

# **The Developmental Science Approach to Measurement in Global Research and Evaluation About Character Development**

**Richard M. Lerner,  
Elizabeth M. Dowling,  
Jonathan M. Tirrell,  
Marc H. Bornstein**



Institute for Applied Research  
in Youth Development

***“There are more things in Heaven and Earth, Horatio, than are dreamt of in your philosophy.”***

**W. Shakespeare, *Hamlet*, Act 1, Scene 5, lines 173-174**

***“There are no shortcuts to the top of the palm tree.”***

**Cameroonian Proverb**

**... likewise, several steps are required to use valid, invariant, and reliable measures within and across different global/cultural settings.**

# Measurement Challenges to Creating a Global Platform in Character Development Research and Evaluation

- Many approaches to measurement within and across settings use a relatively simple approach.
- Arguably the most common approach involves using a measure developed in one cultural/language settings (e.g., Spanish) within another cultural setting (e.g., Zulu).
- Often, this approach begins by translating items from the language of origin to the language of the to-be-studied setting, then back translating, and sometimes engaging in double-back translation.
- This method has errors of both omission and commission.
- So, there is interest in other measurement approaches.

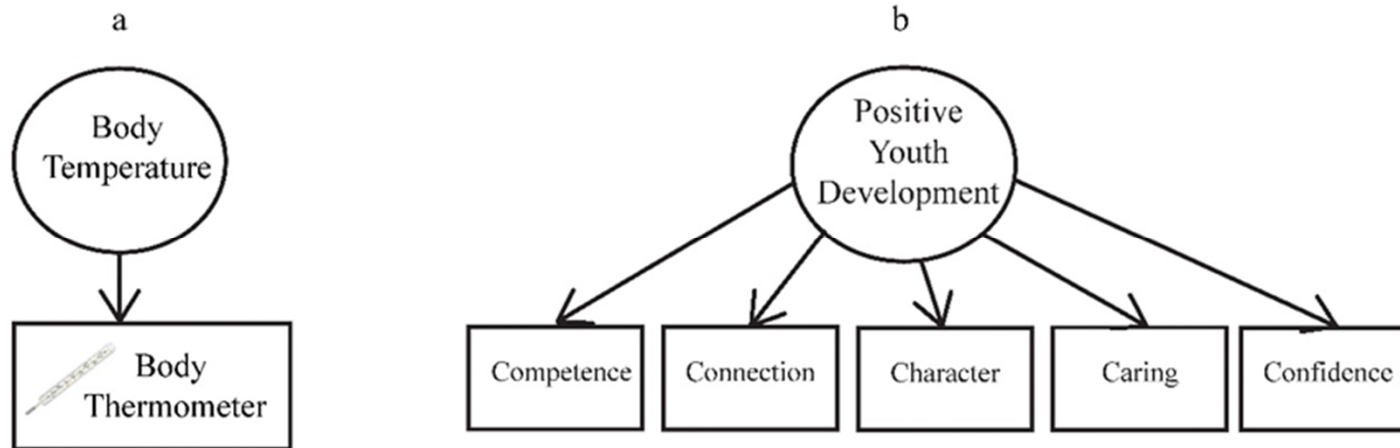
# Measurement Challenges to Creating a Global Platform in Character Development Research and Evaluation

- One alternative approach is framed by foundational ideas in developmental science.
- Our key point is that many more steps must take place to develop and use valid measures.
- For a measure to be valid, possess measurement invariance, and be reliable, the measure must be transformed in a more complex multi-step approach.

## The Philosophical Foundation of Developmental Measurement

- The developmental science approach is built on several steps that, collectively, involve philosophy, theory, and empirical procedures, only one of which involves comparing items from one language to another.
- The developmental science approach is framed by considering the **ontology** of the construct being measured. Ontology addresses the nature or essential characteristics of being and of things that exist.
- Developmental science begins measurement by asserting that what we are measuring **actually exists**. What we believe to exist can be things as common as body temperature or as abstract as the character virtue of generosity, or flourishing, or positive youth development [PYD].
- Not everyone involved in measurement believes that a measure is an index of something that **actually exists**.
- However, because we are measuring something that exists, any variation that we observe in the things we measure (e.g., the mercury level of a thermometer or item scores from a survey) *are caused by the construct*.

# Ontology (Continued)\*



\*Note the direction of the arrows in the figure.

## Ontology (Continued)

- The focus on ontology is based on the view that, for measurement to be meaningful (for it to be about something of importance for human development), the “something” has to be real, even if we cannot directly see it (even if it is a “latent” variable).
- If the something does exist, then variation in what we can see (a “manifest” variable) is caused by variation in the presence or strength of the latent variable.
- To the extent that variation in the measurement of an observed, manifest variable is caused by the latent variable, a measure can be said to be valid.

## Ontology (Continued)

- This approach to measurement is foundational in developmental science because development itself is a latent variable. That is, not all ever-present variations or fluctuations that exist second-by-second in life reflect development.
- Theory is used to extract from the ebb and flow of the incessant changes in life only those changes that reflect *developmental* change, be it cognitive development (à la Piaget), identity development (à la Erikson), moral reasoning development (à la Damon or Kohlberg), or the development of parenting (à la Bornstein).



# Ontology (Continued)

- This approach to measurement enables researchers and evaluators to develop and use tools that measure the development and functioning in a person of any psychological or behavioral construct that reflects their specific understanding, or definition, of development.
- In contemporary, cutting-edge theory, development occurs through specific, mutually influential (i.e., dynamic) relations between the individual and the specific features of their social and cultural setting, as it exists within time (i.e., specific portions of an individual's life (ontogeny) and specific periods of history; Bornstein, 2019).

## Ontology (Continued)

- Thus, valid measurement of latent constructs must be framed by interrogating the meaning of a specific latent construct's existence and development across specific (ontogenetic and historical) times and specific places through specific individual-cultural setting coactions in specific national and cultural settings.

## The Concept of Validity

- The fundamental attribute desired in a measure is validity: that is, the measure *should measure what it is supposed to measure*.
- Validity is not primarily an empirical matter; validity cannot be substantiated by covariation among measures in a *nomological net* (that is a set of correlations among scores of the variable a researcher wants to measure and scores for other, supposedly related. variables) (Borsboom and colleagues, 2004).
- Validity is, then, exclusively an ontological matter.
- To assert that a measure indexes what it is supposed to measure means that the scores generated by the measure reflect a latent variable of interest (e.g., PYD). To reiterate, one must assume that the scores derived from the measure are caused by the existence of the latent variable.

# The Concept of Validity

- The developmental science approach stands in contrast to the Logical Positivist approach of the early-to-mid 20<sup>th</sup> century and to its remnants today which may be termed neo-Positivist.
- The Logical Positivist approach is often manifested in the contention that the psychometric aspects of a measure—reliability, validity, or even measurement invariance/equivalence—reside in items or scales themselves.
- That is:
  - (1) latent variables are just summaries of empirical relations among manifest variables (and thus they have no real existence) and
  - (2) measurement properties exist in the words of a survey or measure that are printed on a page or that appear on a screen.
- **However, in developmental science, valid, invariant, and reliable measures per se do not exist!**

# The Concept of Validity

- What exists are valid, invariant, and reliable ***behaviors*** of a specific sample of participants coacting within the micro-context of a specific set of items or indicators of a latent variable, as this coaction occurs at a specific time and in a specific place.
- Simply, reliable and valid behavior resides in humans and not in words printed on a page or shown on a screen.

# From Validity to Measurement Invariance

- There has been growth in the last two decades of adding to presentations about reliability and validity a discussion of a measure's *invariance* across specific groups of people, specific periods of development, specific periods of historical time, and specific places (ranging from families, through communities, to nations and cultures).
- This interest springs from the growing recognition among developmental scientists that the behavior (performance) of a specific individual or a specific group of individuals in regard to the micro-context of items or surveys varies because of intraindividual changes in the functioning of specific latent variables across the life span and across contexts.
- Across the history of measurement, the Logical Positivist perspective has valued the predictive use of a measure more than the truth of human behavior and development being revealed by a measure.
- In other words, valid knowledge about the holistic behavior and development of an individual has been of less value than being able to predict variance in another construct or of the same construct at a subsequent time.

## From Validity to Measurement Invariance

- Measurement undertaken in developmental science is done in the service of describing, explaining, and optimizing the holistic development **of an individual**, that is, of intraindividual change within a specific person across time and place.
- As such, validity and not (correlational) utility is the primary value of measurement in developmental science.
- Consequently, the bedrock of valid measurement, when seen in the Specificity Principle of developmental science, is measurement invariance – understanding if and when a measure of a specific individual, group, time, or place can be generalized to another individual, group, time, and place.

# Reliability: The “Third Leg” of the Measurement Stool

- Prioritization of validity in measurement in developmental science is congruent with prioritization of ontology, as compared to epistemology (the branch of philosophy dealing with the study of knowledge).
- In discussions of measurement of latent variables in human behavior and development, reliability is often prioritized over validity.
- One reason for emphasizing reliability is that there are known and often-cited quantitative “rules of thumb” regarding values that constitute acceptable levels of reliability (e.g., Cronbach alpha coefficients of 0.7) and good levels of reliability (e.g., Cronbach alpha coefficients of 0.8).
- However, there are no “rules of thumb” levels for quantifying acceptable or good validity.



# Reliability: The “Third Leg” of the Measurement Stool

- A key reason for the difference between reliability and validity is that validity is a question of the ontology of a construct, whereas reliability is a question of epistemology, of what is known about the dependability, repeatability, or internal consistency of a measure.
- The often-used approaches to enhance reliability (e.g., in developmental research, trimming an item set to increase the magnitude of Cronbach’s alpha coefficient and doing so without regard for evidence of *essential tau equivalence*) may increase the reliability estimate by retaining items that are not valid indicators of the latent variable.

# Trade-Offs between Reliability and Validity

- Some well-known tradeoffs between (increased) reliability and (decreased) validity are:
  - Administering items in unidimensional blocks
  - Administering items in fixed orders
  - Administering no or few opposite-scored items
  - Using items with similar language
  - Deleting items with diverse means
  - Deleting all items pertaining to a less-related facet of the latent construct
  - Ignoring response biases
- Using items or scale scores with compromised validity also compromises the likelihood of establishing evidence for any level of measurement invariance (configural-, metric- or weak-, scalar- or strong-, or strict-invariance).

# Examples of the Developmental Science Approach to Measurement in Research and Evaluation in Global Approaches to Character Development

## 1. Adapting and contextualizing existing measures (of Spirituality, Hope, and Thriving) for use in El Salvador

**Measuring Spirituality, Hope, and Thriving Among Salvadoran Youth: Initial Findings from the Compassion International Study of Positive Youth Development**

Jonathan M. Tirrell<sup>1</sup> · G. John Geldhof<sup>2</sup> · Pamela Ebstyn King<sup>3</sup> · Elizabeth M. Dowling<sup>1</sup> · Alistair T. R. Sim<sup>4</sup> · Kate Williams<sup>4</sup> · Guillermo Iraheta<sup>4</sup> · Jacqueline V. Lerner<sup>5</sup> · Richard M. Lerner<sup>1</sup>

## 2. Generating and testing a new measure (of program quality) for use in Rwanda

**Toward a Measure for Assessing Features of Effective Youth Development Programs: Contextual Safety and the “Big Three” Components of Positive Youth Development Programs in Rwanda**

Jonathan M. Tirrell<sup>1</sup>  · Elizabeth M. Dowling<sup>1</sup> · Patricia Gansert<sup>1</sup> · Mary Buckingham<sup>1</sup> · Caitlin A. Wong<sup>2</sup> · Sara Suzuki<sup>3</sup> · Catherine Naliaka<sup>4</sup> · Patience Kibbedi<sup>4</sup> · Emmanuel Namurinda<sup>4</sup> · Kate Williams<sup>4</sup> · G. John Geldhof<sup>5</sup> · Jacqueline V. Lerner<sup>3</sup> · Pamela Ebstyn King<sup>6</sup> · Alistair T. R. Sim<sup>4</sup> · Richard M. Lerner<sup>1</sup>

# Examples of the Developmental Science Approach to Measurement in Research and Evaluation in Global Approaches to Character Development

## 1. Adapting and contextualizing existing measures for use in El Salvador: Sample case of the Hopeful Future Expectations (HFE) scale

Think about how you see your future. What are your chances for the following?	
1.	Graduating from high school.
2.	Going to college.
3.	Having a job that pays well.
4.	Having a job that you like.
5.	Having a happy family life.
6.	To live wherever you want.
7.	To buy the things you need.
8.	To do the things you would like to do.
9.	Be respected in your community.
10.	Have trustworthy friendships.
11.	Be healthy.
12.	Be safe.

1. We started with 12 items from the original HFE scale.
2. We translated and back-translated the items and engaged local stakeholders (practitioners and data collectors) in focus group discussions about item language, content, relevance, meaningfulness.
3. Once data were collected, a confirmatory factor analysis (CFA) demonstrated poor fit to the data. We reviewed modification indices (chi-square reduction values) for points of misfit.
4. Follow-up exploratory factor analyses (EFAs) suggested several factors emerging from the 12 items (e.g., about education, financial success, and happy life) as well as extraneous items not loading on any factor, but with varying fit indices and factor loadings.
5. Seeking a one-factor solution that was most relevant and useful with the sample, we iteratively tested models across subgroups present in the data set, trimming items accordingly, and ended with 6 items (highlighted) that were parsimonious and robust across groups.
6. The single factor demonstrated strong invariance across subgroups.

## **Examples of the Developmental Science Approach to Measurement in Research and Evaluation in Global Approaches to Character Development**

### **2. Generating and testing a new measure for use in Rwanda: Sample case of the “Big Three” measure of effective PYD programs**

- 1. No measure existed in the literature to assess program quality, particularly for the theory-predicated “Big Three” features of effective PYD programs (mentorship, life-skill-building activities, and opportunities for youth contribution and leadership; and safety as a prerequisite).**
- 2. With our stakeholders (local researcher and practitioner partners) we collaboratively generated a pool of 29 items crafted to reflect the 4 factors specific to the context.**
- 3. We again translated and back-translated the items, and continued stakeholder discussions to further refine item language, content, relevance, and meaningfulness after translation.**
- 4. Once data were collected, an initial CFA demonstrated poor fit to the data.**
- 5. Iterative EFAs and CFAs were replicated across subgroups to trim items that were not loading well onto constructs, and to correlate errors of related items.**
- 6. We refined the item pool to 24 items reflecting the 4 items and included an item of Contribution to check for validity.**
- 7. We tested for and established strong invariance across the subgroups.**

# Selected References

- Bornstein, M. H. (2019). Fostering optimal development and averting detrimental development: Prescriptions, proscriptions, and specificity. *Applied Developmental Science, 23*(4), 340-345.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*(4), 1061-1071.
- Card, N. A. (2017). Methodological issues in measuring the development of character. *Journal of Character Education, 13*(2), 29-45.
- Clifton, J. D. W. (2020). Managing validity versus reliability trade-offs in scale-building decisions. *Psychological Methods, 25*(3), 259-270.
- Flora, D. B. (2020). Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates. *Advances in Methods and Practices in Psychological Science, 3*(4), 484-501.
- Rose, T. (2016). *The end of average: How we succeed in a world that values sameness*. New York, NY: HarperCollins Publishers.
- Tirrell, J. M., Geldhof, G. J., King, P. E., Dowling, E., Sim, A., Williams, K., Iraheta, G., Lerner, J. V., & Lerner, R. M. (2019). Measuring spirituality, hope, and thriving among Salvadoran youth: Initial findings from the Compassion International Study of Positive Youth Development. *Child & Youth Care Forum, 48*(2), 241-268.
- Tirrell, J. M., Dowling, E. M., Kibbedi, P., Namurinda, E., Iraheta, G., Dennis, J., Malvese, K., Abbasi-Asl, R., Williams, K., Lerner, J. V., King, P. E., Sim, A. T. R., & Lerner, R. M. (in press). Measuring youth perceptions of being known and loved and positive youth development: Cross-national findings from Rwanda and El Salvador. *Child and Youth Care Forum*.