Autumn 2016, VOL. 30, NO. 2; ISSN 1051-0796

RMT

Special Theme: Measurement in Sports

Research Notes and Editorials Include:

- **Measuring College Football Team Ability – 2015 Season – Patrick Fisher**
- **Exploration of Baseball 'Ability' with FACETS – Michael Peabody**
- **Structural Holes with Rasch PCAR for Detecting Judge's Biased Scores of Olympic Figure Skating Ratings – Tsair-Wei Chien**
- **Applying the Rasch Model to the NBA 2015 Season Games and Comparing MLE, EAP, and MAP Algorithm Results - Tsair-Wei Chien**
- **Note from Rasch SIG Chair – Leigh Harrell-Williams**
- **Profiles in Rasch Measurement – Stefanie Wind**

Transactions of the Rasch Measurement SIG
American Educational Research Association

Measuring College Football Team Ability - 2015 Season

Major college football has gone through several attempts at determining the best team each season. Until 1992 there were several polls - coaches, sportswriters, etc. - where each group would vote for their top 20 - and later 25 teams - in the nation. This frequently resulted in a "split national championship" because the groups' polls would not always agree and two teams would have to share the "mythical national championship."

Beginning in 1992, there was the Bowl Coalition, then the Bowl Alliance in 1995 and, then after the 1997 season, the Bowl Championship Series (BCS) was formed when the Big Ten, Pac-10 and Rose Bowl realized they were missing out and joined the party. The BCS, it was believed, would put to rest the annual argument of which team was the best team. Unfortunately, the methodology to choose the two teams who would face off for the national championship was still flawed. The BCS used a hodge-podge of polls, the ethereal "computer rankings," and a mixture of questionable math to create a percentage of sorts. It is unknown if anyone really knows what that decimal number really meant. There are many problems with any of these methodologies but that is not what this article is about. Let's discuss the 2015 major college football season, currently known as the Football Bowl Subdivision (formerly known as Division IA; the other subdivision is the Football Championship Subdivision, the FCS, formerly known as Division IAA).

The final regular season rankings reveal a level of disagreement between the College Football Playoff Rankings and the reality of scientific measurement. Unfortunately, the traditional rankings tend to follow the along with those teams with the best winning percentage, hence Clemson is #1. Unless there is an obvious and glaring strength-of-schedule disparity to the voters in the polls, the team with the best record nearly always gets the benefit of the doubt, especially among teams from the "BCS Conferences." Those are the Atlantic Coast, Big Ten, Big 12, Pac-12, and Southeastern

conferences. Another consistent issue with the polls is that the lone undefeated team will remain #1 as long as they remain undefeated regardless of how well a one-loss team is playing and despite to whom they lost. These two issues are symptoms of the problem of college football not understanding the benefits of scientific objective measurement.

The College Football Playoff Rankings are used to determine the four teams which will make up the college football playoffs. Prior to those decisions, most of the "BCS conferences" have conference championship games which serve as de facto "play-in" games for the playoffs, unless one of the teams has no chance of making the top four in the final regular season rankings. For example, Michigan State (#5, at the time of the game) and Iowa (#4) played for the Big Ten Conference Championship last season. The winner would be in the CFP. At the other end of the continuum, Texas (unranked, with a 4-7 record entering the Big 12 Conference Championship game) defeated #12 Baylor 23-17 to effectively knock Baylor out of a more prestigious and financially lucrative bowl game.

Table 1 shows the rankings (Top 10 teams) which were used to determine the playoff pairings as well as help create the bowl game match-ups. Because subjectivity proliferates the CFP ranking process, unfortunately, the four best teams are not necessarily chosen for the playoffs. The College Football Playoffs are comprised of four teams seeded one through four and paired off for semifinals and then the winners meet in the National Championship game.

Looking at the Diamond Rankings side of the table, there is one team - Ohio State - which was not involved in the their conference's championship game because they were beaten by Michigan St to spoil their unbeaten season after starting 10-0. Aside from that loss, OSU was one of the top teams in the nation and accordingly, should have been the fourth team in the playoffs instead of Michigan St, despite losing to them. Another unfortunate fact of the college football season is 'when' you suffer your only loss as well as to whom you lost. A team can recover from an early season loss (Alabama lost to Ole Miss early in the season and was able to recover and get into the CFPs).

With the conference championships serving as a bit of a "play-in" game for the playoffs, it is a bit difficult to criticize the process. I'm sure OSU coach Urban Meyer accepts accountability for his team's play and knows if they just win the MSU game, they're in the Big Ten Conference Championship game and then have a chance to get into the playoffs. That notwithstanding, OSU was one of the four best teams at the end of the regular season.

2015 Final Regular Season Rankings (Prior to the Bowl Games)

Rank	College FB Playoff Rankings		Diamond Rankings			
1	Clemson	13-0	Alabama	12-1	96.05	4.95
2	Alabama	12-1	Clemson	13-0	94.41	5.75
3	Michigan St	12-1	Oklahoma	11-1	93.87	5.14
4	Oklahoma	11-1	Ohio State	11-1	89.42	5.52
5	Iowa	12-1	Michigan St	12-1	86.96	5.21
6	Stanford	11-2	Stanford	11-2	86.49	5.04
7	Ohio State	11-1	Florida St	10-2	85.46	5.70
8	Notre Dame	10-2	North Carolina	11-2	85.03	5.23
9	Florida St	10-2	Houston	12-1	85.01	5.42
10	North Carolina	11-2	Iowa	12-1	84.59	5.25
			Mean of 128 teams		59.17	5.33

So, the CFP rankings were used to set the pairings for the playoffs with #1 playing #4 and #s 2 and 3 playing each in the national semifinals. In a stroke of irony, despite the ranking differences between the CFP and Diamond Rankings, the national semifinal pairings would have been the same: Clemson (#1) played Oklahoma (#4) and Alabama (#2) played #3, in this case, Michigan St, but would have been Ohio St (#4, Diamond Rankings). Had Diamond Rankings been used for the pairings, Alabama (#1) would have faced #4, (Ohio State in the Diamond Rankings) Michigan St and Clemson (#2) would have played Oklahoma (#3).

Now, in Table 2, we see the final rankings after the College Football Playoffs were complete. Alabama and Clemson met for the championship

and Alabama slipped past Clemson 45-40 to win the inaugural CFP.

2015 Final Rankings (After the Bowl Games)							
Rank	College FB Playoff Rankings			Diamond Rankings			
1	Alabama	14-1		Alabama	14-1	99.39	4.78
2	Clemson	14-1		Clemson	14-1	94.44	5.08
3	Stanford	12-2		Oklahoma	11-2	93.78	4.75
4	Ohio State	12-1		Ohio State	12-1	91.32	5.26
5	Oklahoma	11-2		Stanford	12-2	88.14	4.91
6	Michigan St	12-2		Michigan St	12-2	86.96	4.89
7	TCU	11-2		Houston	13-1	86.26	5.32
8	Houston	13-1		Florida St	10-3	84.78	5.28
9	Iowa	12-2		Navy	11-2	84.38	5.37
10	Mississippi	10-3		North Carolina	11-3	84.37	5.75
				Mean of 128 teams		61.02	5.18

One of the first things that jumps out when analyzing and comparing the rankings is the volatile nature of the human polls. There are three new teams in the CFP rankings and just one new team in the Diamond Rankings and Navy was #12 prior to the bowl games and playoffs. Another immediate recognition is that Alabama was accurately assessed as the best team in the country before the playoffs by Diamond Rankings.

Similar to other applications of objective measurement, we see much less volatility in the Diamond Rankings than what is seen in any ranking methodology where humans are the primary ranking mechanism. Diamond Rankings also have measures and errors of measurement to allow for legitimate comparisons of ability. Without objective measurement, there will always be subjective arguments when attempting to determine how much better "my team" is than "your team." There are other benefits of using objective measures in sports rankings. In the next issue of RMT, I will write about the playoff outlook for the 2016 season. After that, there will be another article about strength of schedule and conference strength using these measures in a subsequent issue.

Patrick Fisher
SportsMeasures.com

Exploration of Baseball 'Ability' with FACETS

There are a lot of bad baseball statistics. Batting average is Hits/At-Bats, but Walks, HBP, sacrifice, obstruction, inning ending, or substitutions do not count as At-Bats. Earned Run Average (ERA) is 9 * Earned Runs Allowed/Innings Pitched. Runners on base because of defensive errors do not count for earned run calculation and runs are charged only for those batters who that pitcher let on base. These are simply descriptive statistics that aggregate and summarize past performances.

Sabermetricians have been trying to challenge these traditional statistics by introducing new ideas such as Wins Above Replacement (WAR), which is a composite statistic of batting, base-running, fielding, and pitching that attempts to quantify how much better one player is than another using wins as the standard metric.

I decided to attempt to create a measure of baseball "ability" using the Rasch Rating Scale Model (Andrich, 1978; Wright and Masters, 1982) in FACETS (Linacre, 2014). FACETS was used because a batter would encounter the same pitcher multiple times in a game, which necessitated the use of FACET's long-form data structure. The data collection method utilized a 4-facet design: Batter, Pitcher, Inning (dummy anchored at 0), and Outcome. I used data from a 3-game series between the Red Sox and Orioles (September 14-16, 2015) that is freely available from Baseball-Reference.com. I chose these games because I had recently watched all 3 games in their entirety and could provide some

context for the analysis. Adding more games or using an entire season's worth of data would require additional dummy variables to be added to the model.

fielding, only batting ability vs pitching ability. I decided that reaching base on an error was just as good as a hit since batter speed may have influenced the fielder's behavior to cause the

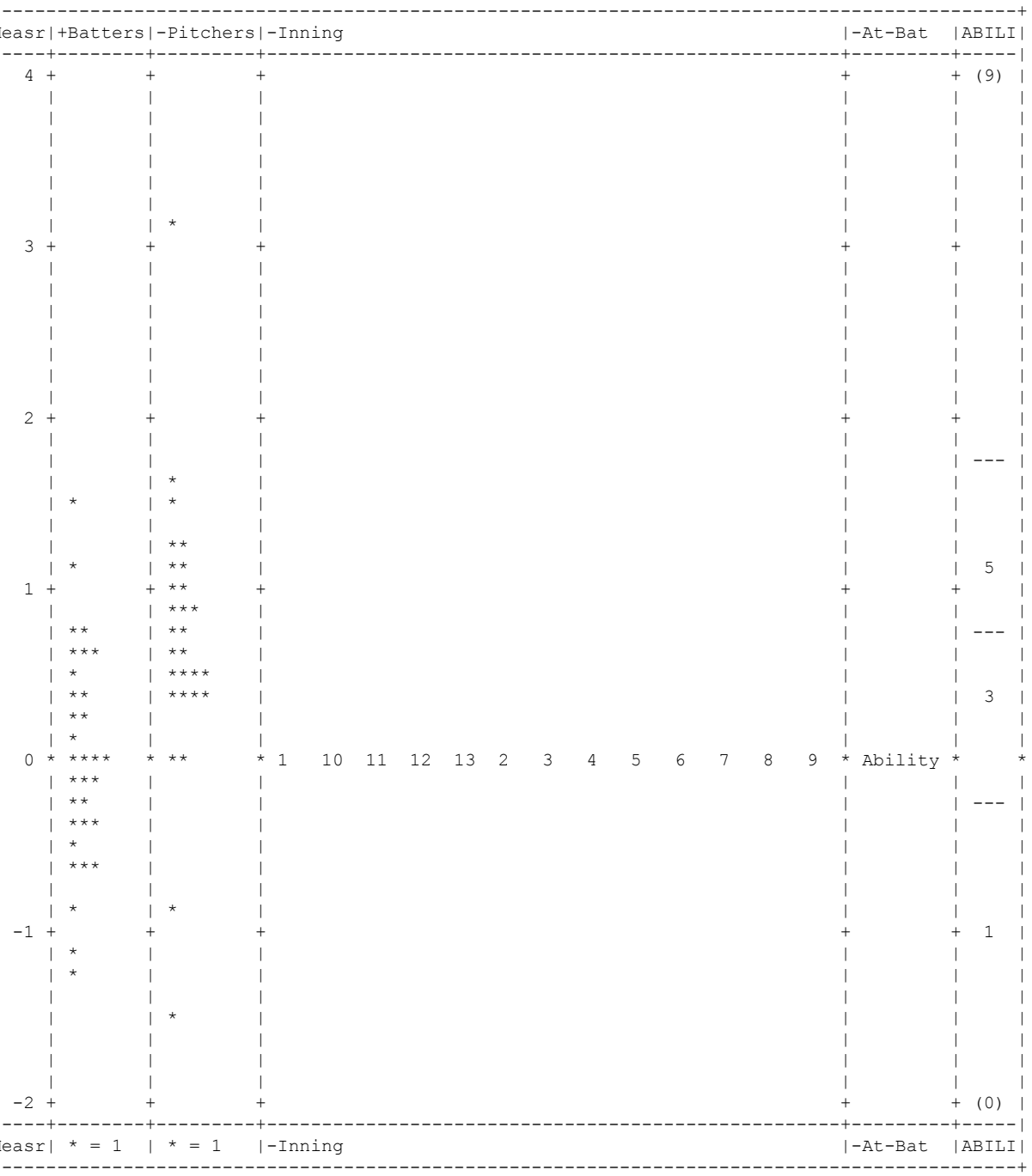


Figure 1. Batter-Pitcher Map from FACETS.

As with anything, I needed to devise some scoring rules for the Outcome variable. I used a theoretical framework which did not consider

error. Similarly, reaching on a fielder's choice is counted as an out. This assumes the fielder would have gotten the batter out had there been

no one on base. I decided to count sacrifice flies and bunts as outs because in these situations batters usually want a hit, but will settle for a successful sacrifice. The final rating scale was: (0) Strikeout, (1) Out, (3) Single/Walk, (5) Double, (6) Triple, and (9) Home Run. This model prioritizes strikeouts over batted outs and getting on base is more important than power.

ORIOLES

Num	Pitching	IP	H	R	ER	BB	SO	HR	meas
16	Britton	2	1	0	0	0	3	0	1.58
17	Drake	1	0	0	0	0	1	0	1.55
20	Givens	1.2	1	0	0	0	3	0	1.2
25	ODay	2	2	0	0	1	4	0	1.16
19	Gausman	6	2	0	0	4	7	0	1.03
15	Brach	3	1	0	0	1	4	0	0.98
22	Johnson	1	0	0	0	0	0	0	0.87
18	Garcia	1	1	0	0	1	0	0	0.63
21	Jimenez	5	6	4	4	3	4	0	0.45
26	Roe	1	1	0	0	0	0	0	0.44
27	Rondon	2	3	3	3	1	3	1	0.42
24	McFarland	1	2	1	1	1	0	0	0.38
28	Wright	3	6	6	6	1	1	2	0.04
23	Matusz	2.1	1	1	1	0	0	1	-0.03

Num	Batting	AB	R	H	RBI	BB	SO	Other	meas
23	Joseph	6	0	4	0	0	0	2B	1.1
21	Janish	1	0	1	0	0	0		0.79
24	Lake	1	1	0	0	0	0		0.79
16	Alvarez	3	0	1	0	0	0		0.65
31	Walker	1	0	0	0	1	0		0.38
25	Machado	11	2	3	1	2	3	HR	0.37
20	Hardy	10	1	2	0	1	1		0.23
27	Parra	10	1	2	0	1	2		0.05
28	Pearce	12	1	1	3	1	3	HR	0.01
22	Jones	12	0	2	2	0	4	2B	-0.31
17	Clevenger	7	0	1	0	0	1		-0.42
30	Schoop	12	0	3	0	0	4		-0.42
32	Wieters	3	0	0	1	1	1		-0.42
19	Flaherty	1	0	0	0	0	0		-0.55
26	Paredes	1	0	0	0	0	0		-0.68
29	Reimold	8	1	2	0	0	5		-0.93
18	Davis	13	2	3	1	0	2		-1.14

RED SOX

Num	Pitching	IP	H	R	ER	BB	SO	HR	meas
13	Ramirez	1.1	0	0	0	0	3	0	3.16
2	Barnes	1.1	1	0	0	0	1	0	1.25
9	Machi	2.1	2	0	0	0	3	0	1.17
5	E.Rod	5.1	5	1	1	3	9	0	0.89
3	Breslow	2.1	1	1	1	0	3	1	0.83
12	Owens	7.2	6	0	0	0	4	0	0.77
11	Ogando	1.1	1	0	0	0	1	0	0.76
6	Hembree	1.1	0	0	0	0	0	0	0.62
4	Cook	1	1	1	0	1	0	0	0.47
1	Aro	2	1	1	1	2	1	0	0.44
10	Mendez	0.1	1	0	0	0	0	0	0.33
7	Kelly	2.1	5	4	3	1	4	1	0.31
8	Layne	0.1	1	1	1	0	0	0	-0.89
14	Ross	0	1	0	0	0	0	0	-1.45

Num	Batting	AB	R	H	RBI	BB	SO	Other	meas
6	Craig	1	0	1	1	0	0	2B	1.55
12	Pedroia	10	3	5	5	1	2	2B, HR, HR	0.67
13	Rutledge	2	0	0	0	1	0		0.56
10	Marrero	5	2	4	1	0	0		0.49
2	Bogaerts	9	1	3	1	1	3	2B	0.2
11	Holt	6	2	3	2	0	0		0.18
9	Ortiz	10	1	2	2	3	3	HR	-0.01
8	Swihart	9	1	1	0	0	1	2B	-0.04
4	Castillo	15	1	2	0	0	2	2B	-0.12
14	Sandoval	12	0	3	1	1	2		-0.13
1	Betts	9	2	2	0	1	3		-0.14
15	Shaw	11	1	1	2	3	6	HR	-0.29
7	Hanigan	6	0	0	0	0	1		-0.57
5	Leon	1	0	0	0	0	0		-0.63
3	Bradley	10	1	0	0	2	7		-1.21

Figure 2. Summary of Baseball "ability" results ordered by logit measure.

Figure 1 shows the FACETS ruler with Batters and Pitchers, while Figure 2 shows a summary of the results as a box score. Ramirez (Red Sox) had easily the highest measure because he recorded 4 outs, 3 of which were strikeouts. Craig (Red Sox) had the highest batter measure because he went 1-for-1 with a double. Interesting to note the Pedroia (Red Sox) had a modest measure of .67 even though he went 6-for-11 with a double and 2 home runs, but had 2 strikeouts. Owens (Red Sox) had an interesting game. By traditional methods he pitched well allowing 0 runs over 7 2/3 innings and striking out 4. However, he did let up 6 hits in those 7 innings and I can say, having watched the game, he was often in danger of giving up runs. The model seems to suggest that perhaps he doesn't have a lot of "ability", but rather just got lucky.

Acknowledgment: Special thanks to Mike Linacre for his helpful comments.

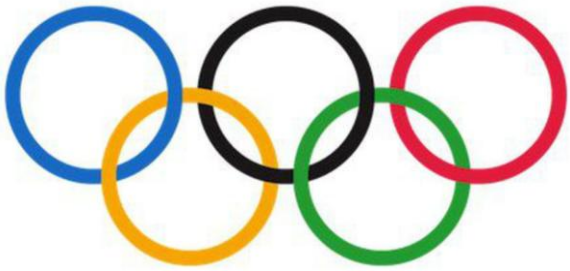
Michael R. Peabody

References

Andrich, D. (1978). A Rating Formulation for Ordered Response Categories. Psychometrika, 43(4), 561-573.

Linacre, J. M. (2014) Facets computer program for many-facet Rasch measurement, version 3.71.4. Beaverton, Oregon: Winsteps.com

Wright, B. D., & Masters, G. N. (1982). Rating scale analysis. Chicago: MESA Press.



Structural Holes with Rasch PCAR for Detecting Judge's Biased Scores of Olympic Figure Skating Ratings

The Pairs Skating competition at the 2002 Winter Olympics in Salt Lake City was contentious. It resulted in the awarding of gold medals to both a Russian and a Canadian pair, after the French judge admitted to awarding biased scores (Linacre, 2009). Rasch fit statistics and principal components analysis of residuals (PCAR) are suggested ways for detecting judge's biased scores in literature (Looney, 1996; Linacre, 2009). However, PCA results didn't prove any concrete collusion evidence (Linacre, 2002). I am interested in applying structural holes of social network analysis (SNA) (Burt, 1995/2004) with Rasch PCAR (McDonald, 1985) for detecting judge's biased scores regarding Olympic figure skating ratings in 2002. It is required to discard the most suspiciously biased judge through statistical methods prior to awarding gold and silver medals.

I extracted data of the top 3 would-be contestants (i.e., the gold medal was awarded to Russia and the silver to Canada or the bronze to China) and obtained results of Rasch PCAR using Winsteps software in Figure 1a and 1b. Results indicate an "Eastern" factor at top and a "Western" factor at the bottom are separated apart, the Japanese judge disliked China with low scores, and the Chinese judge favored China with high scores. Two possible collusions were evident in them (one for USA and Japan, another for Russia, France, and Ukraine) with a correlation coefficient of 1.0 in Figure 1c because they aberrantly endorsed equivalent scores across all eight items (i.e., four skating performances for each two pairs). Two judges of Russia and France earn the highest correlation when analyzing 12 items using data of the top 3 pairs in Figure 1d.

After combining structural holes (SH) with Rasch PCAR, I obtained the highest SH (=172) on judges Russia, France, and Ukraine due to their equivalent scores across all eight items in Figure 1e, and the greatest SH (=196) on judge France in Figure 1f.

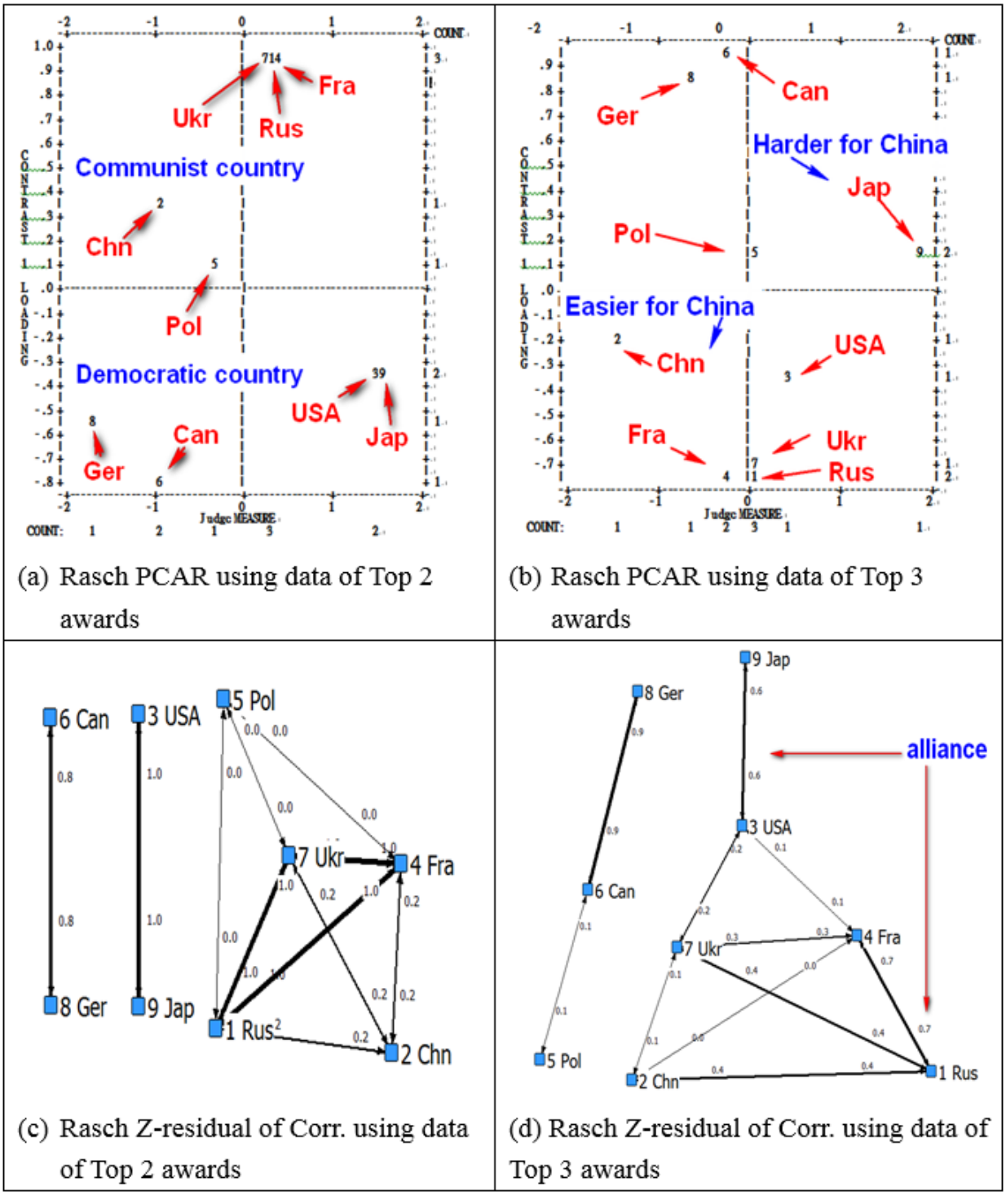
If we discarded the most suspiciously biased judge of France, the three final scores were summed for contestants Russia (58.03), Canada (58.09), and China (56.90). For the final placement, the gold medal should be awarded to Canada, the silver to Russia, and the bronze to China in Olympic figure skating.

The calculation of structure holes (considering both fixed and interaction effects) is shown in Equation (1):

$$C_{ij} = (P_{ij} + \sum_{q} p_{iq} P_{qj})^2 \quad , (1)$$

In which, Cij stands for the structure hole of an interactive cell (like tie lines in Figure 1e and 1f), summing up all cells in a row is the composite SH like values (like the size of bubble circles in Figure 1e and 1f), P_{ij} is the proportion of Rasch PCAR associated to other judges, $\sum_{q} p_{iq} P_{qj}$ is a matrix multiplication of P_{ij}. For more information on this study, see the link at YouTube https://youtube/emrb86GIE, and the MS Excel demonstration at http://www.healthup.org.tw/Olympic2002.zip.

I describe a technique, using structural holes and Rasch PCAR, for identifying an inconsistent judge and apply the method to competitions from the 2002 Winter Olympic Games and hopefully examine the extent to which the International Olympic Committee (IOC) will improve fairness in Olympic judging.



(a) Rasch PCAR using data of Top 2 awards

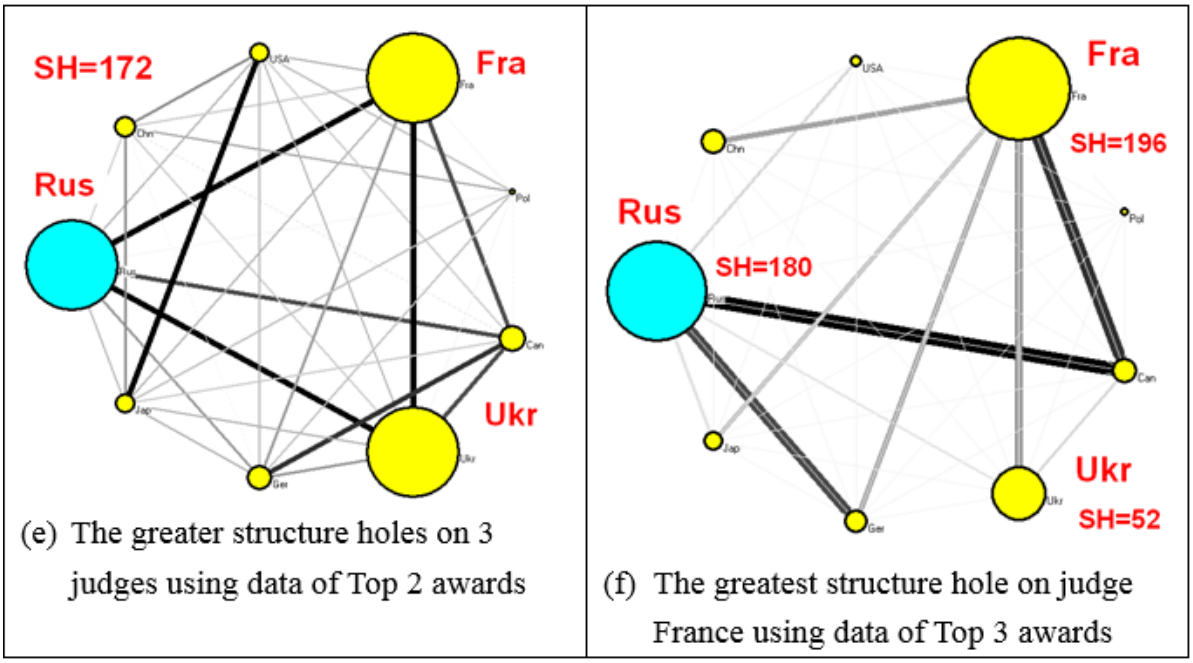
(b) Rasch PCAR using data of Top 3 awards

(c) Rasch Z-residual of Corr. using data of Top 2 awards

(d) Rasch Z-residual of Corr. using data of Top 3 awards

Tsair-Wei Chien
Chi Mei Medical Center, Taiwan

References:



(e) The greater structure holes on 3 judges using data of Top 2 awards

(f) The greatest structure hole on judge France using data of Top 3 awards

Burt, R. S. (1995). Structural Holes: The Social Structure of Competition. Cambridge: Harvard University Press.

Burt, R. S. (2004). Structural holes and good ideas. *American Journal of Sociology, 110*, 349–399.

Linacre, J. M. (2002). Judging Debacle in Pairs Figure Skating. *Rasch Measurement Transactions, 15*(4), 839-840Linacre, J. M. (2009). Local independence and residual covariance: a study of Olympic figure skating ratings. *Journal of Applied Measurement, 10*(2), 157-69.

Looney, M. A. (1996). Figure skating fairness. *Rasch Measurement Transactions, 10*(2), 500.

McDonald, R. P. (1985). Factor analysis and related methods. Hillsdale, NJ: Lawrence Erlbaum, p. 212.

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

Rasch Measurement Theory and Applications Conference

The University of Western Australia is pleased to announce the Seventh International Conference on Probabilistic Models for Measurement, to be held at The University Club on the Matilda Bay of the Swan of River. The conference will cover the range of areas where Rasch measurement theory is applied: education, psychology, health, marketing and social science. Full details of fees and processes for submission of abstracts will be available in June 2017 at: http://www.education.uwa.edu.au/ppl/raschconference/.

Also of note, preceding the conference there will be a five day *Advanced Course in Rasch Measurement Theory and the Application of RUMM2030* from Wednesday 10 January 2018 to Tuesday 16 January 2018. Please see (http://www.education.uwa.edu.au/ppl/raschconference/course) for more information about this course.

Applying the Rasch Model to the NBA 2015 Season Games and Comparing MLE, EAP, and MAP Algorithm Results

Many competitive activities, such as sports or games, often rely on ratings systems. Rating systems provide an estimation of the strength of competitors. This strength estimation makes it possible to deal with balanced matches, motivate competitors by providing them with a measurement of their progress, and make predictions about the outcomes of future competitions. I applied the Rasch Model to the NBA 2015 season games through sequentially calibrating dynamic item difficulties (i.e., estimating each ability after every competition) using three algorithms: MLE (Birnbaum,1968; Linacre, 2016), EAP (Bock & Aitken, 1981; Bock & Mislevy, 1982), and MAP (Samejima, 1968).

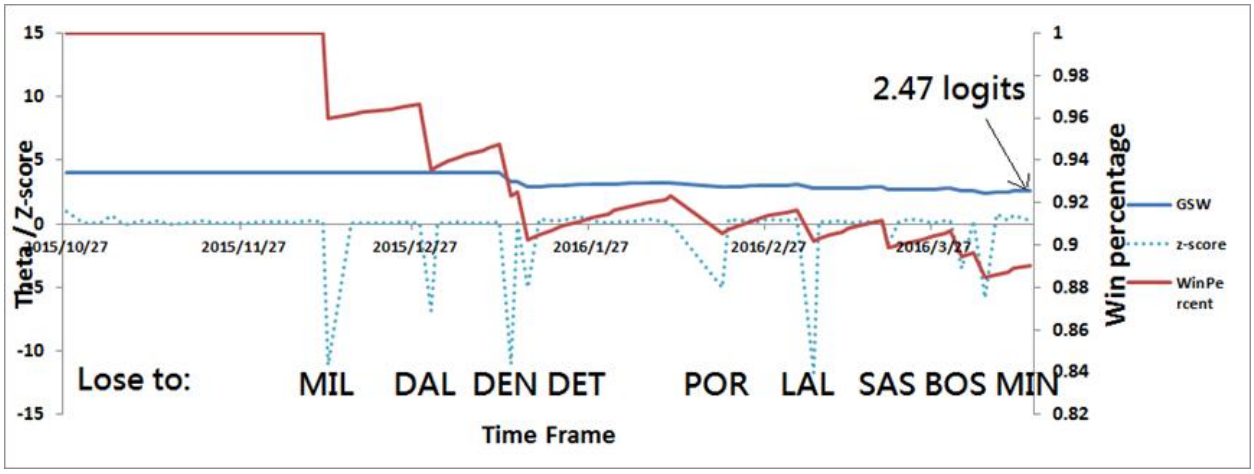


Figure 1. Golden State Warriors with sequential estimated measures had lost nine games with Z-score symbols

The Golden State Warriors (2.47 logits) with a series of sequential estimations had lost nine games with Z-score symbols shown in Figure 1.

Rasch interval theta scores are closely associated with the win percentages in Figure 2. The theta using MAP algorithm has lower correlation coefficients (0.88) compared to their counterparts (0.99 between each other) in Figure 3. The MLE has relatively large standard errors (SE) relative to the Bayesian methods of MAP and EAP in Figure 4, which is consistent to the argument of Birnbaum (1968). The MAP has large Outfit MNSQ relative to counterparts in Figure 5.

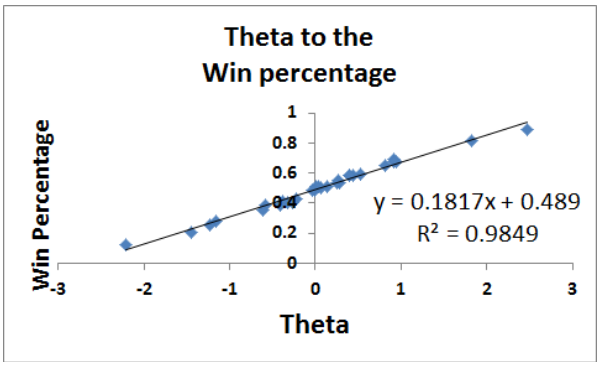


Figure 2. Raw scores of the win percentage compared to the Rasch interval theta scores

Readers interested in learning more are recommended to see the demonstrations at
https://youtube/fWNx4MNc9-E,
https://youtube/-VcBlFJ3q2o, and
https://youtube/d4f_ZLNX3Dc

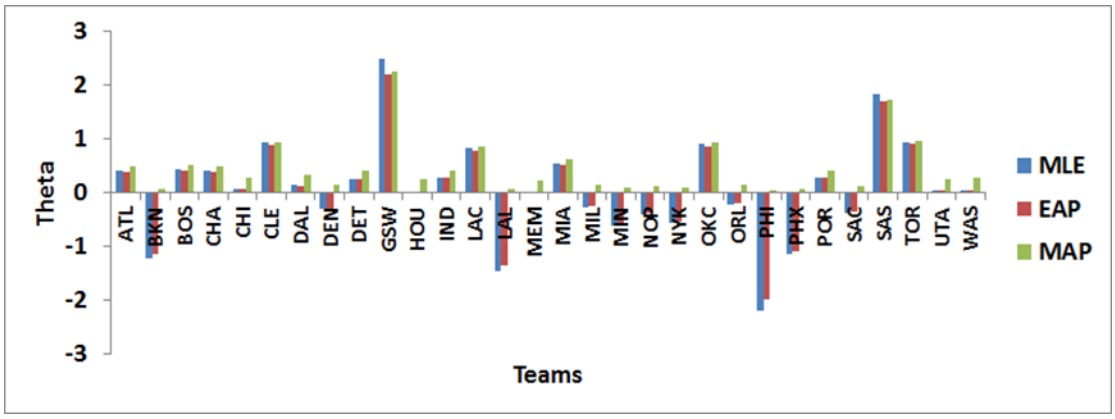


Figure 3. MAP has lower correlation coefficients to counterparts

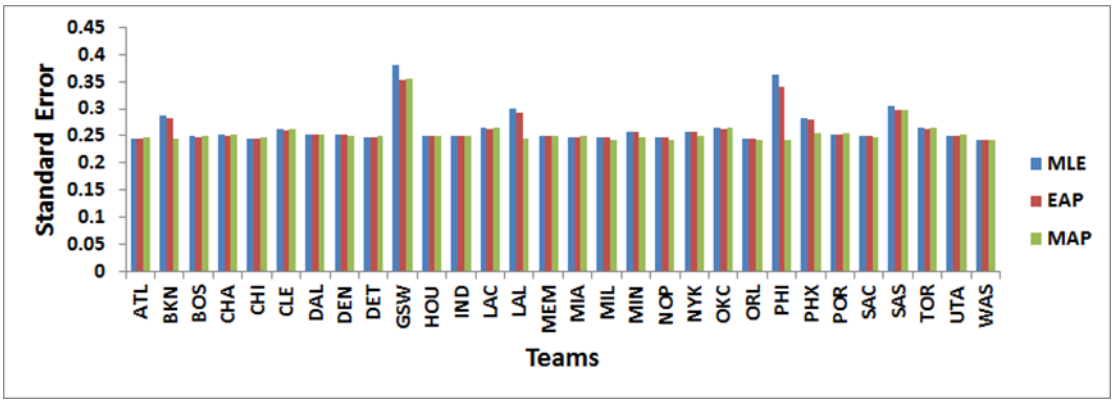


Figure 4. The MLE has large standard error relative to counterparts

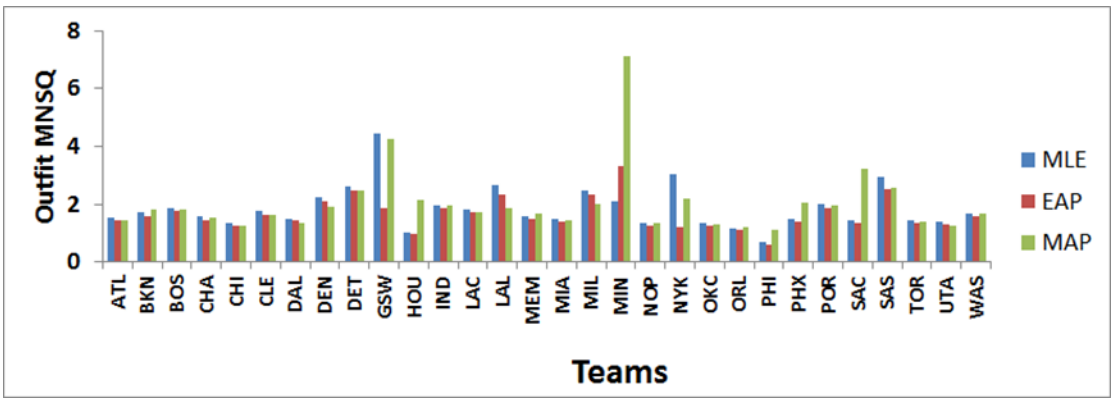


Figure 5. The MAP has large Outfit MNSQ relative to counterparts

Tsair-Wei Chien
Chi Mei Medical Center, Taiwan,

References

Birnbaum, A. (1968). Some latent ability models and their use in inferring an examinee's ability.

In: Lord FM, Novick MR, Birnbaum A. *Statistical Theories of Mental Test Scores.* Reading, MA: Addison-Wesley.

Bock, R. D., & Aitken, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*, 443-459.

Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement,6*, 431-444.

Linacre, J. M. (2016). Is JMLE Really that Bad? No, it's Actually Rather Good! *Rasch Measurement Transactions, 29*(4),1548-1550.

Samejima, F. (1968). Estimation of latent ability using a response pattern of graded scores. *ETS Research Bulletin*, Available at: http://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1968.tb00153.x/epdf.

Journal of Applied Measurement
Vol. 17, No. 2, 2016

Creating a Physical Activity Self-Report Form for Youth Using Rasch Methodology, *Christine DiStefano, Russell Pate, Kerry McIver, Marsha Dowda, Michael Beets, and Dale Murrie*

Examining the Psychometric Quality of Multiple-Choice Assessment Items using Mokken Scale Analysis, *Stefanie A. Wind*

A Practitioner's Instrument for Measuring Secondary Mathematics Teachers' Beliefs Surrounding Learner-Centered Classroom Practice, *Alyson E. Lischka and Mary Garner*

Using the Rasch Model and ***k***-Nearest Neighbors Algorithm for Response Classification, *Jon-Paul Paolino*

Exploring Aberrant Responses Using Person Fit and Person Response Functions, *A. Adrienne Walker, George Engelhard, Jr., Mari-Wells Hedgpeth, and Kenneth D. Royal*

Evaluation of the Bifactor Nominal Response Model Analysis of a Health Efficacy Measure, *Zexuan Han and Kathleen Suzanne Johnson Preston*

Measurement Properties of the Nordic Questionnaire for Psychological and Social Factors at Work: A Rasch Analysis, *C. Røe, K. Myhre, G. H. Marchand, B. Lau, G. Leivseth, and E. Bautz-Holter*

Ben Wright: A wisp of greatness, *Nikolaus Bezruczko*

Richard Smith, Editor, www.jampress.org

Note from Rasch SIG Chair

Greetings Rasch SIG Colleagues!

This is my first RMT note as Chair of the Rasch SIG. When my then-advisor and Rasch SIG Chair Ed Wolfe asked me to be a program co-chair at my first Rasch SIG business meeting, I didn't imagine that I'd being holding the same position anywhere in the near future. However, it is a responsibility and privilege that I am grateful to experience at this point in my career.

There are three important things that I'd like to discuss with you in my note. The first is The Georg William Rasch Early Career Publication Award. This coming April in San Antonio at our 2017 Business meeting will mark the 2nd time that we hand out this award. I would like our SIG members to start thinking about potential nominations for this award. The requirements for the nominee are as follows:

- The nominee should be have earned a doctoral degree no earlier than 5 years prior to the nomination deadline (i.e., degrees must have been earned in December 2011 at the earliest)
- The nominee should be sole or lead author in a paper published in the 12 months prior to the SIG annual business meeting. (Papers that have been accepted and published online in advance of receiving an actual journal issue number will be considered.)
- The publication must be Rasch-related and have been published in a peer-reviewed format.

More details regarding the exact nomination process and deadline (January 2017) will be sent to SIG members via email.

The second topic is membership. It will soon be time to renew your AERA membership. Please make sure that your AERA renewal includes membership in the Rasch SIG. When you renew online, if you select "Renew with changes", the Rasch SIG box may not remain checked so your SIG membership will be dropped. Please double check before you complete the checkout that you still have the Rasch SIG membership indicated. I emphasize doing this because membership in the SIG is important for multiple reasons. The existence of the SIG provides all those who employ Rasch methodologies in their work a place to present their research at AERA. Most of our dues ($10 currently) are used to cover the monetary portion of the Rasch SIG Awards (The Georg William Rasch Early Career Publication Award and the Benjamin Drake Wright Senior Scholar Award), hosting of the Rasch SIG website (raschsig.org), and costs associated with the annual business meetings. We need your support in the Rasch SIG in order to support Rasch-related research and acknowledge the contributions of those in our field. Along with making sure that you are renewing your membership, please encourage your graduate students, colleagues, and/or collaborators to join the SIG when they renew their membership. AERA provides the SIG with names of recent members who have not renewed in the past two years, so I will be reaching out to them shortly to encourage them to join us again.

Lastly, I would like to extend the opportunity for Rasch SIG members to reach out to me with questions, concerns or suggestions regarding the SIG. One of the tasks that I hope to tackle is getting the Rasch SIG website up-to-date. If you have any books or papers or new software that we should list, or are willing to share course syllabi or other resources, please contact me. My email is leigh.williams@Memphis.edu. I look forward to hearing from you and continue serving the SIG over the next two years.

Sincerely,

Leigh M. Harrell-Williams
Rasch SIG Chair

Profiles in Rasch Measurement

I am currently an Asst. Professor of Educational Measurement at The University of Alabama. Although I have known that I wanted a career in educational research for most of my life, I did not become interested in educational measurement specifically until I took an assessment class as an education major during my undergraduate studies. My interest in the field was solidified when I worked as a fourth-grade remedial writing teacher in Florida, and was tasked with preparing students for a high-stakes writing assessment. I wanted to know how the scores were assigned, how to interpret them, and the degree to which their interpretation was consistent across students, years, and contexts. I am grateful to Dr. George Engelhard, Jr. for his willingness to help me begin to explore these questions as I worked with him on my Master's and PhD in Educational Measurement at Emory University.

Although my research questions have become more specific since my initial conversations with Dr. Engelhard, my research continues to focus on performance assessments, raters, and ratings. Specifically, my work focuses on methods for evaluating rating quality within sound measurement frameworks. Most recently, I have been working on developing nonparametric rating quality indices using adaptations of Mokken scaling models (Mokken, 1971) that are conceptually aligned with polytomous Rasch models (Wind, 2016). These nonparametric rating quality indices can be used to evaluate raters in terms of the requirements for invariant measurement prior to the application of parametric models, such as the Rasch model. A key feature across both Mokken and Rasch approaches to evaluating rating quality is the use of graphical displays to communicate fundamental measurement properties that speak to a wide range of audiences. As I continue to explore these and other techniques in my research, I hope to contribute to the development of methods that can be used to collect diagnostic information about rating quality that will lead to improvements in research, theory, policy, and practice in educational measurement.

A defining characteristic of the Rasch measurement community is the generosity and dedication of senior scholars in their efforts to invite and mentor new researchers in the field of measurement. I hope to continue this tradition even as I continue to learn from my many mentors.

Stefanie A. Wind

References

Mokken, R. J. (1971). *A Theory and Procedure of Scale An*alysis. The Hague: Mouton/Berlin: De Gruyter.

Wind, S. A. (2016). Adjacent-categories Mokken models for rater-mediated assessments. *Educational and Psychological Measurement*, Advance online publication, DOI: 10.1177/0013164416643826.

Rasch-related Coming Events

Sept. 28-30, 2016, Wed.-Fri. In-person workshop: Introductory Rasch (M. Horton, RUMM), Leeds, UK, www.leeds.ac.uk/medicine/rehabmed/psychometric

Oct. 3-5, 2016, Wed.-Fri. Intermediate Rasch (M. Horton, Tennant, RUMM), Leeds, UK

Oct. 6-7, 2016, Thur.-Fri. In-person workshop: Advanced Rasch (M. Horton, RUMM), Leeds, UK

Oct. 14-Nov. 11, 2016, Fri.-Fri. Online workshop: Practical Rasch Measurement – Core Topics (E. Smith, Winsteps), www.statistics.com

Nov. 11, 2016, Fri. In-person workshop: 11th International Workshop on Rasch Models in Business Administration, Tenerife, Spain, www.ull.es

Dec. 7-9, 2016, Wed.-Fri. In-person workshop: Introductory Rasch (M. Horton, RUMM), Leeds, UK,