

SPRING 2023, VOL. 36 NO. 1; ISSN 1051-0796



RMT

RASCH MEASUREMENT

- ▶ Overview of this Issue of RMT -- Stefanie A. Wind
- ▶ Invited Speech from 2022 SIG Business Meeting: *Developing a Rasch-Guttman Scenario Scale to Capture the Complexity of Equity-Centered Teaching Practice: A Reflective Essay* -- Wen-Chia Claire Chang
- ▶ Research Note: *Modèle de Rasch/modelo de Rasch/Rasch model/Rasch-Modell...* -- Trevor Bond, Sébastien Béland, and Hudson Golino
- ▶ Book Announcement: *Appliquer le modèle de Rasch- Défis et pistes de solution* -- Sébastien Béland, and Éric Dionne
- ▶ Conference Announcement: IOMW 2023
- ▶ Rasch-Related American Educational Research Association (AERA) Presentations and events

Transactions of the Rasch Measurement SIG
American Educational Research Association

Overview of The Issue

The Spring 2023 issue of RMT includes a mix of research notes and announcements that may be interesting to Rasch measurement researchers.

First, we have provided a copy of Dr. Wen-Chia Claire Chang's presentation from the 2022 Rasch SIG Business meeting on her award-winning paper related to the Rasch/Guttman-based Scenario scale development approach. We included a link to Dr. Chang's slides for reference.

Next, we included a commentary piece from Trevor Bond, Sébastien Béland, and Hudson Golino related to publication languages for Rasch measurement research.

Third, we have included an announcement about a new book on Rasch measurement that was recently published in the French language.

We end the issue with information about upcoming Rasch-related conferences and events.

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 June 1, 2023. Please contact Stefanie at swind@ua.edu to submit something for inclusion.

Sincerely,
Stefanie A. Wind

Rasch Measurement Transactions

www.rasch.org/rmt

Copyright © 2021 Rasch Measurement SIG, AERA

Permission to copy is granted.

Editors: Leigh M. Harrell-Williams
& Stefanie A. Wind

Email submissions to: swind@ua.edu

RMT Editors Emeritus: Richard M. Smith,
John M. Linacre, &
Ken Royal

Rasch SIG Chair: Jue Wang

Secretary: Eli Jones

Treasurer: Dandan Liao

Program Chairs: Manqian Lia &
Audrey Conway Roberts

Developing a Rasch-Guttman Scenario Scale to Capture the Complexity of Equity-Centered Teaching Practice: A Reflective Essay

The text below was taken from Wen-Chia Claire Chang's invited speech given at the 2022 Rasch Measurement SIG Business Meeting in April 2022. The PDF file of the slides can be found at https://www.rasch.org/AERA2022_RaschSIGPresentation.pdf

Note: The bold blue text indicates the slide numbers in Dr. Chang's presentation.

(Slide 1) This reflective essay largely draws from the invited speech for the Georg William Rasch Early Career Award given at the 2022 Rasch SIG business meeting. The nominated article (Chang et al., 2019) presents the work of developing the Teaching Equity Enactment Scenario (TEES) Scale that measures the complexity of equity-centered teaching practice using the Rasch/Guttman-based Scenario (RGS) scale development approach (Ludlow et al., 2014; Ludlow et al., 2020, 2021). The work of developing the TEES Scale was initially built on a doctoral project and has been ongoing and evolving for several years. As I was preparing for the speech, I decided to approach it from an inquiry stance

(Cochran-Smith & Lytle, 2001), borrowing insights from practitioner research in the field of teacher education that critiques and challenges inequities in school, society, and hierarchies in knowledge and knowledge construction. My intention was to reflect on the key questions, theories, and perspectives that drive this work forward and challenge me to rethink notions of measurement validity and validation along the way. From an inquiry stance, I also intended to speak back to the conventional practices of instrument development and validation through an equity- and culture-centered lens.

To this end, I first discuss recent scholarships on validity and validation that seek to challenge traditional validity theories and methods and/or to expand the notions of validity by attending to issues of culture and equity. These theories and perspectives together formulate the reflective lens through which I examine my own instrument development experience. Second, I elaborate on the step-by-step processes of developing the TEES Scale following the RGS methodological framework (Ludlow et al., 2020, 2021) and the psychometric results. Third, I discuss insights and previously unidentified blind spots as I reflect and inquire into the process of developing the TEES Scale.

Critical Reflection and Inquiry: Theories and Perspectives

(Slide 2) My journey of critically examining the notions of measurement validity has been guided by multiple seemingly unrelated milestones, landmarks, and events. As I reflected on the work of developing and validating the TEES Scale—how it began and has become—the initially scattered puzzles eventually became interconnected and formed a sensible picture. In the space below, I invite others into the major turns and milestones of this journey.

(Slide 3) When I was pursuing my doctoral study in measurement, evaluation, statistics, and assessment, I was also working closely with teacher educators who are committed to preparing teachers to teach with a goal to sustain cultural pluralism, challenge systemic inequities, and advocate for social and racial justice. I constantly encountered the disconnection between the two worlds within which I was operating. That is, within quantitative traditions and the associated positivist or postpositivist paradigms, notions of culture, race, and context are rarely dealt with in depth beyond the use of categorical variables. Within the critical traditions, qualitative methods are often used to center the experiences of the

marginalized; quantitative methods and measurement have played little role in advancing equity and social justice and worse, have often been used to perpetuate marginalization and deficit discourse (Flynn, 2015; Russell et al., 2022). Juggling between the two worlds, developing an instrument that can capture the complex nature of enacting equity-centered teaching practice for teacher learning and research/evaluation purposes seemed to be the answer at that time.

(Slides 4, 5) In social science, self-report surveys or instruments often use one-statement, discrete, Likert-type items that are highly correlated with each other to capture complex social and psychological constructs. While a long-standing practice that aims to avoid double-barrel and to increase item clarity, they are not without critiques. For example, the short-stemmed items might not provide sufficient context or complexity to engage participants in processing and responding to the survey tasks, and they are easily subject to social desirability or result in participants automatically giving positive responses regardless of items (Friborg et al., 2006; Furnham, 1985; Rossi et al., 2013; Wallander, 2009). Dissatisfied with the conventional items, I discovered the RGS

approach, first developed by Ludlow and colleagues (2014) to measure the construct of work engagement in later-life activities. As shown in Slide 5, the RGS approach uses comparative survey tasks to engage participants in a deeper and reflective response process. Most importantly, the scenario-style items provide rich, qualitative interpretations of participants' scores along the hierarchical progression of the construct (i.e., variable), which can in turn facilitate actionable next steps that help move participants' status along the variable (Ludlow et al., 2020, 2021).

(Slide 6) As I embarked on the journey of developing an instrument to capture the complexity of enacting equity-centered practice for teacher professional learning and research/evaluation purposes, the question of validity and validation looms. Beyond the beautiful psychometric properties that we are often very excited about, what does it mean by a scale being “validated”?

(Slide 7) I came across the work by Andrew Maul (2017) on rethinking traditional methods of survey validation. Maul (2017) used a series of three studies to illustrate how traditional survey validation procedures do not necessarily provide rigorous evidence for the quality and rigor

of measures as we think they should.

Through the three studies, Maul illustrated that even when the items were revised to be noninterpretable or not understandable by the survey participants, the results from these validation procedures could still be favorable looking. Maul argued that the problem may be more conceptual than technical. One of the issues may be that “the process of ‘validating’ a measure seems to be thought of by many as separate from the process of defining the attribute to be measured and articulating hypotheses concerning the nature of the connection between variation in the attribute and variation in the outcomes of the proposed testing procedures...In the absence of such theory...the results of such analyses are difficult to confidently interpret” (Maul, 2017, pp. 59–60). My encounter with this work and others, such as Kane’s argument-based validity (2013), broadened my understanding of validity and validation and offered me a renewed understanding of the RGS approach I took in developing the TEES Scale, as I will elaborate further in the subsequent section.

(Slides 8, 9) As I delved into the literature on measurement validity and validation, I was once again reminded that the work of instrument development and

validation, in which researchers make choices of construct theories/literature that inform the test content, determine the appropriate methods used and who can participate and at which stages, and present the interpretations of test scores for some purposes, is never value- and culture-free (Flynn, 2015; Salazar, 2018). Dixon-Roman and Gergen (2013) rightly pointed out the value- and culture-laden nature of measurement:

“What is counted as ‘fact’ is always located within a community tradition or cultural history and reflects its particular values. This is also to point out that measurement too is the product of sociocultural process... We must ask, then, whose values are inherent in the measurement, and why should they be privileged over the many other values that circulate within the society?” (p. 17).

Nevertheless, conventional validity frameworks and validation studies have predominantly focused on methodological issues with little attention to culture—the inclusion/exclusion of diverse standpoints, epistemologies, methods, and perspectives, in both the content and process of measurement development, as well as the social consequences resulting from such inclusion and exclusion. Recently, there has

been a growing movement that pushes researchers and practitioners to grapple with what it means and takes to attend to culture and to make equity a central focus in the field of measurement and evaluation (e.g., see Dixon-Roman & Gergen, 2013; Kirkhart, 2013; Randall et al., 2022; Schwandt & Gates, 2016).

For instance, in the field of program evaluation, Kirkhart (1995, 2013) proposed the notion of multicultural validity to focus on issues related to pluralism and diversity in evaluation studies. Multicultural validity is not a new or different kind of validity; rather, Kirkhart (2013) argued that it expands the notion of validity by putting validity or validity argument-making back in culture. When validity is understood as culturally located, legitimate validity evidence needs to recognize and include diverse standpoints, interests, and perspectives of evaluation stakeholders, particularly from those previously marginalized. Kirkhart (2013) defines multicultural validity as “the accuracy and trustworthiness of understandings and actions across multiple, intersecting dimensions of cultural difference” (p. 2). By balancing the interests and voices of diverse communities, multicultural validity provides a necessary but insufficient condition to

achieve social justice in society. The multicultural validity framework includes five dimensions—theoretical (selection and interpretation of underlying theoretical foundations), methodological (use and cultural inclusivity of methods), relational (relationships among the researchers and the researched), experiential (recognition and inclusion of diverse, lived experiences of stakeholders), and consequential (social consequences in terms of promoting justice or perpetuating marginalization), and each calls for a different type of evidence to support or challenge the dimension. Altogether, this framework not only addresses the conventional aspects of validity but also provides normative critique of the culture- and value-laden aspect of validity argument-making in evaluation.

From within the field of measurement/assessment, some scholars (e.g., see Cushman, 2016; Moss, 1992; Slomp et al., 2014) have challenged the narrow conceptions of validity and validation methods and called for expansion and decolonization of validity. For instance, in reviewing the changing conceptions of validity criteria, Moss (1992) concluded that the dominant validity criteria, which emphasize the minimization of factors that produce construct irrelevant noises to

maximize the truthfulness and comparability of the resulting test scores, favor standardized forms of assessment. However, such epistemological stance of viewing validity and validation turns a blind eye toward the power and inherently political nature of assessment—who can determine the construct, select methods to collect evidence and from whom, and have the authority over the interpretations and decisions. Moss (1992) argued that developers and users of assessment need to expand the “repertoire of epistemological strategies and consider alternative models for warranting validity conclusion” (p. 253). Moreover, there is a need to expand the conception of validity beyond a technical, methodological focus by asking normative inquiries, such as why certain methods are privileged in validation and how such privileging has effects on assessment participants and communities.

Similarly, in discussing the imperial legacy of the prevailing concept of validity, Cushman (2016) laid out how validity is an instrumental tool that determines what counts as evidence and what inferences can be made by such evidence, which, in turn, is used to create and maintain the hierarchies in knowledge and knowledge production, manage peoples, differentiate rights,

contribute to social stratification, and uphold institutional structure. To address the hierarchies, injustices, and inequitable opportunities and outcomes created and perpetuated by assessment outcomes that are deemed “valid”, Cushman (2016) proposed that there is a need to “see validity evidence tools not as a way to maintain, protect, conform to, confirm, and authorize the current systems of assessment and knowledge making, but rather as a way to better understand difference in and on the terms of the peoples who experience them” (p. 4).

Built on these earlier scholarships and informed by justice-oriented perspectives and critical race theory, Randall and colleagues (2022) proposed a justice-oriented, antiracist validity (JAV) framework, challenging validation practices that assume race neutrality yet “continue to (re)produce racism through the uncritical promotion of white supremacist assessment practices” (p. 3). Going beyond the traditional view of validity, the JAV framework encompasses a set of critical questions that guide assessment development from an anti-racist standpoint. The set of key questions are organized in terms of five elements: (a) construct articulation, (b) test content, (c) response

processes, (d) internal structure and relations to other variables, and (e) consequences. These questions examine whether and how lived experiences, cultural and linguistic perspectives of marginalized stakeholders are recognized and included and whether/how anti-racist perspectives or approaches are integrated pertaining to each of the five elements. Examples of questions are as follows: are marginalized stakeholders involved at every stage of the construct definition and refinement stage? (Construct element); and have a wide range of interpretations been considered that acknowledge the different ways of knowing, thinking, and experiencing of Black students? (Response processes). Randall et al. (2022) suggests that the framework can be used to guide the collection of validity evidence from the initial stage of assessment development, as well as to critically inquire into the assessment development process retrospectively, to unearth racist ideas embedded in the assessment design and development process.

(Slide 10) Together, these critical perspectives form my “reflective lens” through which I inquire into the work of developing the TEES Scale described below.

Applying the RGS approach to Develop the Teaching for Equity Enactment Scenario Scale

(Slide 11) The Rasch/Guttman-based scenario (RGS) approach integrates Rasch measurement (Rasch, 1960/1980) and the design component of Guttman's facet theory (Guttman & Greenbaum, 1998). Rasch measurement principles guide the scale development. Specifically, items are developed to measure the *variation* of a clearly defined *unidimensional* construct. Items are intentionally developed to capture the *hierarchical progression* of the construct from the easier to the more difficult level with sufficient *spread* among them. Additionally, items are assumed to have an *equal relationship* to the construct, and one's response to an item is not dependent on the response to another item (i.e., *local independence*). A construct theory-informed hypothesis on the expected item locations must be specified as the a priori, and empirical data are used to check the *data/theory match*.

Guttman facet theory design facilitates the systematic construction of scenarios. Most importantly, it enhances the conceptual clarity of the construct. Using Guttman's facet theory design, one must clearly define the construct by identifying a

set of distinct and interrelated elements/characteristics (i.e., facets). Together, the facets define the content universe of the construct. Within each facet, there are variations or ranges to be defined, which is referred to as facet "levels". Guttman's mapping sentence technique is then used to connect facet level descriptions and to construct the scenarios. In the RGS approach, the "facets" are not separate factors/predictors/dimensions to be investigated. Rather, facets are interconnected like an interwoven cable in a complex manner that form the single construct.

(Slide 12) Integrating Rasch measurement and Guttman's facet theory design, an RGS scale has the features of *specific objectivity* in Rasch-based scales (Rasch, 1960/1980; Wright, 1967; Wright & Masters, 1982) and *interpretability* and *actionability* enhanced through scenario-style items (Ludlow et al., 2021). That is, not only does an individual's scale score represent something within them that is not dependent on the difficulty levels of a given set of items, but such a score can also have a richer, qualitative interpretation in relation to the construct via scenarios. Ludlow and colleagues (2020) present the RGS methodological framework that specifies the

seven-step, systematic, yet nonlinear, iterative process of developing an RGS Scale. Importantly, the bidirectional arrows illustrate how a scale developer may move back and forth among a particular set of steps before moving forward. Below, I elaborate on the step-by-step procedure of developing the TEES Scale.

Development of the Teaching Equity Enactment Scenario Scale

(Slides 13, 14) The first step is to define the construct of teaching for equity by systematically documenting the identified literature and/or lived experience manifesting the construct. To identify the relevant literature, theories of equity-centered teaching (Cochran-Smith et al., 2016) that are grounded in a critical sociohistorical view of equity informed the selection of five international syntheses about teaching practices that contribute to broadly defined student learning, have a positive impact on diverse learners, and reflect a complex view of teaching.¹

(Slide 15) The second step requires researchers to fully immerse themselves in the literature, determine the facets, and generate thick narrative descriptions that

portray teaching practice associated with each facet. In this step, I conducted an iterative content analysis of the five international syntheses and clearly documented the coding process. In the case of the TEES Scale, a preliminarily identified six principles of practice for equity (Grudnoff et al., 2017) deductively guided the coding process. Further and more importantly, the content analysis process inductively clarified, revised, and strengthened the understandings of the six principles, which became the six facets of practice for equity. The six facets are (1) selecting worthwhile content and designing and implementing learning opportunities aligned with valued outcomes; (2) connecting to students as learners and their lives and experiences; (3) creating learning-focused, respectful, and supportive learning environments; (4) using evidence to scaffold learning and improve teaching; (5) taking an inquiry stance for further professional engagement and learning; and (6) recognizing and challenging classroom, school, and societal practices that reproduce inequity (Chang et al., 2019).

¹ The international literature include three New Zealand Best Evidence Syntheses (Aitken & Sinnema, 2008; Alton-Lee, 2003; Anthony & Walsaw, 2007), the Teaching and Learning Research Project in the UK (James & Pollard, 2006), the Measures of Effective Teaching in the U.S. (MET Project, 2013), Te

Kotahitanga Effective Teaching Profile (Bishop, Berryman, Cavanagh, & Teddy, 2009), and the Center for Research on Education, Diversity, and Excellence's five standards for effective pedagogy (Dalton, 2007).

(Slides 16, 17) The step of defining a construct and describing its characteristics is often only briefly discussed in scale development studies. Here, I would like to address who was involved, how the construct mapping was done, and what was ultimately measured (and valued) in further detail. Keeping the preliminarily identified six principles in mind, I systematically organized pedagogical practices drawn from each of the five syntheses into one and only one principle. For instance, in reviewing Danielson's framework as one of the selected syntheses, using a spreadsheet software tool, I categorized Domain 3's (Instruction) subdomains 3d and 3e regarding using assessment to facilitate and monitor student learning and adjust instructional approaches under Facet 4, *using evidence to scaffold learning and improve teaching*. During this coding/sorting process, I constantly worked with five experienced teacher educators to review and confirm the analysis. We clarified and negotiated our understandings of each facet, made collective decisions on aspects where disagreement occurred, and documented our decisions and rationales. This iterative process of analysis allowed me to have an in-depth understanding of the facets and to develop rich narratives of two

to three pages in length describing each of the six facets. In addition, again, I continued to engage the teacher educators in reviewing, critiquing the rich narratives I developed and documented the decisions and choices we made.

(Slide 18) With an in-depth understanding of the construct and facets, in Step 3, I began to develop descriptions to capture variations within each facet. Specifically, built on the rich description of each facet, I developed shorter, but still rich, descriptions to portray low, moderate, and high levels of practice associated with each facet. Using a similar peer review and critique processes, I shared the early drafts of the narratives with the five teacher educators for them to review and provide feedback. We worked closely in a dialogical manner, and this iterative, extensive review and feedback processes generate a negotiated, in-depth understanding of facets and within facet variations. In the case of the TEES Scale, the first three steps took approximately one and half years.

(Slides 19, 20) With these rich facet-level descriptions, in Step 4, I encountered multiple challenges: How do I combine the facet-level descriptions to construct scenarios so that they represent the complex construct, are theoretically reasonable, and

are not overwhelming to participants? Given the six facets and three levels, I made several key decisions here in consultation with the five teacher educators and a measurement scholar. First, I did not combine the facet-level descriptions in every possible way (six facets, each with 3 levels, resulting in 729 possible combinations). Rather, I only selected facet descriptions from just the highest, the moderate, and the lowest to construct plausible scenarios that capture three distinct levels along the construct continuum. This decision was based on the understanding of the construct and a need to obtain a proof of concept first before trying to capture the subtleties in between. Second, it is impossible to include all six facets in one scenario because participants would be overwhelmed. I decided that three facets would be feasible and that Facet 6, *recognizing and challenging inequities*, must be present in all scenario combinations, aligning with the construct theory. As such, Facets 1 to 5 are to be systematically selected so that some of the five facets overlap between scenarios. This approach resulted in five scenarios capturing each of the three levels along the construct continuum. For instance, Scenarios F126HHH (i.e., Facets 1, 2, 6, “H”ighest level), F236HHH, F346MHH, F456HHH,

and F156HHH capture the high level of enactment of practice for equity. Scenarios based on the same facet combinations were developed to capture the Moderate (M) and the Lower (L) levels of enactment of practice for equity. Altogether, these decisions resulted in a total of 15 scenarios to be developed next.

(Slides 21, 22, 23) In Step 5, Guttman’s mapping sentence technique provides a tool to systematically connect facet-level descriptions and construct scenarios. The mapping sentence structure includes the formal parts that are to be filled with facet descriptions and informal parts to connect and give context to the facet descriptions. Slide 21 presents an example of mapping sentences and a scenario developed by using the mapping sentences. In this scenario, the teacher is named “Maria”. The high-level description for Facet 1 is “high expectations for all and clear communication of the learning goals”, which I inserted into the first sentence. Using the same mapping sentence structure, I developed the low-level scenario by replacing it with the lower-level facet descriptions for Facets 1, 2, and 6 as shown in Slide 22. Although the mapping sentence technique provides a systematic way to construct scenarios, it can also make

scenarios repetitive and not engaging. To improve the scenarios, I engaged educators in rewriting the scenarios through their lived experiences, reviewing and providing feedback on the scenarios through focus groups and individual written responses, pilot testing, and think aloud exercises. In Slide, 23, the images showing broken (before) and functioning (after) robots illustrate how I engaged key stakeholders and participants in this process to improve the scenarios.

(Slide 24) Next, Step 6 is concerned with scale instructions and scenario response options. The RGS scale uses a comparative response approach. To respond to the scenarios, participants are guided to reflect on their own practice and compare it against the teacher's practice in the scenario. They are then asked to select a response that is closest to their self-assessment. Choosing *About the same* means that participants align their practice to the practice of a specific scenario; choosing the *Slightly lower* or *Much lower* ratings means that participants consider themselves do not fulfill, do less, or are not there yet compared to the scenario; and, choosing the *Slightly higher* or *Much higher* ratings means that participants consider themselves do more and beyond the practice compared to the practice in the

scenario. I expect that scenarios representing higher levels of enactment will be harder for respondents to reply with about the same or higher ratings than scenarios of lower-level enactment. I expect that the calibrated items will loosely form 3 clusters along the hierarchical continuum of the construct with five high-level scenarios on the top, five roughly in the middle, and five at the bottom. This is the theory-informed hypothesis.

(Slides 25, 26, 27, 28) Step 7 checks whether the data confirm the Rasch model expectations, the hypothesized construct continuum specified earlier. Slide 26 presents the variable map based on the pilot study (N = 73), which shows that the high-level items are more difficult, and the low-level items are easier for participants to identify with. The data confirm the a priori theory. However, the variable map also shows that the items are not evenly spread along the construct continuum, which is as expected. Based on this result, it would be useful to have some items that capture the subtleties in between the three distinct levels. In the first variable map of Slide 27, the items in red circles are items that I identified and revised the specific facet-level descriptions to make the items more or less difficult. The arrows indicate the expected

changes. The revised variable map shown on the right based on a new sample of participants ($N = 52$) confirmed the expected changes with a better spread of items along the continuum. In this revised scale, item redundancy, i.e., items measuring a similar difficulty level with one or two standard errors apart from each other, remained. Additionally, multiple applications of the TEES Scale reported respondent fatigue when responding to the full-length scale given the novelty and cognitive demand of the scenario-style, comparative items (Chang, 2021). In the subsequent scale validation and reduction study ($N = 350$), items were identified and omitted based on item-level statistics (i.e., difficulty measures, fit statistics, and significant/large differential item functioning contrast) and content coverage (i.e., the reduction would avoid eliminating multiple scenarios that have the same facet combination). In Slide 28, the items on the second variable maps with red strikethrough were selected and deleted, which led to the final ladder-like scale of 10 items, as shown in the third variable map. The RGS approach enables targeted scale reduction while maintaining the content coverage and the complexity of the construct that I intend to capture.

(Slide 29) The scenario scale is particularly productive in providing rich interpretations and generating reflections/actions that help individuals move along their journey of learning to enact equity-centered teaching practice. To enable score interpretations, on the left of this variable map, both logit and raw scores are plotted, as well as the numeric numbers that correspond to the response options. Once participants' scores along the continuum are identified, scenarios on the right provide a qualitative interpretation of the person's score. In addition, based on my experience of working with teachers, the comparative, reflective nature of the survey tasks can facilitate teachers in identifying areas for improvement that help them grow along the journey of learning to enact equity-centered teaching practice.

Reflections and Implications

(Slide 30) Thus far, I have discussed the work of developing, validating, and enhancing the TEES Scale by using the RGS approach. As I look back and reflect on this journey, there are some lessons learned regarding matters of validity, culture, and equity.

(Slide 31) The work of developing the TEES Scale took approximately three years. For the first two years, all I did was

developing an in-depth understanding of the construct, the facets, and variation within the facet as demanded by the RGS approach. Most importantly, I was not working alone. I came to realize how important it was to work with and engage key stakeholders (in my case, five experienced teacher educators) in developing a shared, in-depth understanding of the construct. This process was undoubtedly time-consuming and at times painful, but a process that helped each of us clarify confusions, discuss disagreements, recognize blind spots, and together reach a negotiated understanding of the construct. Without the work to develop the understanding of the construct along with assessment stakeholders, I would not feel grounded with the psychometric results and know that I can interpret them with a certain level of confidence. I would also like to acknowledge that, despite the theory-informed, purposeful selection of the international literature and the prolonged content analysis process involving key stakeholders, I did not reach out to individuals and communities who tend to be less heard and included in the initial stage of instrument development, such as teachers of color, parents, students, and communities. As a result, the construct captured by the TEES Scale might not reflect the cultural

and racialized experiences and contexts of these stakeholders.

(Slide 32) Along the way, I made many research decisions and choices as discussed earlier. They were guided by the construct theories, measurement principles, and, most importantly, survey participants' and stakeholders' experiences, knowledge, and perspectives. I had educators craft a few scenarios through their experiences and their ways of deliberating their experiences. I also involved teachers, teacher educators, and measurement researchers in cycles of item review, revisions, and implementation for a prolonged period. I recognize that these participants' meaningful and substantive involvement in the process of instrument development and validation contributes to the soundness of validity claims of the proposed interpretations and uses. Through the critical perspectives elaborated earlier, I have also realized that some of my research decisions and practices, although seemingly neutral and concerned primarily with "practical" matters, were nevertheless culturally and racially implicated. For instance, I did not make efforts to seek more teachers of color for my study's participants, perpetuating the norm of "overwhelming presence of whiteness" (Sleeter, 2001) in teacher education in the U.S. context.

(Slide 33) Altogether, this leads to my last reflective thought, which is that instrument development and validation work is inherently located in context and in culture. Therefore, legitimate validity evidence must recognize and include the diverse perspectives and standpoints of assessment stakeholders who are being assessed and/or who will use the information for some purposes. To attend to context and culture in measurement and validation work, I recognize that, through a systematic framework (i.e., the RGS approach), as well as a dialogical approach, to work with assessment stakeholders, instrument development and validation work can achieve a greater level of validity as well as equity. Validity and equity should not be seen as dichotomy, and I, following other scholars, would argue that the notion of validity should be expanded to attend to issues of equity and culture.

(Slides 34, 35) Therefore, if I were to draw my experience and my critical inquiry on using the seven-step RGS development process presented earlier, it would be something like this as shown on the slide, i.e., the RGS approach need to be considered and taken up as situated in sociocultural, historical, and political contexts. As I reflected on this journey, I

recognize that continuous, meaningful engagement with a variety of stakeholders makes me humble as a scale developer, helps me check on my assumptions and blind spots, and makes this work stronger. I encourage assessment developers and users to engage in ongoing critical reflection through the expanded critical validity frameworks that I introduced earlier. In addition, as the movement toward equity grows in the field of evaluation and measurement, this sharing intends to contribute to the continuous dialog on what it means and what it might take to attend to equity and culture in measurement.

References

Chang, W.-C. (2021). Measuring the complexity of equity-centered teaching practice: Enhancement and validation of a Rasch/Guttman scenario scale. *Journal of Applied Measurement, 22*(1/2), 35-59.

Chang, W.-C., Ludlow L. H. , Grudnoff, A., Ell, F., Haigh, M., Hill, M., & Cochran-Smith, M. (2019). Measuring the complexity of teaching practice for equity: Development of a scenario-format scale. *Teaching and Teacher Education, 82*, 69-85.

Cochran-Smith, M., Ell, F., Grudnoff, L., Haigh, M., Hill, M. & Ludlow, L. (2016). Initial teacher education: What does it take to put equity at the center? *Teaching and Teacher Education, 57*, 67-78.

- Cochran-Smith, M., & Lytle, S. L. (2001). *Inquiry as stance: Practitioner research for the next generation*. Teachers College Press.
- Cushman, E. (2016). Decolonizing validity. *Journal of Writing Assessment, 9*(1), 1-6.
- Dixon-Roman, E. J., & Gergen, K. J. (2013). *Epistemology and measurement: Paradigms and practices*. The Gordon Commission on the Future of Assessment in Education.
- Flynn, J. E. (2015). Racing the unconsidered: Considering whiteness, rubrics, and the function of oppression. In M. Tenam-Zemach & J. E. Flynn (Eds.), *Rubric nation: Critical inquiries on the impact of rubrics in education* (pp. 201-221). Information Age Publishing.
- Friborg, O., Martinussen, M., and Rosenvinge, J. H. (2006). Likert-based vs. semantic differential-based scorings of positive psychological constructs: A psychometric comparison of two versions of a scale measuring resilience. *Personality and Individual Differences, 40*, 873-884.
- Furnham, A. (1985). Response bias, social desirability and dissimulation. *Personal and Individual Difference, 7*, 385-400.
- Guttman, R., & Greenbaum, C. W. (1998). Facet theory: Its development and current status. *European Psychologist, 3*(1), 13-36.
- Grudnoff, L., Haigh, M., Hill, M., Cochran-Smith, M., Ell, F., & Ludlow, L. (2017). Teaching for equity: Insights from international evidence with implications for a teacher education curriculum. *The Curriculum Journal, 28*(3), 1-22.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1-73.
- Kirkhart, K. (1995). Seeking multicultural validity: A postcard from the road. *Evaluation Practice, 16*(1), 1-12.
- Kirkhart, K. (2013, April 21-23). Repositioning validity [Invited plenary panel session]. Culturally Responsive Evaluation Association Inaugural Conference, Chicago, IL, United States.
- Ludlow, L. H., Matz-Costa, C., Johnson, C., Brown, M., Besen, E., & James, J. B. (2014). Measuring engagement in later life activities: Rasch-based scenario scales for work, caregiving, informal helping, and volunteering. *Measurement and Evaluation in Counseling and Development, 47*(2), 127-149.
- Ludlow, L., Baez-Cruz, M., Chang, W.C.C., & Reynolds, K. (2020). Rasch/Guttman scenario (RGS) scales: A methodological framework. *Journal of Applied Measurement, 21*(4), 361-378.
- Ludlow, L., Reynolds, K., Baez-Cruz, M., & Chang, W.C.C. (2021). Enhancing the interpretation of scores through Rasch/Guttman scenario scales. In A. Harbaugh & U. Luhanga (Eds.), *Basic elements of survey research in education: Addressing the problems your advisor never told you about*. Information Age Publishing.
- Maul, A. (2017). Rethinking traditional methods of survey validation. *Measurement: Interdisciplinary Research and Perspectives, 15*(2), 51-69.
- Moss, P. A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research, 62*(3), 229-258.

Randall, J., Slomp, D., Poe, M., & Oliveri, M. E. (2022). Disrupting white supremacy in assessment: Toward a justice-oriented, antiracist validity framework. *Educational Assessment*, doi: 10.1080/10627197.2022.2042682.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B. D. Wright. Chicago, IL: The University of Chicago Press.

Rossi, P. H., Wright, J., and Anderson, A. (Ed.) (1983). *Handbook of survey research*. Academic Press.

Russell, M., Oddleifson, C., Kish, M. R., Kaplan, L. (2022). Countering deficit narratives in quantitative educational research. *Practical Assessment, Research, and Evaluation*, 27(14).
<https://scholarworks.umass.edu/pare/vol27/iss1/14/>

Salazar, M. del C. (2018). Interrogating teacher evaluation: Unveiling Whiteness as the normative center and moving the margins. *Journal of Teacher Education*, 69(5), 463-476.

Schwandt, T. A., & Gates, E. (2016). What can evaluation do? An agenda for evaluation in service of an equitable society. In S. Donaldson & R. Picciotto (Eds.), *Evaluation for an equitable society*. Information Age Publishing.

Sleeter, C. E. (2001). Preparing teachers for culturally diverse schools: Research and the overwhelming presence of whiteness. *Journal of Teacher Education*, 52(2), 94-106.

Slomp, D. H., Corrigan, J. A., Sugimoto, T. (2014). A framework for using consequential validity evidence in evaluating large-scale writing assessments: A Canadian study. *Research in the Teaching of English*, 48(3), 276-302.

Wallander, L. (2009). 25 years of factorial surveys in sociology: A review. *Social Science Research*, 38(3), 505-520.

Wright, B. D. (1967). Sample-free test calibration and person measurement. In *Proceedings of the 1967 Invitational Conference on Testing Problems* (pp. 85-101). Princeton: Educational Testing Service.

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

Wen-Chia Claire Chang
National Institute of Education/Nanyang
Technological University, Singapore

**Modèle de Rasch/modelo de Rasch/Rasch
model/Rasch-Modell...**

We begin with a recent anecdote from Trevor:

“I have just arrived back from Hanoi where I conducted an introductory workshop at PROMS for participants of that meeting and to a group of colleagues from VNIES – the Vietnamese Institute for Educational Sciences. As usual, I was assured that those colleagues who had Vietnamese as their first language would be able to follow my presentation in English. And, as usual, I expressed my doubts; I had been in similar situations before. Needless to say, about 20 minutes into the workshop, I paused and asked my Vietnamese colleague Dr Cuong to involve himself actively in the simultaneous translation of my presentation into the local language. The difference was amazing: bored, detached faces were replaced by eager smiles, nods of comprehension and

questions, in Vietnamese, which required further explanations. Almost immediately, local participants with good English comprehension and some knowledge of the Rasch model became involved in answering those questions, even before they got to me or to Cuong. I have had similar experiences in Chinese with Yan Zi, Bahasa with Zali, Pilipino with Michelle Raquel, amongst others. First language experience in learning these key Rasch concepts seems crucial for active workshop participation and understanding.”

This should not be a big surprise for us, because we know a child’s first language and an adult’s second language are both related to the learning process (Hamrick, Lum & Ullman, 2018). We contend that lifting the number of non-English language publications is a key to diffuse the Rasch model more broadly. We also observe that *introduction* to such a complex topic in a first language is useful before diving more deeply in the broader psychometric

literature, which is – essentially – primarily in English.

Luckily, this democratization of the Rasch model has already started. For example, in French, Penta, Arnould & Decruynaere published a book about the Rasch model, in 2005, and, more recently, Dionne & Béland (2023) have edited another. A German text was published by Strobl (2012), and a Portuguese translation of *Applying the Rasch model*- 3 edition (ARM3), was shepherded by Hudson Golino from conception to publication following on from his edited Rasch volume from 2019. Unfortunately, the process is not simple. Enrico Gori and his graduate students translated ARM1 into Italian, and Bond had a doctoral student edit that manuscript, but the project went no further; perhaps because ARM2 was already at the printers. ARM3 in Bahasa Melayu has reached its penultimate manuscript state, but changes in the organisation sponsoring that work have meant the project is also now in limbo. We are told that ARM2 will appear in Mandarin Chinese in the very near future; but we have been told that many times over many years: the manuscript has been with the high-status publisher for about 6 years now.

The publication experience of Hudson is quite interesting. Portuguese is spoken by approximately 258 million people; the ninth most spoken language in the world. The reason for writing books and papers and helping to produce/translate books in/to Portuguese is to allow those who do not have English as their second language to engage with educational/scientific material that might positively impact their careers. The first edition of the book written with Dr. Cristiano Mauro Gomes about Measurement Theory and the Rasch Models sold out in less than two years. It was also well received in 2015. Although there is a significant market in Portuguese-speaking countries for psychometric books, publishers do not seem to promote them well. This is a continuing problem! The more published books about measurement theory and Rasch models, the more likely people of different cultures are willing to engage with this kind of content.

Given that bibliometrics already show that publishing in English - the *lingua franca* of science - is the best way to obtain more citations (Larivière, 2018), non-English publication is not cost free for authors. However, publishing in the national language remains important if we consider

important local issues more closely. When experienced authors choose to publish in English, both the number and the quality of first language publications, are, consequentially, decreased. Interestingly, citations of ARM reported via ResearchGate show considerable publication activity in two Bahasa languages. It is reasonable to contend that first languages better represent our sophisticated constructions of the world, and that first languages better transmit our understandings to others in our context. Surely, science, more broadly, would be in a much better place if more research was published in other languages as well.

Our proposal

Promulgating the Rasch model remains a priority for our community. So, how might we support this for the non-Anglophone world? Of course, we continue to preach: Publish in your national language. Could we encourage authors to also publish a translated abstract / summary of their accepted article in RMT? The French-language journal *Mesure et évaluation en éducation* publishes a special issue every year with some already published articles translated into English. Another idea might be to create an RMT sub-section about the

publications /workshops /anecdotes about the Rasch model in other languages. What are your thoughts about these? What are your suggestions?¹ In any case, we strongly encourage both diversity and flexibility when it's time to promote the Rasch model.

"In diversity there is epistemic strength!"

Oreskes (2019, p. 4)

References

Dionne, É. et Béland, S. (Éds, 2023). *Mettre à profit la modélisation de Rasch: problèmes et pistes de solution*. Québec, Québec : Presses de l'Université du Québec.

Bond, T.G. & Fox, C. M. (2001/2007/2015). *Applying the Rasch model: Fundamental measurement in the human sciences*. (1st/2nd/3rd ed.) Mahwah, N.J.: Erlbaum / New York: Routledge.

Bond, T.G. & Fox, C. M. *Aplicação do modelo de Rasch* (Portuguese Edition) São Paulo : Hogrefe. (Bartalotti, C. Translator, Golino, H.F. Technical Advisor)

Bond, T.G., Yan, Z. & Heene, M. (2020). *Applying the Rasch model: Fundamental measurement in the human sciences*. (4th ed.) New York: Routledge.

Trevor Bond, Sébastien Béland, and Hudson Golino

Golino, H. F., Gomes, C. M., Amantes, A. & Coelho, G. (dir.). *Psicometria contemporânea* :

¹ Please email your ideas and suggestions to Stefanie A.

Wind: swind@ua.edu

Compreendendo os Modelos Rasch. Casa do Psicólogo/ Pearson.

Hamrick, P., Lum, J. A. G., & Ullman, M. T. (2018). Child first language and adult second language are both tied to general-purpose learning systems. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *115*(7), 1487–1492.

Larivière, V. (2018). Le français, langue seconde? De l'évolution des lieux et langues de publication des chercheurs au Québec, en France et en Allemagne. *Recherches sociographiques*, *59*(3), 339–363.
<https://doi.org/10.7202/1058718ar>

Oreskes, N. (2019). *Why trust science?* Princeton: Princeton University Press.

Penta, M., Arnould, C. et Decruynaere, C. (2005). *Développer et interpréter une échelle de mesure: applications du modèle de Rasch*. Mardaga.

Strobl, C. (2012). *Das Rasch-Modell – Eine verständliche Einführung für Studium und Praxis*. Rainer Hampp Verlag.

*Trevor Bond
Sébastien Béland
Hudson Golino*

Book Announcement: *Appliquer le modèle de Rasch- Défis et pistes de solution*



A new book about the Rasch model was published in January 2023. Edited by Éric Dionne (University of Ottawa, Canada) and Sébastien Béland (Université de Montréal, Canada), *Appliquer le modèle de Rasch- Défis et pistes de solution* focus on application and solutions. The idea of a new French language book rises because many professionals and students are more inclined to read in their own language to understand the fundamentals of complex topics such as the Rasch model.

The first chapter is dedicated to the man behind the model: Georg Rasch.

Chapters 2 to 6 are devoted to the theoretical foundations of Rasch modeling, while Chapters 7 to 11 present examples of application. Here is the translated table of contents:

Introduction

I: Theoretical foundations

Chapter 1: Historical look at Rasch's model: A man and its model (Loye)
Chapter 2: Arguing the reliability (of scores and latent trait) of a measurement scale (Béland, Leclerc & Dionne)

Chapter 3: Didactic analysis of the Infit and outfit statistics (Dionne)

Chapter 4: Unidimensionality of measurement scale and good practices on the field (Prosperi)

Chapter 5: What to do in the presence of locally dependent items? (Grondin, Dionne & Béland)

Chapter 6: Exploration of scores for a script concordance test under the microscope of the Rasch model (Dionne, Grondin & Latreille)

II: Technical aspects and applications

Chapter 7: Calibration with Rasch sauce: essential ingredient pre/post design (Chénier & Pilote)

Chapter 8 : Methods for estimating item parameters from the Rasch dichotomous model and their implementation in R (Raïche)

Chapter 9: How to reduce
subjectivity in the assessment of
skills? (Casanova, Aw & Demeuse)

Chapter 10: Analysing items with
ordinal polytomous answer (Béland
& Bourassa)

Chapter 11: A (short) introduction to
the Bayesian Rasch model (Béland,
Chénier & Arias)

Conclusion

Sébastien Béland
Éric Dionne

Conference Announcement: International Objective Measurement Workshop



The International Objective Measurement Workshop (IOMW) 2023 Conference will be held in Chicago on April 11 and 12, just before the AERA/NCME annual conferences, with an additional day of workshops on April 10. The conference will take place in-person at the American Dental Association building, located at [211 E Chicago Ave.](#), a short walk from the AERA/NCME headquarters hotels.

IOMW presents an opportunity for scholars interested in the theory and practice of objective measurement in the human sciences to present research, learn about the most recent developments, and meet with colleagues who share similar interests in an intimate setting.

There will be two exciting keynote speakers: Derek Briggs will talk about his recent book on the history of educational

measurement and will comment on the links to Rasch models. Kirk Becker will speak to issues in automatic item generation and the challenges being faced in this rapidly evolving area.

There will be several software workshops focusing on programs such as Conquest, RUMM, the BEAR Assessment System Software, and others.

The deadline for submitting presentation proposals is past. We have about 40 talks in review and are looking forward to putting together an exciting program of roundtables, posters, and podium presentations.

Information on conference registration, fees, etc. will soon be made available at <https://www.iomw.net/>

Rasch-Related American Educational Research Association (AERA) Presentations and events

Rasch Measurement SIG Business Meeting

Time: Friday, April 14, 11:40 a.m. to 1:10 p.m. CDT
InterContinental Chicago Magnificent Mile, Floor:
Lobby Level, Avenue

Thursday, April 13, 2023

- Current Research in Learning Environments
Impact of Remote Learning, Learning Spaces,
and Student Perception Tools Paper Session
 - *Time:* Thursday, April 13, 8:00 to 9:30 a.m.
CDT Swissôtel Chicago, Floor: Event
Centre, 1st Floor, Vevey 3
 - *Paper:*
 - The Validation of Classroom Emotional
Climate Questionnaire and Gender
Differences in STEM Classrooms
-R. B. Koul, Curtin University, F.
McLure, Charles Darwin University, B.
J. Fraser, Curtin University
 - AERA Poster Session 1 Closed – 54
 - *Time:* Thursday, April 13, 8:00 to 9:30 a.m.
CDT Hyatt Regency Chicago, Floor: East
Tower - Exhibit Level, Riverside West
Exhibition Hall
 - *Paper:*
 - Assessing 11th Graders' Ability to
Make Causal Explanations in Solving
Chemistry Problems
-Z. Wang, S. Shi, S. Chi, S. Chen, East
China Normal University
 - Assessment Paper Session
 - *Time:* Thursday, April 13, 11:40 a.m. -1:10
p.m. CDT Swissôtel Chicago, Floor: Event
Centre, 1st Floor, Montreux 2
 - *Paper:*
 - Developing a Validated Teacher
Disposition Instrument for Teacher
Preparation Programs
-M.R. Eades-Baird, A.J. Wagle, Empire
State College- SUNY
 - Rasch Analysis in K-12 Settings Roundtable
Session
 - *Time:* Thursday, April 13, 2023, 11:40 a.m.
to 1:10 p.m. CDT Chicago Marriott
- Downtown Magnificent Mile, 7th Floor,
Grand Ballroom Salon III
- *Papers:*
 - A Rasch Analysis of BASC-3 Flex
Progress Monitoring Forms With
Preschool Students
-C. DiStefano, R. Gao, F. Wang, H.
Wang, J. Go, F. Greer, University
of South Carolina
 - Assessing Students' Learning
Progression in Stability and Change
Across Middle School Grades
-S. Chi, Z. Wang, Y. Zhu, East
China Normal University
 - Using Rasch Measurement to
Develop 3D Assessment Tasks to
Measure Students' Understanding
of Energy
-C.F. Herrmann-Abell, BSCS
Science Learning, G.E. DeBoer,
American Association for the
Advancement of Science
- AERA Poster Session 2 – 48
 - *Time:* Thursday, April 13, 9:50 to 11:20
a.m. CDT Hyatt Regency Chicago, Floor:
East Tower - Exhibit Level, Riverside West
Exhibition Hall
 - *Papers:*
 - Comparison of R Packages for a
Mixture Rasch Model
-B. Lee, S. Ahn, Jeongwha Arts
College, Y.-J. Choi, Ewha Womans
University, A.S. Cohen, University
of Georgia
 - Concept Mapping as a Quantitative
Measure of Reading
Comprehension: Construct Validity
and Internal Validity
-J. Galisky, University of
California- Santa Barbara
- Language-Minoritized Students and Assessment:
Social Consequences of Test Use Symposium
 - *Time:* Thursday, April 13, 11:40 a.m. to 1:10
p.m. CDT Swissôtel Chicago, Floor: Event
Centre, 1st Floor, Vevey 4
 - *Paper:*
 - The Consequences of Migration
Tests on Low-Literate Adult
Migrants: Teacher Opinions in 20
European Countries
-C.H. Carlsen, Western Norway
University of Applied Sciences, B.
Deygers, Ghent University

- Constructing Learner Assessments Paper Session
 - *Time:* Thursday, April 13, 4:40 to 6:10 p.m.
CDT Hyatt Regency Chicago, Floor: West Tower - Concourse Level, Gold Coast
 - *Paper:*
 - Investigating Rater Bias in Objective Structured Clinical Examinations Under a Many-Facet Rasch Model Framework
-M.R. Peabody, National Association of Boards of Pharmacy, S.O. Sampson, K.D. Bradley, J. Chadha, A. Hall, H. Garces, A. Ayoob, University of Kentucky
- One Instrument, Four Contexts: International Explorations of Pedagogical Language Knowledge
 - *Time:* Thursday, April 13, 4:40 to 6:10 p.m.
CDT Swissôtel Chicago, Floor: Event Centre, 1st Floor, Vevey 4
 - *Paper:*
 - Measuring Preservice Teachers' Pedagogical Language Knowledge in the United States: A Test Instrument
-S. Hammer, Norwegian University of Science and Technology, K.M. Viesca, University of Nebraska-Lincoln

Friday, April 14, 2023

- AERA Poster Session 6 - 55
 - *Time:* Friday, April 14, 8:00 to 9:30 a.m.
CDT Hyatt Regency Chicago, Floor: East Tower - Exhibit Level, Riverside West Exhibition Hall
 - *Paper:*
 - Classroom Coach: Validation of a New Preschool Classroom Quality Measurement
-F. Andrade-Adaniya, J.A. Beal, J. Claxton, B. Hardin, A. Nielsen, HighScope Educational Research Foundation
- Strategies to Improve the Quality of Data and Model Fit in Survey Research
 - *Time:* Friday, April 14, 8:00 to 9:30 a.m.
CDT Chicago Marriott Downtown Magnificent Mike, Floor: 6th Floor, Great America II

- *Paper:*
 - Introducing ViSA: An Initial Report on Video Survey Administration
-A. Alkhalawi, C. Amisshah, T. Rollins, T. Haines, Morgan State University

Saturday, April 15, 2023

- Exploring First-Year Experiences Paper Session
 - *Time:* Saturday, April 15, 9:50 to 11:20 a.m.
CDT Sheraton Grand Chicago Riverwalk, Floor: Level 2, Colorado
 - *Paper:*
 - Measuring Foundational Inquiry Skills in First-Year College Composition Courses
-J.R.M. Parsons, J.C. McConnell Parsons, K.Kohls, S.O. Sampon, J. Ridolfo, University of Kentucky
- AERA Poster Session 9 Closed - 49
 - *Time:* Saturday, April 15, 8:00 to 9:30 a.m.
CDT Hyatt Regency Chicago, Floor: East Tower - Exhibit Level, Riverside West Exhibition Hall
 - *Paper:*
 - Library Usage and Student Success: Library space changes increased student usage which increased student success
-M.C. Ralph, Multistudio, M. Hegeman
- Rasch Models: Considerations and Novel Uses Paper Session
 - *Time:* Saturday, April 15, 8:00 to 9:30 a.m.
CDT Chicago Marriott Downtown Magnificent Mile, 3rd Floor, Kane
 - *Papers:*
 - Rasch Model and Linear Logistic Test Model Practice Effects
-R. E. Schumacker, University of Alabama
 - Using Person Fit to Explore Misreporting in Surveys
-J. Li, G. Engelhard, University of Georgia
- Designing and Validating Games to Measure Social and Emotional Learning: Validation Results and Lessons Learned Symposium
 - *Time:* Saturday, April 15, 9:50 to 11:20 a.m.
CDT Swissôtel Chicago, Floor: Event Centre, 1st Floor, Vevey 3

- *Paper:*
 - Application of Advanced Psychometric Models to Population Game Play Data From Arthur Games
-K. Choi, Y. S. Suh, T. Feng, C. Parks, G.K.W.K. Chung, E. Redman, University of California-Los Angeles
- Critical aspects of Computer-Supported Collaborative Learning Symposium
 - *Time:* Saturday, April 15, 11:40am to 1:10 p.m. CDT Hyatt Regency Chicago, East Tower - Concourse Level, Michigan 1C
 - *Paper:*
 - Validation of a preliminary ATTOCOLE scale for measuring attitude toward (online) group learning using the Rasch measurement model
-K. Kreijns, M. Henderikx, Open Universiteit Nederland
- AERA Poster Session 13 – 40
 - *Time:* Sunday, April 16, 9:50 to 11:20 a.m. CDT Hyatt Regency Chicago, Floor: East Tower - Exhibit Level, Riverside West Exhibition Hall
 - *Paper:*
 - Refinement of a Validation Theory Survey for College Undergraduates: A Quantitative Field-Testing Study
-T.A. May (Sondergeld), D.N. Bright, Y. Fan, C.J. Fornaro, K.L.K.Koskey, T. Heverin, Drexel University
- Measuring Mathematics and Science Instruction Using Classroom Observations Roundtable Session
 - *Time:* Sunday, April 16, 11:40 a.m. -1:10 p.m. CDT Sheraton Grand Chicago Riverwalk, Floor: Level 4, Sheraton Ballroom IV and V
 - *Paper:*
 - A Systematic Approach to Measure Reformed Teaching in Science Classrooms
-Y.Chen, Y.Yin, University of Illinois at Chicago, S. Werner, University of Illinois System, M. Stieff, University of Illinois at Chicago

Sunday, April 16, 2023

- AERA Poster Session 12 – 41
 - *Time:* Sunday, April 16, 8:00 to 9:30 a.m. CDT Hyatt Regency Chicago, Floor: East Tower - Exhibit Level, Riverside West Exhibition Hall
 - *Paper:*
 - Validating a new Academic Acculturation Scale (AAS): A Preliminary Report
-K.V. Middleton, Howard University, C. Amisah, Morgan State University
- The Parameter Estimation Algorithm, the Analytical Tool, and the Psychometric Properties of Continuous Rating Scales Symposium
 - *Time:* Sun, April 16, 9:50 to 11:20 a.m. CDT Chicago Marriott Downtown Magnificent Mile, 6th Floor, Ohio State
 - *Paper:*
 - Re-examining the intervalness of visual analogue scale: The theoretical and simulated perspectives
-I.-H. Liu, Y.-T. Sung, National Taiwan Normal University
- Oracy, Reading, and Pronunciation Proficiencies
 - *Time:* Sunday, April 16, 11:40 a.m. -1:10 p.m. CDT Sheraton Grand Chicago Riverwalk, Floor: Level 4, Sheraton Ballroom IV and V
 - *Paper:*
 - ACTFL Chinese Reading Proficiency Guidelines: Verifying the Difficulty Hierarchy
-J.Lin, Howard University, X. Gu, Chongqing University, T. Huang, College of William and Mary
- Achieving Equity Through Inclusivity: Measuring the Perspectives of Students, Staff, and Parents Paper Session
 - *Time:* Sunday, April 16, 11:40 a.m. -1:10 p.m. CDT Chicago Marriott Downtown Magnificent Mile, Floor: 3rd Floor, Kane
 - *Paper:*
 - Measuring Perceptions of Equity: A Pilot Validity Study of

Instruments Used in an Equity Audit

-*S.O. Sampson, J.R.M. Parsons, S.E. LaCour, University of Kentucky*

- Preparing Special Educators: Recruitment, Roles, Support, Knowledge, and Development
 - *Time:* Sunday, April 16, 2:50 to 4:20 p.m. CDT Sheraton Grand Chicago Riverwalk, Floor: Level 2, Superior B
 - *Paper:*
 - Project Coordinate: Impact of content-focused lesson study on teacher knowledge, collaboration, and MTSS instruction
-M.T. Brownell, H. Sohn, University of Florida, A.E. Benedict, Arizona State University, J. Williams, University of Florida, G. Koziarski, Arizona State University, B. Kelcey, University of Cincinnati

Thursday, May 4, 2023

- Division C Virtual Poster Session
 - *Time:* Thursday, May 4, 8:00 a.m. -Fri, May 5, 6:00 p.m. CDT Division C Virtual Sessions, Division C - Learning and Instruction Virtual Poster Room
 - *Paper:*
 - Refinement of a questionnaire on Science Teachers' Knowledge of Language as Epistemic Tool
-C. Ding, University of Iowa, C.A. Lammert, Texas Tech University, G.W. Fulmer, National Science Foundation, B. Hand, University of Iowa, J. K. Suh, University of Alabama
 - Validation of an Instrument for Assessing Engineering Teaching Efficacy Beliefs: Rasch and Confirmatory Factor Analyses
-E. Yesilyurt, Weber State University, E. Kaya, George Mason University, H. Deniz, University of Nevada-Las Vegas
- Equity in Assessment in the Health Professions Virtual Symposium
 - *Time:* Thu, May 4, 11:30 a.m. -1:00 p.m. CDT Division I Virtual Sessions, Division I

- Education in the Professions Virtual Symposium Room

- *Paper:*
 - Assessing and accounting for differential rater functioning in oral examinations
-R. Chadha, American Board of Obstetrics & Gynecology

Friday, May 5, 2023

- Rasch Measurement Innovations Virtual Paper Session
 - *Time:* Friday, May 5, 11:30 a.m. -1:00 p.m. CDT SIG Virtual Rooms, Rasch Measurement SIG Virtual Paper Room
 - *Papers:*
 - A Rasch Analysis of a Measure of Graph Selection and Reasoning for Dynamic Situations
-C. Donovan, H. L. Johnson, L. Bechtold, University of Colorado-Denver, R. Knurek, K.A. Whitmore, University of Colorado-Denver
 - A Rasch Analysis of the Psychological Well-Being Measure
-P. Klem, R. Allan, C. Donovan, University of Colorado-Denver
 - Detecting Rater Bias in Mixed-Format Assessments
-S.A. Wind, University of Alabama, Y. Ge, The College Board
 - Differential Item Functioning in Student Success When Modeled as a Latent Variable
-C. Donovan, University of Colorado-Denver, H. Huvar, New Mexico State University, J. Snyders
 - Rasch Analysis on Reliability and Fit Statistics of a Self-Efficacy Measure with Different Rating Categories
-C.-L. Tsai, University of Northern Colorado, S. Estrada, University of Texas at Tyler, J. Fulmore, University of Dallas
- Measurement Methodology Innovation Virtual Paper Session
 - *Time:* Friday, May 5, 9:45 to 11:15 a.m. CDT Division D Virtual Sessions, Division D - Section 1: Educational Measurement, Psychometrics, and Assessment Virtual Paper Room

- *Paper:*
 - Measuring the Impact of Peer Interaction in Group Oral Assessments
-K.-Y. Jin, Hong Kong Examinations and Assessment Authority, T. Eckes, TestDaF Institute, University of Bochum

- Multiple Lenses on Spirituality in Education
Virtual Paper Session
 - *Time:* Friday, May 5, 2:30 to 4:00 p.m. CDT
SIG Virtual Rooms, Spirituality & Education SIG Virtual Paper Room
 - *Paper:*
 - The Early Childhood Educators' Spiritual Practices in the Classroom (ECE-SPC) Instrument: Validation Study using Rasch
-J.Mata-McMahon, University of Maryland, Baltimore County, L. Kruse, North Carolina State University, M. J. Haslip, Drexel University