

# *Validity of Subjective and Objective Measures of Wellbeing*



**Bruno D. Zumbo, Ph.D.**  
**Professor**  
**University of British Columbia**  
**Vancouver, Canada**

**Presented at the Training Course on Measuring the Progress of Societies "Statistics, Knowledge and Policy: Understanding Societal Change"**

**July 14th to 17th 2009, Florence (Firenze), Italy**

1

## Preface

- **This lecture was prepared for the 4-day training course on "Statistics, Knowledge and Policy: Understanding Societal Change", an Organisation for Economic Co-operation and Development (OECD) hosted Global Project in association with the Joint Research Centre (JRC) of the European Commission and the International Society for Quality of Life Studies (ISQOLS).**
- **I would like to thank Jon Hall and Barbara Iasiello of the OECD and Filomena Maggino of the Università degli Studi di Firenze for the invitation to speak and for the support and feedback in the formative stages of this presentation.**
- **I would also express my gratitude to my longtime friend Alex Michalos who has been a wonderful sounding board for my ideas about validity, and wonderful collaborator on many research projects. An additional note of gratitude to Alex because he was also instrumental in my invitation to speak today.**

2

## Outline of Session

- 1. Foundational and Conceptual Issues.**
- 2. A Statistical Framework to help think about validation, and implications of the explanatory view for validation practice.**
- 3. Concluding remarks and discussion.**

3

## Section 1

### Foundational and Conceptual Issues

- What is meant by “measurement validity” in various areas related to wellbeing and quality of life.
- Validity as contextualized and pragmatic explanation of variation.
- Show how this applies to both objective and subjective indicators
  - Show the interchange of ideas between objective and subjective indicators ... how ideas in one area can inform the other.

4

## *Validity/validation*

- **Matters of measurement or assessment validity and validation are quite common across the social, behavioral, and health sciences.**
- **We need to acknowledge that there is a great deal of variation within disciplines but, as a whole, the various disciplines within the social sciences have dealt with matters of measurement validity somewhat differently (with lots of communality).**

5

## *Validity/validation*

- **My purpose today is to describe for you a novel approach to validity and validation that:**
  - allows us to cross the disciplinary bounds,
  - exploits the wonderful developments in the various social sciences sub-fields, and
  - allows us to think about validity and validation irrespective of whether we are dealing with the so-called “objective” or “subjective”, and single-indicator or multi-indicator composite indicators of quality of life and wellbeing.
- **We will close with a discussion of how this may be useful for your work.**

6

## *Validity/validation*

- **I wish to take advantage of the statistical and conceptual similarities between objective and subjective indicator systems:**

- Although there are unique features there are some communalities I wish to exploit.
- Some exchangeable language is key:
- I am going to use the language:
  - Objective Indicators: indicators => indices
  - Subjective indicators: items/questions => scale scores
  - Therefore, I would like to use “indicators” and “items” exchangeably.

7

Validity/validation: A few words about how validation is *explicitly* considered, if at all.

- **Because there is so much within discipline variation, it is probably more accurate to discuss “research topics” rather than disciplines. For example:**
  - The topics influenced by the psycho-social, behavioral and health sciences models of measurement usually talk about content, face, construct, predictive, etc. validity.
  - Topics influenced by the economic traditions will talk about utility and representational (axiomatic) views of measurement.
  - Also, some scholarly traditions see measurement validity only as statistical aspects of bias and convergence in estimation. For example, see MacPhail’s (1998) discussion of the economic orientation to validity.

8

## *Validity/validation*

- **Over the course of the last decade I have been developing a theory of validity and articulating its implications for validation practice.**
- **We will provide an overview of the most recent work on this topic and how it highlights a foundation for all of our statistical work.**
  - What follows comes from Zumbo (2009).

9

## *Validity/validation*

- **The term “construct validity” has evolved to be a short-hand for the expression “an articulated argument in support of the inferences made from scores”.**
- **I have argued that construct validity has, from its introduction (Cronbach & Meehl, 1955), been focused on providing an explanation for observed measurement or test scores; that is, the argument in support of the inferences is a form of an explanation.**

10

## *Validity/validation*

- **As we all know, there are strong and weak forms of construct validity (Borsboom, Mellenbergh, & Van Heerden, 2004; Kane, 2001, 2006).**
  - The weak form is characterized by any correlation of the test score with another variable being welcomed as evidence for another “validity” of the test.
  - That is, in the weak form, a test has as many “validities” and potential uses as it has correlations with other criterion (or convergent) variables.
  - In contrast to the weak form of construct validity, the strong form is based on a well-articulated theory and well-planned empirical tests of that theory.
    - In short, the strong form is theory-driven whereas the weak form implies that a correlation with some criterion (or convergent measure) is sufficient evidence to use the test as a measure of that criterion.

11

## *Validity/validation*

- **In my view (e.g., Zumbo, 2005, 2007, 2009), the strong form of construct validity should provide an *explanation* for the test scores, in the sense of the theory having explanatory power for the observed variation in test scores.**
  - I share the view with other validity theorists that validity is a matter of inference and the weighing of evidence; however, in my view, explanatory considerations guide our inferences. Explanation acts as a regulative ideal; validity is the explanation for the test score variation, and validation is the process of developing and testing the explanation.

12

## *Validity/validation*

- **In essence, I see validation as a higher order integrative cognitive process involving everyday (and highly technically evolved) notions like concept formation and the detection, identification, and generalization of regularities in data whether they are numerical or textual.**

13

## *Validity/validation*

- **From this, after a balance of possible competing views and contrastive data, comes understanding and explanation.**
  - **What I am suggesting is a more technical and more data-driven elaboration of what we do on a day to day basis in an open (scientific) society; we are constantly asking why the things are the way we find them to be, answer our own questions by constructing explanatory stories, and thus come to believe some of these stories based on how good are the explanations they provide.**

14

# Validity/validation

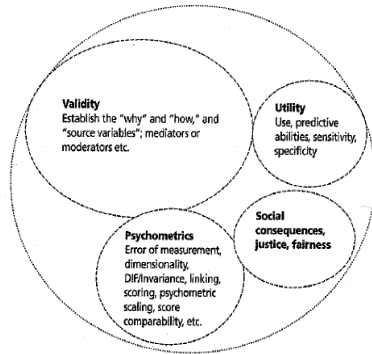


Figure 1 From Zumbo (2009).

Figure 1 depicts the four core elements of the integrative cognitive judgment of validity and the process of validation: validity, statistics, social consequences, and matters of utility – all of which are tightly packed in the Figure close to each other and hence influence, and shape, each other.

We can see that validity is separate from utility, social consequences, and the statistics, but validity is shaped by these.

Furthermore, the inferences are justified by the statistics, social consequences, and utility but validity is something more because it requires the explanation.

# Validity/validation

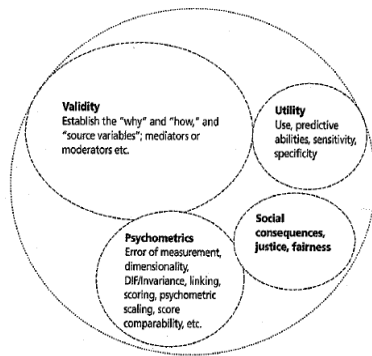


Figure 1 From Zumbo (2009).

The Figure shows that explanation is the defining feature of validity and hence supports the inferences we make from test scores.

In terms of the process of validation, we can see that the process of validation is distinct but is, itself, shaped by the concept of validity.

The process of validation involves consideration of the statistical methods, as well as the psychological and more qualitative methods of statistics, work to establish and support the inference to the explanation – i.e., validity itself; so that validity is the explanation, whereas the process of validation involves the myriad methods of statistics to establish and support that explanation.

## Validity/validation

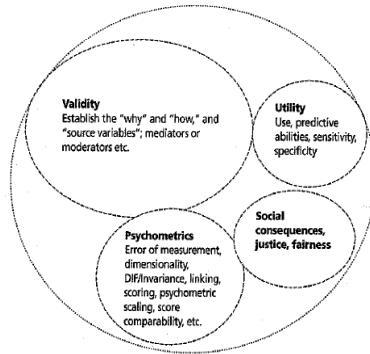


Figure 1 From Zumbo (2009).

The process of validation also includes the utility and evidence from test use such as sensitivity and specificity of the decisions (e.g., pass/fail, presence/absence of disease) made from test scores and predictive capacity (e.g., predictive regression equations); as well as the fourth element of social consequences.

This latter element in the cognitive process depicted in Figure 1 has me clearly aligned with Messick (e.g., Messick, 1998) in that empirical consequences of test use and interpretation constitutes validity evidence in the validation process.

17

## Validity/validation

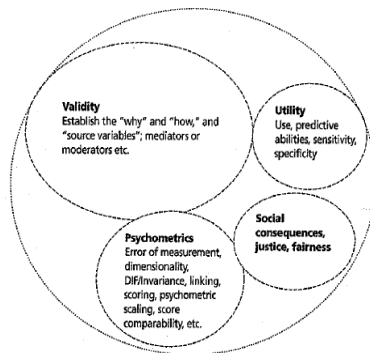


Figure 1 From Zumbo (2009).

The basic idea underlying my explanatory approach is that, if one could understand the variation in an indicator, then that would go a long way toward bridging the inferential gap between subjective (or objective indicators) and the constructs.

According to this view, validity per se, is not established until one has an explanatory model of the variation in indicators and the variables mediating, moderating, and otherwise affecting that observed variation

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.

18

## **An example of how an explanatory view can help validate a widely used index: The HDI**

- **The Human Development Index (HDI) is a creation of the United Nations Development Programme and represents the practical embodiment of their vision of human development as an alternative vision to what they perceive as the dominance of economic indicators in development.**
  - Economic development had the gross domestic product (GDP) so human development had to have the HDI. In essence the HDI represents a measure of the 'quality of life'.

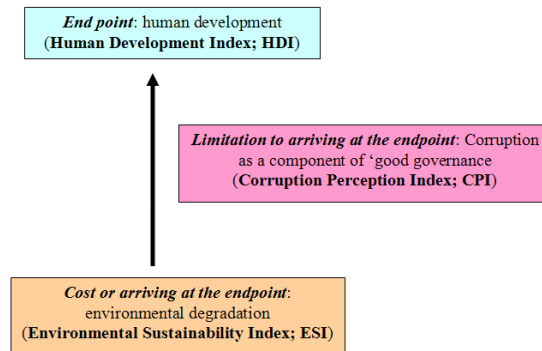
19

## **An example of how an explanatory view can help validate a widely used index: The HDI**

- **Since its appearance in 1990 the HDI comprises three components:**
  1. life expectancy (a proxy indicator for health care and living conditions).
  2. adult literacy combined with years of schooling or enrollment in primary, secondary and tertiary education.
  3. real GDP/capita (\$ PPP; a proxy indicator for disposable income).
- **From my view of validity, one needs to understand the variation in HDI scores across countries, and eventually, over time.**
  - For example, what are the reasons for the variation in HDI across countries: explanatory stories, as well as potential confounding sources of variation.

20

## An example of how an explanatory view can help validate a widely used index



Note: The explanatory hypothesis (to the left) described by Morse, if it holds up, will go a long way to telling us what are the sources of variation, across countries, of the HDI. As such, this hypothesis is key to validating the HDI.

Hypothesised causal chain with three development indices.

Source of material on these slides about the HDI: *Development indicators and indices*, Lead Author: Stephen Morse  
[http://www.eoearth.org/article/Development\\_indicators\\_and\\_indices](http://www.eoearth.org/article/Development_indicators_and_indices)

21

## *Validity/validation*

- **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, or (for subjective indicators) item response modeling, or differential item or test functioning, are irrelevant or should be stopped.

22

## *Validity/validation*

- **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 statistical analyses: that is, to support our claim of the validity of our inferences from a given measure.**
  - **For subjective indicators: 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 statistical theory-building.**

23

## Section 2

**Zumbo's Draper-Lindley-deFinetti (DLD) framework as an over-arching framework for objective or subjective indicators that have a defined domain.**

**Implications of the DLD framework for modeling, model choice, invariance, spotlight on person sampling, expecting heterogeneity, etc..**

24

## Zumbo's DLD Framework: Desired types of inferences

- Zumbo (2001, 2007) presented the following framework modeled on Draper's (1995) approach to classifying causal claims in the social sciences and, in turn, on Lindley's (1972) and de Finetti's (1974-1975) predictive approach to inference.
- Unlike Draper, I focused on the inferences about indicators and persons.

25

## DLD Framework: Desired types of inferences

- The foundation of the approach is the exchangeability of:
  - Sampled and un-sampled respondents (i.e., data sources and sampling units; for example, respondents to surveys); this could be based on the selection function for sub-populations.
  - Realized and unrealized indicators.
- Exchangeable sub-populations of respondents and indicators.

26

## DLD Framework: Desired types of inferences

- **By exchangeability you can think of it in the purely mechanical sense.**
- **I have found this useful to help me think of the various possibilities, whether they happen regularly or not.**
- **This also helps me detail the range of conditions under which invariance is expected to hold.**

27

## DLD framework

- **I present the DLD framework as a part of a methodological unfolding to help think about validity evidence and frame our thinking and work on validation.**
- **The Draper-Lindley-de Finetti (DLD) framework of validity provides a useful overview of the assumptions that must be tested to validate the use of an indicator or index for specific research purposes.**

28

## Zumbo's DLD framework

- This framework describes the relationship between validity and various forms of measurement inference. The form of inference is dependent upon:
  - 1) the degree to which the indicators are exchangeable, and
  - 2) the degree to which the sampling units are exchangeable.

29

### An application to subjective indicators

The various forms of measurement inference.

Exchangeability of Sampled and Unsampled Items in the Target Construct / Domain  
(i.e., sampled tasks or items)

Exchangeability of Sampled and Unsampled Units in Target Population  
(i.e., sampled individuals)

		Exchangeability of Sampled and Unsampled Items in the Target Construct / Domain (i.e., sampled tasks or items)	
		<i>EXCHANGEABLE</i>	<i>NOT EXCHANGEABLE</i>
Exchangeability of Sampled and Unsampled Units in Target Population (i.e., sampled individuals)	<i>EXCHANGEABLE</i>	General Measurement Inference	Specific Sampling Inference
	<i>NOT EXCHANGEABLE</i>	Specific Domain Inference	Initial Calibrative Inference

In terms of inferential strength,  
(Initial Calibrative, Specific Sampling) < Specific Domain < General Measurement Inference

30

## DLD: Various forms of inferences

**A distinction between four forms of inference is made:**

- 1) Initial calibrative inference:** this form of inference does not justify inferences beyond the particular sample from which the data are obtained and the particular indicators that were used.
- 2) Specific sampling inference:** allows for claims about the specific sample in which the measurement took place.
- 3) Specific domain inference:** allows for claims about what is being measured.
- 4) General measurement inference:** allows for comparisons across measures and across different samples.

31

## Zumbo's DLD Framework: Desired types of inferences

### Linking Strength of Inference to Invariance

In terms of inferential strength,

(Initial Calibrative, Specific Sampling) < Specific Domain < General Measurement

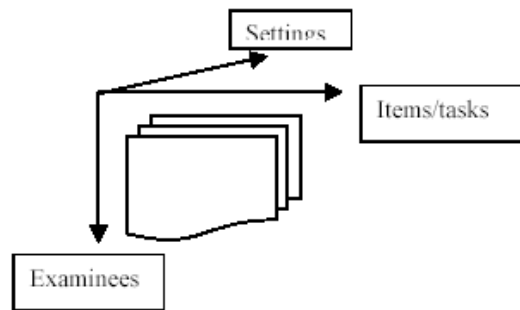


No Concern for Invariance

Invariance of Primary Concern

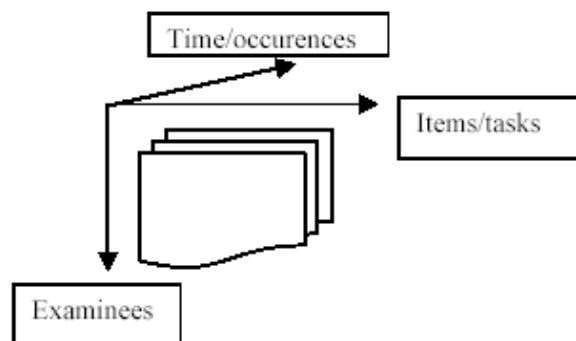
32

## A three-dimensional variant of DLD framework with Settings



33

## A three dimensional variant of DLD with Time or Occurrences



34

## A recent application of the DLD framework

- In terms of subjective indicators, at the 2009 ISQOLS conference (Firenze) Rick Sawatzky, Jacek Kopec, Bruno Zumbo, and Eric Sayer applied DLD in the context of latent variable mixture modeling of measures of emotional wellbeing. The application highlights that wellbeing indices may rely on general measurement inference, especially when comparing groups, because:
  - 1) The items are assumed to be interchangeable. This is needed to ensure that the scores of individuals who answered different questions are comparable on the same scale.
  - 2) The items parameters are assumed to be invariant. This is needed to ensure that the scores of individuals are comparable irrespective of any differences that might exist between individuals.

35

### The Draper-Lindley-de Finetti (DLD) framework of measurement validity (Zumbo, 2007)

		Indicator homogeneity / unidimensionality	
		EXCHANGEABLE	NOT EXCHANGEABLE
Sample homogeneity / parameter invariance	EXCHANGEABLE	General Measurement Inference <b>Required for the comparison of different groups and different instruments</b>	Specific Sampling Inference
	NOT EXCHANGEABLE	Specific Domain Inference	Initial Calibrative Inference

Zumbo, B. D. (2007). Validity: Foundational issues and statistical methodology. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics* (Vol. 26: Psychometrics, pp. 45-79). Amsterdam: Elsevier Science.

## DLD and Wellbeing measures

- The DLD brings to the forefront the matter of sample homogeneity.
  - This is an important issue for all model based measurement (and particularly factor analysis and IRT).
  - In essence this highlights that model driven applications require that the sample is homogeneous with respect to the measurement model.
    - For model based measurement practices, the model assumptions (such as unidimensionality and sample homogeneity) are part of the validity concerns.

37

## Section 3

### Some concluding remarks

38

## What the view of validity and validation implies ... *Some summary remarks*

- **An important issue:**
  - When can we start using a measure or index? Or do we need to establish the “validity” (i.e., the explanation for the index and indicator variation) before we can use the index to make inferences and research conclusions?
    - Answer: Explanation is a regulative ideal.
- **What I am suggesting is that psycho-social, policy, and health studies research take on a robust and integrative research agenda in which the bounds and limitations of the inferences we can make from index and indicator scores (and hence ferreting out invalidity) becomes a core task of the research agenda.**

39

## What the view of validity and validation implies ... *Some summary remarks*

- **The demands are high but I believe that they are in line with the desires spelled out in the seminal paper by Cronbach and Meehl (1955), read as a strong program of construct validity research.**
- **One thing that gets highlighted by the DLD framework is that, in general, in statistics do not unthinkingly assume homogeneity.**
  - Work, where possible, with multi-level and latent class models.
  - Although I do not discuss it herein, Zumbo’s (2009) view of validity has implications for “multilevel measures” (See Zumbo & Forer, in press).
- **In the tradition of inference to the best explanation (or abductive methods) the latent variables of factor analysis may take on an explanatory role.**

40

# Thank You For Your Attention!

**Bruno D. Zumbo**  
**University of British Columbia**  
[bruno.zumbo@ubc.ca](mailto:bruno.zumbo@ubc.ca)

<http://www.educ.ubc.ca/faculty/zumbo/zumbo.htm>



41

## Bibliography

- Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*, 1061-1071.
- Cronbach, L. J., & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 4, 281-302.
- de Finetti, B. (1974-1975). *Theory of probability (Vols. 1 -2)*. New York: Wiley.
- Draper, D. (1995). Inference and hierarchical modeling in the social sciences. *Journal of Educational and Behavioral Statistics*, *20*, 115-147.
- Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, *38*, 319-342.
- Kane, M. (2006). Validation. In R. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). Washington, DC: American Council on Education and National Council on Measurement in Education.
- Lindley, D. V. (1972). *Bayesian statistics, a review*. Philadelphia: Society for Industrial and Applied Mathematics.
- MacPhail, F. (1998). Moving Beyond Statistical Validity in Economics (pp. 119-149). In Zumbo, B. D. (Ed.) (1998). *Validity theory and the methods used in validation: perspectives from the social and behavioral sciences*. Special issue of the journal *Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement*, *45* (1-3), 1-359. Netherlands: Kluwer Academic Press.

42

### Bibliography (cont'd.)

- Messick, S. (1998). Test validity: A matter of consequence. *Social Indicators Research*, 45, 35–44.
- Zumbo, B. D. (Ed.) (1998). Validity theory and the methods used in validation: perspectives from the social and behavioral sciences. Special issue of the journal *Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement*, 45 (1-3), 1-359. Netherlands: Kluwer Academic Press.
- Zumbo, B. D. (2005). Reflections on validity at the intersection of psychometrics, scaling, philosophy of inquiry, and language testing (July 22, 2005). Samuel J. Messick Memorial Award Lecture, LTRC 27th Language Testing Research Colloquium, Ottawa, Canada.
- Zumbo, B.D. (2007). Validity: Foundational issues and statistical methodology. In C.R. Rao and S. Sinharay (Eds.) *Handbook of statistics*, Vol. 26: Psychometrics, (pp. 45-79). The Netherlands: Elsevier Science B.V..
- 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.

43

### Bibliography (cont'd.)

- Zumbo, B. D., & Forer, B. (in press). Testing and Measurement from a Multilevel View: Psychometrics and Validation. In James Bovaird, Kurt Geisinger, & Chad Buckendahl (Editors). *High Stakes Testing in Education - Science and Practice in K-12 Settings [Festschrift to Barbara Plake]*. American Psychological Association Press, Washington, D.C..
- Zumbo, B. D., & Gelin, M.N. (2005). A matter of test bias in educational policy research: Bringing the context into picture by investigating sociological / community moderated (or mediated) test and item bias. *Journal of Educational Research and Policy Studies*, 5, 1-23.
- Zumbo, B. D., & Rupp, A. A. (2004). Responsible modeling of measurement data for appropriate inferences: important advances in reliability and validity theory. In David Kaplan (Ed.), *The SAGE Handbook of Quantitative Methodology for the Social Sciences* (pp. 73-92). Thousand Oaks, CA: Sage Press.

44

This document was created with Win2PDF available at <http://www.win2pdf.com>.  
The unregistered version of Win2PDF is for evaluation or non-commercial use only.  
This page will not be added after purchasing Win2PDF.