0.06	0.01	0.01	0.00	0.07
	900	0.19	0.05	0.05	0.40	0.09	0.02	0.03	0.20	0.03	0.01	0.01	0.07	1.00	0.47	0.08	4.06	1.11	0.53	0.09	4.31	0.01	0.01	0.00	0.06	0.01	0.01	0.00	0.06
	1000	0.18	0.05	0.05	0.41	0.09	0.02	0.03	0.20	0.03	0.01	0.01	0.06	0.99	0.47	0.09	3.89	1.10	0.53	0.10	4.85	0.01	0.01	0.00	0.05	0.01	0.01	0.00	0.06
20	100	0.60	0.11	0.27	1.12	0.30	0.05	0.15	0.55	0.08	0.01	0.04	0.14	1.02	0.33	0.22	2.73	1.04	0.33	0.24	2.76	0.02	0.03	0.00	0.13	0.03	0.05	0.00	0.13
	200	0.42	0.08	0.16	0.81	0.22	0.04	0.08	0.39	0.06	0.01	0.03	0.11	1.02	0.33	0.18	2.06	1.05	0.34	0.18	3.10	0.02	0.02	0.00	0.10	0.02	0.02	0.00	0.10
	300	0.33	0.06	0.14	0.65	0.18	0.03	0.08	0.33	0.05	0.01	0.02	0.08	1.01	0.33	0.19	2.72	1.05	0.34	0.20	2.82	0.01	0.02	0.00	0.08	0.02	0.02	0.00	0.08
	400	0.29	0.05	0.14	0.53	0.15	0.03	0.07	0.29	0.04	0.01	0.02	0.08	1.00	0.33	0.20	2.84	1.05	0.34	0.20	3.06	0.01	0.01	0.00	0.07	0.01	0.02	0.00	0.07
	500	0.26	0.05	0.11	0.48	0.14	0.02	0.06	0.26	0.04	0.01	0.02	0.06	1.00	0.33	0.18	2.46	1.05	0.34	0.18	2.52	0.01	0.01	0.00	0.05	0.01	0.01	0.00	0.06
	600	0.23	0.04	0.10	0.44	0.12	0.02	0.05	0.23	0.04	0.01	0.02	0.06	1.00	0.33	0.21	2.61	1.05	0.34	0.21	2.72	0.01	0.01	0.00	0.05	0.01	0.01	0.00	0.05
	700	0.22	0.04	0.09	0.39	0.11	0.02	0.05	0.20	0.03	0.01	0.01	0.06	1.00	0.33	0.19	3.00	1.05	0.34	0.20	3.17	0.01	0.01	0.00	0.05	0.01	0.01	0.00	0.06
	800	0.20	0.04	0.05	0.36	0.11	0.02	0.03	0.18	0.03	0.01	0.01	0.05	1.01	0.33	0.07	2.87	1.06	0.34	0.07	2.94	0.01	0.01	0.00	0.05	0.01	0.01	0.00	0.05
	900	0.19	0.03	0.07	0.31	0.10	0.02	0.04	0.17	0.03	0.01	0.01	0.05	1.00	0.32	0.13	2.76	1.05	0.34	0.14	2.87	0.01	0.01	0.00	0.04	0.01	0.01	0.00	0.05
	1000	0.18	0.03	0.06	0.32	0.10	0.02	0.03	0.17	0.03	0.00	0.01	0.05	1.00	0.32	0.12	2.64	1.05	0.34	0.13	2.91	0.01	0.01	0.00	0.04	0.01	0.01	0.00	0.06
30	100	0.60	0.09	0.32	1.07	0.31	0.04	0.18	0.51	0.08	0.01	0.03	0.12	1.02	0.27	0.30	2.38	1.02	0.26	0.31	2.31	0.02	0.03	0.00	0.12	0.02	0.03	0.00	0.11
	200	0.41	0.06	0.22	0.72	0.22	0.03	0.12	0.36	0.06	0.01	0.03	0.10	1.01	0.27	0.29	2.57	1.03	0.27	0.31	2.58	0.02	0.02	0.00	0.09	0.02	0.02	0.00	0.09
	300	0.33	0.05	0.18	0.55	0.18	0.03	0.10	0.29	0.05	0.01	0.03	0.08	1.02	0.26	0.30	2.51	1.04	0.27	0.31	2.52	0.01	0.01	0.00	0.07	0.01	0.02	0.00	0.07
	400	0.29	0.04	0.17	0.46	0.16	0.02	0.09	0.25	0.04	0.01	0.03	0.07	1.01	0.26	0.35	2.35	1.04	0.27	0.36	2.37	0.01	0.01	0.00	0.06	0.01	0.01	0.00	0.06
	500	0.26	0.04	0.14	0.41	0.14	0.02	0.07	0.22	0.04	0.01	0.02	0.06	1.01	0.26	0.30	2.52	1.04	0.27	0.30	2.55	0.01	0.01	0.00	0.06	0.01	0.01	0.00	0.06
	600	0.23	0.03	0.12	0.37	0.13	0.02	0.07	0.21	0.04	0.00	0.02	0.06	1.00	0.26	0.28	2.24	1.03	0.27	0.29	2.31	0.01	0.01	0.00	0.05	0.01	0.01	0.00	0.05
	700	0.22	0.03	0.12	0.34	0.12	0.02	0.06	0.18	0.03	0.00	0.02	0.05	1.01	0.27	0.33	2.32	1.04	0.27	0.33	2.36	0.01	0.01	0.00	0.04	0.01	0.01	0.00	0.04
	800	0.20	0.03	0.11	0.31	0.11	0.02	0.06	0.17	0.03	0.00	0.02	0.05	1.01	0.26	0.31	2.34	1.04	0.27	0.33	2.45	0.01	0.01	0.00	0.04	0.01	0.01	0.00	0.04
	900	0.19	0.03	0.10	0.30	0.10	0.01	0.05	0.16	0.03	0.00	0.02	0.04	1.01	0.27	0.29	2.31	1.04	0.27	0.30	2.35	0.01	0.01	0.00	0.04	0.01	0.01	0.00	0.04
	1000	0.18	0.02	0.09	0.29	0.10	0.01	0.05	0.16	0.03	0.00	0.02	0.04	1.00	0.26	0.28	2.29	1.03	0.27	0.27	2.32	0.01	0.01	0.00	0.04	0.01	0.01	0.00	0.04
40	100	0.60	0.08	0.31	1.03	0.32	0.04	0.17	0.50	0.08	0.01	0.03	0.11	1.02	0.23	0.29	2.32	1.02	0.22	0.29	2.11	0.02	0.02	0.00	0.12	0.02	0.02	0.00	0.11
	200	0.41	0.05	0.25	0.72	0.22	0.03	0.13	0.37	0.06	0.01	0.04	0.09	1.02	0.23	0.38	2.26	1.03	0.23	0.39	2.26	0.01	0.02	0.00	0.08	0.02	0.02	0.00	0.08
	300	0.33	0.04	0.20	0.56	0.18	0.02	0.11	0.29	0.05	0.01	0.03	0.07	1.01	0.23	0.37	2.10	1.03	0.23	0.37	2.10	0.01	0.01	0.00	0.06	0.01	0.01	0.00	0.06
	400	0.29	0.04	0.17	0.45	0.16	0.02	0.09	0.24	0.04	0.01	0.03	0.06	1.01	0.23	0.34	2.11	1.02	0.23	0.35	2.21	0.01	0.01	0.00	0.05	0.01	0.01	0.00	0.06
	500	0.26	0.03	0.15	0.38	0.14	0.02	0.08	0.21	0.04	0.00	0.02	0.06	1.01	0.23	0.38	2.03	1.03	0.23	0.39	2.13	0.01	0.01	0.00	0.05	0.01	0.01	0.00	0.05
	600	0.23	0.03	0.14	0.35	0.13	0.02	0.08	0.19	0.04	0.00	0.02	0.05	1.00	0.23	0.36	2.10	1.02	0.23	0.37	2.10	0.01	0.01	0.00	0.04	0.01	0.01	0.00	0.04
	700	0.21	0.03	0.13	0.36	0.12	0.01	0.07	0.20	0.03	0.00	0.02	0.06	1.00	0.23	0.36	2.33	1.02	0.23	0.37	2.40	0.01	0.01	0.00	0.04	0.01	0.01	0.00	0.04
	800	0.20	0.02	0.10	0.30	0.11	0.01	0.06	0.17	0.03	0.00	0.02	0.05	1.00	0.23	0.29	2.09	1.02	0.23	0.29	2.14	0.01	0.01	0.00	0.04	0.01	0.01	0.00	0.04
	900	0.19	0.02	0.11	0.29	0.10	0.01	0.06	0.16	0.03	0.00	0.02	0.05	1.00	0.22	0.33	2.14	1.03	0.23	0.34	2.16	0.01	0.01	0.00	0.04	0.01	0.01	0.00	0.04
	1000	0.18	0.02	0.10	0.28	0.10	0.01	0.06	0.15	0.03	0.00	0.02	0.04	1.00	0.23	0.32	2.12	1.03	0.23	0.32	2.14	0.01	0.01	0.00	0.03	0.01	0.01	0.00	0.05
50	100	0.60	0.07	0.37	0.87	0.32	0.04	0.20	0.46	0.08	0.01	0.05	0.11	1.03	0.21	0.42	1.91	1.01	0.20	0.44	1.83	0.02	0.02	0.00	0.10	0.02	0.02	0.00	0.09
	200	0.41	0.05	0.23	0.66	0.23	0.02	0.13	0.35	0.06	0.01	0.03	0.08	1.02	0.21	0.38	2.01	1.02	0.20	0.38	2.05	0.01	0.02	0.00	0.07	0.01	0.02	0.00	0.07
	300	0.33	0.04	0.19	0.51	0.18	0.02	0.11	0.27	0.05	0.01	0.03	0.07	1.01	0.21	0.39	2.04	1.02	0.21	0.40	2.04	0.01	0.01	0.00	0.06	0.01	0.01	0.00	0.06
	400	0.29	0.03	0.19	0.47	0.16	0.02	0.10	0.25	0.04	0.00	0.03	0.06	1.01	0.21	0.41	2.13	1.02	0.21	0.42	2.10	0.01	0.01	0.00	0.05	0.01	0.01	0.00	0.05
	500	0.26	0.03	0.15	0.38	0.14	0.02	0.08	0.21	0.04	0.00	0.02	0.06	1.01	0.20	0.38	1.99	1.02	0.20	0.39	1.99	0.01	0.01	0.00	0.04	0.01	0.01	0.00	0.04
	600	0.23	0.02	0.14	0.34	0.13	0.01	0.08	0.19	0.04	0.00	0.02	0.05	1.00	0.20	0.39	1.98	1.02	0.20	0.39	2.06	0.01	0.01	0.00	0.04	0.01	0.01	0.00	0.04
	700	0.22	0.02	0.13	0.31	0.12	0.01	0.07	0.17	0.03	0.00	0.02	0.05	1.01	0.21	0.39	2.01	1.02	0.21	0.40	2.05	0.01	0.01	0.00	0.04	0.01	0.01	0.00	0.04
	800	0.20	0.02	0.13	0.29	0.11	0.01	0.07	0.16	0.03	0.00	0.02	0.04	1.00	0.20	0.40	1.87	1.02	0.20	0.41	1.89	0.01	0.01	0.00	0.03	0.01	0.01	0.00	0.03
	900	0.19	0.02	0.11	0.28	0.11	0.01	0.06	0.16	0.03	0.00	0.02	0.04	1.00	0.20	0.41	1.93	1.02	0.21	0.41	1.94	0.01	0.01	0.00	0.03	0.01	0.01	0.00	0.03
	1000	0.18	0.02	0.11	0.27	0.10	0.01	0.06	0.15	0.03	0.00	0.02	0.04	1.01	0.21	0.38	2.06	1.02	0.21	0.38	2.09	0.01	0.01	0.00	0.03	0.01	0.01	0.00	0.04

References

Andersen, E.B. (1973). A goodness of fit test for the Rasch model.*Psychometrika, 38*, 123–140.

Baghaei, P., Yanagida, T., & Heene, M. (2017). Development of a descriptive fit statistic for the Rasch model. *North American Journal of Psychology*, *19*, 155-168.

Mair, P. (2015). Package 'eRM'. https://cran.r-project.org/web/packages/eRm/eRm.pdf

Steiger, J. H. and Lind, J. (1980) Statistically-based tests for the number of common factors. Paper presented at the Annual Spring Meeting of the Psychometric Society, Iowa City.

Table 2. . Simulation Results of the Alternative Hypothesis Conditions with 8 DIF Items

			Andersen χ^2/df				S Statistic χ^2/df			Andersen RMSEA				S Statistic RMSEA			
k	*N*	*M*	*SD*	*Min*	*Max*	*M*	*SD*	*Min*	*Max*	*M*	*SD*	*Min*	*Max*	*M*	*SD*	*Min*	*Max*
10	100	1.86	0.77	0.17	6.49	2.08	0.88	0.17	7.18	0.08	0.05	0.00	0.24	0.09	0.05	0.00	0.25
	200	2.73	0.99	0.17	7.76	3.19	1.17	0.19	9.15	0.09	0.03	0.00	0.18	0.10	0.03	0.00	0.20
	300	3.54	1.16	0.39	8.81	4.23	1.41	0.42	10.95	0.09	0.02	0.00	0.16	0.10	0.02	0.00	0.18
	400	4.38	1.31	0.99	10.57	5.29	1.60	1.13	12.54	0.09	0.02	0.00	0.15	0.10	0.02	0.02	0.17
	500	5.21	1.45	1.09	11.28	6.32	1.78	1.19	14.28	0.09	0.02	0.01	0.14	0.10	0.02	0.02	0.16
	600	6.08	1.57	1.80	14.71	7.41	1.95	2.04	17.91	0.09	0.01	0.04	0.15	0.10	0.02	0.04	0.17
	700	6.90	1.67	1.78	14.57	8.45	2.08	2.13	18.75	0.09	0.01	0.03	0.14	0.10	0.01	0.04	0.16
	800	7.76	1.78	2.69	16.90	9.54	2.22	3.23	21.47	0.09	0.01	0.05	0.14	0.10	0.01	0.05	0.16
	900	8.62	1.90	3.03	18.48	10.61	2.37	3.86	22.99	0.09	0.01	0.05	0.14	0.10	0.01	0.06	0.16
	1000	9.45	1.98	3.44	18.28	11.66	2.48	4.18	23.20	0.09	0.01	0.05	0.13	0.10	0.01	0.06	0.16
20	100	1.41	0.45	0.35	3.77	1.43	0.45	0.35	3.76	0.06	0.04	0.00	0.17	0.06	0.04	0.00	0.17
	200	1.66	0.50	0.37	4.27	1.70	0.50	0.38	4.20	0.05	0.02	0.00	0.13	0.05	0.02	0.00	0.13
	300	1.97	0.55	0.37	4.63	2.03	0.57	0.37	4.76	0.05	0.02	0.00	0.11	0.06	0.02	0.00	0.11
	400	2.28	0.61	0.48	5.65	2.35	0.62	0.50	5.90	0.05	0.01	0.00	0.11	0.06	0.01	0.00	0.11
	500	2.59	0.66	0.84	5.57	2.67	0.68	0.86	5.62	0.06	0.01	0.00	0.10	0.06	0.01	0.00	0.10
	600	2.91	0.71	0.76	6.16	3.00	0.73	0.78	6.23	0.06	0.01	0.00	0.09	0.06	0.01	0.00	0.09
	700	3.23	0.75	1.06	6.83	3.33	0.77	1.08	6.98	0.06	0.01	0.01	0.09	0.06	0.01	0.01	0.09
	800	3.55	0.80	1.11	8.02	3.65	0.82	1.13	8.50	0.06	0.01	0.01	0.09	0.06	0.01	0.01	0.10
	900	3.88	0.84	1.25	7.82	3.99	0.86	1.28	8.11	0.06	0.01	0.02	0.09	0.06	0.01	0.02	0.09
	1000	4.18	0.88	1.32	7.84	4.30	0.90	1.35	7.93	0.06	0.01	0.02	0.08	0.06	0.01	0.02	0.10
30	100	1.38	0.34	0.40	3.16	1.38	0.34	0.41	3.14	0.06	0.03	0.00	0.15	0.06	0.03	0.00	0.15
	200	1.72	0.41	0.55	4.12	1.77	0.42	0.57	4.09	0.06	0.02	0.00	0.13	0.06	0.02	0.00	0.12
	300	2.07	0.47	0.71	4.27	2.13	0.48	0.72	4.43	0.06	0.01	0.00	0.10	0.06	0.01	0.00	0.11
	400	2.42	0.51	0.79	4.84	2.51	0.53	0.81	4.96	0.06	0.01	0.00	0.10	0.06	0.01	0.00	0.10
	500	2.79	0.57	1.14	5.56	2.90	0.59	1.21	5.58	0.06	0.01	0.02	0.10	0.06	0.01	0.02	0.10
	600	3.12	0.60	1.19	6.22	3.24	0.62	1.21	6.49	0.06	0.01	0.02	0.09	0.06	0.01	0.02	0.10
	700	3.48	0.64	1.50	6.34	3.62	0.67	1.56	6.87	0.06	0.01	0.03	0.09	0.06	0.01	0.03	0.09
	800	3.82	0.66	1.77	6.37	3.98	0.69	1.86	6.58	0.06	0.01	0.03	0.08	0.06	0.01	0.03	0.08
	900	4.18	0.71	1.86	7.48	4.35	0.74	1.94	7.73	0.06	0.01	0.03	0.08	0.06	0.01	0.03	0.09
	1000	4.53	0.74	1.95	7.81	4.72	0.77	2.01	8.05	0.06	0.01	0.03	0.08	0.06	0.01	0.03	0.10
40	100	1.30	0.29	0.47	2.51	1.29	0.28	0.45	2.51	0.05	0.03	0.00	0.12	0.05	0.03	0.00	0.12
	200	1.33	0.29	0.55	2.80	1.34	0.29	0.56	2.89	0.04	0.02	0.00	0.10	0.04	0.02	0.00	0.10
	300	1.47	0.32	0.60	3.12	1.49	0.32	0.60	3.08	0.04	0.02	0.00	0.08	0.04	0.01	0.00	0.08
	400	1.63	0.34	0.73	3.28	1.65	0.34	0.75	3.29	0.04	0.01	0.00	0.08	0.04	0.01	0.00	0.08
	500	1.78	0.36	0.64	3.25	1.81	0.37	0.65	3.31	0.04	0.01	0.00	0.07	0.04	0.01	0.00	0.07
	600	1.94	0.39	0.83	3.69	1.97	0.39	0.84	3.69	0.04	0.01	0.00	0.07	0.04	0.01	0.00	0.07
	700	2.09	0.41	0.77	4.02	2.12	0.41	0.79	4.01	0.04	0.01	0.00	0.07	0.04	0.01	0.00	0.07
	800	2.24	0.42	0.89	4.22	2.28	0.42	0.91	4.24	0.04	0.01	0.00	0.06	0.04	0.01	0.00	0.06
	900	2.41	0.44	1.06	4.31	2.44	0.44	1.09	4.32	0.04	0.01	0.01	0.06	0.04	0.01	0.01	0.06
	1000	2.56	0.47	1.00	4.70	2.60	0.47	1.03	4.75	0.04	0.01	0.00	0.06	0.04	0.01	0.01	0.07
50	100	1.15	0.23	0.48	2.23	1.14	0.22	0.48	2.21	0.03	0.03	0.00	0.11	0.03	0.03	0.00	0.11
	200	1.27	0.25	0.56	2.63	1.27	0.25	0.56	2.65	0.03	0.02	0.00	0.09	0.03	0.02	0.00	0.09
	300	1.39	0.27	0.61	2.62	1.40	0.27	0.61	2.61	0.03	0.01	0.00	0.07	0.03	0.01	0.00	0.07
	400	1.51	0.29	0.68	3.08	1.52	0.29	0.68	3.08	0.03	0.01	0.00	0.07	0.03	0.01	0.00	0.07
	500	1.63	0.30	0.67	2.97	1.64	0.30	0.68	3.00	0.03	0.01	0.00	0.06	0.03	0.01	0.00	0.06
	600	1.75	0.32	0.78	3.30	1.77	0.32	0.79	3.29	0.03	0.01	0.00	0.06	0.04	0.01	0.00	0.06
	700	1.89	0.34	0.78	3.31	1.91	0.34	0.79	3.31	0.03	0.01	0.00	0.06	0.04	0.01	0.00	0.06
	800	2.00	0.35	0.91	3.42	2.02	0.35	0.92	3.45	0.03	0.01	0.00	0.06	0.04	0.01	0.00	0.06
	900	2.13	0.36	0.96	3.84	2.15	0.37	0.99	3.84	0.03	0.01	0.00	0.06	0.04	0.01	0.00	0.06
	1000	2.26	0.37	0.97	3.99	2.28	0.37	0.97	3.99	0.04	0.01	0.00	0.05	0.04	0.01	0.00	0.06

Table 3
Modified cut-off values for Andersen's χ^2/df value for different test lengths

k	Andersen's χ^2/df
10	<1.90
20	<1.60
30	<1.50
40	<1.45
50	<1.40

Rasch-related Coming Events

July 31-Aug. 3, 2017, Mon.-Thurs. Joint IMEKO TC1-TC7-TC13 Symposium, Rio de Janeiro, Brazil, www.imeko-tc7-rio.org.br

Aug. 7-9, 2017, Mon-Wed. In-person workshop and research colloquium: Effect size of family and school indexes in writing competence using TERCE data (C. Pardo, A. Atorressi, Winsteps), Bariloche Argentina.

Aug. 7-9, 2017, Mon-Wed. PROMS 2017: Pacific Rim Objective Measurement Symposium, Sabah, Borneo, Malaysia, www.proms.promsociety.org/2017/

Aug. 10, 2017, Thurs. In-person Winsteps Training Workshop (M. Linacre, Winsteps), Sydney, Australia. www.winsteps.com/sydneyws.htm

An Online Multidimensional Computerized Adaptive Testing (MCAT) Module Using APP

The development of item response theory (IRT) in conjunction with the advances in computer technology has made computerized adaptive testing (CAT) feasible and applicable (Wang & Chen, 2004). Many unidimensional CATs (UCAT) have been discussed in literature (Chien & Djaja, 2015; Halkitis, 1993/1996; Linacre, 2006; 1998; Lunz & O'Neill, 1998; Raîche, Blais,& Riopel, 2006).

Furthermore, the Multidimensional Random Coefficients Multinomial Logit Model (MRCMLM) has been proposed to capture the complexity of modern assessment (Adams, Wilson, & Wang, 1997; Wang & Chen, 2004). The merger of these two (i.e., MRCMLM and CAT) is called multidimensional computerized adaptive testing (MCAT; Segall, 1996). Thus, we can consider using MCAT to simultaneously estimate person measures for an inventory consisting of multiple subscales.

In tradition, we do CAT for each subscale separately, not like MCAT on all subscales jointly. In general, MCAT is more efficient than separate unidimensional CAT in terms of reducing test length (Wang, 2010).

As with all forms of web-based technology, advances in mobile communication technology are rapidly increasing. So far, however, no online MCAT assessment has been published in journals.

An online MCAT using maximum likelihood (ML) estimation with the Newton-Raphson iteration method was programmed by authors to administer the 3-domain Maslach Burnout Inventory (Lee, Chien, Yen, 2013) here. By scanning a QR-code, the first item randomly selected appears on the smartphone (Figure 1). Person domain scores can be estimated via MCAT (Figure 2). In MCAT process, the measurement of standard error (MSE) for each subscale decreased when the number of the items increased (Figure 3). A snapshot of an unexpected response with an asterisk (*) with $|Z| \geq 2.0$ are shown on a smart phone (Figure 4). The link to the MCAT video demonstration is https://youtube/Wc_9Tov-__w for interested readers.

Tsair-Wei Chien, *Chi Mei Medical Center, Taiwan*
Wen-Chung Wang, *The Hong Kong Institute of Education, Hong Kong, China*

References

Adams, R. J., Wilson, M. R., & Wang, W.-C. (1997). The multidimensional random coefficients multinomial logit model. Applied Psychological Measurement, 21, 1-23.

Chien, T.-W. ,Djaja,N., Janda, M., Olsen, C., Whiteman,D. (2015). Mobile Computer-Adaptive Tests (CAT): Skin Cancer Risk Scale and Standard Errors. Rasch Measurement Transactions, 29:2, 1519-1521.

Halkitis, P.N. (1993). Computer-adaptive testing algorithm. Rasch Measurement Transactions, 6:4, 254-255.

Halkitis, P.N. (1996). Computer-adaptive testing: CAT with a limited item bank. Rasch Measurement Transactions, 9:4,471.

Lee, H.F., Chien, T.W., Yen, M. (2013). Examining factor structure of Maslach Burnout Inventory among nurses in Taiwan. Journal of Nursing Management, 21:4, 648-656.

Linacre, J.M. (2006).Computer-Adaptive Tests (CAT), Standard Errors and Stopping Rules. Rasch Measurement Transactions, 20:2, 1062.

Lunz, M.E., O'Neill, T.R. (1998). Computer-adaptive testing: CAT: Taking items twice? Rasch Measurement Transactions, 12:3, 656-657.

Linacre, J.M. (1998).Computer-adaptive testing: CAT: Maximum possible ability. Rasch Measurement Transactions, 12:3, 657-658.

Raîche, G., Blais, J.-G., Riopel, M. A. (2006). SAS Solution to Simulate a Rasch Computerized Adaptive Test. Rasch Measurement Transactions, 20:2, 1061.

Segall, D. O. (1996). Multidimensional adaptive testing. Psychometrika, 61, 331-354.

Wang, W. -C., Chen, P. -H. (2004).Implementation and measurement efficiency of multidimensional computerized adaptive testing. Applied Psychological Measurement, 28:5,295–316.

Wang, W. -C. (2010). Recent Developments in Rasch Measurement. Hong Kong: The Hong Kong Institute of Education Press; 2010.

Figure 1. By scanning a QR-code, the first item randomly selected appears on the smartphone.

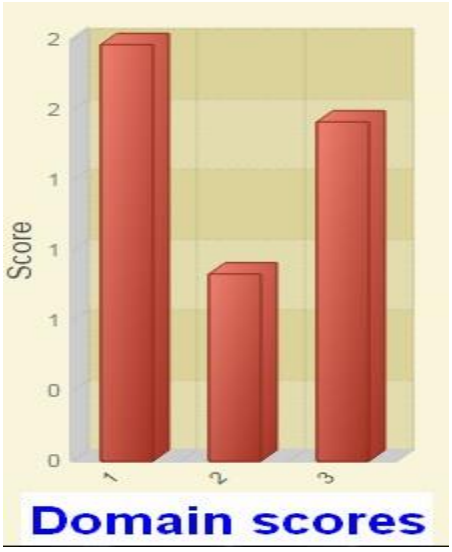

Figure 2. Person domain burnout scores are estimated via MCAT

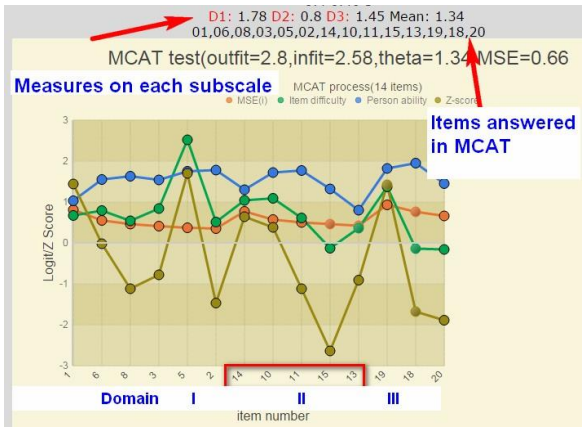

Figure 3. Snapshots of MCAT is shown on a smart phone.

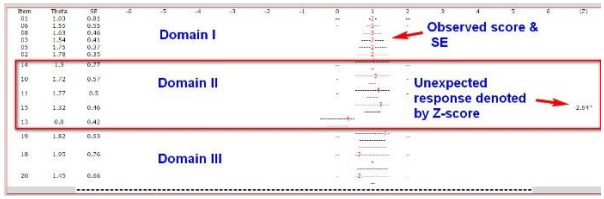

Figure 4. Snapshots of an unexpected response with an asterisk (*) when $|Z| \geq 2.0$ shown on a smart phone.

International Metrology Conference in Paris, 19-21 September

A number of measurement presentations of interest will be given at the upcoming 18th International Metrology Conference in Paris, France, this September. The full program is posted at http://cim2017.com/files/pdf/CIM2017-Programme-GB-Web_2.pdf.

Rasch's probabilistic models for measurement offer a basis for metrological traceability to new unit standards in education, health care, and other fields (Fisher, 2009; Fisher & Stenner, 2016; Mari, Maul, Torres Irribara, & Wilson, 2016; Mari & Wilson, 2014, 2015; Pendrill, 2014; Pendrill & Fisher, 2013, 2015; Wilson, 2013). Several CIM 2017 presentations will expand on this theme.

Leslie Pendrill, past chair of the European Association of National Metrology Institutes, will speak on person-centered health care quality at

15:50 on Wednesday 20 September in session S10.

A poster on psychometric metrology by William Fisher, and a poster by Matt Barney and Fisher on a psychometric metrology approach to AI data analytics, will be presented 13:45 to 15:15 on Wednesday 20 September.

That same poster session also includes work on emotion measures by R. Taymanov and K. Sapoznikhova from the Mendeleyev Institute in St. Petersburg; who, along with Barney, Pendrill, Fisher, and Jack Stenner, participated in the IMEKO Joint Symposium in Berkeley last year.

Fisher and Stenner have a podium presentation on uncertainty in metrology and psychometrics in CIM session S12 at 10:40 AM on Thursday 21 September.

Also of note is a poster session on metrology education on Tuesday 19 September. Conversations on integrated assessment and instruction with the authors of these presentations could open up productive new collaborations.

Note from SIG Chair

Greetings Rasch Enthusiasts!

I'd like to take a moment to remind you of the upcoming AERA Annual Meeting proposal deadline: July 24, 2017, at 11:59 PM Pacific Time. We look forward to your proposals to the Rasch SIG as we would love to have you present in one of our sessions. Remember that we are graduate-student friendly. Encourage those students that you are working with in an advising, teaching or mentoring capacity (including summer internship or post-doc experiences) to submit their work. Please give a round of thanks in advance to our 2018 Rasch SIG Program Co-Chairs, Liru Zhang and Eli Jones, and to those of you who volunteered to be proposal reviewers.

Please join me in congratulating the recipient of the 2017 Georg William Rasch Early Career Publication Award, Adrienne Walker. Her nomination noted the contribution of her 2016 *Journal of Applied Measurement* paper, "Exploring Aberrant Responses Using Person Fit and Person Response Functions". The picture below shows Adrienne (left) and Leigh (Rasch SIG Chair) posing with the award, which was announced during the Rasch SIG Business Meeting.

In addition to the presentation of the award to Adrienne Walker and Stefanie Wind's presentation at the Rasch SIG Business Meeting, the attendees discussed how to remain relevant and visible within AERA and to the outside world. Suggestions for improving the visibility of the SIG and Rasch Measurement and adding to what we offer to members when they attend AERA included the possibility of proposing half-day workshops through AERA on various Rasch models and data analysis. Discussions also centered around working more closely with 2018 IOMW organizers, Andrew Maul and Ronli Diakow. Several key financial initiatives were discussed. Pricing for the current hosting company for the Rasch SIG website (raschsig.org) has been renegotiated, resulting in significant cost savings. We are working with AERA management to evaluate using their SIG website hosting feature, a service already included in the yearly management fees that we pay to the SIG. Decreasing the current amount of the honorariums that go with the Georg William Rasch Early Career Publication Award and the Benjamin

Drake Wright Award was discussed, with agreement that these honorariums should be reduced to make them more aligned with the amount that IOMW and other SIGs award. Combining with other SIGs, such as the Survey Research SIG or Assessment in Higher Education SIG, to have an off-site social during the Annual Meeting instead of part of the Business Meeting was also suggested. Additionally, if your institution or company might be interested in partially sponsoring the awards or socials, please let me know.

As always, I would like to extend the opportunity for Rasch Measurement SIG members to reach out to me with questions, concerns or suggestions regarding the SIG. I look forward to hearing from you.

Sincerely,

Leigh M. Harrell-Williams
Rasch Measurement SIG Chair

Profiles in Rasch Measurement

My name is Sarah Thomas and I recently received my PhD in quantitative psychology from the University of Virginia. I first became interested in measurement while taking a psychology course as an undergraduate. The idea that a person could be represented as a series of scores fascinated me. These interests led me to pursue a career in psychology, specializing in applications of measurement models to common problems and, more recently, the detection of cheating on high-stakes standardized tests.

The contamination of measurement data through cheating is a shockingly common problem that has led to an explosion of research in this area. My introduction to the field of cheating detection occurred in 2013 when I agreed to participate as a team member in the "Test Fraud Detection Challenge." For this challenge, I was part of a team that analyzed data for evidence of test fraud and presented our findings at the Conference on Test Security. Our objectives were to identify items that were leaked online and examinees who may have accessed those items. We decided to investigate Rasch model estimates, Classical Test Theory statistics, and cluster analyses. After the conference, we were given information on which items were discovered in the online leak. We discovered that we had correctly classified 64 of the 65 items. I was hooked. The detection of cheating was a perfect place for me to apply my knowledge of psychometrics to an important, real-world problem.

My recent work combines techniques from machine learning with estimates from Rasch and Item Response Theory models to detect leaked test items. This project uses data from a testing company in which some items were discovered to be compromised in screenshots and notes. However, the quality of these suspected item categories is unknown, so I designed, and am currently running, an experiment to mimic an item leak in the lab. The greatest benefit of this experiment is that the correct status of items and examinees will be known, unlike in most cases of test fraud, allowing a direct evaluation of the accuracy of various statistical methods for identifying test fraud to be assessed. I am particularly interested in assessing the usefulness of Rasch model estimates in detecting test fraud, as situations of test fraud represent a potential violation of the assumptions and the misfit of the data to the model may be indicative.

I think it is imperative that we, as a field, continue to focus on achieving good, objective measurement and informing others about the Rasch model. I also think that increased attention to test security issues, particularly for measures that are associated with high-stakes decisions, should be a focus as we move forward.