Winter 2017, VOL. 31, NO. 3; ISSN 1051-0796

RMT

RASCH MEASUREMENT TRANSACTIONS

Notes:

Transactions of the Rasch Measurement SIG
American Educational Research Association

What Overall Indices can Successfully Check for Rasch Model Fit?

There are many different fit indices, some of which check for overall model fit (Maydeu-Olivares, 2013), some check for specific violations of the model (Hattie, 1985; McDonald and Mok, 1995), and some are for checking the violation of the assumption of local independence between items (Chen and Thissen, 1997; Liu & Maydeu-Olivares, 2013).

Wu et al. (2017) stress that items in a test will show good fit (i.e., fit mean-squares around 1) if the items have similar discrimination, even if the discrimination power is poor. That is, if all items are equally “bad”, the items will still show good fit, because they have equal discrimination. Consequently, when there is no misfitting item, we might conclude that the response data fit the Rasch model, we cannot conclude that we have the best test. The test reliability may still be low.

In this paper, a simulation was conducted (Linacre, 2007) to generate 13 datasets with an equal discrimination parameter from 0 to 2.0 across 20 items and 200 persons. Another two included random responses and two domains interlaced with an equal item length and a correlation of 0 were combined for comparison using Winsteps.

The results in Table 1 show that all Infit MNSQ are acceptable (within 0.5 and 1.5). Four overall indices in comparison are shown in Figure 1. We can see only Dimension Coefficient (DC) (Chien, 2012) presents congruent with Rasch feature (i.e, the higher value is around 1.0 for the discrimination). Average variance extracted (AVE $= \frac{\sum \lambda_i^2}{(\sum \lambda_i^2)+(\sum \varepsilon_i)}$, where λ denotes factor loading) fits into 2-PL IRT model (i.e., the higher discrimination earns the greater value). If the reliability criterion is set at 0.7, we can see that both (i.e., Cronbach’s alpha and Rasch real person separation reliability) indices are merely validated in term of the discrimination value less than 0.4.

Table 1. The result of the study on individual and overall indices.

	Discrimination parameter											2 factors	
	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0	interlaced	Random
Cronbach Rel.	0.06	0.58	0.76	0.86	0.88	0.89	0.9	0.93	0.91	0.91	0.92	0.80	0.10
Rasch Rel.	0.05	0.57	0.74	0.86	0.89	0.91	0.92	0.85	0.94	0.94	0.95	0.78	0.06
Validity(AVE)	0.00	0.12	0.18	0.27	0.32	0.33	0.35	0.38	0.39	0.39	0.40	0.24	0.08
DC	**0.54**	**0.54**	0.70	0.75	0.73	**0.78**	**0.68**	**0.67**	**0.66**	**0.62**	**0.59**	**0.23**	**0.48**
Infit MNSQ(L)	0.87	0.74	0.73	0.82	0.83	0.84	0.73	0.74	0.74	0.71	0.62	0.84	0.91
Infit MNSQ(H)	1.13	1.13	1.20	1.35	1.51	1.36	1.30	1.51	1.36	1.18	1.30	1.16	1.12
Outfit MNSQ(L)	0.87	0.68	0.80	0.79	0.81	0.79	0.71	0.17	0.71	0.63	0.22	0.85	0.90
Outfit MNSQ(U)	1.13	1.18	1.18	1.59	1.42	**2.48**	**2.53**	**1.45**	**3.13**	**3.99**	**1.78**	1.26	1.13
MEAN(Infit)	1.00	1.00	1.00	1.01	1.02	1.04	1.02	0.98	1.00	0.99	0.98	1.03	1.00
P.SD (Infit)	0.06	0.08	0.12	0.13	0.15	0.15	0.35	0.19	0.15	0.36	0.16	0.07	0.06
MEAN(Outfit)	1.00	1.00	1.00	1.05	1.03	1.07	1.05	1.01	1.11	1.06	0.84	1.00	1.00
P.SD (Outfit)	0.06	0.09	0.12	0.21	0.13	0.37	0.19	0.38	0.53	0.54	0.34	0.09	0.06
PTMEASURE(L)	0.06	0.28	0.30	0.3	0.28	0.22	0.26	0.29	0.24	0.2	0.28	0.28	0.14
PTMEASURE(U)	0.38	0.46	0.56	0.65	0.74	0.74	0.76	0.84	0.8	0.84	0.86	0.56	0.33
Log Likelihood(prob.)	0.48	0.49	0.42	0.42	0.44	0.40	0.42	0.57	0.6	0.61	0.63	0.99	0.49

Tennant & Pallant (2006) stated that the Rasch model fit statistics performed poorly where dimensions were interlaced and where the correlation between factors was ~ 0.7. Two contrast parts are separated by Rasch principle component analysis in Figure 2. As we cannot know in advance whether two interlaced dimensions may exist and how far for the departure is allowed for detecting unidimensionality, this analysis for DC should be undertaken as a matter of routine in future.

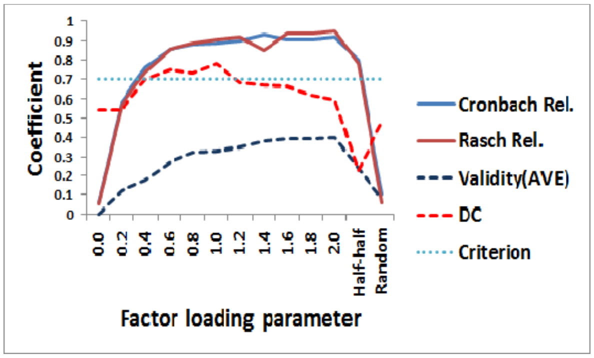


Figure 1. Four overall indices in comparison

Rasch Measurement Transactions
www.rasch.org/rmt
Editor: Kenneth Royal
Email submissions to: Editor \at/ Rasch.org
Copyright © 2017 Rasch Measurement SIG, AERA
Permission to copy is granted.
RMT Editor Emeritus: John M. Linacre
Rasch SIG Chair: Leigh Harrell-Williams
Secretary: Mikaela Raddatz
Treasurer: Matt Schulz
Program Chairs: Liru Zhang & Eli Jones

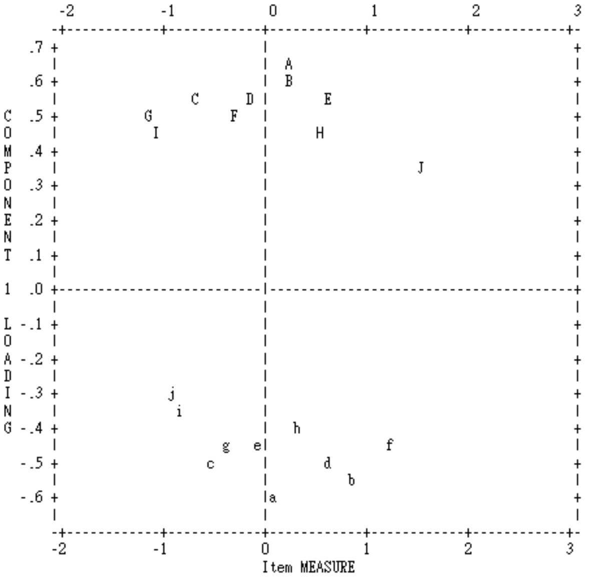


Figure 2. Two contrast parts are separated by Rasch principle component analysis in Table 23 from Winsteps

Tsair-Wei Chien
Chi-Mei Medical Center, Taiwan

References

Chen, W. H, & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. Journal of Educational and Behavioral Statistics, 22:265–289.

Chien, T. W. (2012). Cronbach's Alpha with the Dimension Coefficient to Jointly Assess a Scale's Quality. Rasch Measurement Transactions, 26:3, 1379.

Hattie, J. (1985). Methodology review: assessing unidimensionality of tests and items. Appl Psychol Meas 9:139–164.

Linacre, J. M. (2007). How to Simulate Rasch Data. Rasch Measurement Transactions 21:3, 1125.

Liu, Y., & Maydeu-Olivares, A. (2013). Local dependence diagnostics in IRT modeling of binary data. Educational & Psychological Measurement, 73:254–274.

Maydeu-Olivares, A. (2013). Goodness-of-fit assessment of item response theory models. Measurement 11:71–101.

Tennant, A., Pallant, J.F. (2006). Unidimensionality Matters! (A Tale of Two Smiths?). Rasch Measurement Transactions, 20:1, 1048-51.

McDonald, R., & Mok, M. M. (1995). Goodness of fit in item response models. Multivariate Behavioral Research, 30(1):23–40.

Wu, M.., Tam, H. P., & Jen, T. H. (2017). Educational Measurement for Applied Researchers. Springer Nature: Singapore.

Medical Students Fail Blood Pressure Measurement Challenge: Implications for Measurement

Rakotz and colleagues (2017) recently published a paper describing a blood pressure (BP) challenge presented to 159 medical students representing 37 states at the American Medical Association's House of Delegates Meeting in June 2015. The challenge consisted of correctly performing all 11 elements involved in a BP assessment using simulated patients. Alarmingly, only 1 of the 159 (0.63 %) medical students correctly performed all 11 elements.

According to professional guidelines (Bickley & Szilagyi, 2013; and Pickering et al, 2005), the 11 steps involved in a proper BP assessment include: 1) allowing the patient to rest for 5 minutes before taking the measurement; 2) ensuring patient's legs are uncrossed; 3) ensuring the patient's feet are flat on the floor; 4) ensuring the patient's arm is supported; 5) ensuring the sphygmomanometer's cuff size is correct; 6) properly positing cuff over bare arm; 7) no talking; 8) ensuring the patient does not use his/her cell phone during the reading; 9) taking BP measurements in both arms; 10) identifying the arm with the higher reading as being clinically more important; and 11) identifying the correct arm to use when performing future BP assessment (the one with the higher measurement).

All medical students involved in the study had confirmed that they had previously received training during medical school for measuring blood pressure. Further, because additional skills are necessary when using a manual sphygmomanometer, the authors of the study elected to provide all students with an automated device in order to remove students' ability to use the auscultatory method correctly from the testing process. The authors of the study reported the average number of elements correctly performed was 4.1 (no SD was reported).

While the results from this study likely will raise concern among the general public, scholars and practitioners of measurement may also find these results particularly troubling. There currently exists an enormous literature regarding blood pressure measurements. In fact, there are even academic journals devoted entirely to the study of blood pressure measurements (e.g., *Blood Pressure Monitoring*), and numerous medical journals devoted to the study of blood pressure (e.g., *Blood Pressure, Hypertension, Integrated Blood Pressure Control, Kidney & Blood Pressure Research, High Blood Pressure & Cardiovascular Prevention*, etc.) Further, a considerable body of literature also discusses the many BP instruments and methods available for collecting readings, and various statistical algorithms used to improve the precision of BP measurements. Yet, despite all the technological advances and sophisticated instruments available, these tools likely are of only limited utility until health care professionals utilize them correctly.

Inappropriate inferences about BP readings could result in unintended consequences that jeopardize a patient's health. In fact, research (Chobanian et al, 2003) indicates most human errors when measuring BP result in higher readings. Therefore, these costly errors may result in misclassifying prehypertension as stage 1 hypertension and beginning a treatment program that may be both unnecessary and harmful to a patient. This problem is further exacerbated when physicians put a patient on high blood pressure medication, as most physicians are extremely reluctant to take a patient off the medication, as the risks associated with stopping are extremely high. Further, continued usage of poor BP measurement techniques could result in patients whose blood pressure is under control to appear uncontrolled, thus escalating therapy that could further harm a patient. Until physicians can obtain accurate BP measurements, it is unlikely they can accurately differentiate those individuals who may need treatment from those that do not.

So, I wish to ask the measurement community how we might assist healthcare professionals (and those responsible for their training) to correctly practice proper blood pressure measurement techniques? What lessons from psychometrics can parlay into the everyday practice of healthcare providers? Contributing practical solutions to this problem could go a long way in directly improving patient health and outcomes.

Kenneth D. Royal
North Carolina State University

References

Pickering T, Hall JE, Appel LJ, et al. Recommendations for blood pressure measurement in humans and experimental animals part 1: blood pressure measurement in humans – a statement for professionals from the Subcommittee of Professional and Public Education of the American Heart Association Council on High Blood Pressure Research. *Hypertension*. 2005;45:142-161.

Bickley LS, Szilagyi PG. Beginning the physical examination: general survey, vital signs and pain. In: Bickley LS, Szilagyi PG, eds. *Bates' Guide to Physical Examination and History Taking*, 11th ed. Philadelphia, PA: Wolters Kluwer Health/ Lippincott Williams and Wilkins; 2013:119-134.

Chobanian AV, Bakris GL, Black HR, et al. Seventh report of the Joint National Committee on prevention, detection, evaluation and treatment of high blood pressure. *Hypertension*. 2003;42:1206-1252.

Rakotz MK, Townsend RR, Yang J, et al. Medical students and measuring blood pressure: Results from the American Medical Association Blood Pressure Check Challenge. *Journal of Clinical Hypertension*. 2017;19:614–619.

Journal of Applied Measurement
Vol. 18, No. 4, 2017

The Effects of Item Placement in the Young Schema Questionnaire - *Victoria K. Moir, Christopher W. Lee, and Ida Marais*

Stability of INFIT and OUTFIT Compared to Simulated Estimates in Applied Setting - *Kari J. Hodge and Grant B. Morgan*

Approximate Functional Relationship between IRT and CTT Item Discrimination Indices: A Simulation, Validation, and Practical Extension of Lord's (1980) Formula - *John T. Kulas, Jeffrey A. Smith, and Hui Xu*

Social Desirability Amongst Chinese Teachers - *Randall E. Schumacker and Cathy Ka Weng Hoi*

I'm scared to go to School! Capturing the Effects of Chronic Daily Fears on Students' Concept of Self - *Rense Lange, Cynthia Martínez-Garrido, and Alexandre Ventura*

Confidence to Perform in the Global Marketplace: Constructing and Validating a Survey Instrument for Community College Students - *Snejana Slantcheva-Durst and Mingyang Liu*

Measuring Anger Types among Malaysian Adolescents using the Rasch Model - *Ahmad Zamri Khairani, Nor Shafrin Ahmad, and Mohd Zahuri Khaira*

Richard Smith, Editor, www.jampress.org

IOMW 2018

The International Objective Measurement Workshop (IOMW) meeting will be held in New York City on April 10 and 11, with an additional half-day of workshops on April 12. IOMW presents an opportunity for scholars interested in the theory and practice of objective measurement in the human sciences to present research, learn about the most recent developments, and meet with colleagues who share similar interests in an intimate setting. For more information about the meeting, please see www.iomw.org.

An Application of “Comparison vs. Preferences”

David Andrich (2002) explains that it is important to distinguish between “comparison” (where an objective answer is expected), versus “preference” (where the subjective perception of the person is requested). Inspired by Andrich’s example for coffee and sugar, we proposed a questionnaire of 9 rating scale items to get the opinion of 14 judges on 4 different new bottled smoothies produced with natural ingredients (fruits and vegetables), developed by a small emerging company in Mexico. The set of responses was analyzed with Winsteps®.

The nine rating scale items are organized as follows:

Part 1- Preference. Rate your preference of the smoothie (5 items with 5 categories from (1) less favorite to (5) more favorite):

1) Color (col)
2) Taste (sab)
3) Texture (tex)
4) Sweetness (dul)
5) Natural flavor (nat)

Part 2-Comparison. Compare the product (4 items with 5 categories from (1) very poor to (5) very good):

6) Information contained on label (inf)
7) Packaging image or aspect (emp)
8) Bottle cap (tap)
9) Pricing (pre)

The names of the four smoothies are: Berry Blast (bb), Caribbean Fruit (ca), Taro 1 (t1) and Taro 2 (t2). The combination of product and attribute describes each item, for instance tex-bb corresponds to the texture of the berry blast smoothie.

All judges participated in the focus group, but they did not exchange opinions of impressions during the tasting session. At the end, they rated the 36 items (4 products x 9 items) and had some time to write feedback. Their comments are not reflected in this analysis.

Answers were analyzed using the Rasch model and measures were calculated on a single scale. The main results are: person mean = 0.75 logits, standard deviation = 0.39, separation = 1.5. The Wright map shows the importance of the correct meaning of comparison and preference as suggested by Andrich.

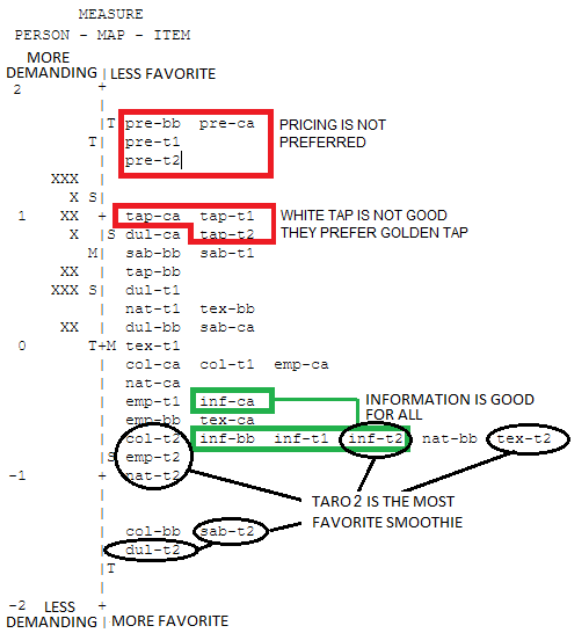
Considering “Preferences”, the Taro 2 smoothie (t2) is the more favorite in seven attributes (except the bottle’s tap and price), this product has the higher potential to be accepted among future consumers. Each product may be preferred or rejected in specific attributes: color is the best attribute of Berry Blast (col-bb = -1.5 logits) while its flavor or texture are not favorite (sab-bb = 0.64 logits, tex-bb = 0.58 logits).

Considering “Comparison”, the attribute of “Information” shows similar good measures (mean = -0.64 logits) followed by “Product package” that is good for all judges. Both attributes are considered better than the “Golden tap” or “Pricing” for all the products. In fact, “Pricing” is the most difficult attribute to be favorite with mean = 1.6 logits, that is higher than the measure of the most demanding judge (1.32 logits).

Misfit to the Rasch Model and low values on point-biserial correlation on certain items have an interesting interpretation because they imply that judges may be confused setting the attributes of a product (they cannot classify Taro 1 smoothie as a good or bad product) of an item (it is difficult to identify that the four products are made with natural ingredients). Confusing answers of consumers are certainly not expected for a suitable product!

The conclusions of this analysis are: Taro 1 will not pass to a commercial stage. All the smoothies must reinforce their flavor to improve the perception as a natural product. All the products will use the golden tap. Prices should be recalculated to become more acceptable.

The Rasch-Andrich rating scale model provides a suitable tool to identify the main attributes to improve the acceptance of a product in food industry.



```
      MEASURE
PERSON - MAP - ITEM
 2          +
            |
            |T pre-bb   pre-ca
           T|  pre-t1
            |  pre-t2
     XXX    |
       X S  |
 1    XX    +  tap-ca   tap-t1
       X    |S dul-ca   tap-t2
           M|  sab-bb   sab-t1
      XX    |  tap-bb
     XXX S  |  dul-t1
            |  nat-t1   tex-bb
      XX    |  dul-bb   sab-ca
 0         T+M tex-t1
            |  col-ca   col-t1   emp-ca
            |  nat-ca
            |  emp-t1   inf-ca
            |  emp-bb   tex-ca
            |  col-t2   inf-bb   inf-t1   inf-t2   nat-bb   tex-t2
            |S emp-t2
-1          +  nat-t2
            |
            |
            |  col-bb   sab-t2
            |  dul-t2
            |T
            |
-2          +
```

Agustin Tristan-Lopez & Agustin Tristan-Aldave, Instituto de Evaluacion e Ingenieria Avanzada, Mexico

Reference

Andrich, D. (2002). Comparisons vs. preferences. Rasch Measurement Transactions. 16:1 p.859

IRT is part of Rasch! (No, not really)

The Encyclopedia of Clinical Neuropsychology recently published a definition of Item Response Theory (see https://link.springer.com/content/pdf/10.1007/978-3-319-56782-2_1209-2.pdf).

The definition reads:

"… The mathematical model is known as Rasch modeling, and typically the three-parameter Rasch model is invoked. The three parameters are the guessing parameter, the likelihood that an individual will get an item correct simply by guessing; the discrimination parameter or the probability of a correct response at a given level of difficulty; and the difficulty parameter or the level of skill in the construct where an item has 0.5 discrimination."

Editor's Note:

A "thank you" is in order to Mike Linacre for discovering this gem. On first glance, it is good fun to note more than 50 years of psychometric research has turned upside down. However, this example underscores the need to seek information directly from the source, which in this case is the psychometrics literature. It will be interesting to see how many papers in the field of neuropsychology and other health-related disciplines confuse IRT and Rasch in future works.

PROMS 2018

The Pacific-Rim Objective Measurement Symposium (PROMS) will be held July 25-27, 2018 in Fudan University, Shanghai, China. Preconference workshops will be held from July 23-24. The theme of the meeting is *Applying Rasch Measurement in Language Assessment and across the Human Sciences.*

PROMS 2018 will feature three keynotes:
Tim McNamara, University of Melbourne
Yan Jin, Shanghai Jiao Tong University
George Engelhard, University of Georgia

The PROMS 2018 website is available at: https://proms.promsociety.org/2018/.

PROMS 2018 will also feature workshops on the application of the Rasch model in both English and Chinese. PROMS invites presentations on the theory and practice of applying the Rasch model across the human sciences, including business, education, health and psychology.

Following the usual PROMS practice, accepted papers will be allocated to presentation strands with similar focus. The deadline for abstract submissions is April 1, 2018; Notifications of abstract acceptance will occur before May 5, 2018; Early bird registration is available until May 30, 2018.

Profiles in Measurement

I am currently the Senior Psychometrician at the American Board of Family Medicine and received my bachelor's degree in Classics, Master's in Higher Education, and PhD in Educational Assessment, Evaluation, and Research from the University of Kentucky. I was introduced to Rasch measurement by my dissertation chair Dr. Kelly Bradley at the University of Kentucky. At that time, I was working on UK's SACS accreditation team and was interested in both accreditation and student learning outcomes assessment. After encouraging me to take her Rasch measurement class, Kelly showed me that my research interests, with a focus on statistics and measurement, were much better aligned with the world of professional certification and licensure than that of higher education.

Currently, my research is primarily related to applying Rasch measurement principles in certification and licensure testing – mostly in things like differential item functioning, equating, and aspects of validity. However, at the American Board of Family Medicine my role is not constrained to examinations and I also provide psychometric consultation to a variety of other organizational research partners. This provides an opportunity for me to work on different surveys, scale development, and other interesting projects with people who often have little or no training in psychometric methods or Rasch measurement models. I also enjoy R programming and am working on several packages related to Rasch measurement models and operational exam scoring.

Recently, I have been presented with an opportunity to serve the Rasch community by acting as webmaster for the Rasch.org website. In this capacity I am often found commenting on the Rasch Measurement Forum (http://raschforum.boards.net/board/1/post) trying to assist others in solving a wide-range of measurement-related problems. I have always been impressed with the sense of community among Rasch scholars and I see this as an opportunity to play some small part in helping those new to Rasch measurement have the same access to information and experience that I had.

Michael Peabody (mpeabody@theabfm.org)

Rasch-related Coming Events

April 10-12, 2018, Tues.-Thurs. Rasch Conference: IOMW, New York, NY, www.iomw.org.

Apr. 13-17, 2018, Fri.-Tues. AERA, New York, NY, www.aera.net.

May 25 - June 22, 2018, Fri.-Fri. On-line workshop: Practical Rasch Measurement - Core Topics (E. Smith, Winsteps), www.statistics.com.

June 27 - 29, 2018, Wed.-Fri. Measurement at the Crossroads: History, philosophy and sociology of measurement, Paris, France., https://measurement2018.sciencesconf.org.

June 29 - July 27, 2018, Fri.-Fri. On-line workshop: Practical Rasch Measurement - Further Topics (E. Smith, Winsteps), www.statistics.com.

July 25 - July 27, 2018, Wed.-Fri. Pacific-Rim Objective Measurement Symposium (PROMS), (Preconference workshops July 23-24, 2018) Fudan University, Shanghai, China "Applying Rasch Measurement in Language Assessment and across the Human Sciences" www.promsociety.org