Winter 2016, VOL. 30, NO. 3; ISSN 1051-0796

RMT

RASCH MEASUREMENT TRANSACTIONS

Research Notes and Editorials Include:

Transactions of the Rasch Measurement SIG
American Educational Research Association

Differential Person Functioning in the Rasch Model

Johanson and Alsmadi (2002) introduced differential person functioning (DPF) as the person counterpart to differential item functioning. DPF is when a person performs substantially better or worse on one subset of items after the item subsets have been matched for difficulty level. The purpose of this note is to describe differential person functioning as defined by Johanson and Alsmadi (2002) in the context of Rasch measurement.

A key element of differential person functioning is that it is focused at the individual person level. The objective of a DPF analysis is to detect within-person variability of correct responses. Another key aspect of a DPF analysis is that it looks for within-person variation over two or more item subsets after controlling for subset difficulty.

In the Rasch framework, differential person functioning can be conceptualized as one of many potential causes of person misfit to the model. This follows from the Rasch model property of invariance, which states that when the data fit the model, any subset of items should yield the same achievement estimate for the person within measurement error. The invariance property includes item subsets that are matched on difficulty like those used in DPF analyses.

Figure 1 provides an example of differential person functioning. It is taken from a preliminary analysis comparing the Mantel-Haenszel and UB (Smith, 1985) procedures for examining differential person functioning in simulated test data comprised of two subsets of items (Walker, 2016). In Panels A and B, person response functions for the two subsets of items are illustrated by the plotted 1s and 2s. It is noted that the item subsets cover the same range of item difficulty, located on the x-axis. On each plot, a reference line is drawn where the probability of giving the correct response is 0.50.

In Panel A, a person exhibiting no DPF is shown. The probabilities for giving the correct responses to items in Subset 1 and Subset 2 are the same. The response functions reflect this model-expected performance because there are no major discrepancies between the two functions over the range of item difficulty. In Panel B, a person exhibiting DPF is shown. The probabilities for giving the correct responses to the items in Subset 1 are higher than for the items in Subset 2. The large gap between the plotted response functions for Subset 1 and Subset 2 reflect these discrepancies.

Person misfit caused by DPF has special implications for subscore reporting and use. If DPF is detected, that means that something other than the person's achievement and the difficulty of the items is influencing performance on the item subsets. In these cases, the conclusion that a person is *weaker* or *stronger* in one sub-content area may not be accurate. Conceptualizing DPF as a cause of person misfit provides a way to conduct DPF analysis and substantive follow-up review within existing quality checking procedures.

A. Adrienne Walker
Emory University

References

Johanson, G. & Alsmadi, A. (2002). Differential person functioning. *Educational and Psychological Measurement*, 62(3), 435-443.

Smith, R. M. (1985). A comparison of Rasch person analysis and robust estimators. *Educational and Psychological Measurement*, 45, 433-444.

Walker, A. A. (2016, April). *Comparison of Bfit-P and Mantel-Haenszel for Examining Differential Person Functioning in the Rasch Model*. Paper session presented at the International Objective Measurement Workshop, Washington, DC, USA. Retrieved from https://aadriennewalker.wordpress.com/research/

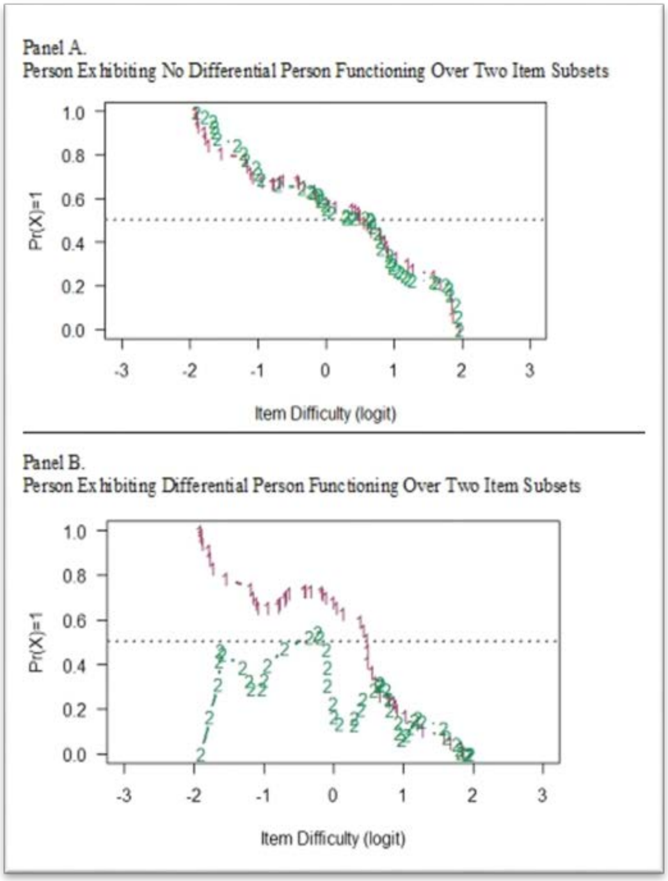


Figure 1. Differential Person Functioning Over Two Item Subsets Using Person Response Functions

An Updated Version of Ben Wright's "Method and Meaning of Measurement" Infographic

Ben Wright presented a one-page summary of Rasch measurement titled *Method and Meaning of Measurement* (reproduced in Wright, 2009; see Figure 1). This document is essentially an *infographic*, or a "graphic visual representation of information, data, or knowledge intended to present information quickly and clearly" (Infographic, n. d.). George Engelhard, Jr. shared a scanned version of this well-known document with the Rasch measurement community in 2009.

In this note, we present an updated version of the infographic that overcomes some visual challenges in reading the original version. We also attempt to present a brief summary of the major ideas captured in the infographic. However, we feel that, as is the general nature of Ben's elegant contributions, the piece stands on its own.

Ben Wright's *Method and Meaning of Measurement* is presented in five major parts, separated with headings in the original version. Our transcription and update to this infographic is presented in Figure 2. We have labeled the five parts as follows in our version: (I) Rasch measurement; (II) Interpretation; (III) Estimation; (IV) Verification; and (V) Residual analyses within item subsets.

I. Rasch Measurement

The first panel in Part I (Data) presents the basic structure of the data matrix that is used as the starting point for Rasch analyses. Persons are presented in the rows, and items are presented in the columns. Dichotomous observations (X_{ni} = 0, 1) are in the cells. These observations are interpreted using a set of *scoring rules* (Wilson, 2005; originally labeled "qualitative pointer" by Ben Wright) that indicate the ordinal directionality of the qualitative responses (1 = *Yes*; 0 = *No*). Sums of zeroes and ones down the columns provide item scores, and sums of zeroes and ones across the rows provide person scores. In the second panel (Measurement Model), the dichotomous Rasch model is presented in exponent and log-odds form, with a reference to Rasch (1960).

II. Estimation

Next, estimation procedures are used to obtain values on the logit scale for items and persons. Various techniques can be employed to estimate the parameters from a matrix of responses. In his infographic, Wright cited Unconditional Maximum Likelihood Estimation (UCON; Wright & Panchapakesan, 1969). Estimation procedures begin with the probability of event (P_{ni}). Then, estimates are improved using item information in order to minimize residuals between the observed probabilities and the expected probabilities. Information is calculated using variances (P*(1-P)). The sum of information across rows is person information (Q_n), and the sum of information across items is item information (Q_i). Residuals (Y_{ni}) are discrepancies between observed and expected responses. Sums of residuals across rows provide the person residuals; sums of residuals across columns provide the item residuals. In this section, Wright has also included a reminder that missing data can be skipped during estimation because estimates are based on row and column totals. The last part of Section II includes notes regarding several useful properties of Rasch model estimates, including their linear and additive nature, and the corresponding inferences that can be made based on the estimates.

III. Interpretation

Once the estimates are achieved, we are now able to interpret the results (Section III). The person's measure (B_n) gives us the person's ability on the logit scale. The item calibration (D_i) gives us the item's difficulty on the scale. With these estimates, the empirical probabilities (P_{ni}) can be examined in terms of our expectations given our theories/understanding, and help refine existing theories or create new theories. It is also important to examine model-data misfit, which points to anomalies that lead to discovery. Our interpretations of person and item measures are qualified based on the magnitude of misfit.

IV. Verification

In section IV, the fit of responses to the model are examined. This section describes fit analysis using a chi-square Infit (weighted fit) and Outfit (unweighted fit) statistic; both of these statistics can be calculated for items and persons. The Infit statistic mean square error statistic is a weighted statistic that can be calculated across items and persons. Infit is a weighted average of squared residuals, where the weight is defined by the variance for items (item Infit) or persons (person Infit). On the other hand, the Outfit statistic is an unweighted Rasch fit statistic. This statistic is an average of the squared standardized residuals (Z_{ni}), and it is calculated as the sum of squared residuals divided by the number of persons (N; item Outfit) or the sum of squared residuals divided by the number of persons (U; person Outfit).

V. Residual Analyses within Item Subsets

In the original infographic, Wright labeled Part V *Bias*. The procedures illustrated in this section reflect residual analyses of responses specific to a subset of items. In current practice, these procedures would be used to identify potential differences in item subset functioning that may help to identify the presence of bias, following qualitative review (AERA, APA, & NCME, 2014). Accordingly, we have renamed this section *Residual Analyses within Item Subsets.* The procedures included in this section are as follows. First, residuals are summed across persons (rows) within Item Subset G, and squared. Then, item information is summed across persons within the same subset (Q_G). These values are used to calculate a measure of potential bias, error, significance, and noise.

Conclusion

Along with other members of the Rasch measurement community, we consider Ben Wright's *Method and Meaning of Measurement* an invaluable summary of the theory and procedures for measurement in the social and behavioral sciences. The succinct presentation of this infographic reflects the inherent simplicity of Rasch measurement theory that Ben Wright often emphasized. We hope that our transcription will serve to continue the use of *Method and Meaning of Measurement* as a tool to communicate the essential principles and procedures associated with Rasch measurement theory.

References

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for Educational and Psychological Testing*. Washington, DC: AERA.

Infographic (n.d.). In *Wikipedia*. Retrieved October 5, 2016 from https://en.wikipedia.org/wiki/Infographic.

Rasch, G. (1960). *Probabilistic models for some intelligence and achievement tests* (Expanded edition, 1980. Chicago: University of Chicago

Press). Copenhagen, Denmark: Danish Institute for Educational Research.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. New York, NY: Taylor & Francis.

Wright, B. D. (2009). Method and meaning of measurement. *Rasch Measurement Transactions*, *23*(1), 1199.

Wright, B., & Panchapakesan, N. (1969). A procedure for sample-free item analysis. *Educational and Psychological Measurement*, *29*(1), 23–48. https://doi.org/10.1177/001316446902900102

Stefanie A. Wind, Cheng Hua, Mitch Porter*, Catanya Stager*, and Sija Zhang** (*denotes alphabetical listing) - The University of Alabama

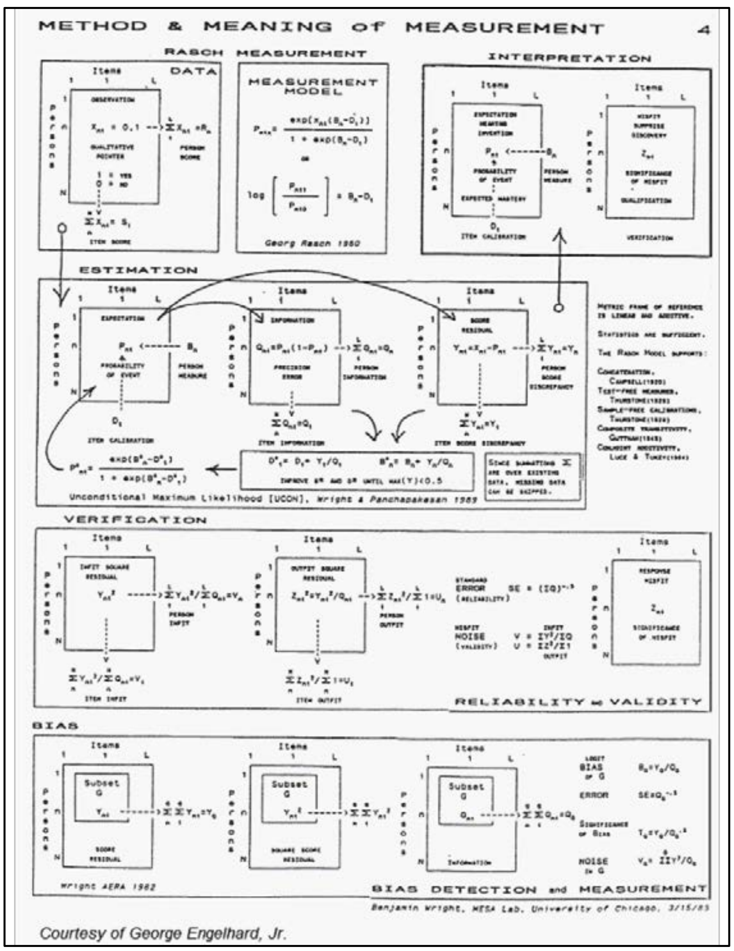


Figure 1. Original Infographic

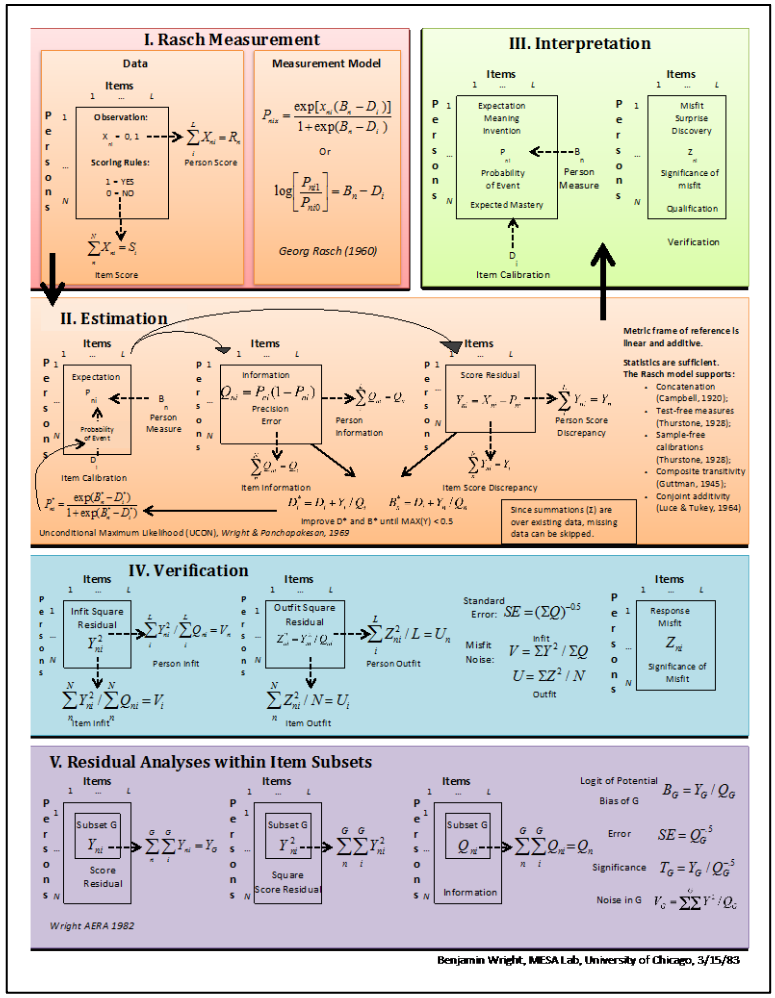


Figure 2. Updated Infographic

Rasch-related Coming Events

Mar. 31, 2017, Fri. Conference: 11th UK Rasch Day, Warwick, UK, www.rasch.org.uk

Apr. 2-3, 2017, Sun.-Mon. Validity Evidence for Measurement in Mathematics Education (V-M2ed), San Antonio, TX.

Apr. 26-30, 2017, Wed.-Sun. NCME, San Antonio, TX, www.ncme.org

Apr. 27-May 1, 2017, Thur.-Mon. AERA, San Antonio, TX. www.aera.net

May 26-June 23, 2017, Fri.-Fri. Online workshop: Practical Rasch Measurement – Core Topics (E. Smith, Winsteps), www.statistics.com

Measurement of Household Food Insecurity: Two Decades of Invariant Measurement

Rasch (1960/1980) described a measurement theory based on the requirements of specific objectivity that support invariant measurement. Invariant measurement (Engelhard, 2013) provides meaningful scores that maintain their meaning over different contexts when appropriate model-data fit is obtained. In particular, invariant measurement yields measurement systems that can be used to track changes over time.

Rasch measurement theory (RMT) has been utilized in a variety of fields, but one of the unsung success stories has occurred in the development and use of an instrument to measure household food insecurity. Food insecurity is defined as *the social and economic condition of having limited access to enough food to lead a healthy, active life* (Anderson, 1990). In 2014, 17.4 million households in the United States (14.0 percent) were food insecure at some point during the year (Coleman-Jensen et al., 2015).

The Household Food Security Survey Module (HFSSM) is the primary instrument used to measure food insecurity in the United States. It is administered by the U.S. Census Bureau on behalf of the Economic Research Service of the U.S. Department of Agriculture as a part of the December Food Security Supplement (FSS). The HFSSM has been used since 1995, and it consists of 18 items (10 items focused on issues related to all households, and an additional 8 items for households with children). The original survey was calibrated based on Rasch measurement theory, and the calibration of items has been maintained since 1995. Model-data fit has been examined each year to monitor the psychometric quality of the HFSSM.

A Wright Map based on the most recent two years is shown in Figure 1 for households with children (Engelhard, Rabbitt, and Engelhard, 2016). The items range from Item 1 (Child(ren) not eat for whole day) that is hard to endorse to Item 18 (Worried food would run out) that is relatively easier to endorse. A categorical index is also used: food secure (0-2 points), low food security (3-7 points), and very low food security (8-18 points).

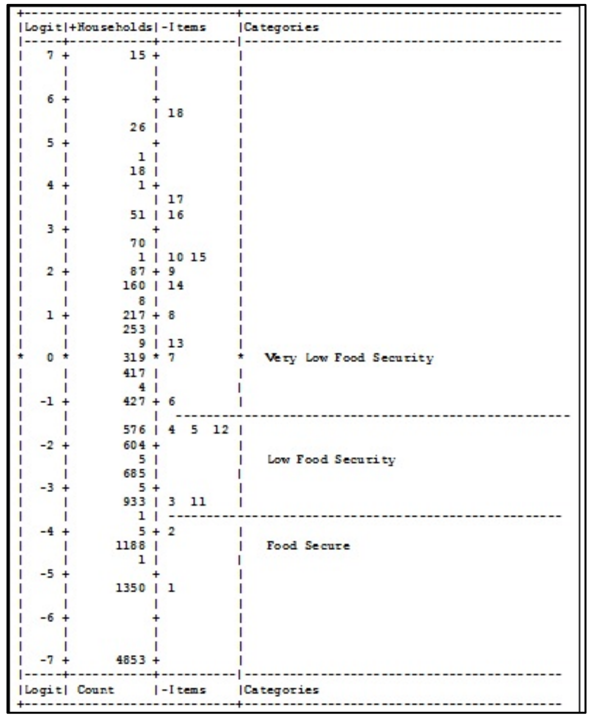
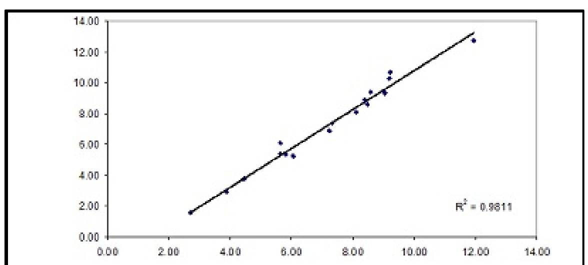


Figure 1. Wright Map

Figure 2 shows in a simple way the stability of the item locations over time. The ordering of the items in the HFSSM have remained invariant from 1998 to the present. This is a remarkable accomplishment given the complexity of a construct, such as food insecurity.



Note. Item locations are scaled to range between 0 and 14 logits with a mean of 7.

Figure 2. Plot of current item locations (2012-2014) on anchored item locations (1998)

We are planning several new studies related to the psychometric quality of the HFSSM. First of all, there is some evidence of DIF in previous studies related to the gender of the respondents. We are also planning to look at the use of scale scores based on the Rasch model, rather than the categorical indices of food security that are used in many substantive studies of factors related to food insecurity (Gundersen, Kreider, & Pepper, 2013; Rabbitt, 2013). Finally, we are exploring a Bifactor Rasch model for examining the structure of the HFSSM related to the measurement of food security in households with and without children.

The HFSSM is an exemplar of the types of measures that are possible in the social sciences, and that can achieve the goal of

One ruler for everyone, every time and everywhere … Wright (1968).

References

Anderson, S. A. (1990). Core Indicators of Nutritional State for Difficult-to-Sample Populations. *The Journal of Nutrition, 120*(11S).

Coleman-Jensen, A. Rabbitt, M.P., Gregory, C. & Singh, A. *Household Food Security in the United States in 2014*, ERR-194, U.S. Department of Agriculture, Economic Research Service, September 2015.

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

Engelhard, E.M., Rabbitt, M.P., & Engelhard, G. (2016). *Examining the Psychometric Quality of Household Food Security Estimates Obtained from the U.S. Household Food Security Survey Module.* Paper presented at the biennial meeting of the International Objective Measurement Workshop in Washington, DC.

Gundersen, C., Kreider, B., & Pepper, J. (2011). The economics of food insecurity in the United States. *Applied Economic Perspectives and Policy, 33*(3), 281-303.

Rabbitt, Matthew P., (2013). *Measuring the Effect of Supplemental Nutrition Assistance Program Participation on Food Insecurity Using a Behavioral Rasch Selection Model*. Unpublished, paper, University of North Carolina at Greensboro.

Rasch (1960/1980). *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Danish Institute for Educational Research. (Expanded edition, Chicago: University of Chicago Press, 1980).

Note: The views expressed in this paper are those of the authors and do not necessarily reflect those of the Economic Research Service or the U.S. Department of Agriculture.

George Engelhard, Jr. - University of Georgia
Emily M. Engelhard - Feeding America
Matthew P. Rabbitt - Economic Research Service, United States Department of Agriculture

Seeking nominations for the Georg Rasch Early Career Publication Award

The award shall be presented to an individual for outstanding Rasch measurement research published within five years of obtaining a doctoral degree. This is the 2nd time the Rasch Measurement SIG will be offering this award.

The award includes a stipend of $1,000 and a plaque that includes the name of the award (The Georg William Rasch Measurement Early Career Publication Award), the winner’s name, the title of the winning article, and the name of the journal or peer reviewed research publication in which the article was published. The award will be given to one person, biannually in odd-numbered years.

The deadline for nominations is January 13, 2017. Nominations are submitted by sending an email to the convener of the Awards Committee proposing the name of the nominee and describing the grounds on which the nominee meets the requirements for the award.

For more information about the award and eligibility criteria please contact Mikaela Raddatz (mraddatz@abpmr.org).

Enhancements of Partial Credit Model (PCM) for the Analysis of Parents' Opinions on Giftedness

Background

Hong Kong Academy for Gifted Education (HKAGE) is a non-governmental organisation, providing research-based information and support to all gifted students aged 10-18, teachers and their parents across Hong Kong. To cater for the needs of parents, a survey on parents' opinions has been undertaken in 2015. One of the aspects concerned is about their opinions of whether a child is gifted when the child possesses a certain characteristic (such as "Be interested in number-related games and solving mathematical problems"). These characteristics fall into seven areas, namely: (i) Inter-personal, (ii) Intra-personal, (iii) Bodily-kinesthetic, (iv) Music, (v) Verbal-linguistic, (vi) Visual-Spatial, and (vii) Logical-mathematical/Uniqueness.

The related responses from 311 parents of HKAGE (**HKAGE parents**) and 111 parents of general population contacted via some social organizations (**SocOrg parents**) were collected. It should be noted that the socio-economic status (SES) of HKAGE parents is, in general, better than that of SocOrg parents.

Partial Credit Model and its Enhancements

In the following, we use Item Response Theory (IRT) modeling to analyze the responses. The IRT model employed is based on **Partial Credit Model (PCM)** with some enhancements so as to explore some group effects in a systematic manner. The basic form of PCM is stated below.

$$P_{\mathrm{k}}(\lambda) \propto \exp\{k\lambda - \sum_{j=1}^{k} \tau_{\mathrm{j}}\}$$

Conventionally, $Pk(\lambda)$ is the probability of a student with ability λ obtaining the score k on an item with minimum mark equal to 0 and maximum mark equal to m, and $\{\tau j\}$ are the non-centralized thresholds (i.e., non-centralized threshold = centralized threshold + item difficulty). In our current setting, a parent acts as a student with ability λ. Each item has its difficulty (the average of τj). The above mentioned model is the standard one, the right hand side of which could be written in WINBUGS programming code as follows:

*$P_K(\lambda)$ α exp(k*lambda[i] - sum(tau[j,1:k]))*

We want to capture systematically the differences in ability across different groups, namely: (i) *HKAGE parent* vs. *SocOrg parent* and (ii) *Primary student parent* vs. *Secondary student parent*. ***Primary student parent*** is a parent whose eldest child is in primary level. Similarly, ***Secondary student parent*** is a parent whose eldest child is in secondary level. Accordingly, the model could be enhanced as follows:

$P_K(\lambda)$ α exp(k(lambda[i]+(a1[parGrp[i]]+ a2[parSta[i]]]]+a1a2[parGrp[i],parSta[i]]))- sum(tau[j,1:k]))*

The enhancements are simply to adjust a parent's ability (lambda[i]) based on his/her groups using the following convention:
parGrp[i] = 1 if the i^{th} parent is a *HKAGE parent*
= 2 if the i^{th} parent is a *SocOrg parent*
parSta[i] = 1 if the i^{th} parent is a *Primary student parent*
= 2 if the i^{th} parent is a *Secondary student parent*

For model identification, certain constraints (similar to the ones used in two-way ANOVA) have to be applied to the coefficients a1 (row effect), a2 (column effect), a1a2 (the interaction effect). The values for a1, a2 and a1a1 could then be estimated using the MCMC method under the Bayesian framework.

Estimation Results

A parent got 2 marks if he/she answers YES (which is supposed to be the correct answer), 0 mark when answering NO and 1 mark when answering UNCERTAIN. If his/her ability is higher than the item's difficulty, he/she got a higher chance of answering the item correctly; vice versa. The estimation results of the coefficients a1, a2 and a1a2 are presented in Table 1.

Table 1. Estimation results of the coefficients a1, a1, and a1a2

Node	M	SD	MC error	5.00%	95.00%
a1[1]	0.3260	0.06966	0.006474	0.2145	0.4448
a1[2]	-0.3260	0.06966	0.006474	-0.4445	-0.2141
a2[1]	0.04337	0.06397	0.006318	-0.05804	0.1525
a2[2]	-0.04337	0.06397	0.006318	-0.1524	0.05854
a1a2[1,1]	-0.1171	0.07724	0.009503	-0.2556	-0.002288
a1a2[1,2]	0.1171	0.07724	0.009503	0.0023	0.2557
a1a2[2,1]	0.1171	0.07724	0.009503	0.0023	0.2557
a1a2[2,2]	-0.1171	0.07724	0.009503	-0.2556	-0.002288

Based on the estimated coefficients in Table 1, the adjustment terms to parent ability (λ) for different groupings could be derived accordingly (see Table 2).

Table 2. Adjustment terms to parents' ability according to his/her groupings

Source/ Level	Primary	Secondary
HKAGE	0.2523	0.3997
SocOrg	-0.1655	-0.4865

Findings

From Table 2, the followings could be observed:

(i) When a *HKAGE parent* transits from *Primary student parent* to *Secondary student parent*, his/her ability got an increase of 0.1474 (i.e., 0.3997 -0.2523). It may be due to that they could learn more as time goes by.

(ii) On the other hand, when a *SocOrg parent* transits from *Primary student paren*t to *Secondary student parent*, his/her ability got a decrease of -0.3210 (i.e., -0.4865 – -0.1655). It may be due to that they are too busy to concern the matter of giftedness.

(iii) From (i) & (ii), it should be noted that the difference between *HKAGE* and *SocOrg parents* becomes bigger when both of them transit from *Primary* to *Secondary student parents*, changing from 0.4178 (i.e., 0.2523 – (-0.1655))) to 0.8862 (0.3997 – (-0.4865)).

According to the estimated item difficulty of each question, **the top three of difficult items** are: (i) Understands and likes oneself; has self-confidence (0.7209), (ii) Listens attentively; shows empathy and respect (0.9961), and (iii) Gets along with peers well and enjoys being with them (1.282). On the other hand, **the top three of easy items** are: (i) Asks many unexpected questions or expresses unique opinions on some topics (-0.8708), (ii) Be interested in number related games and solving mathematical problems (-0.9701), and (iii) Asks many questions and thinks about how things work and the principles behind (-1.428). Besides, we can take average of item difficulties of the questions within the same area, which are tabulated below.

Table 3. Average of item difficulties for each area

Area	Average of Item Difficulties
Inter-personal	0.7919
Intra-personal	0.5327
Bodily-kinesthetic	0.3278
Music	0.1602
Verbal-linguistic	-0.0931
Visual-Spatial	-0.4086
Logical-mathematical/ Uniqueness	-0.9834

The average of abilities of all the parents is 0.1094. Therefore, amongst these seven areas the abilities in Logical-mathematical Uniqueness, Visual-Spatial, and Verbal-linguistic are easily accepted by parents, in general, as a kind of giftedness. On the other hand, the abilities in Music, Bodily-kinesthetic, Intrapersonal, and interpersonal are rather difficult to be accepted by parents.

According to the current theory of multiple intelligences, gifted education should possess a much wider perspective rather focusing on academic excellence alone. Such an attitude should be promoted to general population of parents in Hong Kong. With understanding and appropriate help provided from their parents, the chance of gifted students to develop their talents would be much higher.

Fung Tze-ho
Hong Kong Academy for Gifted Education

Letter to the Editor: Biased against beautiful people: My response to Maul (2016) and Bond (2015)

Dear Editor,

Thank you for the opportunity to respond to the article by Andrew Maul (Maul, 2016). Put simply, I believe that you cannot earn "gotcha", lack of diversity points from the humorous anecdote provided. The claim rests on a fundamental misunderstanding of the anecdote and its relationship to Rasch Measurement. I will ignore the attempt at personal politics, as in the end we are all judged by the sum total of our lifetime contribution to each other in this world. I am confident that Ben's total score is very high. His life's work will last a very long time. Nor can I change the timing of the comments - following the farewell tribute to a colleague and teacher who has passed away, a person who cannot respond for himself. But what is done is done. However, I cannot let the article stand without a reply. Respectfully, I have a different interpretation of the humorous story retold by Trevor Bond (Bond, 2015). As the freedom song says, "I am what I am." We cannot change who we are, and I feel compelled to respond to a story that is well known in the Rasch measurement community in Australia. I mean no offence to anyone, only to add an alternate view about a clever and beautiful story. As the philosopher Madonna said "Beauty is where you find it."

Now jokes can be a bit of a hit and miss affair ... Just ask Professor, Sir Tim Hunt ... in Rasch Measurement parlance, they can be either "on target" or "off target" with their intended audience.

So, far be it for me to defend Trevor Bond (AKA ... "My name is Bond. Trevor Bond.") and his humorous anecdotes. And I know that over interpreting jokes can be boring ("It's a joke, Joyce", as they say in Australia.) But I feel compelled to speak now in order to set the record straight, as I refer to Bond's anecdote in my psychometric work. (I hope Ben's colleagues, friends and family members understand that this is a teachable moment, and this is why I am speaking now, even though it is during this time of passing, during their sad time of grief and mourning.)

In essence, I argue that the story has a very simple but effective joke structure (set-up and punch-line), and it usually produces wry, appreciative smiles from colleagues and students. I believe these smiles are good-natured, and that people can see the Rasch humor. I personally gain much enjoyment and meaning from this humorous story as it rests on a deep understanding about Rasch Measurement. So let me try and deconstruct the story...

To me this is a story about a great and passionate man, Dr. Ben Wright, who dedicated his life to measurement in the human sciences, and even after his massive stroke, he was still trying to teach us, his students, about measurement. Especially, how we should strive for uni-dimensional measurement instruments in our multi-dimensional world. This also happens to be one of themes of Trevor Bond's book which was co-written with Christine Fox. See the following quote from Chapter 3:

"We all are aware that the complexity of human existence can never be satisfactorily expressed as one score on any one test. We can, however, develop some useful estimates of some human attributes, but we can do that one attribute or ability at a time. Confusing a number of attributes into a single generic score makes confident predictions from that score more hazardous and the score a less useful summary of ability or achievement. But carefully constructed tests that make good measurement estimates of single attributes might be sufficient for a number of thoughtfully decided purposes. For special or difficult situations, collecting additional estimates of other appropriate attributes is essential. Of course, qualitative data might be used to complement the quantitative results. Human beings are complex, multidimensional creatures to be sure. But whereas using height as a measure of a person is an obvious and convenient reductionism, in many cases useful predictions can be made about the suitability of doorway heights based on that just one estimate alone. And we would be naive to think that this would be sufficient for every person." (Bond & Fox, 2015, page 40).

Or as Ben once wrote in RMT "Variation in discrimination is also rejected by Rasch as a symptom of item bias, multi-dimensionality. This phenomenon has been followed up empirically many times (e.g. Masters, 1988). The items which vary in discrimination have been demonstrated to be contaminated by item bias or to introduce extra dimensions." (Wright, 1992). Masters (1988) states: "The first step in their identification is the recognition that unusual item discrimination can be an indication that an item is giving some individuals an unintended advantage. The responsibility then lies with the test developer to investigate each unusually discriminating item to determine whether or not it is introducing and giving special weight to differences on a second, undesired dimension." (Masters, 1988, page 29).

So according to my interpretation, Bond's humorous anecdote can easily be rewritten to suit any occasion. Here is my attempt:

Ben said: *Do you know Nick who wrote this paper?*

I replied: *Yes, from Australia. Do you remember him?*

Ben said: *Aah, yes. He's enthusiastic.*

I then said: *And very determined.*

Then Ben said (smiling): *And very enthusiastic.*

Now, can you see that for the last three lines you can substitute any single descriptor, attribute or construct, and place it in comparison with another one (e.g. graceful and stylish, sporty and honest, brave and handsome, ethical and respectful, funny and persuasive) and see how the joke still works in Rasch measurement terms?

To me the anecdote is a very touching and human story – the triumph of the human spirit over neurological disability. It shows an old master still imparting vital knowledge to his disciple, fighting against his own limitations and impending mortality. It is the stuff of Samurai legends, Hollywood westerns and Space operas. Story tellers like Joseph Campbell would be proud. It is truly beautiful.

Finally, I would like to pay my respects to Ben Wright and to his great contribution to educational and psychological measurement. May he rest in peace.

Nick Marosszeky
Macquarie University

References

Bond, T. (2015). A Humorous Side of Ben. *Rasch Measurement Transactions, 29*(3), 1533.

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (Third Edition). New York: Routledge.

Masters, G. N. (1988). Item Discrimination: When More Is Worse. *Journal of Educational Measurement*, *25*(1), 15 - 29.

Maul, A. (2016). Valuing One Another for All the Right Reasons. Rasch Measurement Transactions, 29(4), 1558.

Wright, B. D. (1992). IRT in the 1990s: Which Models Work Best? 3PL or Rasch? *Rasch Measurement Transactions, 6*(1), 196 - 200.

Profiles in Rasch Measurement

Hello Rasch community!

My name is Adrienne, or more formally, A. Adrienne Walker. I will complete my PhD and start a postdoc at Emory University in August 2016. I am excited to introduce myself in this special issue of *Rasch Measurement Transactions*.

My educational background is in psychology, but I was introduced to the field of psychometrics in my work as an assessment specialist at the Georgia Department of Education. I am a data analyst at heart, and in the five years that I spent working on large-scale, high-stakes achievement tests, I recognized that psychometrics, and specifically Rasch measurement theory, represented a framework that is central to inferring substantive meaning (in terms of student achievement) from test data. At the urging of a Georgia Technical Advisory Committee member, who ultimately became my advisor, I returned to school.

My research interests are in the validity of test score inferences and their appropriate uses, and these interests have been shaped by my professional experiences in K-12 educational assessment. My doctoral work examines the use of the person response function as a graphical tool for inspecting and evaluating item responses of students whose achievement may not have been measured well by tests. This work is situated in model-data fit, specifically person fit, which I conceptualize as an idiographic perspective of test performance. I argue that adequate *individual* person fit is necessary for meaningful score inference, and that information about *individual* person fit is necessary for appropriate score interpretation and use. Currently a gap between educational assessment research and practice exists because individual person fit information is not routinely provided (alongside a test score) to help practitioners interpret the score. I believe that research seeking to close this gap is vital in today's climate of student and teacher accountability.

Our duty as psychometricians is to make measures that are meaningful and useful. But an additional challenge that we as Rasch theorists undertake is to make measures that are accessible and clear. Now is a good time for renewed attempts to make the concept of person fit accessible and meaningful to educational stakeholders, and for person fit to assume a role in test score reporting practice. (I envision individual student reports that include not only a score and a standard error of measure, but also an indicator of person fit and a graphical display of fit.) I would like extend my work to promote this aim.

In closing, I'd like to thank Professor George Engelhard (my advisor and colleague), Richard Smith, and other Rasch pioneers including Ben Wright for the groundwork that they laid in the area of person fit (e.g., graphical depictions of residuals, Kid Maps, etc.). I am also grateful for work done by Rob Meijer and his colleagues (e.g., person response functions as graphical tools). These teacher/researchers have influenced my thoughts and ideas. Lastly, I'd like to thank Ken Royal for this opportunity.

Sincerely,

A. Adrienne Walker
Aawalk5@emory.edu