

***Situating Methodological Challenges to Ensure
Fairness And Equity In Testing Within My View
Of Validity As Contextualized And Pragmatic
Explanation***

UBC-Paragon Webinar Series 2021-2

March 7, 2021

Bruno D. Zumbo

Professor & Distinguished University Scholar

Tier 1, Canada Research Chair in Psychometrics and Measurement

Paragon UBC Professor of Psychometrics and Measurement



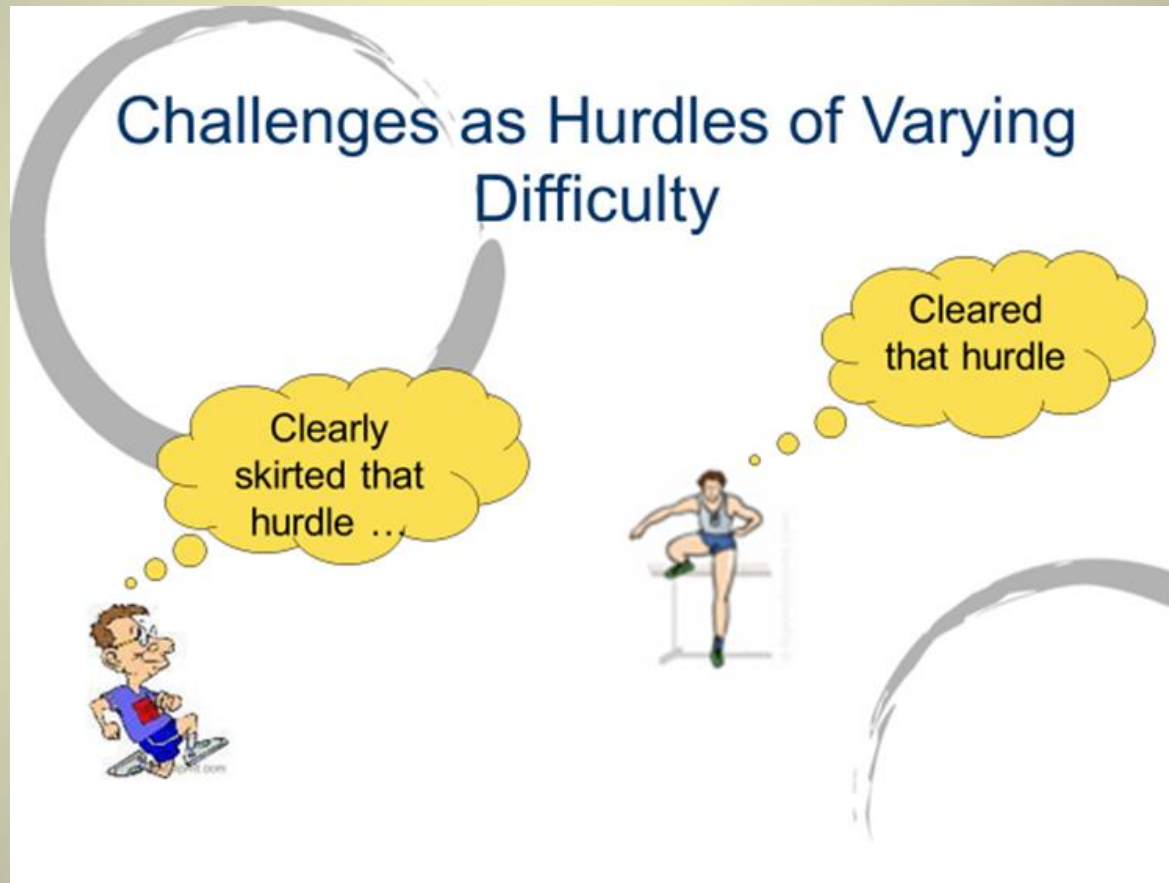
University of British Columbia



Outline

- 1. Three big challenges (hurdles) when considering methodologies used to ensure fairness and equity in testing and assessment**
 - A guiding principle: the many ways to being human
- 2. Situating the challenges within my view of test / assessment validity and validation practices?**
 - Validity as Contextualized and Pragmatic Explanation, and Its Implications for Validation Practice
- 3. Fairness & Equity: Ecological Model of Item and Test Responding**

- I use the metaphor of a hurdler, with some hurdles being more easily cleared. In contrast, others we, collectively as a community of scholars, have (until recently successfully) skirted.



Three Big Challenges (hurdles)

There are other (smaller) challenges, but I thought I would highlight three of the bigger ones.

1. To **what extent** are we **measuring our attribute or competency of interest** with the test, assessment, or survey in use?
2. What are the implicit and explicit conceptions of fairness and equity of the methodology of my choice?

Response: My approaches embody **statistical and psychometric** models, an **ecological model of item and test performance**.

By observing the testing situation, we hope to identify clues about the way the test is constructed, understood and performed as a social occasion.

Three Big Challenges (hurdles)

3. How are my conceptions of fairness and equity informed by the school demographics, logic, or organization of the **Canadian** education system and shaped by **Canadian social values**, as I live them?

Response: To what extent might we be measuring, unintentionally, other (un)important constructs that are not meant to be included in our inferences of our attribute or domain of interest, such as, conformity to expected cultural norms (e.g., related to, for example, multiculturalism, ethnicity, gender identity, and gender roles)?

Zumbo, B. D. (2007). Three generations of differential item functioning (DIF) analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly*, 4(2), 223–233.

Zumbo, B.D., Liu, Y., Wu, A.D., Shear, B.R., Astivia, O.L.O. & Ark, T.K. (2015). A Methodology for Zumbo's Third Generation DIF Analyses and the Ecology of Item Responding. *Language Assessment Quarterly*, 12, 136-151.

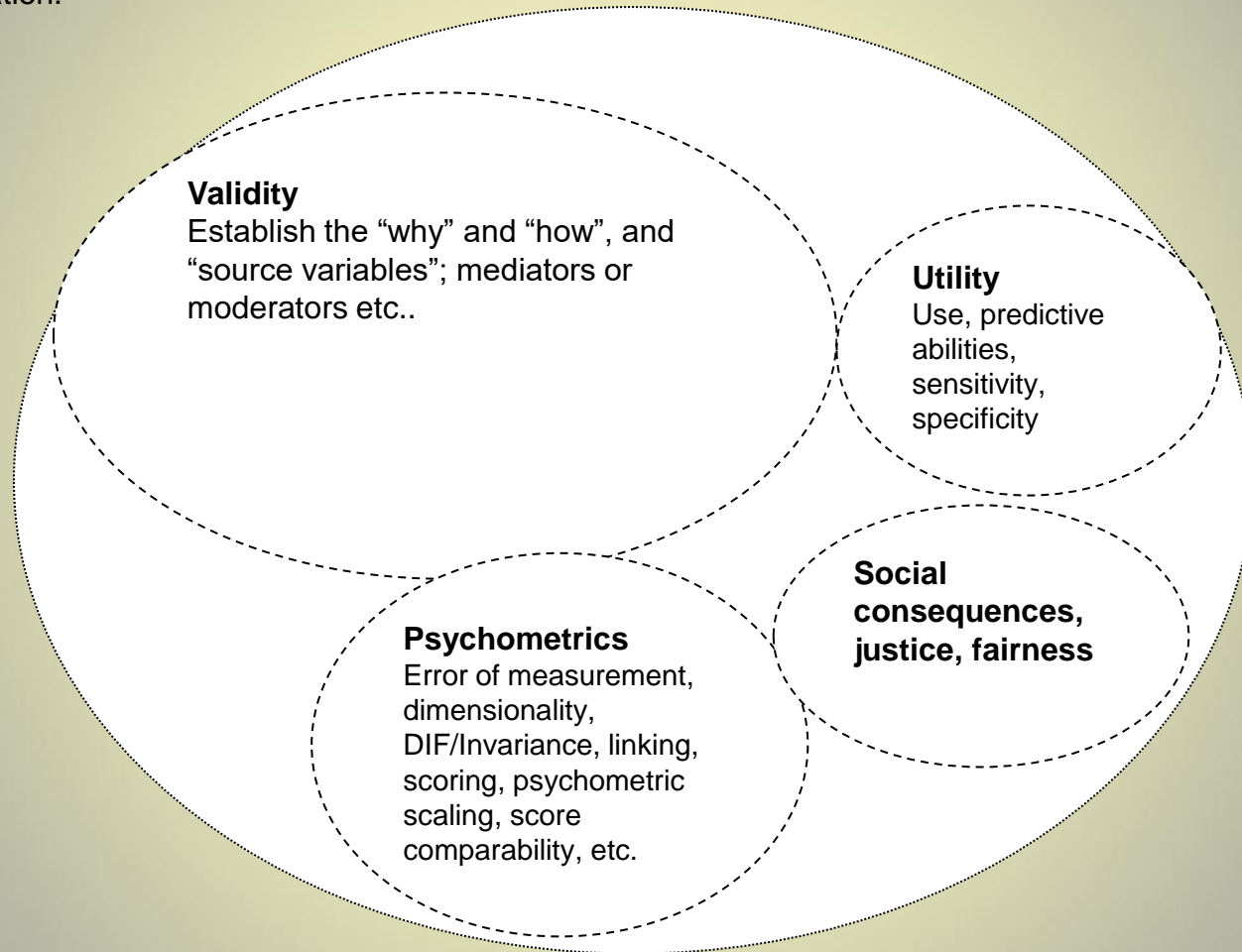
Addey, C., Maddox, B. & Zumbo, B.D. (2020) Assembled validity: rethinking Kane's argument-based approach in the context of International Large-Scale Assessments (ILSAs). *Assessment in Education: Principles, Policy & Practice*, 27:6, 588-606. DOI: 10.1080/0969594X.2020.1843136.

There are many ways to be human

- The principle, as I see it, that there are many ways to be human is **at the core of how I live and theorize** and how I express my **Canadian experience and identity**.
- Over the last 30 years my experience has been that the **field of psychometrics** has tended to go into a **moral panic over gender identity, gender expression, and aspects of cultural expression**.
 - At the core of my theorizing and the methods I develop and/or advocate aim to challenge that view and aim to honour the many ways of being human and capturing the human experience.

Validity as Contextualized and Pragmatic Explanation, and Its Implications for Validation Practice

Figure 1. **A Depiction of the Integrative Cognitive Judgment** in the Contextualized and Pragmatic Explanation View of Validity and Validation.



Validity as Contextualized and Pragmatic Explanation, and Its Implications for Validation Practice

- First, please note that I am separating “validity” from the “process or elements of validation”.
 - Validity involves establishing an explanation for the observed score variation. This is, of course, an old tradition in philosophy of science.
 - In the mid-1950s Cronbach and Meehl brought this idea most clearly to the measurement community by drawing on a form of Hempel-Oppenheim deductive nomological model of explanation for measurement validity in what Cronbach and Meehl called construct validity and a nomological network.

Validity as Contextualized and Pragmatic Explanation, and Its Implications for Validation Practice

- My own work in the last approx. 20 years has challenged that view.
 - Part of the problem with the Cronbach and Meehl approach is that it was rooted in a neo-behaviorist tradition of conflating explanation and confirmation as well as suffering of Michael Scriven's later clearly articulated concern with the Hempel-Oppenheim deductive nomological model of explanation.

Validity as Contextualized and Pragmatic Explanation, and Its Implications for Validation Practice

- I have espoused a different explanatory view. I have my roots more firmly in a pragmatic approach and particularly an inference to the best explanation like strategy.
- Second, in separating “validity”, per se, from the validation process then I have a clearly sense of the role of social consequences, justice, and fairness in the validation process and separate from validity itself. This recognized measurement as a power tool in public debate and public policy.

Validity as Contextualized and Pragmatic Explanation, and Its Implications for Validation Practice

- The basic idea underlying my explanatory approach is that, if one could understand why an individual responded a certain way to an item or scored a particular value on a scale, then that would go a long way toward bridging the inferential gap between test scores (or even latent variable scores) and constructs.

Validity as Contextualized and Pragmatic Explanation, and Its Implications for Validation Practice

- According to this view, validity per se, is not established until one has an explanatory model of the variation in item responses and/or scale scores and the variables mediating, moderating, and otherwise affecting the response outcome.
- This is a tall hurdle indeed. However, I believe that the spirit of Cronbach and Meehl's (1955) work was to require explanation in a strong form of construct validity.

Validity as Contextualized and Pragmatic Explanation, and Its Implications for Validation Practice

- Overlooking the importance of explanation in validity we have, as a discipline, focused overly heavily on the validation process and as a result we have lost our way.
 - This is not to suggest that the activities of the process of validation, such as correlations with a criterion or a convergent measure, dimensionality assessment, item response modeling, or differential item or test functioning, are irrelevant or should be stopped.

Validity as Contextualized and Pragmatic Explanation, and Its Implications for Validation Practice

- Quite to the contrary, the activities of the process of validation must serve the definition of validity.
 - My aim is to re-focus our attention on why we are conducting all of these psychometric analyses: that is, to support our claim of the validity of our inferences from a given measure.

Validity as Contextualized and Pragmatic Explanation, and Its Implications for Validation Practice

- For example, as Zumbo (2007) highlighted conducting test and item bias is not just about protecting a test developer or test user against lawsuits.
 - Conducting test and item bias is also a statistical methodology that ferrets out invalidity that distorts the meaning of test results for some groups of examinees and thus establishes the inferential limits of the test.

Zumbo, B.D. (2007). Three generations of differential item functioning (DIF) analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly*, 4, 223-233.

Zumbo, B. D. (2009). Validity as Contextualized and Pragmatic Explanation, and Its Implications for Validation Practice. In Robert W. Lissitz (Ed.) *The Concept of Validity: Revisions, New Directions and Applications*, (pp. 65-82). IAP - Information Age Publishing, Inc.: Charlotte, NC.

Validity as Contextualized and Pragmatic Explanation, and Its Implications for Validation Practice

- One of the limitations of traditional quantitative test validation practices (e.g., factor-analytic methods, validity coefficients, and multitrait-multimethod approaches) is that they are descriptive rather than explanatory.
 - The aim of my explanatory approach is to lay the groundwork to expand the evidential basis for test validation by providing a richer explanation of the processes of responding to tests and variation in test or items scores and hence promoting a richer psychometric theory-building.

Fairness & Equity: Ecological Model of Item and Test Responding

In studies of fairness **gender** has, in the main, been characterized in the binary as biological sex wherein (binary) biological sex differences on item performance that are eventually explained by item characteristics such as item format and item content.

In what I described in 2007 as my **Third Generation DIF “gender”** more properly should **be considered a social construction**, and gender differences on item performance are explained by contextual or situational variables (ecological variables, if you wish), such as institutionalized gender roles, classroom size, socioeconomic status, teaching practices, and parental styles.

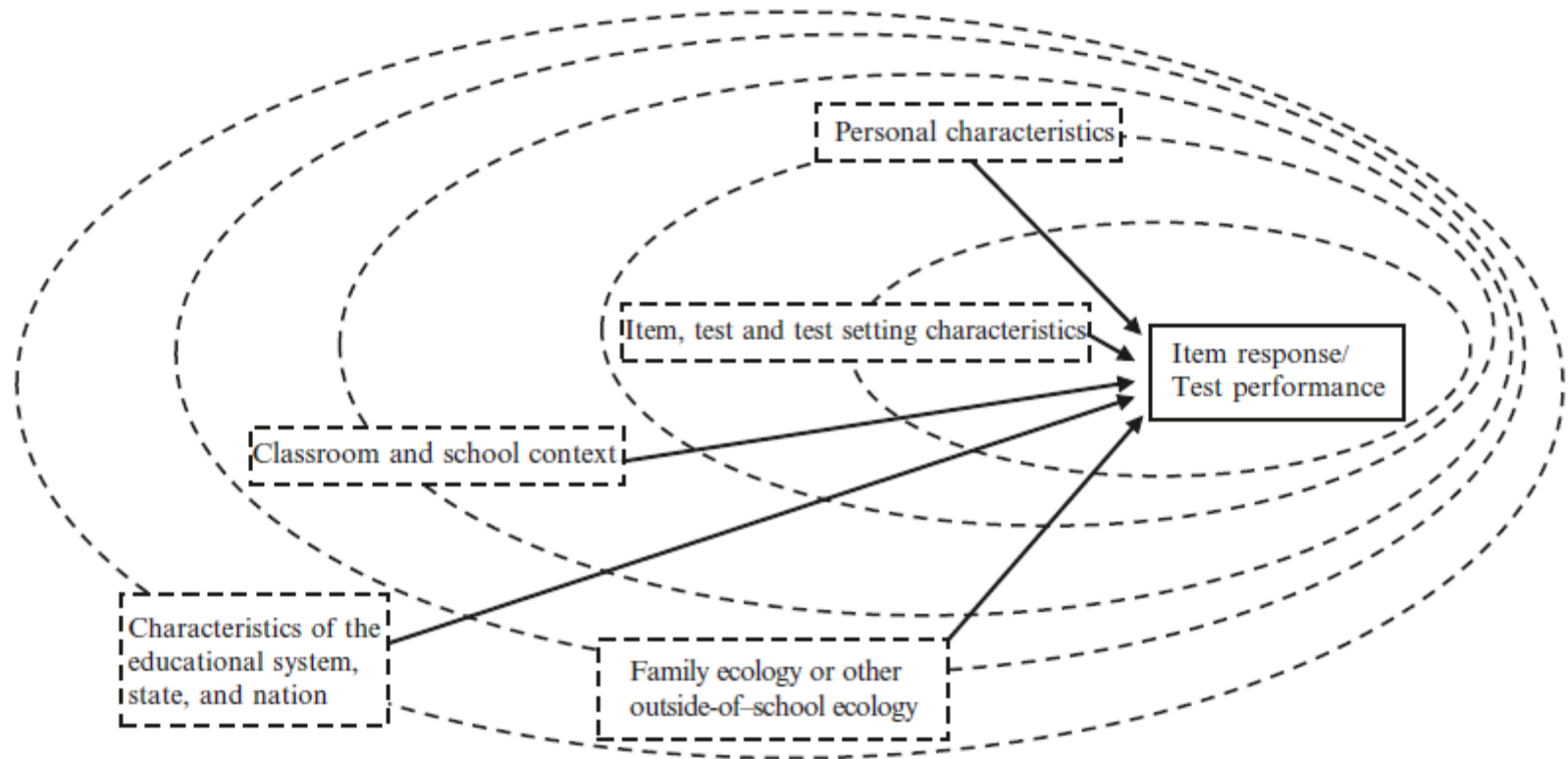
Ecological Model of Item and Test Responding

We believe that these **richer ecological variables** have been **largely ignored** in relation to explanations for (and causes of) DIF because of the focus on test format, content, cognitive processes, and test dimensionality that is pervasive in the second generation of DIF.

Zumbo, B. D. (2007). Three generations of differential item functioning (DIF) analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly*, 4(2), 223–233.

Zumbo, B.D., Liu, Y., Wu, A.D., Shear, B.R., Astivia, O.L.O. & Ark, T.K. (2015). A Methodology for Zumbo's Third Generation DIF Analyses and the Ecology of Item Responding. *Language Assessment Quarterly*, 12, 136-151.

Ecological Model of Item and Test Responding



- Zumbo, B.D., Liu, Y., Wu, A.D., Shear, B.R., Astivia, O.L.O. & Ark, T.K. (2015). A Methodology for Zumbo's Third Generation DIF Analyses and the Ecology of Item Responding. *Language Assessment Quarterly*, 12, 136-151.
- Chen, M.Y., & Zumbo, B.D. (2017). Ecological framework of item responding as validity evidence: An application of multilevel DIF modeling using PISA data. In B. D. Zumbo and A.M. Hubley (Eds.), *Understanding and Investigating Response Processes in Validation Research* (pp. 53-68). New York, NY: Springer.

What my approach to fairness and equity implies ...

- Traditional views follow a “**social address**” model of criterion prediction and group differences.
 - This spills over in to test validation; group differences.
- In using the common “social address” approach to group comparisons, **classification into groups** might be **confused with fixed biological or ethnic classification**.

As John Stuart Mill (1848) wrote:

Of all the vulgar modes of escaping the consideration of the effect of social and moral influences on the mind, the most vulgar is attributing the diversities of conduct and character to inherent natural differences. (p. 319).

What my approach to fairness and equity implies ...

- In a series of chapters and papers from 1998 to 2021, I have made the case that the aim is: **identifying the determinants (or explanatory theory)** of task / item / test score variation ... **the explanation is the basis of any strong validity claims.**
- I take an **ecological systems approach**
- Most research on response processes focuses on cognitive factors.
 - We have taken a **broader view of response processes** proposed by Zumbo & Hubley (2017) and embrace the notion of **assessment 'in vivo'** to shine a spotlight on test-takers' behaviour, stance, gesture, motivation, and affect besides cognition.

What my approach to fairness and equity implies ...

- Political context, consequences of testing, assessment data usage, and matters of social justice are core to equity and fairness.
- Typically, the outcome of a test or assessment serves multiple purposes. One needs to validate each of the various uses of the well-being data by the different stakeholders -- e.g., regional comparisons, time trends, policy evaluation, etc..
- **My approach highlights response processes, explanation of task / item / test variation, and is 'in vivo'** (Zumbo & Huble, 2017)

So... Now What?

- Do not **skirt these hurdles**.
- There are now **sophisticated methodologies to tackle these challenges**. (Complex models of the impact of the ecological model of item and test responding)
- A well thought out research program needs to be established and funded to address these issues.
- We can use assessment outcomes, but we need to get going on this research program to support the inferences we intend to make with them.

What the view of validity and validation implies for studies of fairness and equity in test and assessment

- Examples of issues in a validation research agenda:
 - One needs to **investigate if and how race, gender, and culture as related to the object of measurement** (i.e., individual students or communities) may shape and alter the inferences one makes from testing or assessment outcomes.
 - One needs to **investigate the role of various levels of measurement (e.g., individual, community, neighborhood, city, province, region) in the inferences**. For example, is one measuring learning or perhaps different secondary (or primary) dimensions at the various levels of analysis. Including the predictive nature at these various levels. (see, Zumbo et al, 2017)

Zumbo, B.D., Liu, Y., Wu, A.D., Forer, B., & Shear, B.R. (2017). National and International Educational Achievement Testing: A Case of Multi-Level Validation Framed by the Ecological Model of Item Responding. In B. D. Zumbo and A.M. Hubley (Eds.), *Understanding and Investigating Response Processes in Validation Research* (pp. 341-362). New York, NY: Springer.

THE END!

Citation:

Zumbo, B.D. (2021). Situating Methodological Challenges to Ensure Fairness and Equity in Testing Within My View of Validity as Contextualized and Pragmatic Explanation [Video]. *UBC-Paragon Webinar Series 2021-2*. University of British Columbia, Vancouver, B.C.

<https://www.youtube.com/channel/UCCKvIPbkn15FVDKRplm-fuw>

