CONTEMPORARY VIEW OF VALIDITY: HOW SHOULD WE CHANGE OUR WAY TO DEVELOP AND VALIDATE INSTRUMENTS?

An application to the measurement of the emerging adults' financial well-being

Angela Sorgente & Margherita Lanz
Università Cattolica del Sacro Cuore, Milan, Italy

What most of us know about validity:

- Validity is a property of a test: a test is valid if it measures what it claims to measure.
- Different kinds of validity exist (e.g., content, predictive, concurrent, and construct validity).
- Test developers have to evaluate if an instrument is valid or not.

What most of us know about validity:

- Validity is a property of a test: a test is valid if it measures what it claims to measure.
- Different kinds of validity exist (e.g., content, predictive, concurrent, and construct validity).
- Test developers have to evaluate if an instrument is valid or not.

Are these statements true?

No more!

American Educational Research Association., American Psychological Association., National Council on Measurement in Education., & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

1. VALIDITY AS UNITARY CONCEPT

- "This unified concept of validity **integrates considerations** of content, criteria, and consequences into a construct framework for testing rational hypotheses about theoretically relevant relationships, including those of an applied as well as of a scientific nature" (Messick, 1989).

- "The essence of unified validity is that the appropriateness, meaningfulness, and usefulness of score-based inferences are inseparable and that the unifying force behind this integration is the trustworthiness of empirically grounded score interpretation, that is, **construct validity**" (Messick, 1989).

Many kinds of evidence, one kind of validity

- "Construct validity **cannot** generally be expressed in the form of a **single simple coefficient**. The data often permit one to establish upper and lower bounds for the proportion of test variance which can be attributed to the construct. The integration of diverse data into a proper interpretation cannot be an **entirely quantitative process**" (Cronbach and Meehl, 1955).

- "Whether or not an interpretation of a test's properties or relations involves questions of construct validity is to be decided by examining the entire body of evidence offered, together with what is asserted about the test in the context of this evidence" (Cronbach and Meehl, 1955).

The importance of theory

1. Literature review
2. Nomological net (Cronbach and Meehl, 1955)
3. Testable hypotheses: which outcomes? (criterion-related evidence), which dimensions? (score structure evidence), which differences across groups? (known groups evidence), …
4. Collect the different kinds of validity evidence
5. Make a judgment about the construct validity

2. VALIDITY IS NOT A PROPERTY OF THE TEST

- "What is to be validated is not the test or observation device as such but the **inferences** derived from test scores or other" (Messick, 1989).

  the term "inference" would be reserved for conclusions derived from empirical evidence

- "Trying to define validity as a property of tests quickly leads to absurdities since the validity of a test can vary from population to population" (Wiliam, 2014).

- "Measurement validation is an **ongoing process** wherein one provides evidence to support the appropriateness, meaningfulness, and usefulness of the specific inferences made from scores about individuals from a given sample and in a given context" (Zumbo, 2009)

Test developers VS Test users

- "Making validation a joint responsibility of test developer and test user draws attention to the fact that **test users have a responsibility** to determine whether the validity evidence produced by the developer does in fact support the chosen use of the test." (Wiliam, 2014).

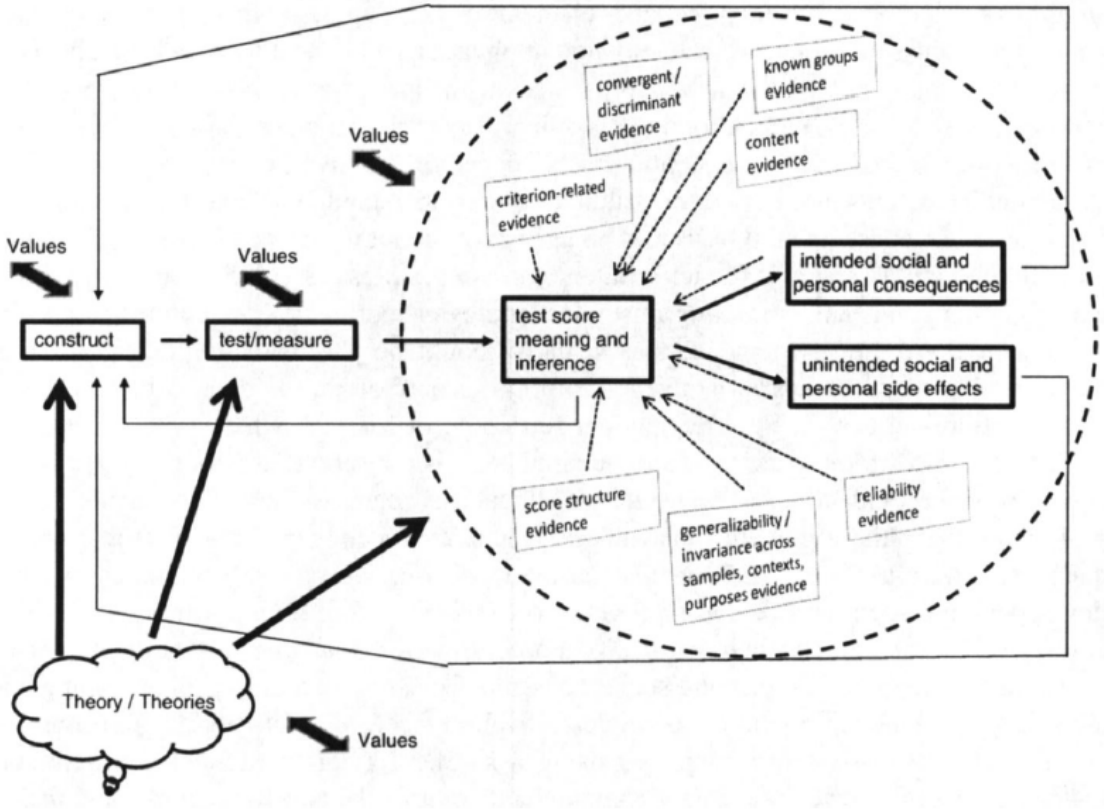
- "Rather than a general requirement regarding the need for the various lines of validity evidence, evidentiary **needs follow from the claims one wishes to make** and the inferences on wishes to draw." (Sackett, 2014).

PRACTICAL IMPLICATIONS

Taking as example

Sorgente, A., & Lanz, M. (2019). The multidimensional subjective financial well-being scale for emerging adults: Development and validation studies. *International Journal of Behavioral Development*, *43*(5) 466–478.

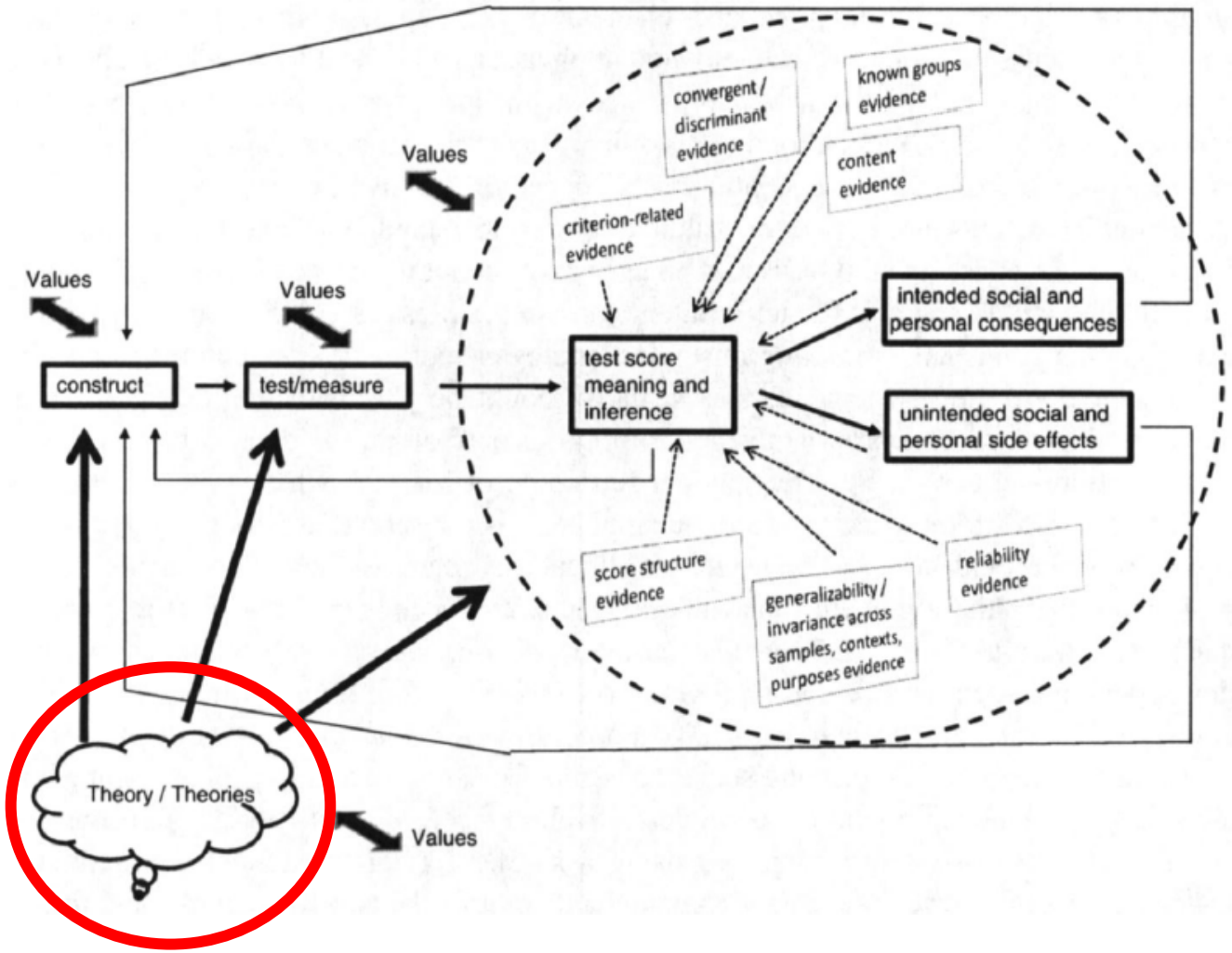
Many steps are necessary



All the kinds of validity evidence have to be collected using the **Structural Equation Models** (Zumbo, 2005).

Hubley, A. M., & Zumbo, B. D. (2011). Validity and the consequences of test interpretation and use, *Social Indicators Research*, 103(2), 219-230.

1. THEORY



The contemporary view of construct validity **is based on a well-articulated theory** and well-planned empirical tests of that theory. It should provide an explanation for the test scores, in the sense of the theory having explanatory power for the observed variation in test scores.

Adolescent Res Rev (2017) 2:255–292
DOI 10.1007/s40894-016-0052-x

CrossMark

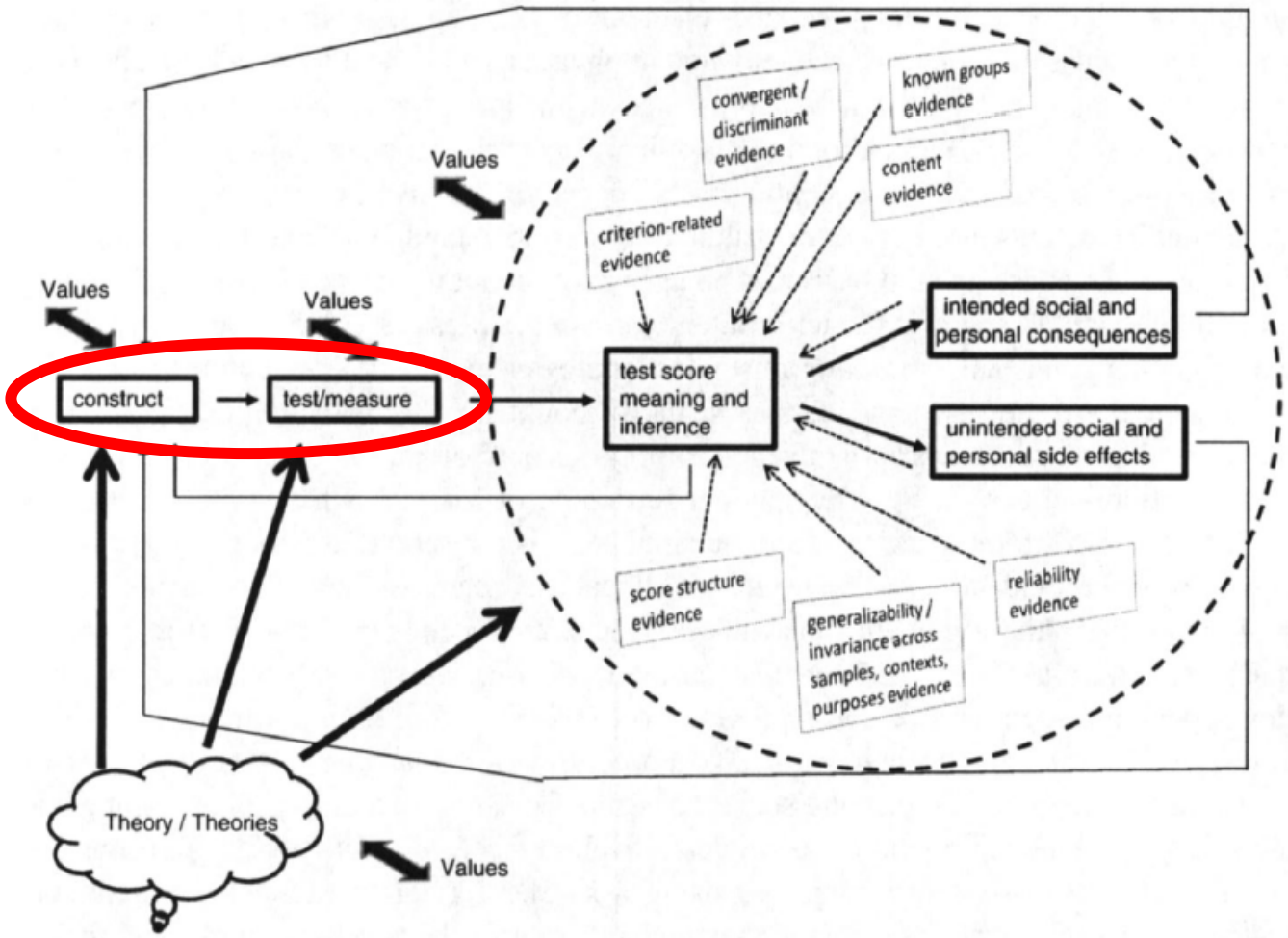
SYSTEMATIC REVIEW

Emerging Adults' Financial Well-being: A Scoping Review

Angela Sorgente[1] · Margherita Lanz[1]

- DEFINITION and COMPONENTS: Financial well-being has both an objective and a subjective side, believing that objective financial well-being consists of individuals' material resources (e.g., income), whereas subjective financial well-being is individuals' perception and evaluation of their own financial condition.
- NOMOLOGICAL NET: predictors, outcomes
- CONTEXT: mainly in US
- MEASUREMENT: Lack of adequate instruments to measure the subjective financial well-being of European *emerging adults*

2. ITEM DEVELOPMENT



WHAT?

Developing items that should measure the construct

HOW?

- Literature search
- Developing items ad hoc in order to saturate the construct definition

Literature search

- 44 papers included in the literature review
- 36 measuring the subjective financial well-being
- 66 items retrieved

Item developed ad hoc

- Structural Interviews with experts (eight bank employees and one full professor of demography; *maximum variation sample*):

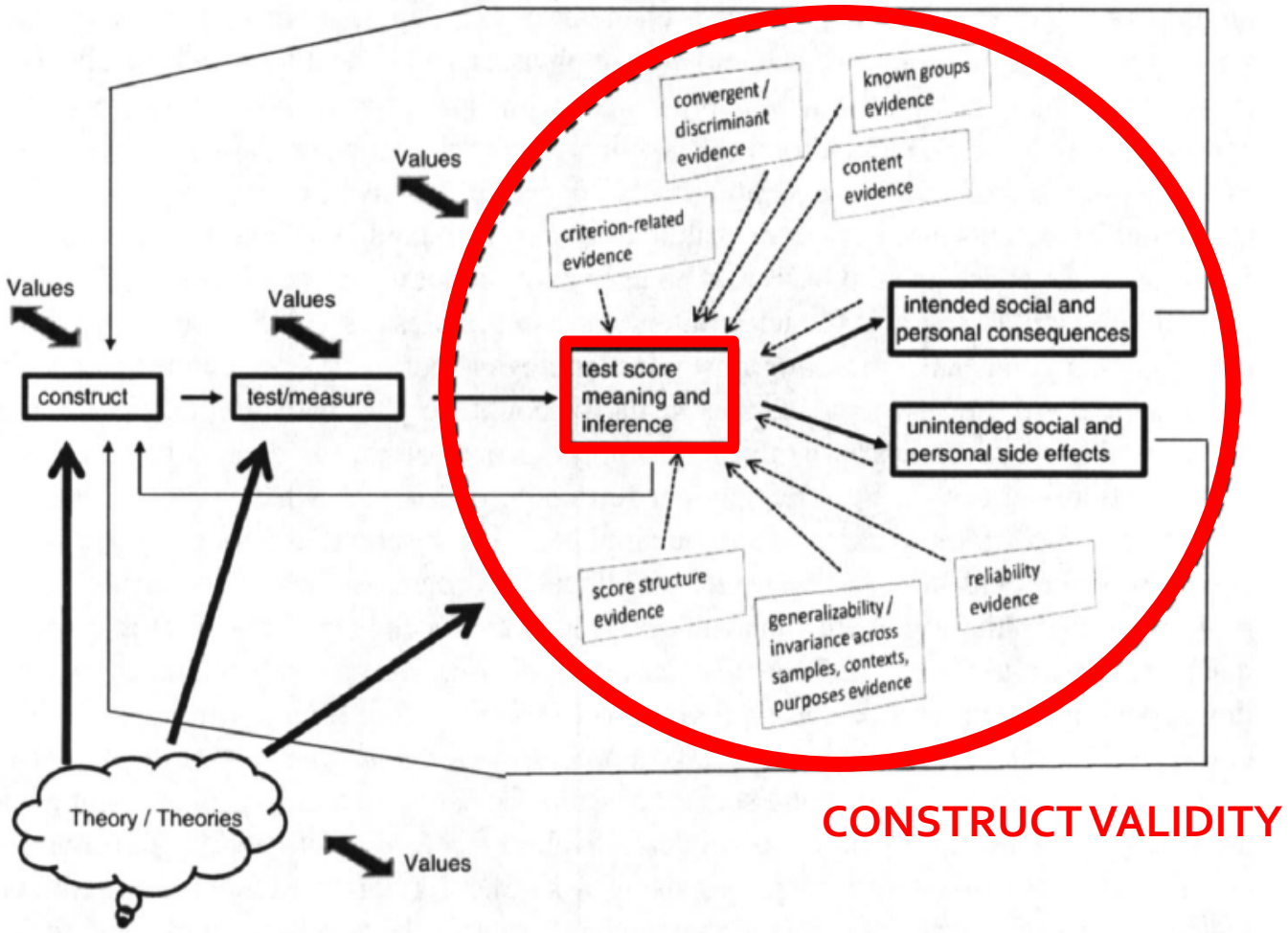
1. Definition of emerging adult client
2. Definition of financial well-being
3. Comparing the country of investigation with other countries

26 out of 66 items not-fitting with the European context

5 items developed ad hoc

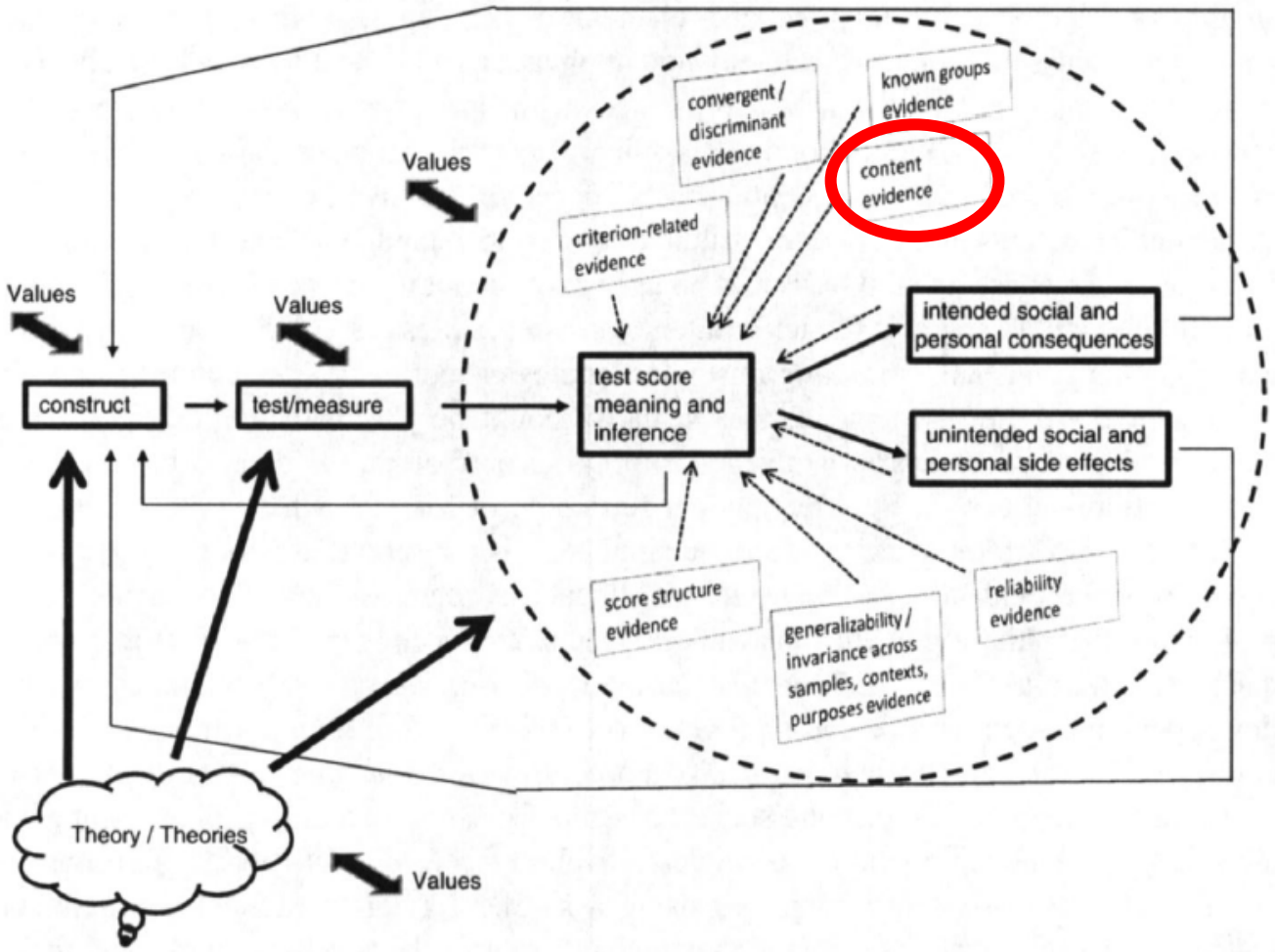
From 66 to 45 items

3. CONSTRUCT VALIDITY



Construct validity is the central and most important feature of validity. Construct validity can be argued for by collecting **different kinds of evidence**.

3.1. Content evidence



WHAT?

Content evidence evaluates whether the content of the instrument accurately represents the concept it was designed to measure.

HOW?

Establishing content evidence of validity may involve a panel of experts reviewing items generated from a literature search of the defined concept being measured.

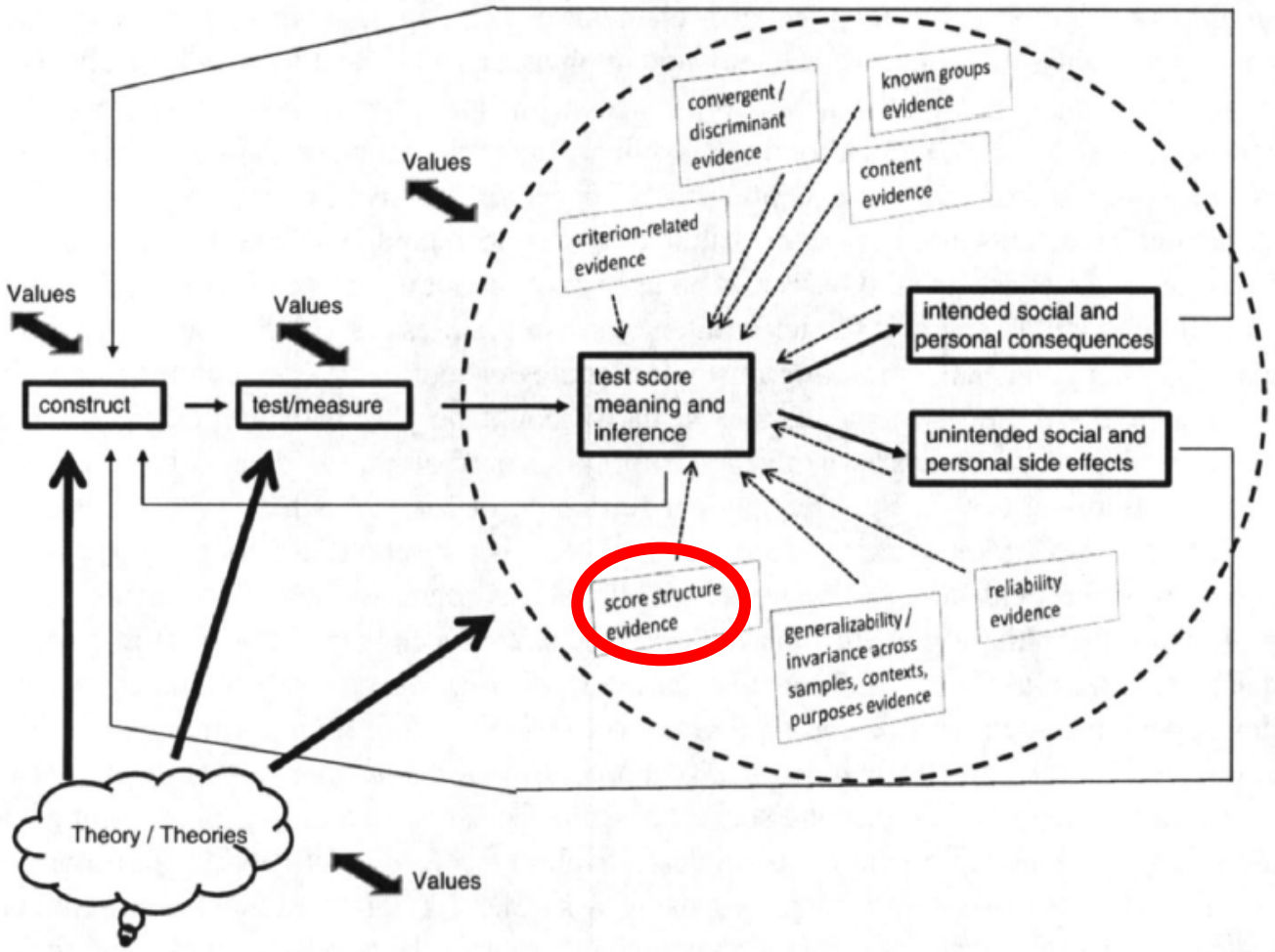
- **Cognitive Interviews with emerging adults** (eight emerging adults; *maximum variation sample*):

Emerging adults had to evaluate, in terms of its comprehensibility and applicability (think-aloud procedures; Ericsson & Simon, 1980):

1. each item
2. the instructions
3. response scale for the instrument.

From 45 to 44 items

3.2. Score structure evidence



WHAT?

Verifying that the test factorial structural is coherent with the theory

HOW?

Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA)

EFA

- 374 emerging adults

Descriptive statistics and Exploratory Factor Analysis (EFA) in order to evaluate each item's:

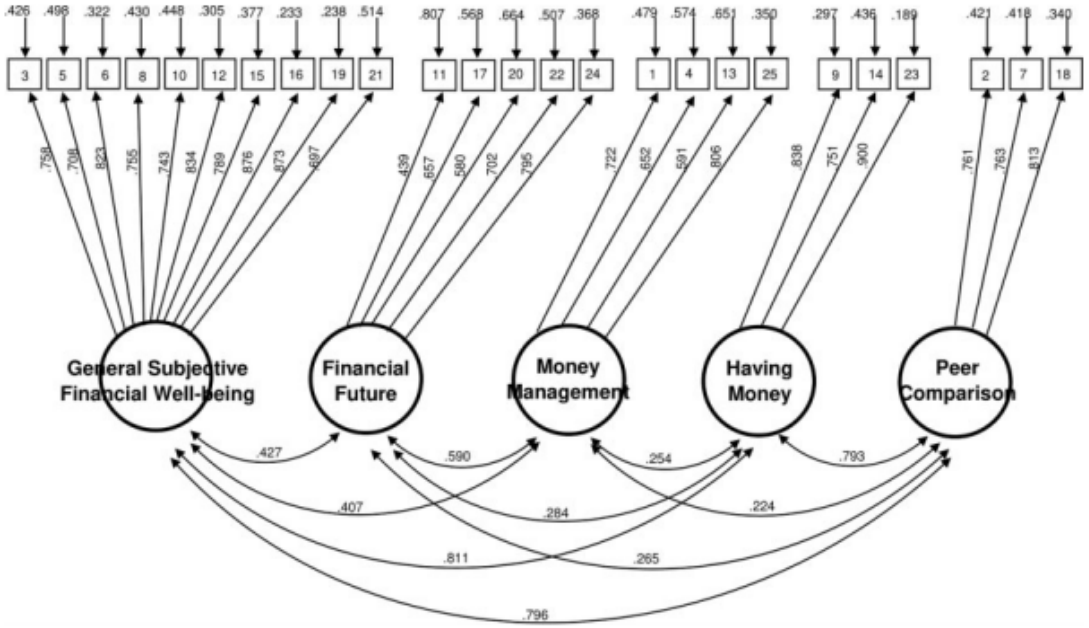
1. response rate
2. normal distribution of answers to that item
3. correlation of the item with other items (communality)
4. whether the item loaded on just one factor
5. and factors' internal consistency

From 45 to 25 items
5 dimensions

CFA

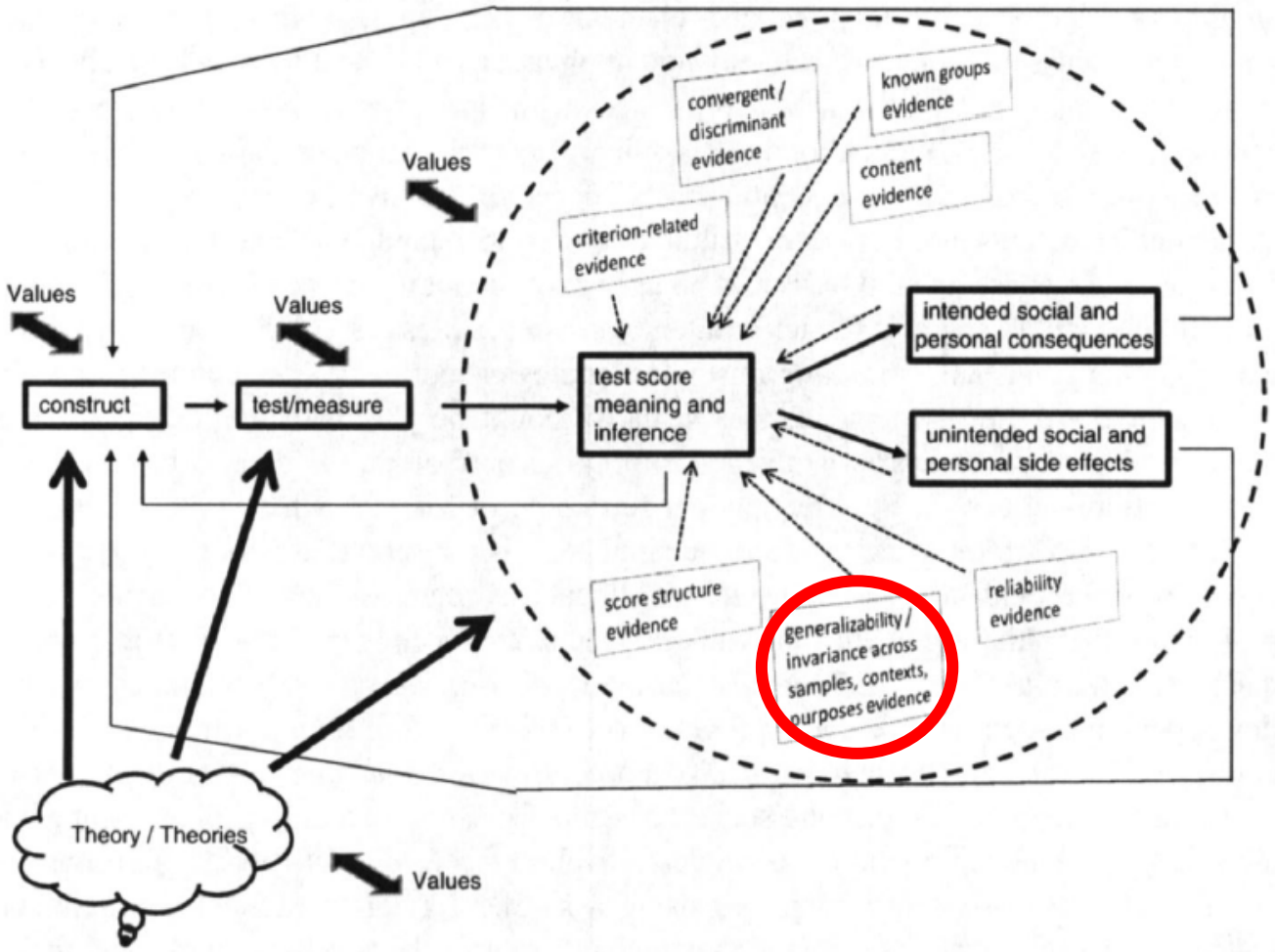
- 516 emerging adults

Confirming EFA results



From 25 to 25 items
5 dimensions

3.3. Generalizability evidence



WHAT?

Investigating uniformities and differences in the test processes and structures over time or across groups and settings—that is, the generalizability of test interpretation and use.

HOW?

Measurement invariance

Gender
123 male VS 393 female

Age
375 aged 18-24 VS 141 aged 25-29

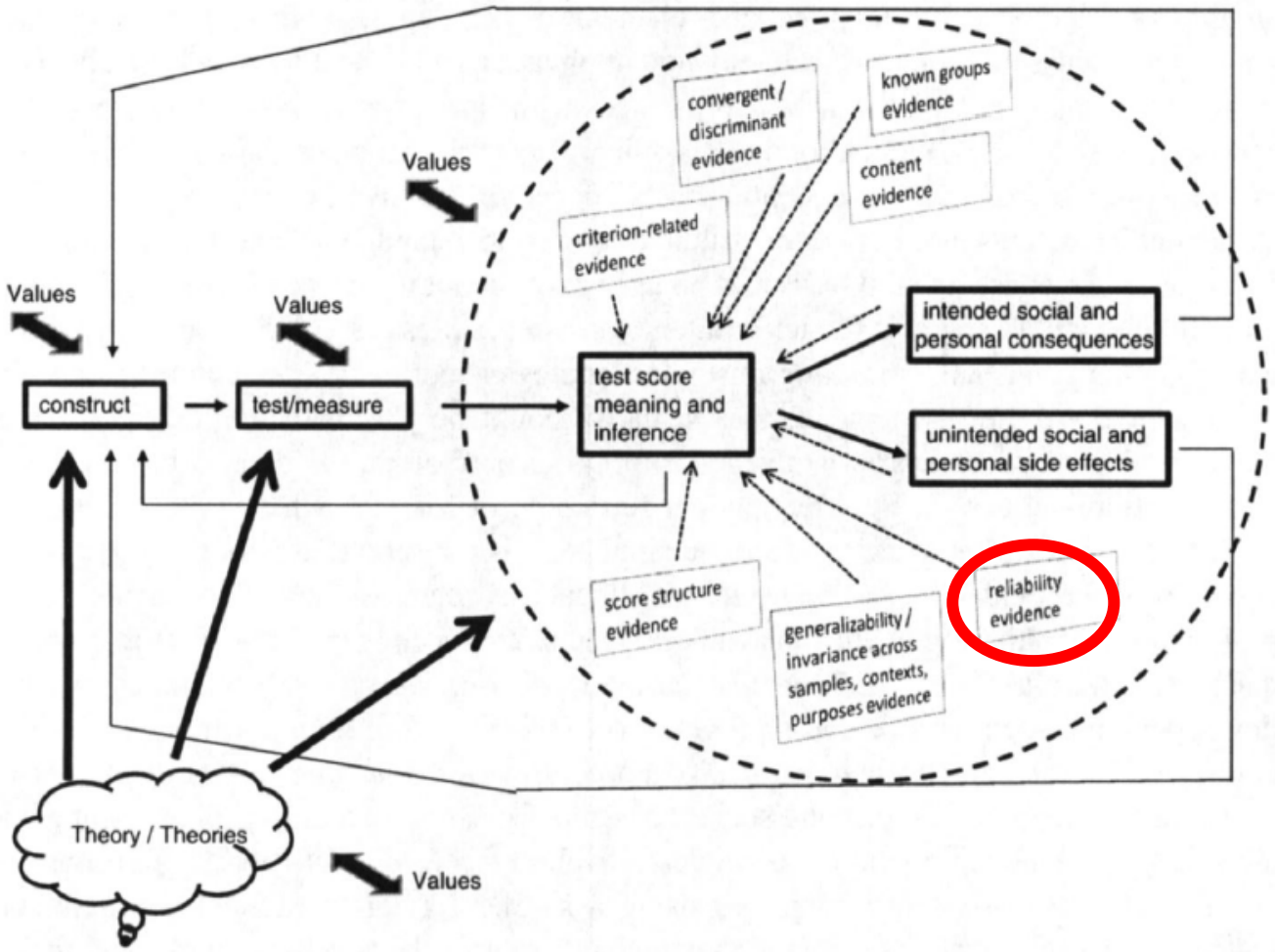
Living Arrangement
231 with VS 276 without parents

Occupational status
149 employed VS 367 unemployed

Table 3. Measurement Invariance for MSFWBS across different socio-demographic conditions.

Invariance	χ2	Df	P	RMSEA (CI)	p	CFI	SRMR	Δ_{CFI}	$\Delta_{\chi 2}$	Δ_{df}	p
Gender (123 male *vs* 393 female)											
Configural	1396.69	535	<.001	.08(.07,.08)	<.001	.892	.07				
Metric	1422.75	550	<.001	.08(.07,.08)	<.001	.891	.07	−.001			
Scalar	1453.96	570	<.001	.08(.07,.08)	<.001	.889	.07	−.002			
Uniqueness	1525.10	595	<.001	.08(.07,.08)	<.001	.883	.08	−.006			
Factors variance	1527.48	600	<.001	.08(.07,.08)	<.001	.884	.08		2.38	5	.794
Factors covariance	1551.36	610	<.001	.08(.07,.08)	<.001	.882	.08		23.88	10	.008
Factors mean	1582.33	615	<.001	.08(.07,.08)	<.001	.879	.09		30.97	5	<.001
Less "Financial Future"	1567.78	614	<.001	.08(.07,.08)	<.001	.881	.09		16.42	4	.002
Age (375 aged 18–24 *vs.* 141 aged 25–29)											
Configural	1379.34	535	<.001	.08(.07,.08)	<.001	.895	.07				
Metric	1406.14	550	<.001	.08(.07,.08)	<.001	.894	.07	−.001			
Scalar	1441.26	570	<.001	.08(.07,.08)	<.001	.892	.07	−.002			
Uniqueness	1468.59	595	<.001	.07(.07,.08)	<.001	.892	.07	0			
Factors variance	1480.91	600	<.001	.07(.07,.08)	<.001	.891	.08		12.32	5	.031
Factors covariance	1487.56	610	<.001	.07(.07,.08)	<.001	.891	.08		6.65	10	.758
Factors mean	1523.50	615	<.001	.08(.07,.08)	<.001	.887	.08		35.94	5	<.001
Less "Peer Comparison"	1502.31	614	<.001	.07(.07,.08)	<.001	.890	.08		14.75	4	.005
Living arrangement (231 living without parents *vs.* 276 living with them)											
Configural	1332.07	535	<.001	.08(.07,.08)	<.001	.899	.07				
Metric	1339.44	550	<.001	.07(.07,.08)	<.001	.900	.07	+.001			
Scalar	1364.69	570	<.001	.07(.07,.08)	<.001	.899	.07	−.001			
Uniqueness	1423.73	595	<.001	.07(.07,.08)	<.001	.895	.08	−.004			
Factors variance	1424.32	600	<.001	.07(.07,.08)	<.001	.895	.08		0.59	5	.988
Factors covariance	1436.00	610	<.001	.07(.07,.08)	<.001	.895	.08		11.68	10	.307
Factors mean	1438.96	615	<.001	.07(.07,.08)	<.001	.895	.08		2.96	5	.706
Occupational status (149 having a job *vs.* 367 not having a job)											
Configural	1396.98	535	<.001	.08(.07,.08)	<.001	.893	.07				
Metric	1419.91	550	<.001	.07(.07,.08)	<.001	.892	.07	−.001			
Scalar	1465.51	570	<.001	.08(.07,.08)	<.001	.889	.07	−.003			
Uniqueness	1494.41	595	<.001	.08(.07,.08)	<.001	.888	.08	−.001			
Factors variance	1496.60	600	<.001	.08(.07,.08)	<.001	.889	.08		2.19	5	.822
Factors covariance	1507.12	610	<.001	.08(.07,.08)	<.001	.889	.08		10.52	10	.396
Factors mean	1531.68	615	<.001	.08(.07,.08)	<.001	.886	.08		24.56	5	<.001
Less "Peer Comparison"	1518.68	614	<.001	.08(.07,.08)	<.001	.888	.08		11.56	4	.021

3.4. Reliability evidence



WHAT?

The ability of an instrument to measure consistently

HOW?

This consistence can be verified at different levels: consistency among items (internal consistency), consistency over time (stability), consistency among judges (inter-rater reliability), and consistency among equivalent versions of the instrument (parallel forms reliability).

British Journal of Psychology (2014), 105, 399–412

The British Psychological Society

www.wileyonlinelibrary.com

From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation

Thomas J. Dunn*, Thom Baguley and Vivienne Brunsden

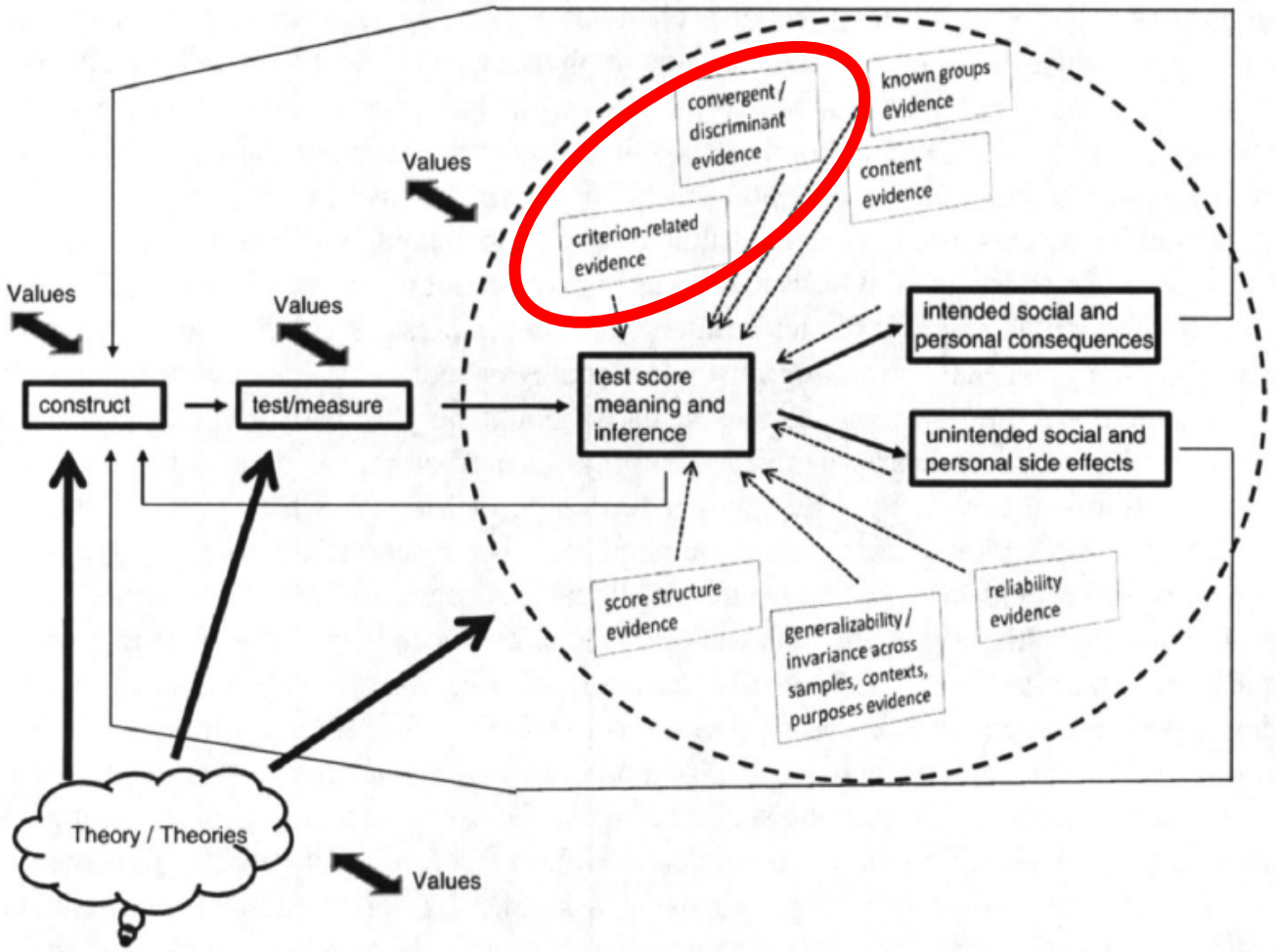
Division of Psychology, Nottingham Trent University, UK

Reliability Evidence

Composite reliability was estimated for each MSFWBS factor. All factors on the scale proved to be highly reliable: General Subjective Financial Well-being (ω=.940), Money Management (ω=.781), Peer Comparison (ω=.823), Having Money (ω=.866), and Financial Future (ω= .771).

3.5. Criterion-related evidence

3.6. Convergent evidence



WHAT?

Criterion-related evidence is based on the degree of empirical correlation between the test scores and criterion scores. Convergent evidence is based on the degree of empirical correlation between two measures that are supposed to be measuring the same construct.

HOW?

Testing the regression/correlation system linking the *latent* variables by Structural Equation Modeling

Psychological Well-being

Physical Well-being

Subjective Well-being

Dimensions of the objective financial well-being

Table 5. Criterion-related evidence for the MSFWBS.

	BIT	SWLS	Physical Well-being
General Subjective Financial Well-being	.327***	.490***	.188***
Money Management	.324***	.359***	.178***
Peer Comparison	.196***	.348***	.092
Having Money	.197***	.323***	.090
Financial Future	.452***	.493***	.182***

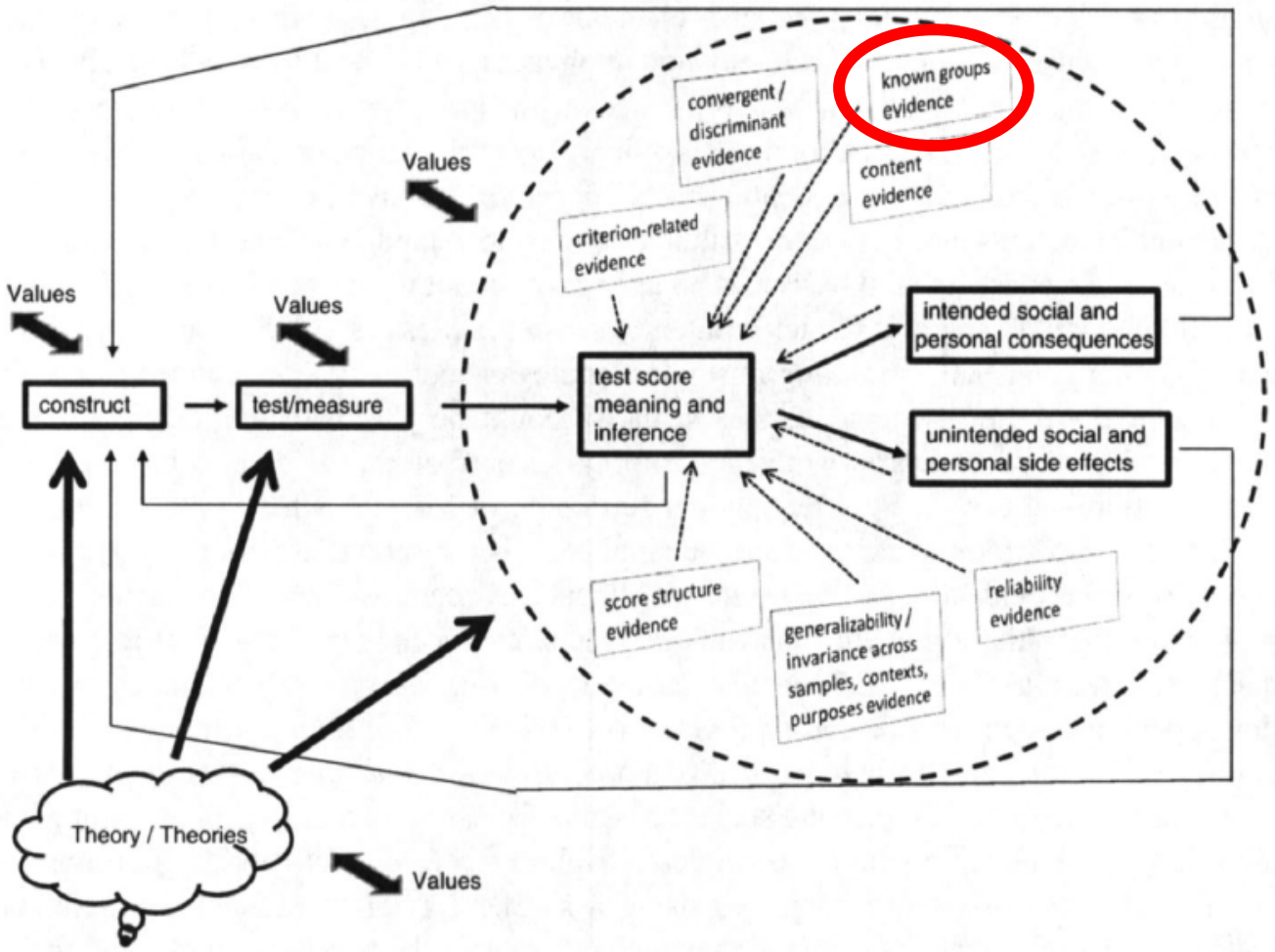
Note. BIT = Brief Inventory of Thriving; SWLS = Satisfaction With Life Scale; ***p<.001

Table 4. Convergent validity evidence for the MSFWBS.

	Personal income	Family of origin income
General Subjective Financial Well-being	.238***	.466***
Money Management	.067	.112*
Peer Comparison	.383***	.512***
Having Money	.285***	.442***
Financial Future	.084	.084

Note. *p<.05; ***p<.001.

3.7. Know group evidence

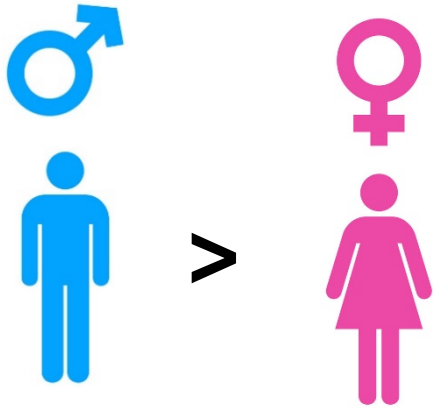


WHAT?

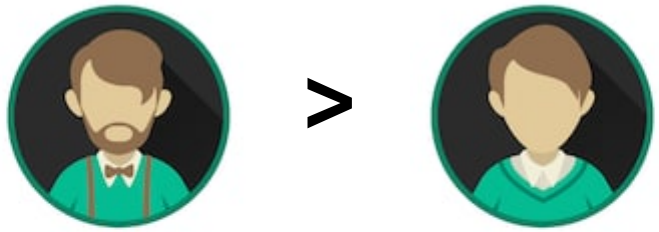
Verifying if the instrument is able to detect difference among groups that are expected according to the theory

HOW?

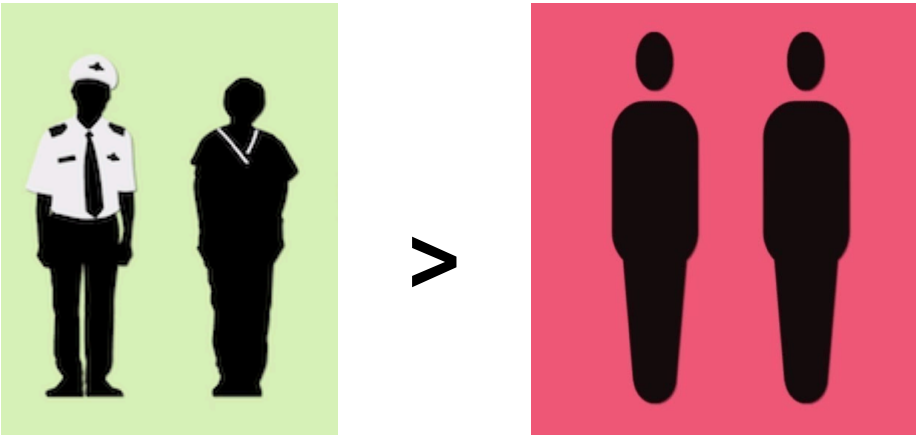
Factorial invariance: checking if the latent mean of factors changes across group



male higher than female



25-29 higher than 18-24



employed higher than unemployed

ijbd

International Journal of Behavioral Development
2019, Vol. 43(5) 466–478

DOI: 10.1177/0165025419851859
journals.sagepub.com/home/jbd
SAGE

The multidimensional subjective financial well-being scale for emerging adults: Development and validation studies

Angela Sorgente and Margherita Lanz

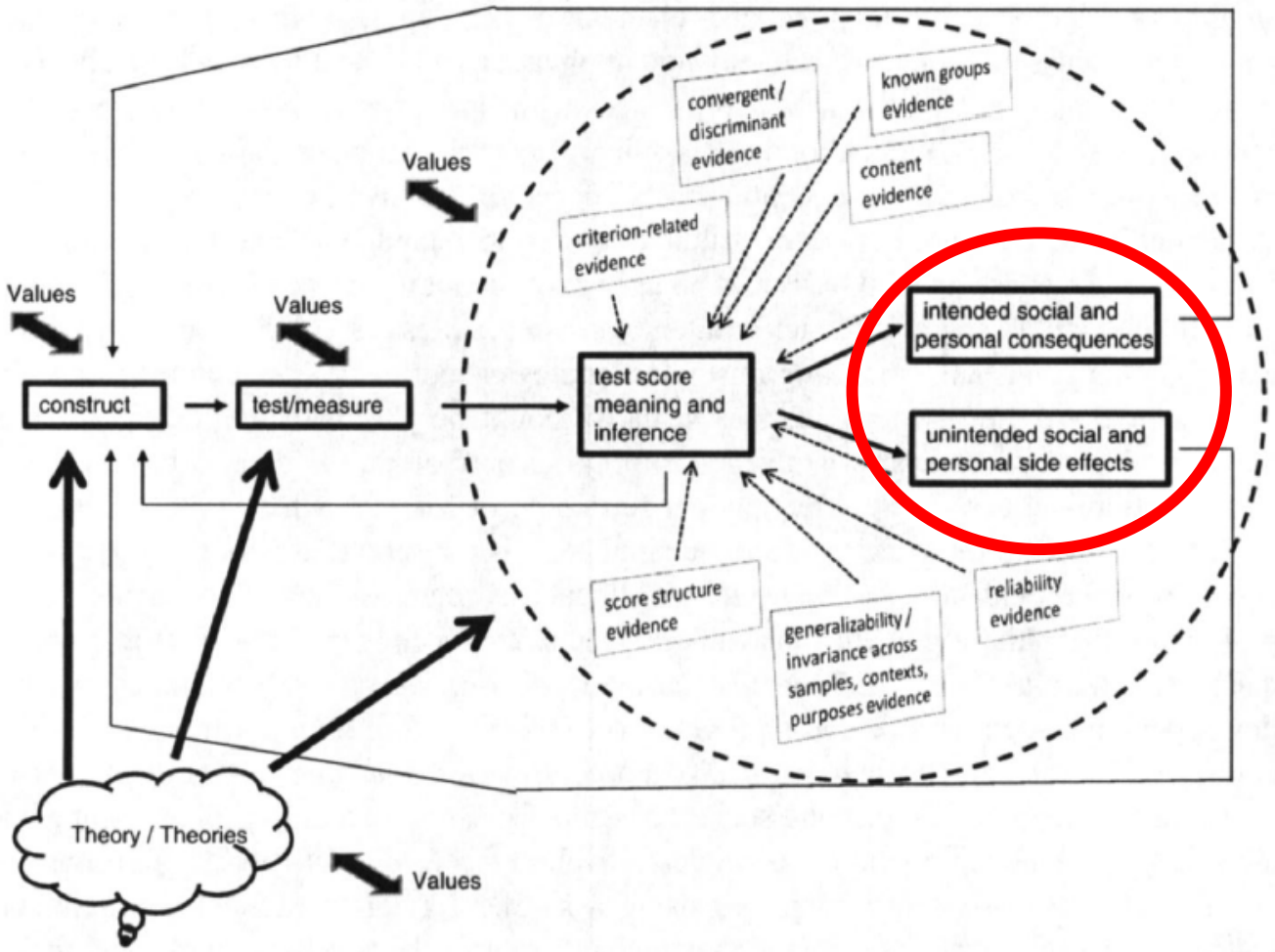
Abstract

After the 2008 economic crisis, the financial condition of youth has become a frequent research topic and the need for an instrument measuring financial constructs relevant for this stage of life is increasing. The current paper consists of four studies aiming to develop and validate an instrument measuring subjective financial well-being in a population of European emerging adults. The first study collected qualitative data – performing interviews with experts and the target population – in order to contextualize the construct. Based on these results, we developed the initial 44-item version of the scale. The second study aimed to test the psychometric characteristics of these items. There were 25 items measuring five different aspects of subjective financial well-being (general subjective financial well-being, money management, peer comparison, having money, financial future) with acceptable psychometric properties. The third study aimed to collect validity evidence about the newly developed scale. Five different kinds of evidence suggested that the scale is a good measure of subjective financial well-being. The last study tested measurement invariance between the Portuguese and Italian versions of the scale. Results suggested that the Multidimensional Subjective Financial Well-being Scale works well in both countries.

Keywords

Emerging adult, financial well-being, scale development, validation, cross-cultural

3.8. Consequential evidence



WHAT?

Intended and unintended social and personal consequences of the use of the instrument (e.g., for a test measuring depression in old people → finding very small or very large numbers of depressed elderly).

HOW?

Verifying if test use produce negative, harmful consequences. Only if linked to a deficiency in the test do consequences affect conclusions about validity. Otherwise, consequences can only affect decisions about test use.

CONCLUDING REMARKS

ABOUT A GAP

There is clearly a gap between the recommended guidelines and validation practice (Menold et al., 2018).

What we have found in "validation" papers published in Emerging Adulthood Journal:

Contemporary view of validity

- Starting from qualitative study
- CFA
- Measurement invariance

Traditional view of validity

- Cronbach's alpha
- Correlation between observed variables
- T-test

HOW TO REDUCE THIS GAP?

- **As scholar**: remember that
  - Validity is unified (construct validity)
  - Validity is a property of the inference made from a test score

- **As test developer**: Collect your test's validity evidence by Structural Equation Modeling

- **As test user**: Check if the validity evidence produced by the test developer supports the claims you wish to make. If instead the sample is different:
  - Perform CFA to be sure that the score structure is the same
  - Test measurement invariance before to compare groups
  - Test the reliability of your factors using the composite reliability

THANK YOU

angela.sorgente@unicatt.it