SUMMER 2025, VOL. 37 NO. 3; ISSN 1051-0796

RMT

RASCH MEASUREMENT TRANSACTIONS

- Overview of this Issue of RMT -- Stefanie A. Wind & Leigh Harrell-Williams
- Assessing the Stability of the Construct of Critical Core Knowledge of Nurse Anesthesia Across Single-Point-in-Time and Longitudinal Assessments - Rajat Chadha, Eric R. Deutz, Timothy J. Muckle
- Announcements from Rasch Measurement SIG
- AERA 2025 Rasch SIG Keynote Address: *The Little Steam Engine that Could* – Trevor BOND

Transactions of the Rasch Measurement SIG
American Educational Research Association

Overview of The Issue

The Summer 2025 issue of Rasch Measurement Transactions (RMT) includes several articles and announcements that may be interesting to our community of Rasch measurement researchers.

The issue begins with a research note authored by Rajat Chadha, Eric R. Deutz, and Timothy J. Muckle.

Next, we present announcements and updates related to the AERA Rasch Measurement Special Interest Group (SIG), including a call for nominations for the 2025 Georg William Rasch Early Career Publication Award.

We end the issue with a note based on the keynote address given at the SIG business meeting in April 2025 by Dr. Trevor Bond, recipient of the 2024 Benjamin D. Wright Senior Scholar Award.

As always, we welcome your contributions to the next issue for RMT. We would appreciate receiving your research note, conference or workshop announcement, etc. by September 30, 2025. Please contact Stefanie at swind@ua.edu or Leigh at leigh.williams@memphis.edu to submit something for inclusion.

Sincerely,
Stefanie A. Wind & Leigh Harrell-Williams

Rasch Measurement Transactions
www.rasch.org/rmt

Assessing the Stability of the Construct of Critical Core Knowledge of Nurse Anesthesia Across Single-Point-in-Time and Longitudinal Assessments

Introduction

The National Board of Certification and Recertification for Nurse Anesthetists (NBCRNA) is a nonprofit entity with responsibility for the certification and recertification of Certified Registered Nurse Anesthetists (CRNA) throughout the United States. Its credentials are required in all jurisdictions for the practice of nurse anesthesia. The Continued Professional Certification Assessment (CPCA), a computer-based examination, is a required component of the maintenance of certification for CRNAs. The CPCA intends to measure the construct of critical core knowledge needed to safely practice nurse anesthesia across four domains: airway management; applied clinical pharmacology; applied physiology and pathophysiology; and anesthesia equipment, technology, and safety (Ferris et al., 2021).

The NBCRNA introduced the Maintaining Anesthesia Certification Check (MAC Check) as a self-paced longitudinal assessment in August 2024 (NBCRNA, 2025) after a randomized controlled trial established several benefits of the longitudinal assessment approach over the single-point-in-time CPCA (Chaudhary et al., 2024). MAC Check, like CPCA, intends to measure the same construct of critical core knowledge of nurse anesthesia.

Both CPCA and MAC Check are performance standard assessments to determine if a nurse anesthetist's knowledge level on core domains of anesthesia is consistent with current, proficient practice, and if not, to identify areas where they may need additional education. CRNAs maintain their certification, even if they do not meet the performance standard, provided they complete additional focused continuing education in the area where the performance standard is not met. MAC Check has now replaced CPCA as a requirement for maintaining nurse anesthesia certification.

The objective of this study is to illustrate the use of item hierarchies to assess the stability of the construct of critical core knowledge of nurse anesthesia across the two modes of administration – single-point-in-time CPCA and longitudinal MAC Check assessment – thus providing validity evidence within the context of certification testing. A comparison of hierarchy of item locations from independent calibrations from the two modes was conducted. This study also examined drift in item calibrations in MAC Check from the calibrations estimated with CPCA using existing criteria used across all assessments managed by the credentialing organization. The American Board of Family Medicine conducted a similar investigation to assess stability of the construct of the Family Medicine Certification Scale between one-day exam and longitudinal assessment (O'Neill et.al., 2024) and found that the items functioned

similarly across the two modes of administration ($r = 0.558$, $p < .001$).

Methodology

CPCA is a fixed-length examination consisting of 130 operational, calibrated items and augmented with 20 pretest items. Items are selected randomly from a pool. Certificants have three hours to complete the assessment at a test center or online with remote proctoring.

MAC Check is a longitudinal assessment spread over 13 quarters and certificants are required to complete 20 operational and 5 pretest items in each quarter. Additionally, certificants are administered up to 5 previously incorrectly answered operational items in each quarter after the first quarter. As in the case of CPCA, items are selected at random from the pool. While the MAC Check is not proctored, the participant is allocated one minute to answer each question, along with the option to add time from a five-minute timebank (in 30-second increments) that is refreshed every quarter.

Both CPCA and MAC Check item types include multiple choice questions with either a single correct option or multiple correct options and all items are scored dichotomously. Items in the CPCA item bank are calibrated to the same measurement scale using the dichotomous Rasch Model (Rasch, 1960). Almost all the CPCA item bank calibrations are based on administration between February 2024 and August 2024, except for calibrations of two items that are based on administration between October 2023 and February 2024. These calibrations were estimated based on administration to between 346 and 2,439 certificants.

MAC Check item pool comprises 782 operational items borrowed from the CPCA item bank. MAC Check data from 23,041 certificants who answered at least 25 items between August 2024 (start of MAC Check) and February 2025 was included in the analysis. Items were calibrated using the MAC Check data based on administration to between 661 and 5,715 certificants. If an item was administered more than once to a certificant in the case of prior incorrect response in an earlier quarter, only the first administration data was included in the analysis.

Stability in the hierarchy of items across MAC Check and CPCA was assessed using the Pearson correlation coefficient between calibrations for the same 782 operational items based on the MAC Check administrations and calibrations from the CPCA item bank. Both sets of calibrations were estimated independently in Winsteps (Linacre, 2022) using the dichotomous Rasch Model.

In a separate analysis of MAC Check data, all operational items were anchored to calibrations from the CPCA item bank. Items were then assessed for significant drift using these criteria: 1) a minimum absolute displacement of 0.75 logits; and 2) absolute displacement divided by standard error of the item calibration estimate greater than 2. These criteria are consistent with the criteria used across all assessments at the

credentialing organization to identify items that drift in calibration over time.

Findings

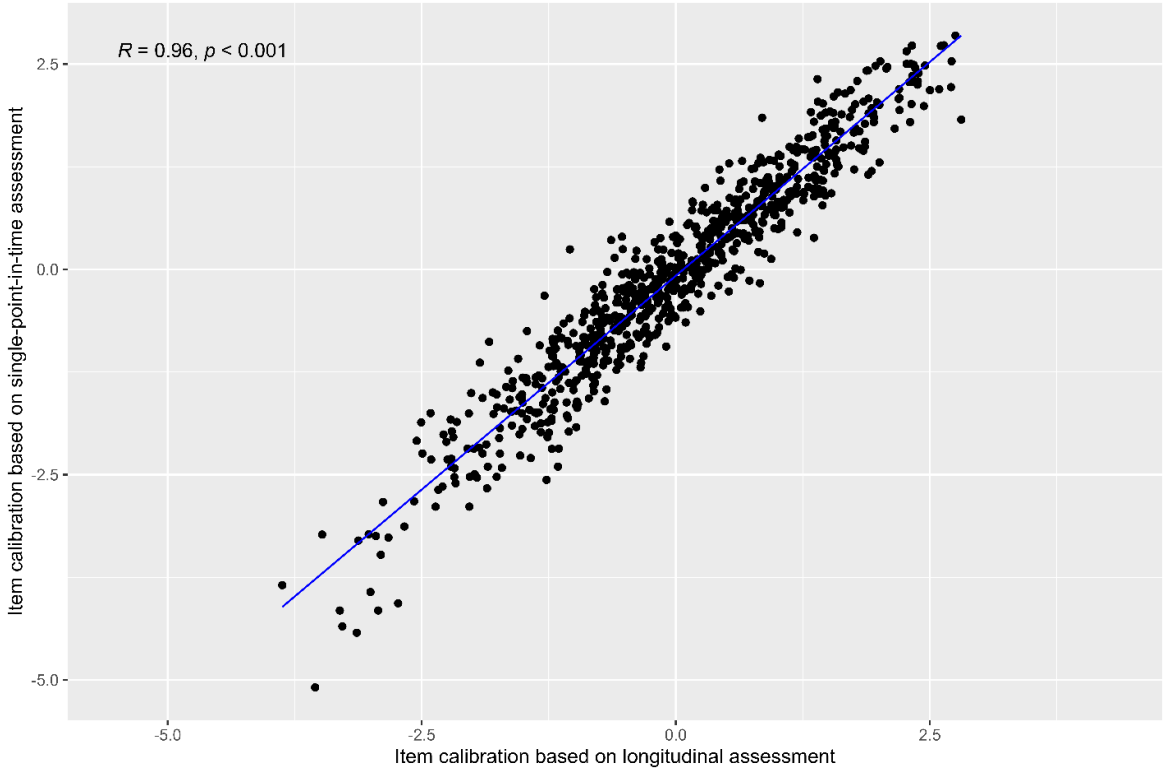
The Pearson correlation coefficient between independent calibrations of 782 items from MAC Check and CPCA administrations was very high (r=0.96, p<.001, Figure 1). This finding suggests that the hierarchy of item calibrations and the construct being measured are stable across the two modes, reinforcing the validity of score interpretations across the two assessment types.

A very small proportion (4.5%) of the items in MAC Check demonstrated drift from the calibrations estimated in the CPCA. Approximately two-thirds (62.8%) of these items became more difficult in MAC Check and the remaining became easier. This number (4.5%) is either similar to or less than the proportion of items that have been observed to drift in calibrations at six-month intervals in CPCA and in other assessments managed by the credentialing organization.

Discussion

A very high linear correlation of 0.96 was observed between independent item calibrations from single-point-in-time and longitudinal assessments. Additionally, this study found a very small proportion of the items (4.5%) demonstrated significant drift in longitudinal assessment when compared to the calibrations from the single-point-in-time assessment.

Figure 1: Scatter plot between independent item calibrations (in logits) from two modes



The correlation between pairs of item calibrations in this study is higher than the correlation of 0.558 that O'Neil et al. (2024) reported in a similar study of comparison of stability of item hierarchy between a one-day exam and longitudinal assessment. The authors reported that a large proportion of items became easier in the longitudinal assessment format, attributing the finding to the availability of extra time (5 minutes in longitudinal versus 1.3 minutes per item in one-day examination) and access to external resources. Notably, the difference in allocated time per item is not that large in the case of the single-point-in-time and longitudinal assessments in this study. The single-point-in-time assessment allows 180 minutes for 150 items and the longitudinal assessment allows 1 minute per item with the option to add time from a five-minute timebank that replenishes every quarter. Candidates in this study are also not allowed to use external resources during their exam, and this time restriction may help to limit their ability to do so.

Conclusion

This study provides evidence to support the stability of the construct of critical core knowledge needed to safely practice nurse anesthesia across the single-point-in-time and longitudinal assessments, and thereby supports demonstrates how item hierarchies can be used to support the validity of test score interpretation across modes of administration as outlined in Standard 9.9 in the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014). In other words, evidence from this study suggests that similar scores from MAC Check and CPCA represent similar knowledge of the construct they intend to measure.

Limitations

One limitation of this study is that the data from longitudinal assessment is based on just two quarters. Further study is needed to assess the stability of the construct when additional data from the longitudinal assessment becomes available in the future.

Rajat Chadha, PhD, ICE-CCP
The National Board of Certification and Recertification for Nurse Anesthetists (NBCRNA)
Corresponding author:
RChadha@nbcrna.com

Eric R. Deutz, M.Ed.
The National Board of Certification and Recertification for Nurse Anesthetists (NBCRNA)

Timothy J. Muckle, PhD, ICE-CCP
The National Board of Certification and Recertification for Nurse Anesthetists (NBCRNA)

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Choudhry, S. A., Muckle, T. J., Gill, C. J., Chadha, R., Urosev, M., Ferris, M., & Preston, J. C. (2024). Transforming Assessments of Clinician Knowledge: A Randomized Controlled Trial Comparing Traditional Standardized and Longitudinal Assessment Modalities. *Practical Assessment, Research, and Evaluation*, *29*(1).

Ferris, M., Gill, C., Preston, J., & Browne, M. (2021). Validity Evidence for the Continued Professional Certification Assessment: Professional Practice Analysis of the National Board of Certification and Recertification for Nurse Anesthetists. *Journal of Nursing Regulation*, *11*(4), 51-62.

Linacre, J. M. (2022). Winsteps® Rasch measurement computer program (Version 5.3.4). Portland, Oregon: Winsteps.com

NBCRNA (2025). *Maintaining Anesthesia Certification (MAC) Program Handbook.* https://www.nbcrna.com/docs/default-source/publications-documentation/handbooks/nbcrna-hb-mac.pdf?sfvrsn=864d5b54_3

O'Neill, T. R., Stelter, K., & Wang, T. (2024). Examining the construct stability of the family medicine certification scale between one-day exam and longitudinal assessment. *The Journal of the American Board of Family Medicine*, *37*(5), 900-908.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Danmarks pædagogiske Institut.

Updates and Announcements from the Rasch Measurement Special Interest Group (SIG) of the American Educational Research Association (AERA)

The AERA Rasch SIG has several important updates to share with the Rasch community.

We appreciate your engagement with the SIG and look forward to connecting with you through SIG activities.

Call for Nominations: 2025 Georg William Rasch Early Career Publication Award

We are currently accepting nominations for The Georg William Rasch Early Career Publication Award, which is an AERA-sanctioned award. This award is presented to an individual for outstanding Rasch measurement research published within ten years of obtaining their doctoral degree and will be presented during the AERA 2026 Annual Conference.

The main purpose of this award is to foster ongoing quality research in the area of Rasch measurement, and to encourage the development of a Rasch measurement focus in the early phases of one's career.

Eligibility Criteria:
Nomination for the Georg William Rasch Measurement Early Career Publication

award should be based on a scholarly publication authored by a nominee that fulfills the following criteria:

- The publication must include a Rasch measurement focus.
- The publication may be based on the dissertation work of the nominee or other recent research the nominee has conducted.
- The nominee should be either the single author or the lead author (in the case of a jointly authored paper) of the article.
- The article must have been published within four (4) calendar years (April 2022 – April 2026) prior to the Rasch Measurement SIG's business session at the 2026 AERA Conference.
- Only peer-reviewed research publications that are published or in press (accepted for publication) are eligible for nomination.
- The nominee should have received his/her doctoral degree no earlier than ten years prior to the nomination deadline (December 31, 2015).
- The nominee must be a member of the Rasch SIG or must become a member by the time the award is presented at the annual meeting.
- Nominators are not required to be SIG members.

The Award:
The award includes a monetary stipend and a plaque that includes the name of the 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.

Nomination Deadline:

The nomination deadline is October 31, 2025.

To Submit:
Nominations are submitted by sending an email to the Rasch SIG Secretary, Dr. Kaiwen Man: kman@ua.edu

Completed nominations should include the following:

- A letter nominating the author of an early career publication. Please include the name of the author, the date he/she received the doctoral degree, and the name of the institution that conferred the degree. The nominator's letter must include reasons that the paper is an example of an outstanding Rasch measurement research publication
- A copy of the published paper, including complete bibliographic information
- A copy of the Table of Contents of the journal or other peer reviewed research publication in which the paper appeared OR an acceptance letter from the journal will be acceptable if it is currently in press
- A current CV for the nominee

Update on SIG Membership:

Urgent Call for Renewal

As some of you know, membership and participation in the Rasch SIG has been steadily declining over the past several years. Despite our efforts to increase membership over the last few months, our current membership count remains below the minimum required to maintain the SIG as part of AERA.

To ensure that our SIG can continue to provide this welcoming space, I am kindly asking you to (re)join the Rasch SIG for the next year. Please also invite your colleagues and students to join.

The current membership due is $7 per year. These dues allow the SIG to exist as an official sub-unit of AERA and provide research awards to junior and senior scholars.

Our current membership roster lists 61 members. We need at least 75 to be considered "viable" according to AERA.

I extend my sincere thanks to those of you who have recently renewed your SIG membership. If you have not done so, please consider renewing your membership to the Rasch SIG as soon as possible.

Opportunities to Contribute to the SIG

The Rasch SIG is actively working to increase engagement within our community of Rasch scholars. Our current efforts are focused on three main activities: (1) mentoring; (2) webinar series; and (3) in-person activities at AERA 2025.

Please complete our survey linked here to let us know how you plan to be involved with the SIG.

If you are interested in contributing to any of these efforts, you may also reach out to me directly via email: swind@ua.edu. I would love to hear from you!

Stefanie A. Wind
Chair, Rasch Measurement SIG

The Little Steam Engine that Could

Trevor G. Bond

Address delivered to the RM SIG meeting on the occasion of the presentation of the Benjamin Drake Wright Senior Scholar Award at AERA 2025 in Denver, Colorado.

At the beginning of 2025 Applying the Rasch Model: Fundamental Measurement in the Human Sciences with over 12,000 citations, published in four editions plus one in Portuguese, was the most influential Rasch measurement publication in the recent half century. In the children's classic, The Little Steam Engine that Could, a long train must be pulled over a high mountain after its locomotive breaks down. Larger locomotives are asked to pull the train; they refuse, being occupied with much more important tasks. Only the little engine is willing to try and, while repeating the mantra "I think I can, I think I can", overcomes a seemingly impossible task. I have adopted this analogy to highlight some key moments along the track to ARM's success.

How it all began

In the 90s, I made several presentations about using Rasch analysis in cognitive developmental settings for the Jean Piaget Society. The oft-repeated question at the end of those presentations was, inevitably: Can you tell me how to do that analysis - in about 10 minutes? Even though I really wanted to write a book about Piaget's formal operational thinking, it was obvious that I needed to think about writing a book about Rasch analysis for colleagues such as these and their research students.

Judy Amsel, senior editor for Lawrence Erlbaum and Associates, and wife of a long-term JPS member, told me that if I ever intended to write a book about Rasch analysis, I should give her publishing group the right of first refusal. "There is a space in our catalogue for that book, and clearly you have the capacity to fill it." The idea sort of developed slowly in my head but nothing was written ever down and no action was taken. On a visit to the University of Toledo, fellow Piagetian Bill Gray introduced me to "new hire" Christine Fox. I read a Rasch-based paper of hers written in the context of nursing education and realised our styles of writing/teaching seem quite compatible.

When I asked Christine to consider being the co-author, she merely said, "Send me a draft of your first chapter to read, and I'll consider it." So, I was obliged to put some of those earlier thoughts in writing. After her acceptance, we divided the bulk of the chapters so that for each, one of us would be the lead author and the other the secondary author. An increasing number of draft chapters were emailed back and forth across the Pacific for consideration and iterative revisions.

In 2000, I was invited by the College of Education to spend a sabbatical period at the University of Toledo where Christine and I would co-teach a master-level Rasch measurement course, based on the ideas we had developed for the book. All went well up until Chapter 6: The Rating Scale Model was up for consideration. In short, Christine emphatically rejected my approach and, full of misplaced confidence, I staunchly defended it - until I realised that the crux of our disagreement was due to the fact that I learned about the Rasch Rating Scale model under the umbrella of the Partial Credit model. Yes, I rewrote my homework and resubmitted it a week later to Dr Fox.

Before I left the US at the end of my sabbatical in mid-2000, I posted a hard copy of what I thought was the penultimate draft along with electronic versions of the same on disk to the LEA offices in Mahwah, New Jersey. What surprised me was that instead of getting some advice about "Rewrite this; Drop that; Do more here; etc.", I learnt that the manuscript had been sent to the copy editor for correction; it wasn't the penultimate draft, it was going into production. A couple of iterations done by hardcopy and email followed, and then the long wait of expectation.

If I remember correctly, AERA 2001 was held in Seattle. I was a little stunned to be approached by Weimou Zhu who told me that he had just purchased a copy of Applying the Rasch Model from the Erlbaum book display. Of course, I laughed at him, but he dug into his bag and pulled out the first copy that I had ever seen. I think I just about broke my neck getting to the appropriate section of the booksellers' displays and asked to see a copy. When I asked, "How can I get a copy of this?", the LEA rep replied, "You just pay for it; you can use your credit card." I pointed out that the name on my AERA badge was "Trevor Bond", the same name as was on the book. He laughed and said, "In that case I will just give you one." Apparently, the book sold out and did so at every subsequent AERA meeting at which I attended and presented. Coincidentally, my colleague Joel Michell, author of *Measurement in Psychology*, told me that he knew when I had made an important presentation or run another Rasch workshop because each time there would be a subsequent little upward blip in the sales of his book.

What happened next

There were three important consequences to the publication of the first edition:

1. In a subsequent Winsteps workshop, Mike Linacre was asked, "Does your software make those steppingstone maps that are in Bond & Fox?" When Mike replied, "No", his participant said, "Don't you think it would be a good idea if it did?" I was quite embarrassed to hear this, but soon, Mike had implemented what he called the "Bubble Maps" which have been a feature of Winsteps ever since.
2. Wim van der Linden wrote a scathing review of our book in the International Journal of Testing. The journal editor, Bruno Zumbo, insisted that I had no right of reply in the matter. As an aside, let me note I have been approached by a number of Wim's colleagues since then who, to a person, have apologised for the intemperate review

stating that Wim was really a nice guy, and that they couldn't imagine what had got him so steamed up about ARM 1.

3. Well, of course, for many, the problem was the book's unbridled success. While all the important Rasch measurement people were writing their technical tomes aimed at students of psychometrics who were mathematically literate, some outsider without any formal training in probabilistic latent trait models, had jumped the gun, publishing the book that, in retrospect, they wished that they had written.

But, was it a case that they didn't write the book, or the more fundamental case that they couldn't write the book?

Quite sheepishly, I confessed to Mike Linacre that I had deliberately kept secret the writing of ARM from him and Ben, because I didn't want them to do a better job with the same idea. He replied, "Oh, don't worry about that. Ben and I have tried to write that book a number of times; but by the time we got to the end of the first chapter we realised that once again our writing was so full of mathematics that our intended audience would not have any access to it."

Van der Linden had said that ARM was totally unnecessary; *Best Test Design* was the most suitable introductory text to Rasch measurement:

> I fear that the readers for which this book is written may have serious difficulty understanding what is modern measurement. Or, worse still, that, if this is the only text on measurement they ever read, they may walk about with serious misconceptions for the rest of their lives. I have difficulty recommending this book as an introductory text to modern measurement. Readers will be much better off with a balanced, elementary text as Hambleton, Swaminathan, and Rogers (1991). If they want to focus solely on the Rasch model, my preference would be Wright and Stone (1979), or, better still, the introductory chapter and chapters 5 and 6 in the original text by Rasch (1960).

Quite frankly, many distinguished published psychometricians just don't get it. The following story might be apocryphal, but it makes the point: A famous European expert in modern measurement was invited to speak to a group of teachers undergoing postgraduate studies in a Perth, WA university. He was advised by his host to make sure that his presentation was appropriately basic for such a psychometrically unsophisticated audience. The sponsoring academic later remonstrated with the expert presenter for starting with the opening, "Let x be a vector." The visiting expert was horrified at the criticism: "Let x be a vector", he reiterated. "What could be more basic than that?"

In this context, an important Rasch colleague likened the unexpected success of ARM to that of one of the classical statistics texts; saying, "The Crocker and Algina *Introduction to Classical*

and Modern Test Theory, from 1987 is regarded as the book that everybody wished they had written, but none had." Jeffrey K Smith's review of that edition is apt on the key point:

> The writing style of the authors is very appealing, especially given the complexity of the material to be communicated. For example, a description of the true score model begins with: "The classical true school model is one of the most significant issues from British psychologist Charles Spearman's fascination with the concept of correlation" (p. 106) instead of the "Let *x* be a random normal deviate defined over . . ." approach that students find so difficult.

For the foreword to ARM 2, Mike Linacre wrote:

> Benjamin Wright, the leading advocate of practical Rasch measurement, would convince audiences of the value of Rasch measurement. But even his books proved perplexing for the neophyte. It seemed that by the time the Rasch practitioner had become competent enough to write a book, those early hurdles that are always encountered in using an unfamiliar methodology were forgotten.
>
> Trevor Bond and Christine Fox remembered, and wrote this book—now revised to excel its former self! Purists may decry the lack of technical sophistication within these pages––but that exactly accords with the needs of the neophyte. This book filled a gaping void. Its success has motivated others to follow its lead, but so far, none has been so ruthless in rejecting the shackles of Greek letters, in-group jargon, statistical niceties, and the constraining historical precedents of a century of psychometrics.
>
> Mike Linacre 2007

Back to Basics

In my very early days as a regular visitor to Chicago, one of the very important RM steam engines complained to Ben Wright: "I don't see you why you are so enamoured with Trevor Bond, he doesn't know very much about the Rasch model." I am told that Ben replied, "True. But we can fix that. Look at how he can teach." The success of ARM 1 apparently upset another very important steam locomotive. Seems that I should have invited him to be the co-author, not another newbie like Christine Fox. In fact, he was so important that I would have never even thought to ask him to participate in such a lowly venture as mine. Christine copped some flack; and you will be hard pushed to find a single citation of my work in his subsequent volumes.

I became an active member in the Rasch Measurement SIG of AERA; presented at MOMS; regularly participated at IOMW, and co-organised with William Fisher the second of the IOMW meetings held in Los Angeles. That workshop and the associated AERA meeting became important cornerstones of *rapprochement* between US and European Rasch measurement practitioners. Ben Wright's combative style of presentation and interaction had alienated quite a

few in the movement, including key European figures. I was able to work with Mattias von Davier to welcome to IOMW XI New Orleans, (April 6-7, 2002, following AERA-NCME) Jürgen Rost (Ger.), Cees Glas (Neth.), Claus Carstensen (Ger.), from Europe as well as Margaret Wu, Christian Monseur, Ray Adams and Joel Michell from Australia. That was another benefit of being a little steam engine. ARM 1 (2001) was yet to make its big splash, so I was irrelevant to the history of the interplays between all the big steam engines. In my apparent naivety I was able to invite both groups to participate.

In the absence of a local organiser for IOMW in San Diego in 2004, I took on the organisation of that workshop in Cairns, Australia. The strong support of many first-time delegates from Southeast Asia convinced me to change my focus from the east of the Pacific Ocean to the west. PROMS – the Pacific Rim Objective Measurement Symposium – was born with its first meeting in Kuala Lumpur in 2005, at about the same time that I was headhunted for a position as Professor and Head of Department in Hong Kong.

That move revealed an unexpected benefit of the Bond & Fox didactic writing style. French is my second language, and I have always been sensitive to the issues that grammar and sentence structure raise for those working in their second language. Most people in Southeast Asia who were interested in learning about Rasch measurement and applying it to their own doctoral or professional research had English as their second language. Many reported that the expositions in ARM were easier to grasp for them.

LEA boss, Larry Erlbaum took me to dinner at AERA in San Diego where I promptly tried to twist his arm for an increase in royalties for the 2007 ARM 2. Larry warned me that in his experience many second editions didn't live up to the promise of the first. He concluded, in that case, he would happily increase my royalties to the top level, knowing that if the second edition bombed as many did, it wouldn't cost him much; and if the second edition was successful he wouldn't mind the payment. In return, I had to promise to spend the money on French Bordeaux. Luckily for my wine cellar, Edition 1 sold 3,698 copies, and Edition 2 sold 6,118. On a later visit to Chicago, I made arrangements to visit Ben Wright at his home sometime after the terrible health incident that retired him from his intellectual life. I was lucky to catch him on what seemed to be a very good day. I reminded him that *Applying the Rasch Model* had been published and that I had sent him a copy. When he said that he didn't remember receiving it, I excused myself to his professional library and found his ARM copy for his inspection. Ben was puzzled as he flipped through the book, but paused each time he looked at one of my new variable maps, he was prompted to say, "Ah, yes. I remember; I do remember." He also claimed to remember co-author Christine Fox very well.

There was an obvious heir apparent waiting in the wings to inherit Ben's mantle as the champion of the Rasch model in the English-speaking world. Then, quite unexpectedly, the

almost completely unknown little steam engine, also from Down Under, produced the Rasch measurement book that would swamp all other publications. Of course, that was never my intention. But, I smile. And, I have paid the price for that intrusion, too.

The Pacific Rim Objective Measurement Symposium - PROMS has been held every year, pandemic years excepted, since 2005. Almost every big Rasch locomotive has been invited as a keynote speaker once. And most of them visited only that once, but others have managed a long-term commitment to changing psychometrics in eastern Asia. The most notable of them are Jack Stenner – MetaMetrics, George Engelhard – US, Mike Linacre – WinSteps, and Wang Wen Chung – Taiwan and HK. Rob Cavanagh, Australia; Zhang Quan John, China; and Yan Zi, Hong Kong have also made long term commitments to RM and PROMS.

George Engelhard is the internationally recognised colleague who has persisted. Given my experience at PROMS with the other important steam engines of Rasch measurement, I wondered why George even bothered to turn up. It soon became obvious. George exemplified my ideals about communication, passion for teaching, and a willingness to meet graduate students and early career academic staff on their own levels. He has proved to be a wonderful mentor at PROMS and has he served as President for the past three years.

> The Aryadoust, Tan and Ng (2019) paper reports on central role of teaching:
>
> Engelhard (2019, personal communication) stated "I believe that my research is being cited because I write as a teacher [. . .] Measurement is viewed as complex and statistical, while I view measurement as essentially a facet of clear thinking about the constructs in our theories [. . .] I have tried to [. . .] introduce the use of meaningful and invariant scales in numerous fields."
>
> This resonated with Bond's (2019, personal communication) idea about the success of Bond and Fox's (2001, 2007, 2015) book, stressing that making an attempt to communicate the properties of the Rasch model to the evergrowing field of psychology, medicine, and social sciences is a key factor in attracting more scholars to this field.

The recent figures I have seen from Taylor & Francis show that ARM 4 sales in Asia outstrip those in the US. As the Little Steam Engine That Could, author of ARM, I have noticed in the past two decades, significant growth in Rasch measurement applications in HRQoL (health related quality of life) research in particular, and in health fields more generally; as well as in educational psychology, and education, more broadly.

Quantifying Success

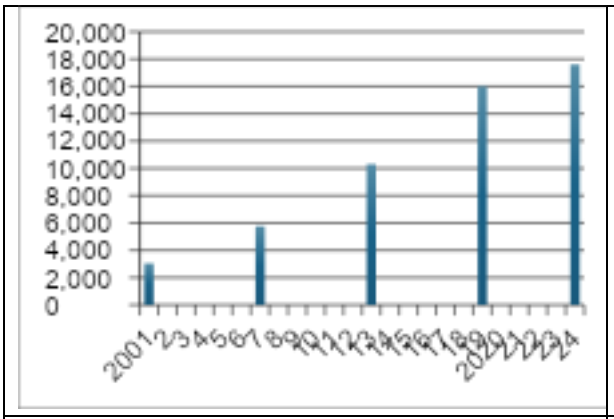
Surely an RM SIG Keynote should include at least some attempt at quantifying the underlying variable, but I won't go further than producing mere counts.

In 2019, Aryadoust, Tan and Ng published "A Scientometric Review of Rasch Measurement: The Rise and Progress of a Specialty" in *Frontiers in Psychology*:

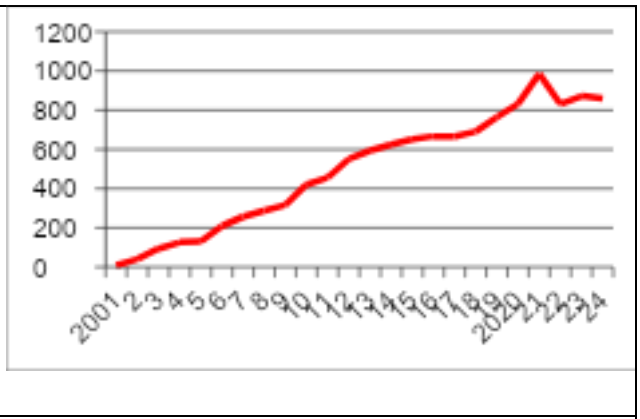
> Publications with high levels of strength signify major milestones in the development of Rasch measurement. Two notable publications by TG Bond (strength = 133.0131 and 88.1181, respectively) had the largest magnitudes of document bursts, with spans lasting 5 and 6 years, respectively. This suggested that works by TG Bond were not only highly influential but had greatly contributed to the development of Rasch measurement.
>
> The most influential publication in Rasch measurement was the volume "Applying the Rasch Model: Fundamental Measurement in the Human Sciences" by Bond and Fox (2007), followed by its earlier edition (Bond and Fox, 2001). The third edition of the book ranks fourth after Tennant and Conaghan's (2007) article in *Arthritis and Rheumatism-Arthritis Care and Research*. From this perspective, Bond and Fox's book was exceptional as its three editions were among the top four influential publications in the Rasch measurement field.

Needless to say, I was more that delighted by that. I had always kept an eye on the growth of references to the Rasch model via Google scholar, along with the corresponding increase in the number of ARM citations by including summary graphs in ARM3 and ARM4.

Figure 1
Graphs of Annual Rasch Model References and Annual ARM Citations



Rasch model references 2001 - 2024



Annual ARM citations

In early 2025, Google scholar citation counts for Rasch model texts looked something like this:
Rasch Models for Measurement 2421
Best Test Design 4281

Probabilistic Models for Some Intelligence and Attainment Tests 15031
Constructing Measures 2981
Invariant Measurement 675
Rasch Analysis in the Human Sciences 1726
Applying the Rasch Model 12080

Another leading Rasch colleague suggested I should make the ARM comparison more broadly and asked: "So where does ARM (12080 citations) sit in the pantheon of the great classics of psychometrics?" I had *never* even thought of making such a comparison. For example, *Statistical Theories of Mental Test Scores* by the late and honoured statisticians and psychometricians, Frederic M. Lord and Melvin R. Novick: "No single book since 1968 when Lord & Novick first appeared has had a comparable impact on the practice of testing . . . " has been cited by 12772. The aforementioned *Introduction to Classical and Modern Test Theory* by Crocker & Algina (1986) has 11472 hits. But, of course, we can't overlook *Statistical methods for research workers*, the masterpiece by Ronald A. Fisher, cited by more than 30000.

The RM SIG Award

I am delighted to have been honoured with an association to Ben Wright who, above all, was a passionate teacher. The dedication in ARM2 reads:

To Ben Wright,
Provocative, persistent, and passionate
TGB & CMF

And me? *The Little Steam Engine that Could*? My sister-in-law describes me, thus:
"You are a legend in your own lunchbox."

Trevor Bond

References

Aryadoust. V, Tan, H.A.H. & Ng, L.Y. (2019) A Scientometric Review of Rasch Measurement: The Rise and Progress of a Specialty. *Frontiers in Psychology*. *10*:2197. doi: 10.3389/fpsyg.2019.02197

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). London: Lawrence Erlbaum.

Crocker, L.M. & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Holt, Rinehart, and Winston.

Fisher, R.A. (1925*) Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.

Lord, F.M. & Novick, M.R. (1968). *Statistical Theories of Mental Test Scores*. Menlo Park: Addison-Wesley.

Smith, J. K. (1987). Review *of Introduction to Classical and Modern Test Theory*, by L. Crocker & J. Algina. *Journal of Educational Measurement*, 24(4), 371–374. http://www.jstor.org/stable/1434888

van der Linden, W. J. (2001). Applying the Rasch Model. *International Journal of Testing, 1*(3–4), 319–326. https://doi.org/10.1080/15305058.2001.9669478