Ten Myths of Methodology for Studying Character Development or for Evaluating Character Education Programs

Richard M. Lerner and Elizabeth M. Dowling
Institute for Applied Research in Youth Development
Tufts University

Some Truths about Scientific Methods

Bachrach's Truths about Research

1. "People don't usually do research the way people who write books about research say that people do research."[a]
2. "Research is not statistics."[b]
3. "Things take more time than they do."[a]

Bachrach, A.J. (1962). ***Psychological research: An introduction*****. Random House.**
a = p. vii, b = p. 1.

Theory is the Foundational Tool of Science

- The specific method used in good scientific research or evaluation must be derived from the theory-based question asked in the specific research.
- **Theory guides:**
  - The selection of what to study (e.g., what character virtues or strengths to measure);
  - Levels to study (e.g., the person, the context, both person and context);
  - Frames the times in life to study (e.g., childhood, young adulthood);
  - Frames how often to study a person and/or context (e.g., cross-sectionally or longitudinally, or both).
  - Determines the design of the study (e.g., interviews, surveys, or randomized control trial [RCT] experiments).
  - Influences the statistical methods used to analyze the data.
  - Impacts the interpretation – **the meaning** – attached to the findings of the study.

The Scientific Domains of Studying Character Development and Evaluating Character Education Programs

- Both domains of character scholarship deal with change in a person across different time periods.

- Development pertains to changes occurring during specific periods of life, for example, childhood, adolescence, or even the entire life span.

- Educational changes pertain to the specific time between the beginning and end of an educational program and, possibly, with longer-term impacts on character.

- However, both domains assess both quantitative (variational) and qualitative (transformational) aspects of change.

The Methods of Developmental Research and of Program Evaluation

- For both research and evaluation, there are three areas of methodology:
  - *Measuring change* (measurement of character)
  - *Research design* (planning the points in time when character should be measured)
  - *Data analysis* (Selecting the quantitative or qualitative tools used to assess change in character)

- Many of the beliefs that researchers possess about these facets of methodology are not correct.

- These beliefs are actually myths!

- There are at least ten (10) such myths.

- The first three derive from theory.

Myth 1

Studying character *trait* development

or

Building character *traits* through character education programs

are useful and important endeavors.

False

The concept of a trait does not align with the concepts of development or education.

Dictionary Definitions of Trait Align with the Definition of Trait in General, Developmental, and Educational Psychology

- The *Oxford Review Encyclopedia of Terms*, https://oxford-review.com/oxford-review-encyclopaedia-terms/the-difference-between-an-state-and-a-trait/, defines a **trait** as something that is part of an individual's personality and therefore a long-term characteristic of an individual that is evident through their behavior and emotions.

- The *Merriam-Webster Dictionary* https://www.merriam-webster.com/dictionary/trait defines **trait** as a distinguishing quality (as of personal character), an inherited characteristic.

Dictionary Definitions of Trait Align with the Definition of Trait in General, Developmental, and Educational Psychology

- The trait approach to understanding human personality or character is exemplified by the **Five Factor Theory** (**FFT**), a conception championed by Costa, McCrae and their colleagues for more than four decades (e.g., Costa & McCrae, 1980; 2006; McCrae, et al., 2000).

- For instance, McCrae, Costa, et al. (2000, pp. 175-176) believe that personality traits reflect "nature over nurture" and that "personality traits are more or less immune to environmental influences…significant variations in life experiences have little or no effect on measured personality traits." They argue that "Barring interventions or catastrophic events, personality traits appear to be essentially fixed after age 30" (Costa, McCrae, & Siegler, 1999, p. 130).

- Therefore, in both common parlance and in some areas of science, character traits are held to be fixed, stable, and biologically-set facets of individual functioning that, just as is the case for eye color, do not change across time, no matter how much development occurs or how many high quality character education programs are experienced.

There is *Very* Weak Evidence that Character Traits Even Exist!

- If an attribute of character exists as a trait akin to eye color, then scores for character at one time in life would account for 100% of the scores for that same facet of character at all later times in life, despite the varying settings and experiences people have across life. Brown eye color remains brown eye color for childhood, through adolescence, and across adulthood.

- However, the percentage agreement of scores for the same character attribute across life rarely exceeds 50% agreement and, ordinarily, agreement is less than 10%!

- The reasons for the lack of agreement cannot be accounted for by genes, because they have remained the same across life!

- Therefore, do not refer to character traits in the study of character development or in the evaluation of character education programs.

- Doing so is tantamount to "shooting oneself in the foot!"

Myth 2

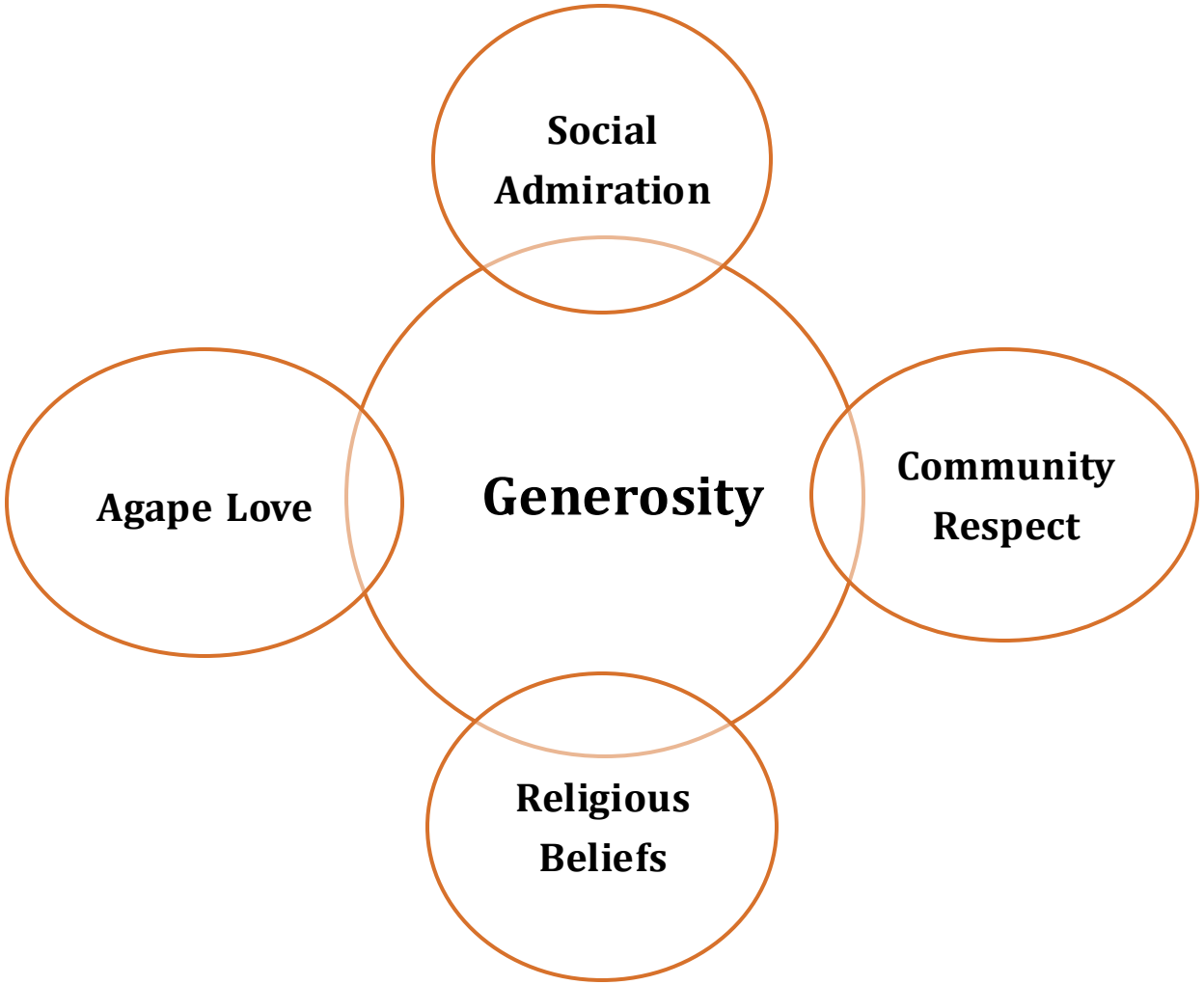
The validity of a character measure can be established by the interrelations (e.g., the correlations) between the scores from the measure with scores from measures of other constructs.

False

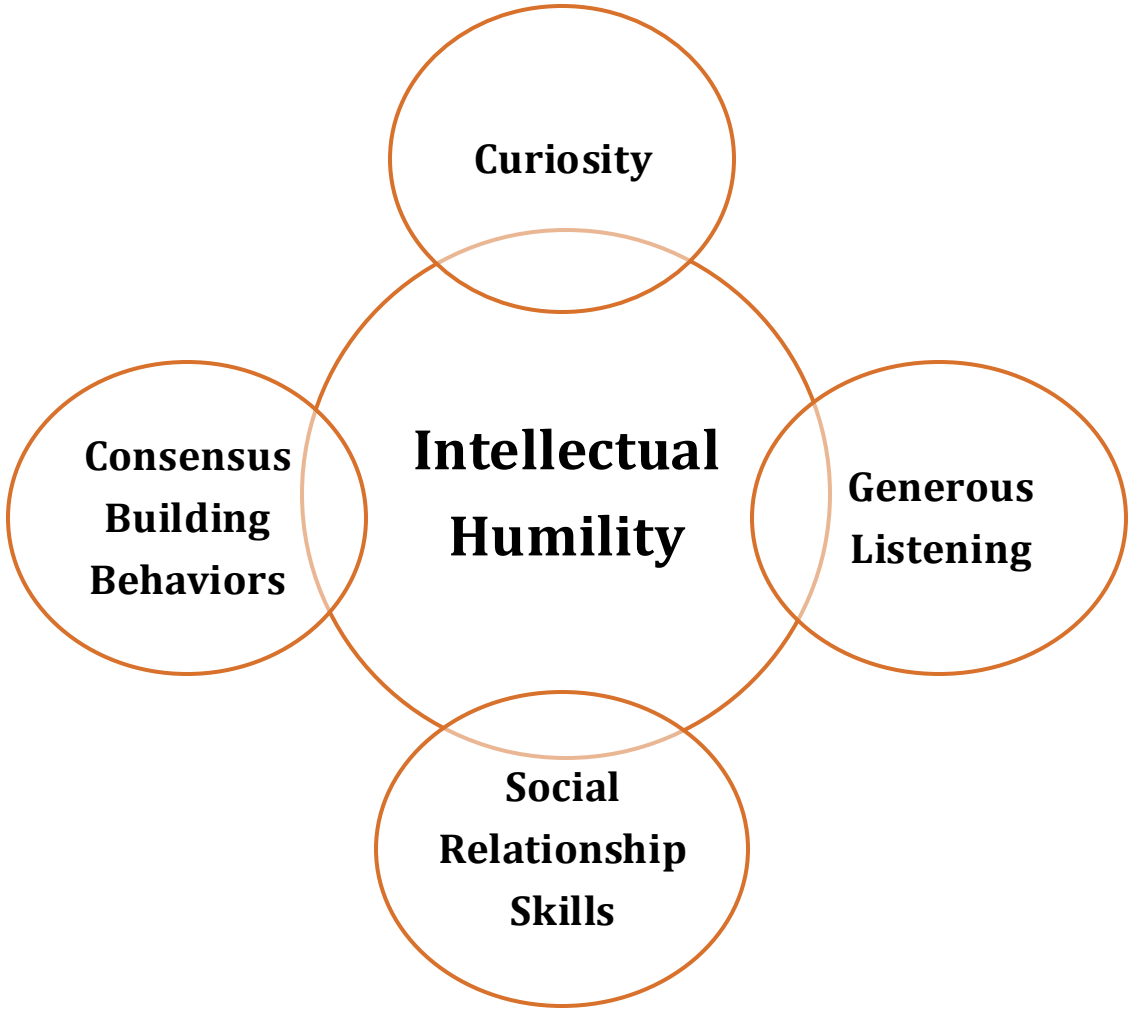
- Validity is a theoretical concept; it is not created because scores from a measure of character and scores of other attributes covary, are correlated, with each other.

- The possession of a character attribute may predict other concepts, but character is not equal to these other concepts.

Possession of the Character Attribute of Generosity May Predict (Be Correlated With) Other Attributes That are Not Generosity



Possession of the Character Attribute of Intellectual Humility May Predict (Be Correlated With) Other Attributes that are Not Intellectual Humility

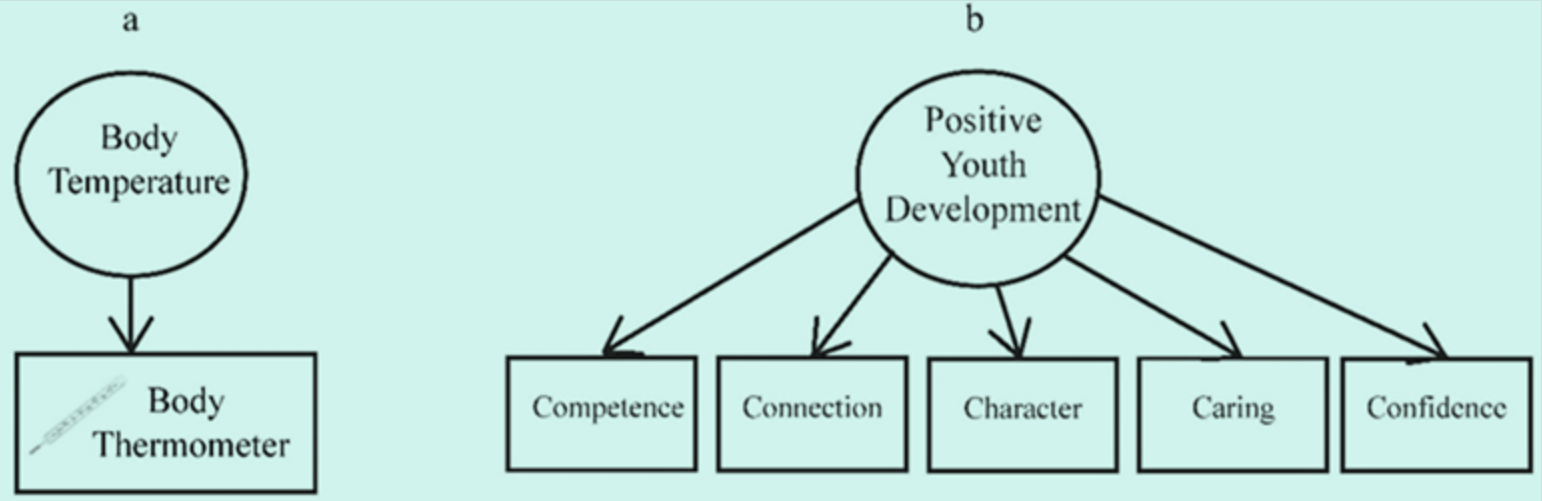


Myth 3

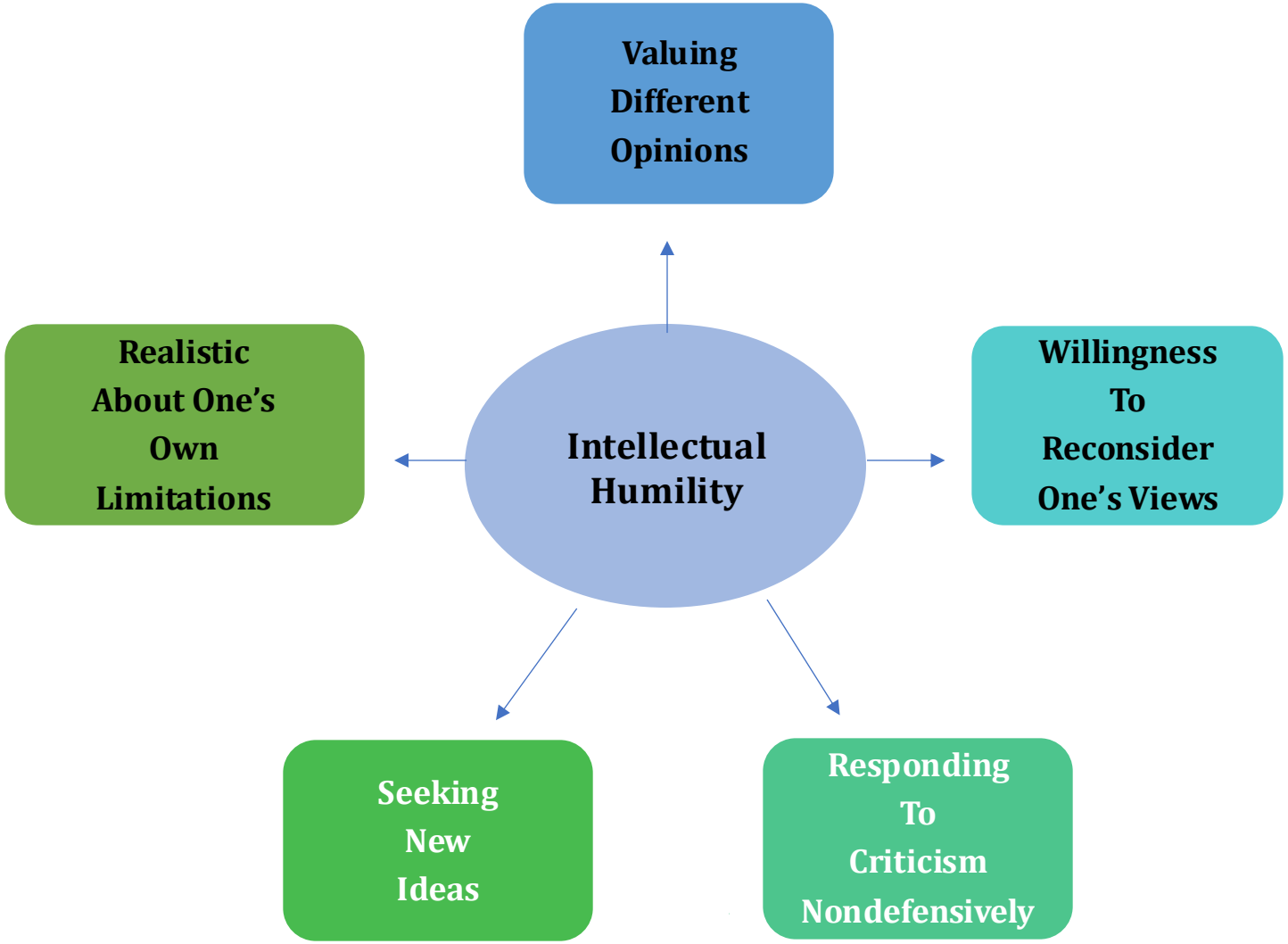
Character is what is measured by the character development items included in the character measure being used.

False

- The validity of a measure of character is not created by scores from observable attributes of thinking, feeling, or behaving.
- Unlike reliability, there is no validity coefficient. There are no numbers (e.g., correlations) that can be used to establish validity.
- A measure of character is valid if and only if it provides scores that would be theoretically held to exist if an attribute of character existed.
- Because validity is a theoretical concept, it should be thought of as causing the scores used to measure the concept.



Intellectual Humility (IH) and Five Attributes Created By Possession of IH



Myth 4

Traditional Randomized Control Trial (RCT)
Design
is the
Gold Standard
in
Character Education Program Evaluations.

False

Traditionally designed RCTs are “fools gold.”

Solomon, R. L., & Lessac, M. S. (1968). A control group design for experimental studies of developmental processes. *Psychological Bulletin, 70* (3, Pt.1), 145–150.

Experimental and Quasi-Experimental Designs

- A traditional randomized control trial (RCT) is designed to prove that a treatment (e.g., participation in a character education program) ***causes*** a change in character.

- The traditional design has two groups, and participants are randomly assigned to Group 1 or Group 2, so as to eliminate through randomization any systematic pre-existing conditions in participants that could impact the results (this step is termed controlling for *endogeneity*).[1]

  - Group 1 is measured (pre-tested), treated, and then remeasured (post-tested).

  - Group 2 is then measured (pre-tested), not treated, and than remeasured (post-tested).

- Causality can be claimed when, on remeasurement (post-testing), Group 1 is significantly different than Group 2.

[1] When random assignment to conditions cannot be achieved because research participants cannot be randomly assigned to conditions, such as age, race, poverty, etc., then the design is not considered a true experiment. It is considered a quasi-experimental design.

RCT Experimental Designs and Internal Validity

- The assertion of the existence of a causal relation is considered true because it is believed that the traditional, two-group RCT is ***internally valid***.

- Internal validity exists when treatment and only treatment can be associated with the remeasurement differences between Group 1 and Group 2.

RCT Designs May or May Not Possess Internal Validity

- Because of the claim of internal validity for the traditional, 2-group RCT design, RCTs have been considered the ***Gold Standard*** for proof of causality in character development research and in evaluations of character education programs.

- However, as implemented through the 20th century and the first three decades of the 21st century, the traditional design RCTs pertinent to the study of character of have been largely "**Fools' Gold**."

The Solomon & Lessac (1968) Four-Group Design:

Control Group 1 Controls Only for Having a Treatment

Experimental Group	**Control Group 1**
X	**X**
O	
X	**X**

The Solomon & Lessac (1968) Four-Group Design:

Control Group 2 Controls for Any Effect of Pre-Testing

Experimental Group	Control Groups 1 and 2	
X	X	X
O		O
X	X	

The Solomon & Lessac (1968) Four-Group Design:

Control Group 3 Controls for any Effects of Maturation, Growth, of Simply Development Between the Times of the Pre-Test and the Post-Test

Experimental Group	Control Groups 1, 2, and 3		
X	X	X	X
O		O	
X	X		

- Without all three control groups there is no legitimate claim of internal validity AND no legitimate claim of causality.

Myth 5

By studying *average changes* across time in a group of individuals involved in a character education program, researchers or evaluators can learn about character development or about the effectiveness of a character education program.

False

- Character education is about changing a specific individual's character through the person's participation in a character education program.

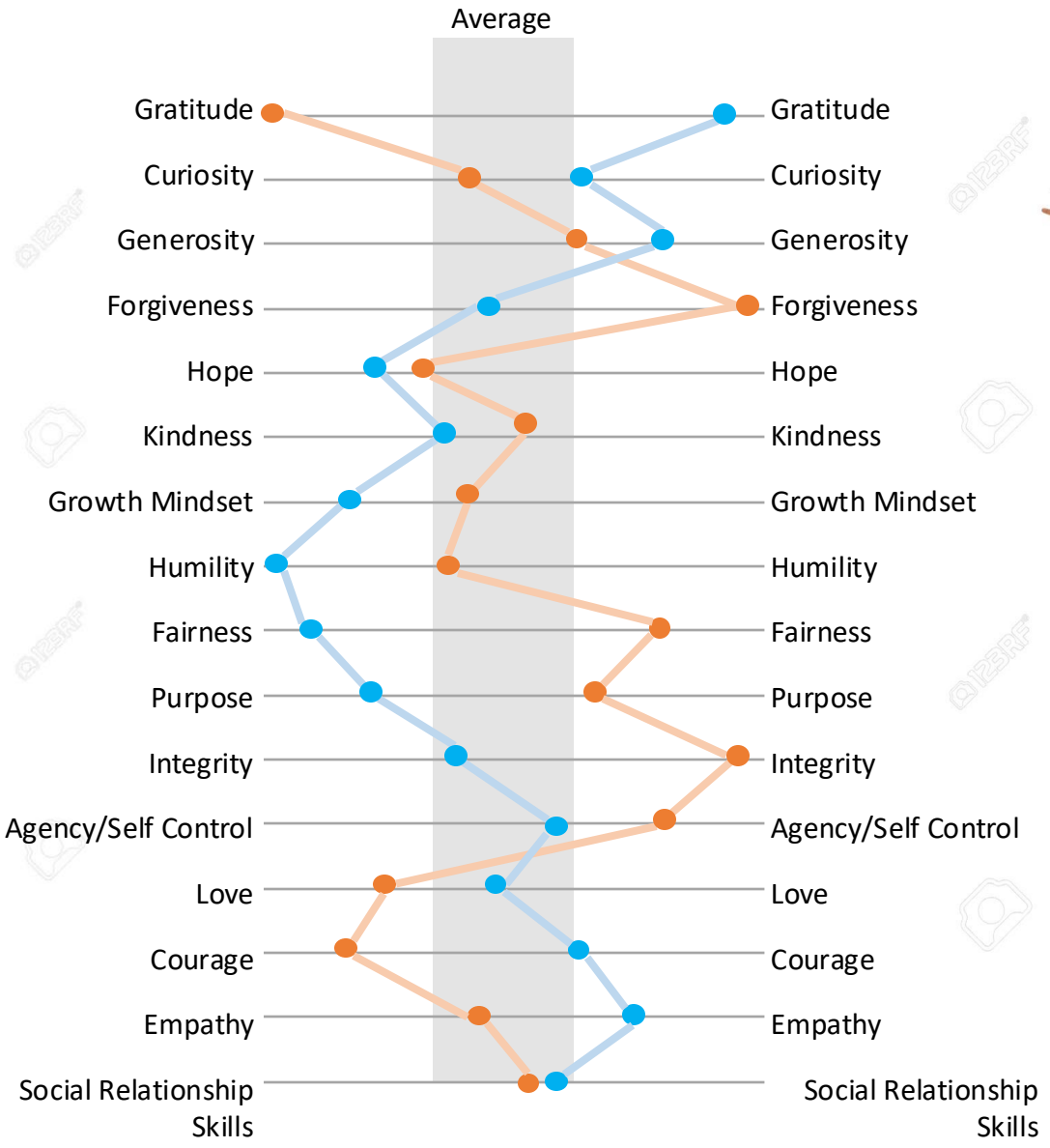
- It is about the change ***within*** an individual as a consequence of a specific individual-context relation.

- It is not about the average changes ***between*** people in a group. Average differences between people within or across time do not equal individual changes across time.

- To understand character, researchers or evaluators must consider ***BOTH differences*** in character scores ***between individuals*** in a sample (between-person differences) ***and the fluctuations*** of an individual's character scores ***within each individual*** across time and place (within-person variations).

- Although measures of character exist for between-person research and evaluation, there is a need for context-specific, valid, and reliable measures of character that can capture both within- and between-person variations.

Figure adopted from Saqr et al. (2024)

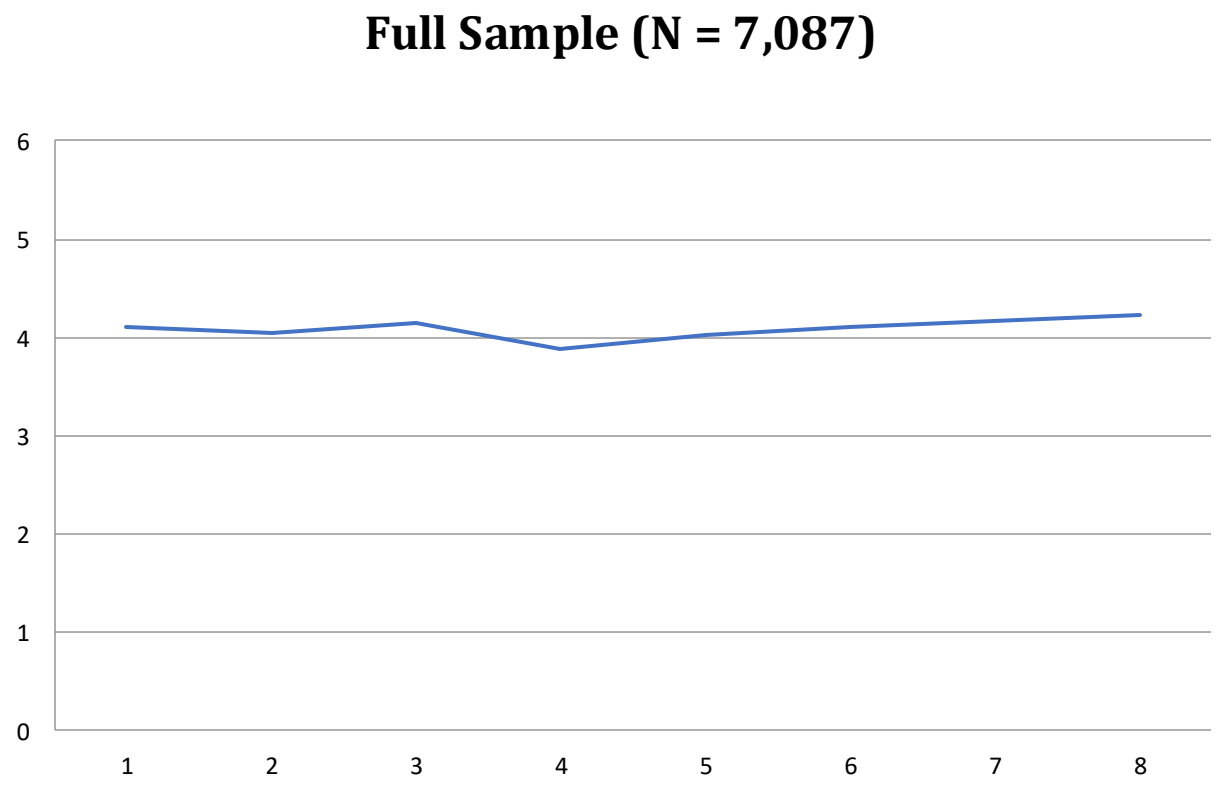

Average
Gratitude
Curiosity
Generosity
Forgiveness
Hope
Kindness
Growth Mindset
Humility
Fairness
Purpose
Integrity
Agency/Self Control
Love
Courage
Empathy
Social Relationship Skills
Gratitude
Curiosity
Generosity
Forgiveness
Hope
Kindness
Growth Mindset
Humility
Fairness
Purpose
Integrity
Agency/Self Control
Love
Courage
Empathy
Social Relationship Skills

The Lerner & Lerner
4-H Study of Positive Youth Development

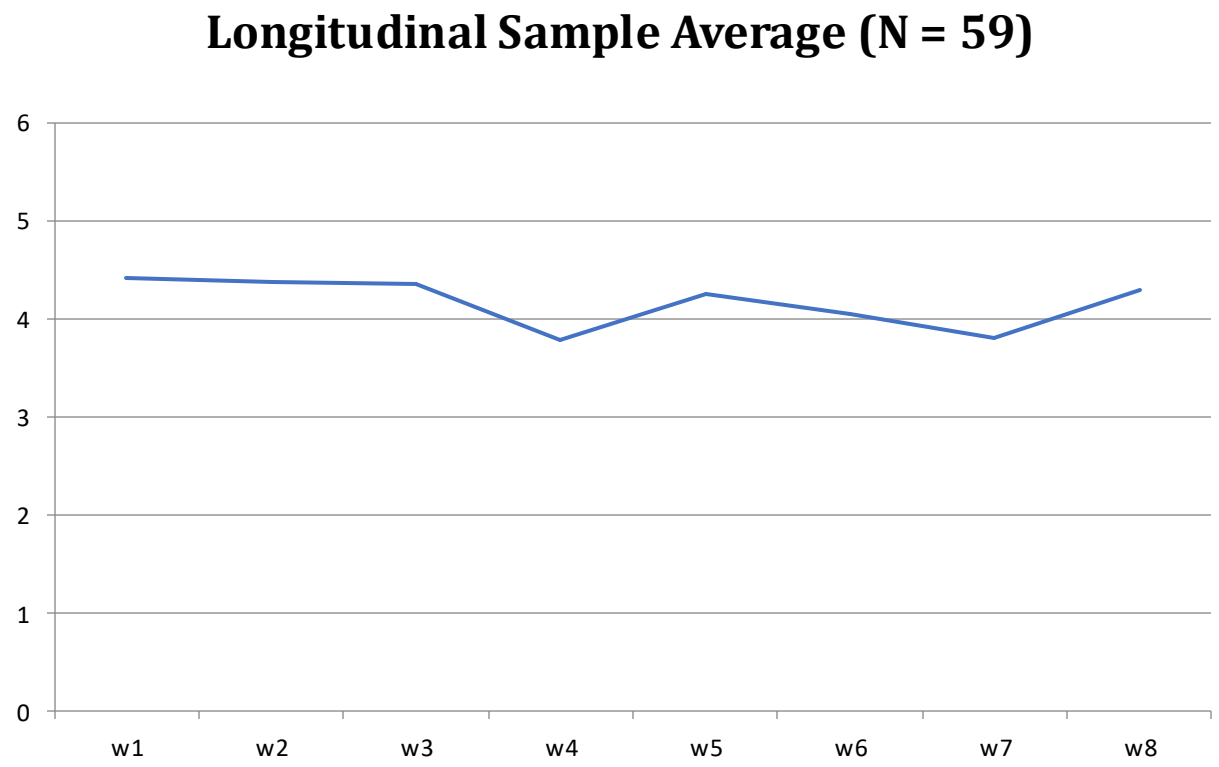
- 8 waves (Grades 5 to 12, respectively), 7087 participants
- Gender: 60.6% Female; 39.4% Male
- Race: 70.7% White; 10.2% Hispanic; 7.9% Black; 3.7% Multiethnic; 3.3% Other; 2.3% Native American; 2.1% Asian
- Mother's education: 33.6% 4-year degree or higher; 37.2% 2-year or technical degree; 20.5% High School; 8.6% less than High School
- Mean per capita income $15,279.26

Note that the longitudinal sample presented in these analyses includes all cases with at least 6 waves of data.

Agency/Self Control: Goal Optimization Skills

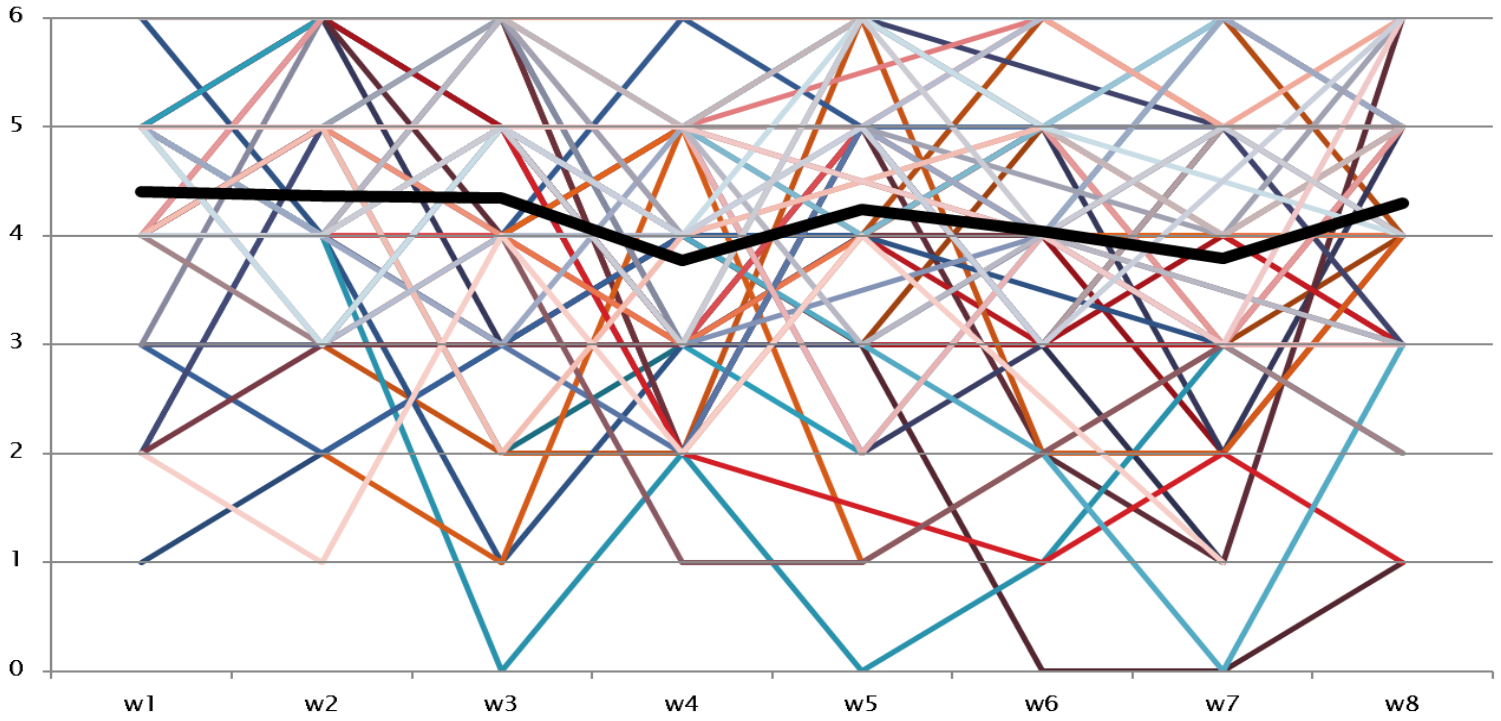


Agency/Self Control: Goal Optimization Skills



Agency/Self Control: Goal Optimization Skills

Person-Specific Pathways for Longitudinal Sample (N = 59)



The Sample Case of Change in Daily Emotional Experience

Data from a Study Conducted by

Nilam Ram, Peter Molenaar, and Colleagues
Developmental Systems Group
@PSU

and

John R. Nesselroade and Colleagues
Center for Developmental & Health Research Methodology
@UVA

Data Set

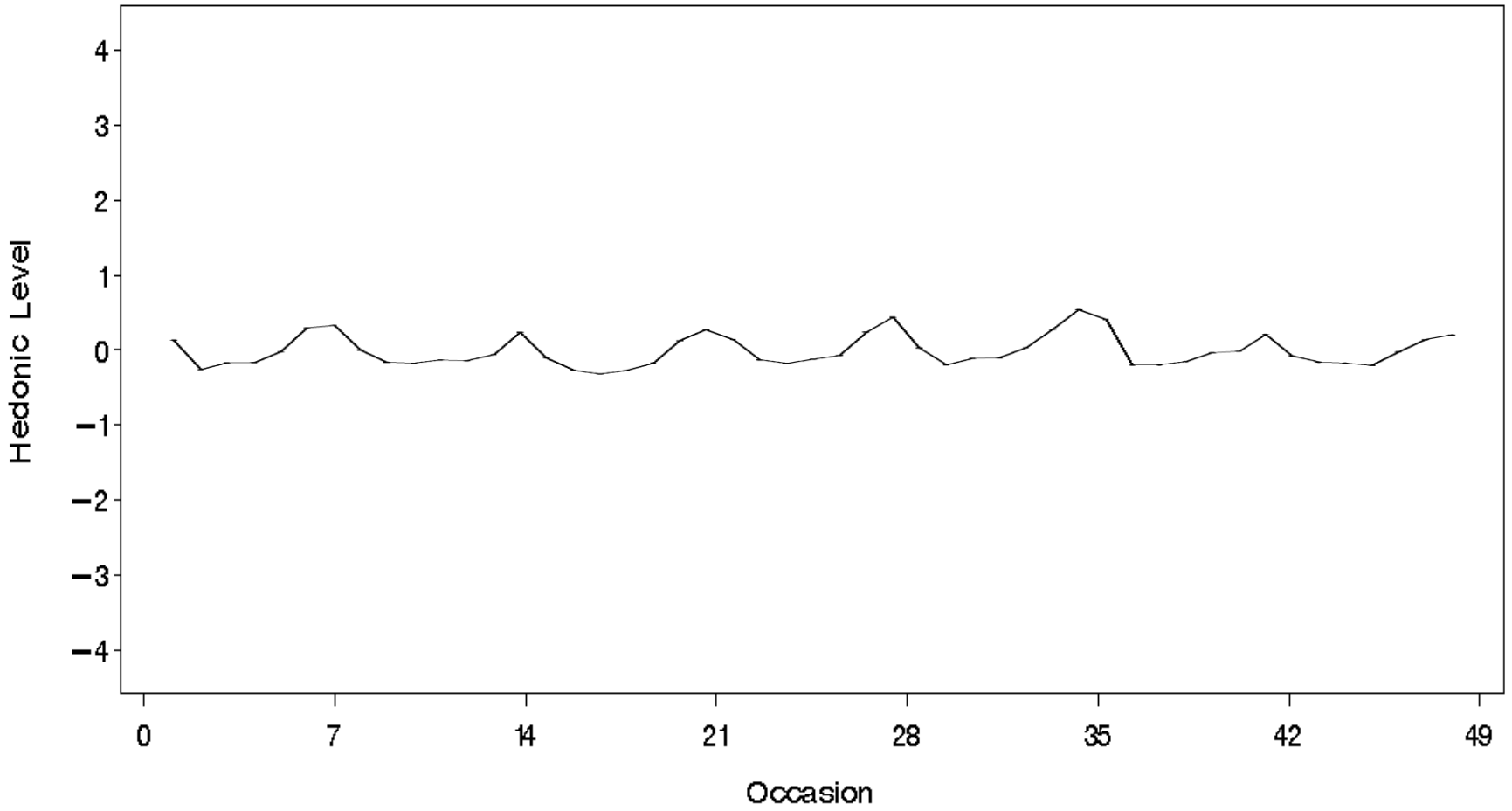
❖***Participants***:

- 180 students enrolled in a semester-long course on subjective well-being at the Univ. of Illinois:
  - Mean age = 20.2 (SD=1.8)

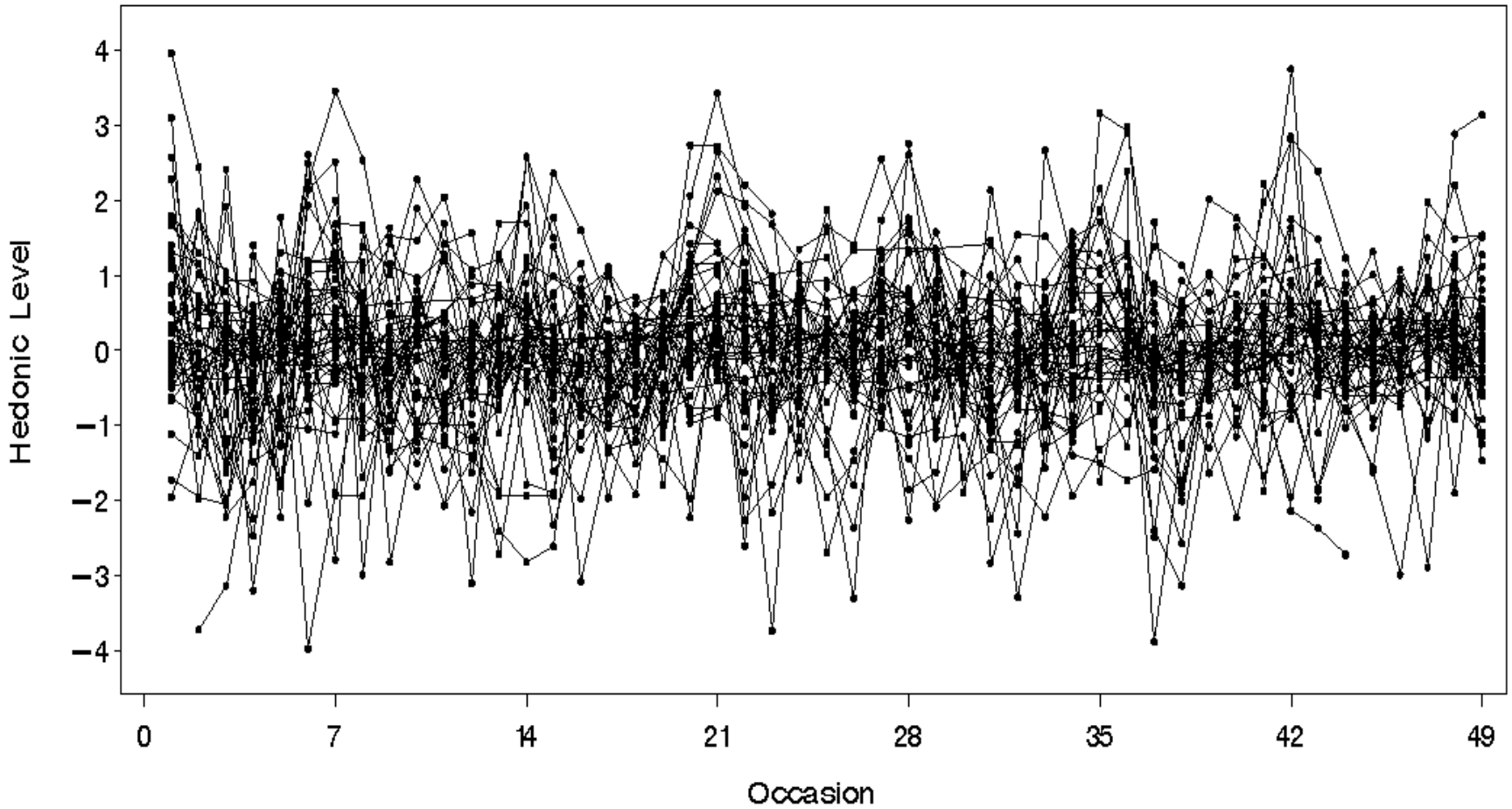
❖***Measures***:

- Daily Diary of Emotion Ratings completed on 50 successive days
  - How often was each emotion felt today?
- Hedonic Level
  - 8 Pleasant and 8 Unpleasant Affect Items

Daily Mood – Averaged Across All Individuals



Individual Daily Mood



The Sample Case of "Measures and Methods Across the Developmental Continuum" Project

Data from a Study Conducted by:

Data Sets

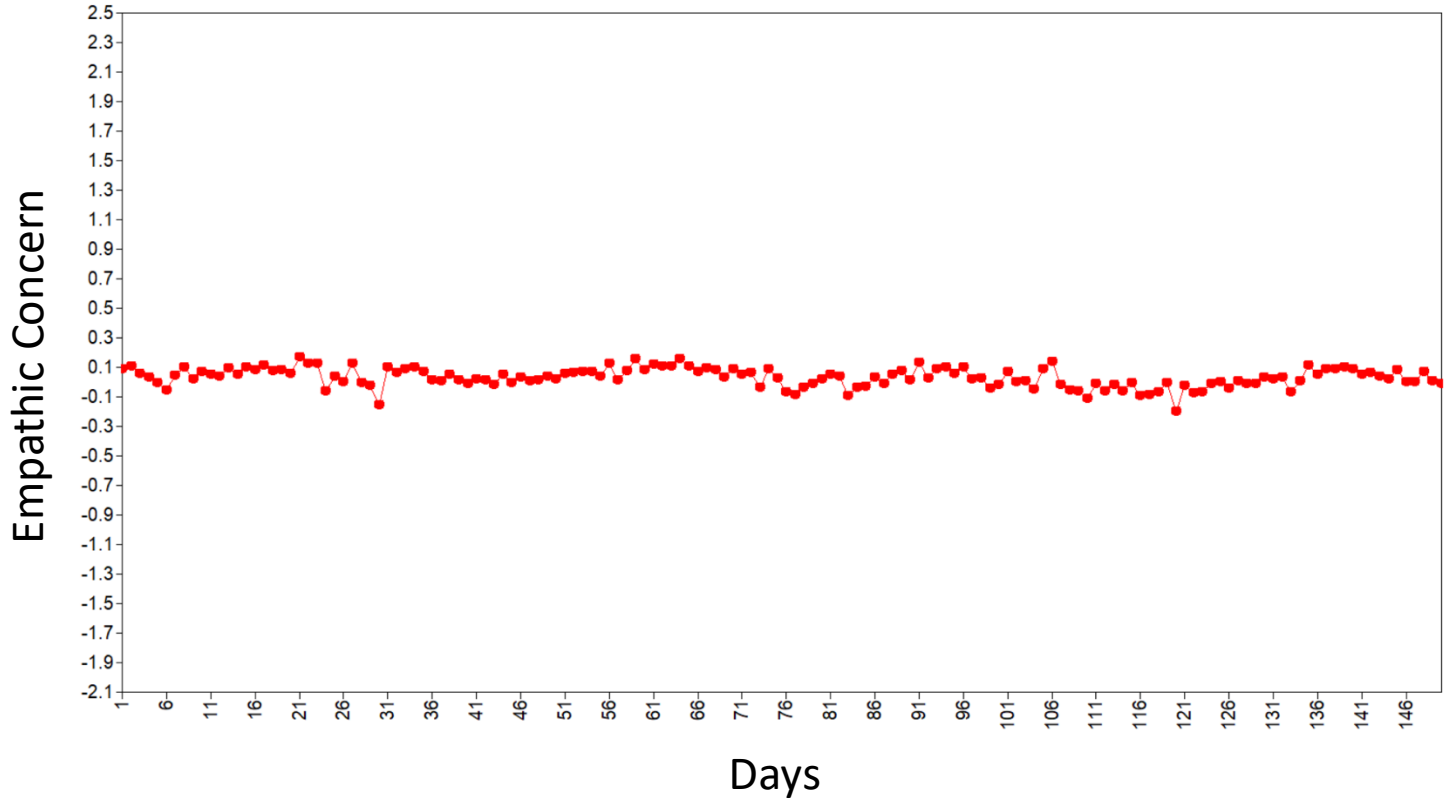
❖ ***Participants*:**

- **Group 1**: 35 students, mean age = 15.91 (SD=1.69).
- **Group 2**: 216 participants, mean age = 16.0 (SD = 0.82).

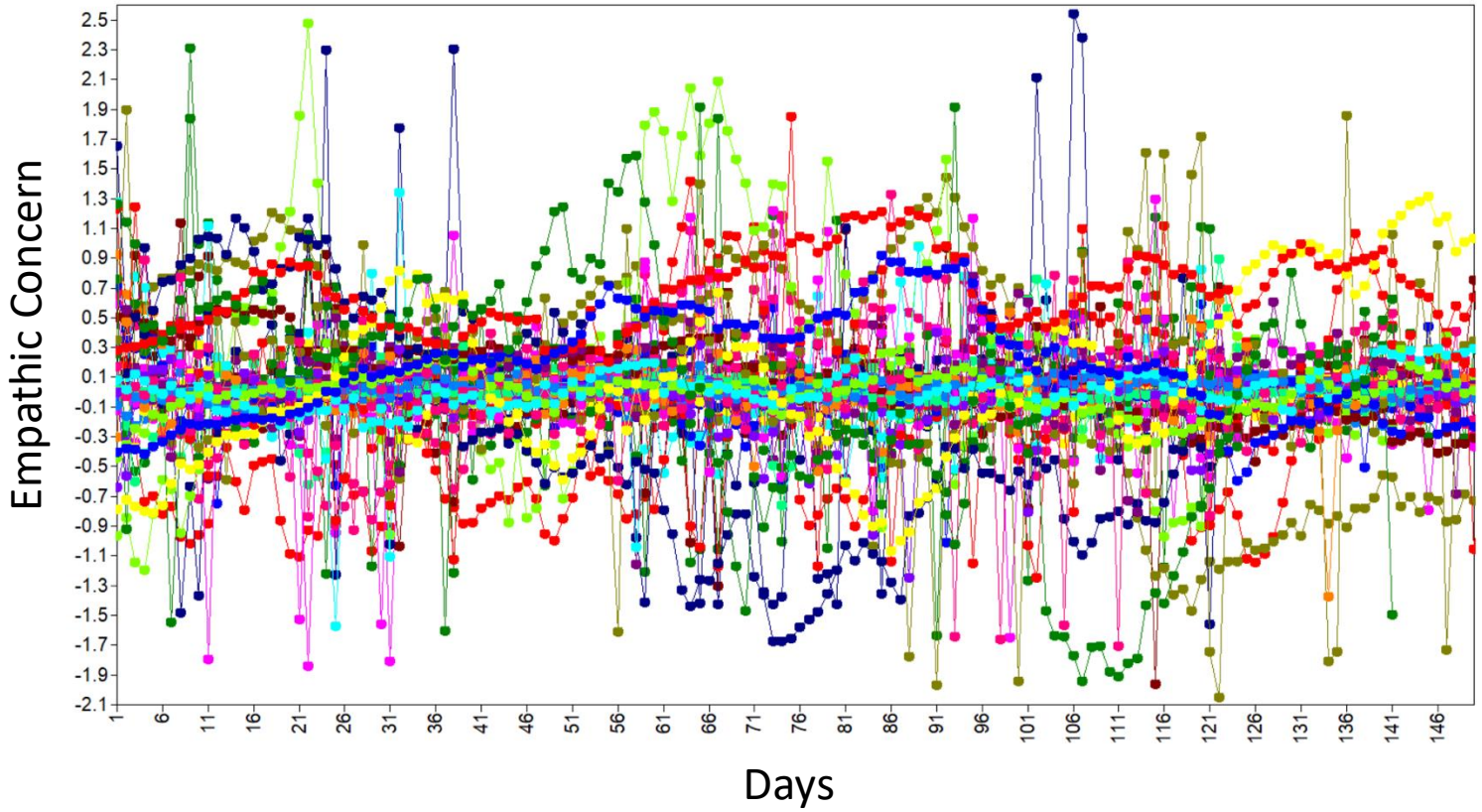
❖ ***Measures***:

- **Group 1 measures**: Self-reports of Empathic Concern for classmates on approximately 36 occasions across one academic semester.
- **Group 2 measures**: Self-ratings of the character attributes of Curiosity, Relationship Skills, and Generosity across approximately 15 weeks of the academic year.

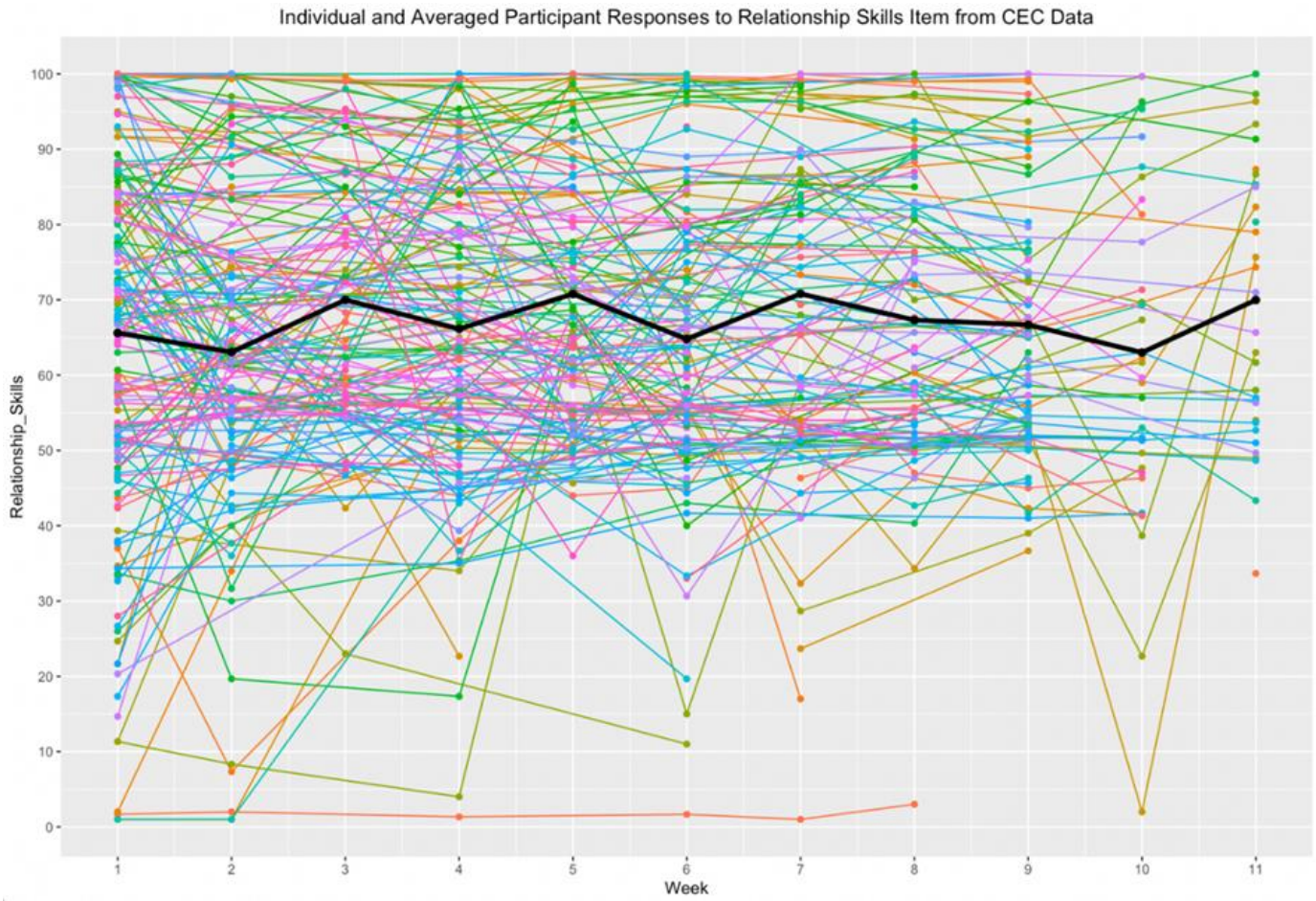
The Average Time-Series Plot of Empathic Concern



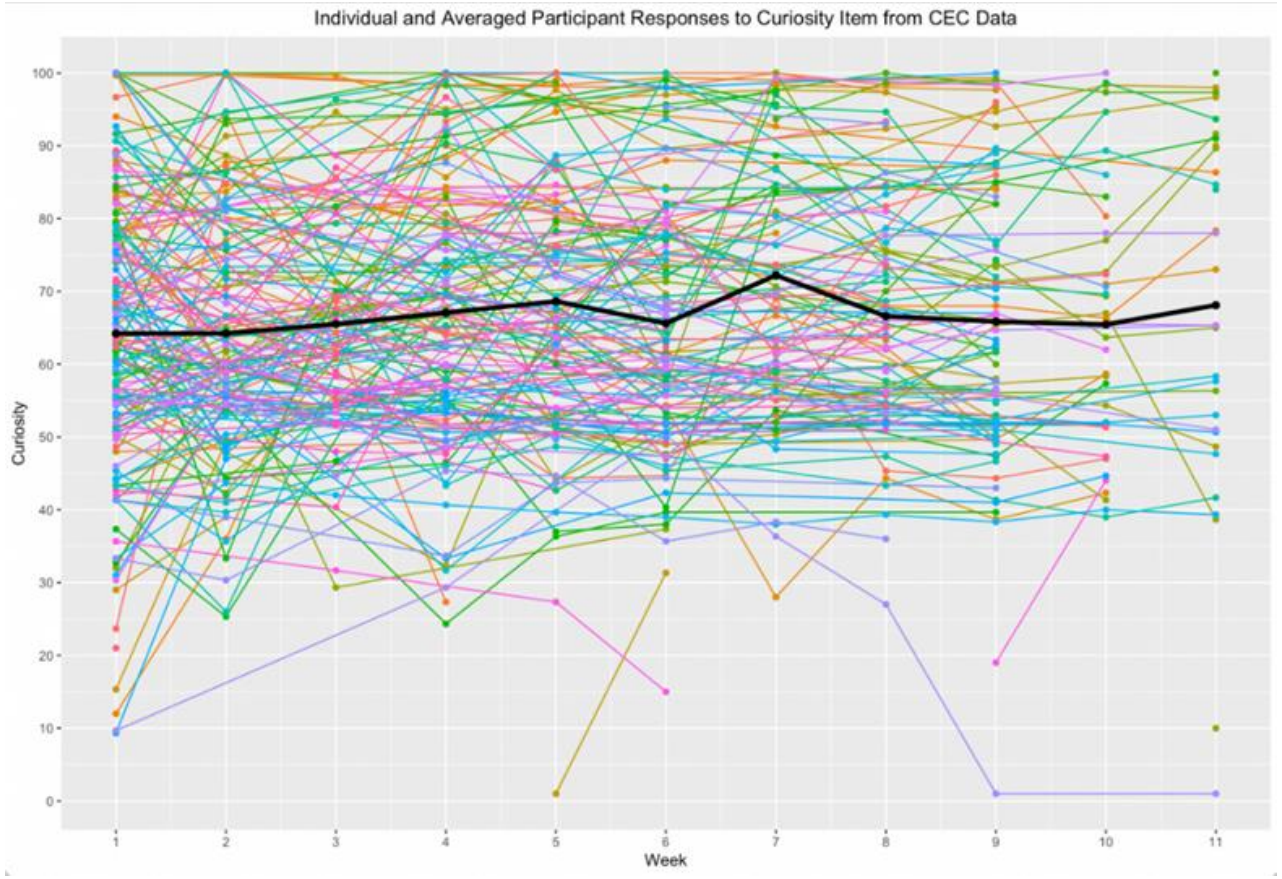
Time-Series Plot for 35 Participants



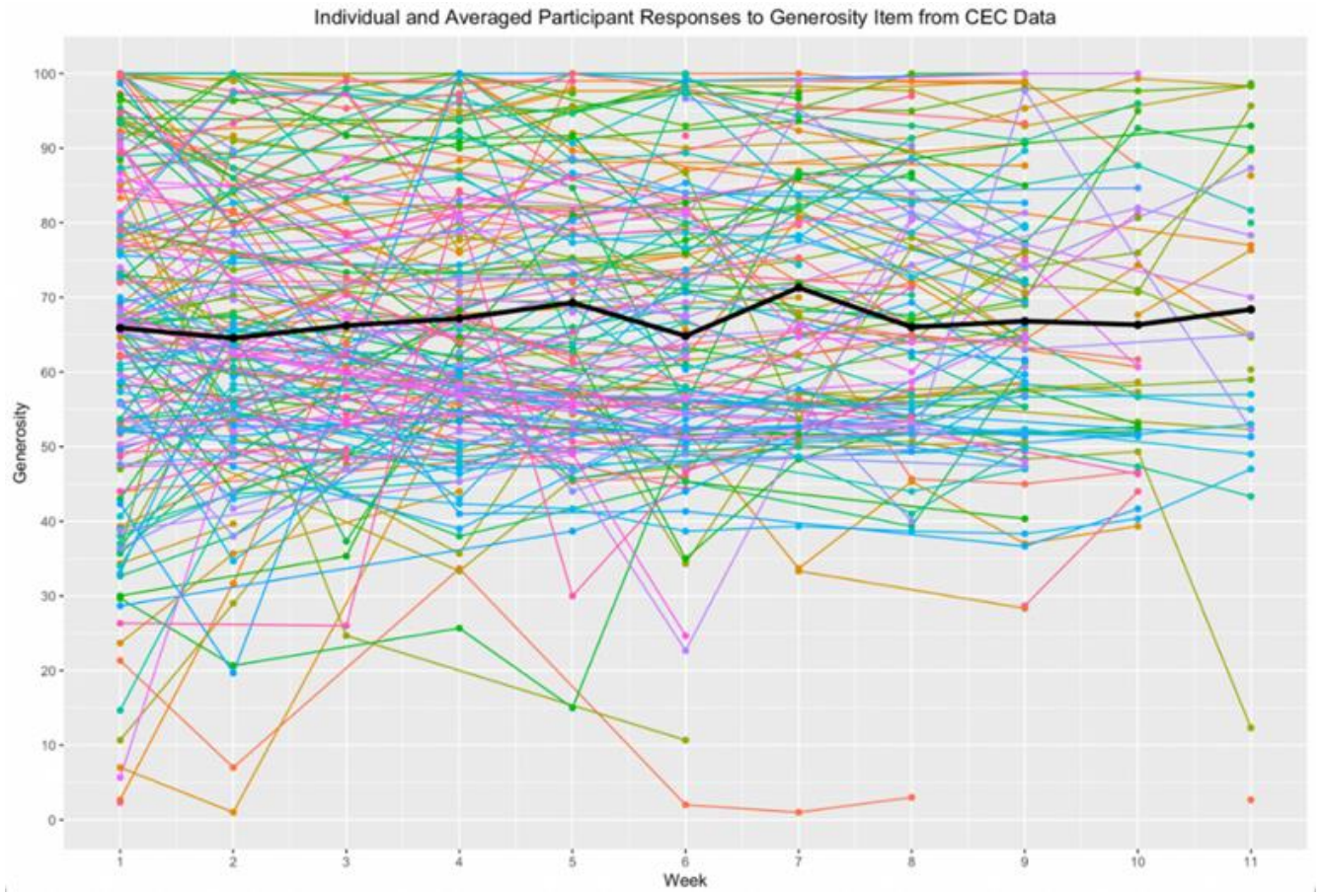
Relationship Skills



Intellectual Character: Curiosity



Generosity


Individual and Averaged Participant Responses to Generosity Item from CEC Data
Generosity
Week
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
11

Myth 6

Changes in an individual's character can be assessed by comparing scores for character attributes across two points in time.

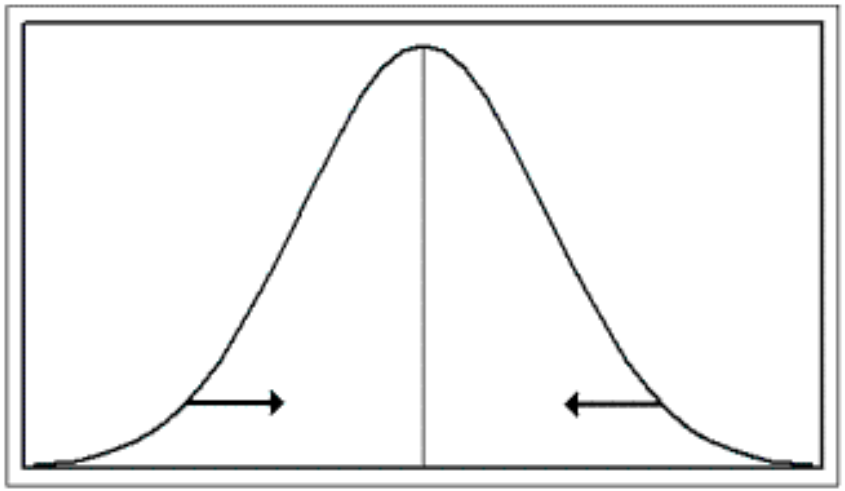
False

At least three (3) measurement occasions are needed (and even more is better) to protect against statistical artifacts (effects of regression to the mean) and to improve estimation of a developmental or change pathway.

The Third Times A Charm

- Regression to the mean is a statistical artifact of repeated sampling of a character attribute.

- If the first time a character attribute is measured the score is far away from the average score for a group, then when the attribute is measured a second time the score will be closer to the mean of the group.



- Therefore, the change in scores from Time 1 to Time 2 is more likely to reflect the effect of regression to the mean than true, meaningful, change resulting from development or the impact of a character education program.

The Third Times A Charm

- However, John Nesselroade and colleagues[1] found that regression to the mean is largely a phenomenon associated with two times of measurement

- It is much less likely to occur with three times of measurement and becomes increasing unlikely with even more times of measurement.

- Therefore, it you want to demonstrate development or the impact of your program, use at least three (3) times of measurement.

- And more is better!

- **[1]Nesselroade, J. R., Stigler, S. M., & Baltes, P. B. (1980). Regression toward the mean and the study of change. *Psychological Bulletin*, *88*(3), 622-637.**

Myth 7

In studying character development or in evaluating a character education program, measures should be used that have been shown to be reliable and valid with individuals as similar as possible to ones participating in the current research or program.

False

❖ Reliability and validity scores for people in past research are not necessarily applicable to the people participating in your study or your program.

❖ Reliability does not reside in the words printed on a page or appearing on a screen.

❖ INDIVIDUALS think, feel, or behave reliably or in a manner that actually reflects the character attribute of interest.

❖ Therefore, new people require new estimates of validity and reliability.

❖ Past information about validity and reliability can, at best, suggest the range of potential scores for your sample of people. But, if you are using a new group, past scores will not be relevant.

False (Continued)

- For example, reliable and valid measures of character in adolescents participating in a character education program in a suburb of Durham, North Carolina, the city of Boston, or rural areas near Corvallis, Oregon may not be relevant to reliability or validity for rural adolescents in the KwaZulu Natal Province of South Africa, in urban areas of Bogota, Colombia, or in the suburbs of Brisbane, Australia.

- The usefulness of measures in all these locations will vary in relation to cultural context. As well, the people in these settings vary in life experiences and in their respective current life circumstances. Putting a person from one point in history and cultural setting into a different time and place will not result in the same outcomes for the person as would have occurred in the individual's original niche.

- **Measures cannot adapt by themselves to being consistent across time and place. Researchers and evaluators need to adapt measures to the time and place – and people – for whom the measure is intended for use.**

Myth 8

Once a test or survey measure has been shown to be reliable and valid, it is okay to make minor wording changes to item stems or response options to make the tool more appropriate for the sample being studied.

False

- When reliability and validity are ascertained, the information pertains to ***specific*** people responding to the ***specific*** items and response options on a measurement tool (a survey or questionnaire).

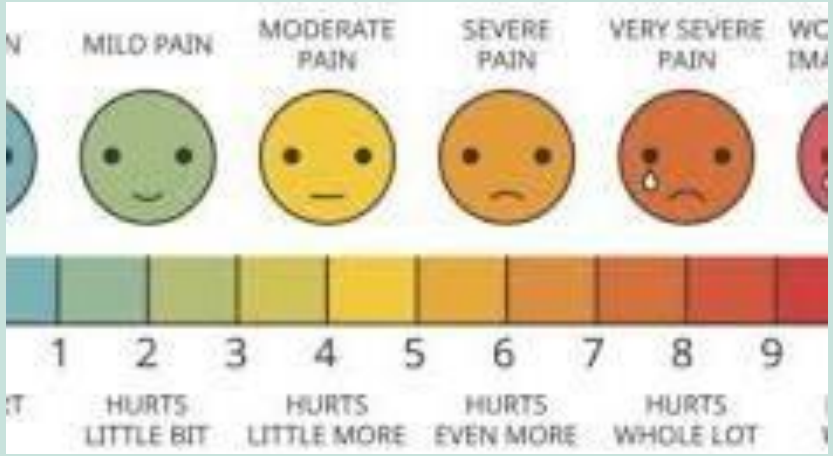
- This relation cannot be generalized to different people ***or*** to different tests.

- Individual ⇔ measure relations ***always*** vary across people, time, and place.

- Changing one word or one facet of the testing procedure means you have created a new measure, and new reliability and validity information is needed for the new individual ⇔ measure relation.

Myth 9

Use of Likert items is an appropriate way to assess the strength or magnitude of a character attribute.



EXAMPLE OF A FIVE-POINT LIKERT SCALE

Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1	2	3	4	5

False

True interval measurement is needed.

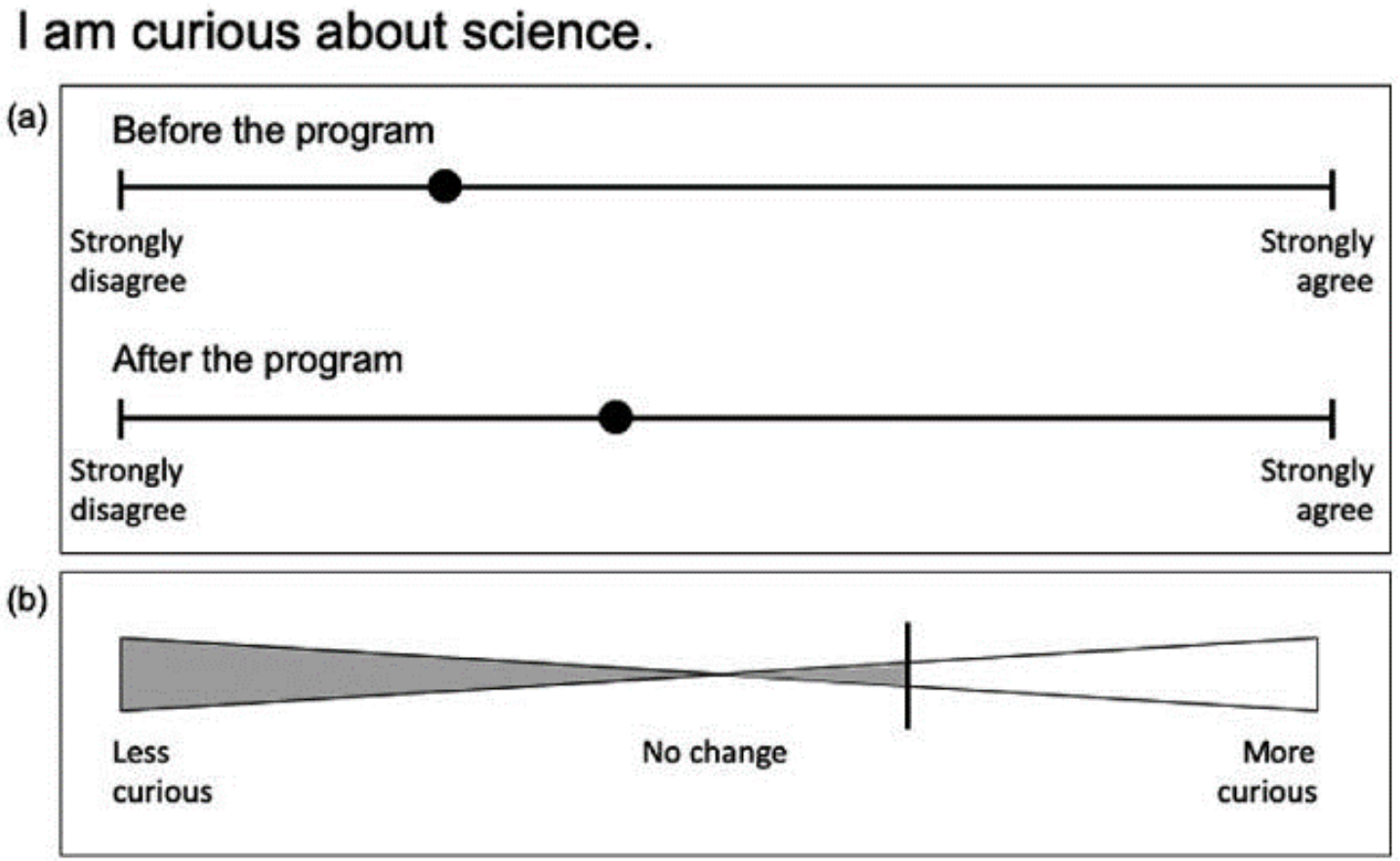
The Visual Analog Scale

- Historically, use of a true interval scale was difficult to implement because it required researchers to manually measure participants' responses using rulers, which made the Likert scale a more efficient option. However, with technological advances, the visual analog scale is now an efficient and often more valid option.

- Today, with automated and accurate electronic collection methods readily available, researchers should discontinue use of Likert-like scaling and implement more effective collection methods that involve true, interval measurement.

- An alternative to the Likert scale is the **visual analog scale**, which consists of a continuous number line with endpoint verbal anchors (and perhaps a midpoint anchor).

Two Versions of the Visual Analog Scale

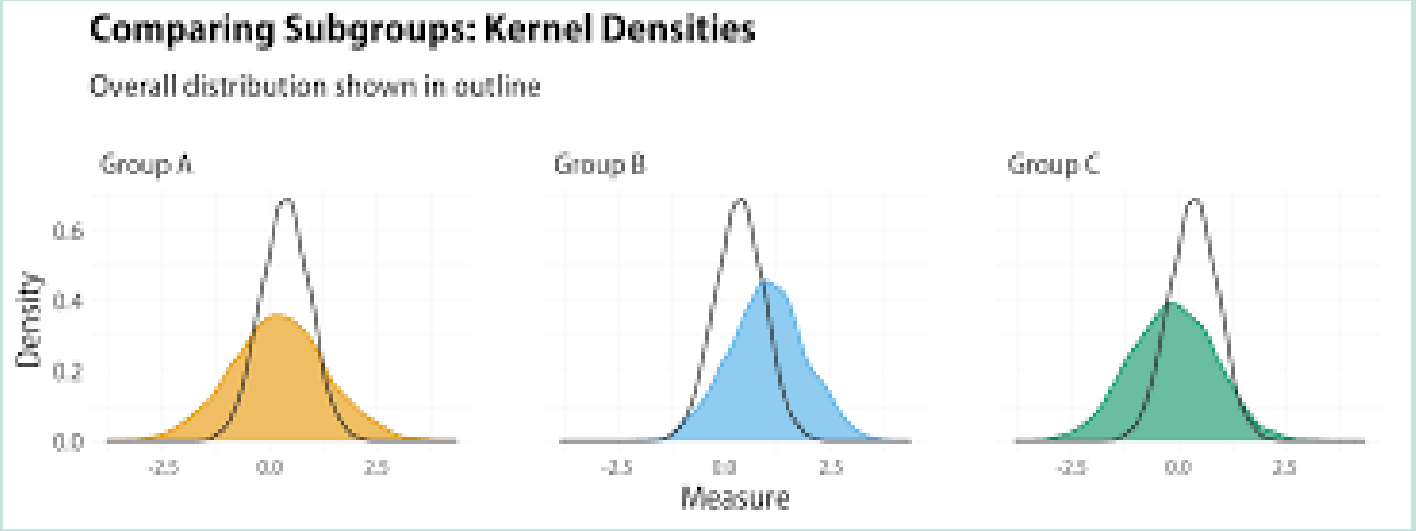


Myth 10

Once reliability and validity are established for a specific group being studied or evaluated, you can use scores from the measures to assess differences within the group, for example involving gender, age, race, religion, etc. subgroups.

False

- An average score for a specific group cannot be assumed to represent all subgroups in the study of evaluation.
- Such equivalence (also termed invariance) must be assessed.
- Today, there are easily accessible statistical and visualization procedures to determine if there is invariance across subgroups.



Implications and Next Steps

From Myths to Meaningful Research and Evaluation

- The examples we have provided of the facts countering the myths provide only an introduction to the methods that can be used to enact meaningful and important research about character development and to conduct rigorous evaluations of character education programs.

- Going beyond understanding why the myths are false and recognizing that there exist corrective paths to counter them are important first steps in acquiring the methodological skills needed to travel these paths well.

- Needless to say, common to all commitments to gain new expertise is, what Malcom Gladwell has estimated, a 10,000 hour investment.

- However, there are “tried and true” collaborative shortcuts that are available.

A Collaborative "Platform" for Enacting Exemplary Character Development Research and Character Education Program Evaluations

- The TWCF Global Innovations in Character Development (GICD) Platform: The Jubilee Center, RYTE Institute, and IARYD

- First of its kind international capacity building platform focused on character.

- Grantees from every region of the world, engage in capacity building support from Platform managers and from their peers

- The "secret sauce" of the GICD success: Fully collaborative, respectful (non-colonialist), and humble practitioner ⇔ researcher collaborations.

TWCF GICD Character Platform

❖ The Platform provides “soup to nuts” support to:

- Strengthen conceptual, evaluation, and methodological skills
- Generate opportunities for peer learning and knowledge exchange
- Plan for sustainability and scale
- Support for creation and dissemination of research outputs

❖ Many lessons learned, particularly around the *describing* and *explaining* aims of developmental science (measures and methods, analyses and reporting)

❖ Return on investment is high—with nimble (small) team of experts, can scale impact of character education programs and the network within higher education

❖Excellent model for a similar platform for practitioner - researcher collaboration, educating for character development in higher education.

Thank You!