The Utility and Feasibility of Metric Calibration for Basic Psychological Research

Etienne LeBel The University of Western Ontario



Inspirational Quotations



JOHN TUKEY

"...being so disinterested in our variables that we do not care about their units can hardly be desirable" (Tukey, 1969, p. 89).



JACOB COHEN

"...psychologists have to start respecting the units they work with, or develop measurement units they can respect enough so that researchers can agree to use them" (Cohen, 1994, p. 1001).



Over-arching Goal

 Both useful and feasible to calibrate the metric of instruments in basic psychological research



<u>Outline</u>

- Definitions and basic concepts
- Metric calibration strategies
- Past metric calibration research
- Utility of Metric Calibration
- Feasibility: 3 Empirical demonstration studies
- Limitations and Future Directions

Definitions and Basic Concepts

 Metric: unit of measurement used to quantify the amount of something

- E.g., Celsius metric (°C)
 - Fridge range = -10 to +50 °C
 - Freezer range = -50 to +70 °C



Definitions and Basic Concepts

• Metric: unit of measurement used to quantify the amount of something

- E.g., Beck's Depression Inventory
 - Metric = 0 to 63 (BDI; Beck & Steer, 1987)
- E.g., Self-report Depression Scale
 - Metric = 25 to 100 (SDS; Zung, 1965)

Definitions and Basic Concepts

- Arbitrary metric:
 - Scores not inherently meaningful, other than relative interpretation

 Formally: Unknown where a given score locates an individual on the underlying psychological dimension

(Blanton & Jaccard, 2006a, 2006b)







Main Strategies of Metric Calibration

- Strategy 1
 - Mapping scores to qualitatively distinct behaviors
- Strategy 2
 - Mapping scores to gradation of behaviors

(Blanton & Jaccard, 2006a, 2006b; Sechrest et al., 1996)

- Strategy 3
 - Experimental approach
 - Manipulate construct to extreme levels





Dimension

Ideal Characteristics of Behavioral Reference Points

- Theoretically-relevant
- Interpretationally clear (e.g., 1 or 0; hrs/day)
- Objective
- Unambiguous construct-wise

• Also, theoretically-configured context

Past Metric Calibration Research

- Specific areas of applied psychology:
 - Clinical psychology

(Kazdin, 1999, 2001; Harman et al., 2001; Sechrest et al., 1996)

Sport psychology

(Andersen, McCullagh, & Wilson, 2007)

• Forensic psychology

(Pirelli et al., 2011; Hanson, 2009; Hanson et al., in press)

• Arbitrary metrics in psychology (Blanton & Jaccard, 2006a, 2006b)

Contraction of the second seco

Utility of Metric Calibration

- 1. Help in the interpretation of data
 - a. Enhance interpretability of statistical effects
 - b. Facilitate extraction of more information from data patterns
 - c. Help overcome limitations of NHST
- 2. Facilitate construct validity research
 - a. Help shed brighter light on psychological constructs
 - b. Help with conceptual challenges (e.g., construct definition)
 - c. Benchmark for detecting problems/improving measures

Utility of Metric Calibration

- 3. Contribute to theoretical development
 - a. Facilitate theoretical debates involving absolute claims
 - Allow more precise theorizing via enhanced scientific language
 - c. Preliminary platform for quantitative testing of theories (Meehl, 1978)
- 4. Facilitate general accumulation of knowledge
 - a. Calibration findings valuable information in their own right
 - Guiding framework for cataloguing magnitude of psychological effects
 - c. Facilitate phenomenon-based research (Rozin, 2001)

Feasibility of Metric Calibration

• Empirical demonstration studies

• **Study 1:** Need for cognition (NFC), task persistence (TP), conscientiousness

• Study 2: Self-enhancement

• Study 3: Risk-taking

Study 1: NFC and TP

- Participants
 - 94 UWO introductory psychology undergraduates
 - 69 females, 25 males (age = 18.5, *SD* = 2.2)
- Procedure & Materials
 - Need for cognition measure
 - Task persistence measure
 - Word association decision task
 - Anagram Persistence task
 - Demographics & Debriefing questions

- Need for cognition (NFC)
 - Tendency to engage in cognitively effortful activities and enjoy thinking in its own right (Cacioppo & Petty, 1982)
 - 18-item scale (Cacioppo, Petty, & Kao, 1984)
 - E.g. item: "I find satisfaction in deliberating hard for long hours."
 - E.g. item: "Thinking is not my idea of fun" (R)

1= Extremely	2 = Somewhat	3 = Uncertain	4 = Somewhat	5 = Extremely
Uncharacteristic	Uncharacteristic		Characteristic	Characteristic

- NFC behavioral reference point
 - Cognitively effortful (vs. simpler) Remotes Association Task (RAT) (Mednick & Mednick, 1967)

Task 1: FRIES KISS TOAST Fourth word: Answer: FRENCH	
Task 2: BOARD MAGIC DEATH Fourth word: Answer: BLACK	

- Task persistence
 - Tendency to persist in an effortful behavior or frustrationinducing activity (Steinberg et al., 2007)
 - 2-item self-report measure (Steinberg et al., 2007)
 - Item 1: "I will keep trying the same thing over again even when I have not had success the first time"
 - Item 2: "I will often continue to work on something, even after other people have given up."

1= Very untrue,	2 = Somewhat	3 = Somewhat	4 = Very true,
not at all like	untrue or not	true or like	very much
me	like me	me	like me

- Task persistence behavioral reference point
 - Anagram persistence task

(Brandon et al., 2003; Quinn et al., 1996)

Anagram #2		
KLYXI		
Press the ENTER key to submit answe	<i>81.</i>	
	sk	IP



Wald's χ^2 = 9.71, *B* = 1.20, *odds ratio* (*OR*) = 3.33, *p* < .002

Underlying Dimension

Study 1: Results: Task Persistence



Study 1: Discussion

- Enhance MMR analyses
 - Re-analysis of O'Hara et al. (2009)



Study 2 Demonstration

• Self-enhancement measures

- Background context
 - Pan-cultural self-enhancement debate

(Sedikides et al., 2003; Heine, 2005)

<u>Study 2</u>

- Participants
 - 97 UWO introductory psychology undergraduates
 - 50 females, 47 males (age = 18.9, *SD* = 1.3)
- Procedure & Materials
 - 2 self-enhancement measures
 - Filler task (RAT)
 - Over-claiming technique
 - Balanced Inventory of Desirable Responding
 - Demographics & Debriefing questions

- Self-enhancement
 - Tendency to view characteristics of oneself in an overly positive manner (Hogan & Nicholson, 1988)
 - Better-than-average judgments

(Alicke et al., 1995; Gaertner et al., 2008)

NEGATIVE:

disobedient

lsnobbish

lazv

gullible

 Rate extent to which each listed trait describes yourself relative to the average Western student of your own age and gender
POSITIVE: dependable intelligent considerate observant polite

Ur y	our own age and	a gender	polite	disrespectful
1 = Much worse than the average university student of my age and gender	4 = As well as the average university student of my age and gender	7 = Much better than the average university student of my age and gender	respectful cooperative reliable friendly creative	mean unforgiving vain uncivil unpleasant

- Self-enhancement behavioral reference point
 - Over-claiming technique variant (OCT; Paulhus et al., 2003)
 - 150 items (10 categories of 15 items)
 - 3 non-existent items (foils) per category; 30 foils total
 - Behavioral index: # of foils claimed as familiar

PLEASE INDICATE FOR EACH ITEM WHETHER YOU ARE FAMILIAR WITH THE ITEM OR NOT, BY CLICKING THE APPROPRIATE RESPONSE OPTION:

0 = Never heard of it

1 = Familiar with it

Historical Names and Events

1. Napoleon
2. Robespierre
3. El Puente*
4. My Lai
5. The Lusitania
6. Ronald Reagan
7. Prince Lorenzo*
8. The Luddites
9. Neville Chamberlain
10. Vichy Government
11. Queen Shattuck*
12. Bay of Pigs
13. Torquemada
14. Wounded Knee
15 Clara Barton

Study 2: Results



Study 3 Demonstration

• Risk-taking measures

- Demonstrate metric calibration for:
 - Measures capturing state-like constructs
 - Behavioral measures

<u>Study 3</u>

- Participants
 - 99 individuals from UWO campus
 - Compensated \$5 + earnings in BART task
 - 39 females, 58 males, 2 non-specified (age = 24.5, *SD* = 5.5)
- Procedure & Materials
 - Balloon Analogue Risk Task (BART)
 - Columbia Card Task (CCT)
 - Risky gambles Lottery task
 - Two self-report risk-taking measures
 - Demographics & Debriefing questions

- Risk-taking
 - Behavior involving possibility of gains but with potential negative consequences (Ben-zur & Zeidner, 2009; Lejuez et al., 2002)
- Balloon Analogue Risk Task (BART)

(Lejuez et al., 2002)

- Ps inflate 30 simulated balloons onscreen
 - Each balloon pump worth 1 cent
 - If balloon explodes, money is lost for that trial
 - Scoring: mean # of pumps (non-exploding trials)



• Columbia Card Task (CCT) – hot version

(Figner et al., 2009)

- Ps sequentially turn over cards in 4 x 8 array
- Accumulate as many points as possible
- Can continue unless loss card turned



- Behavioral reference points
 - Risky gambles in lottery risk task (Hsee & Weber, 1999)

Lottery	Option A	Option B
1	\$6 for certain	Flip a coin. Receive \$10 if heads, receive \$0 if tails.
2	\$2 for certain	Flip a coin. Receive \$10 if heads, receive \$0 if tails.
3	\$8 for certain	Flip a coin. Receive \$10 if heads, receive \$0 if tails.
4	\$5 for certain	Flip a coin. Receive \$10 if heads, receive \$0 if tails.
5	\$4 for certain	Flip a coin. Receive \$10 if heads, receive \$0 if tails.

- If Option B selected, experimenter would actually flip a coin
- Risky gambles on lotteries with larger sure bets reflective of higher risk-taking reference point



Wald's χ^2 = 4.85, *B* = .03, odds ratio (OR) = 1.03, *p* < .03



\$4 safe bet: Wald's χ^2 = 3.24, *B* = .08, odds ratio (OR) = 1.08, *p* < .07 \$6 safe bet: Wald's χ^2 = 5.78, B = .30, odds ratio (OR) = 1.35, p < .02

Study 3: Results: CCT

Study 3: Discussion

• BART & CCT calibrated to common \$4 reference point

- Implication:
 - Enhanced interpretation of data patterns
 - Proposed benefit 1. b) extraction of more information



Underlying Dimension

Limitations & Caveats

• Small sample sizes

• Consensus re: reference points

Future Directions

- Richer behavioral reference points
 - E.g., EAR (Mehl et al., 2002)
 - E.g., Eye-tracking

- Experimental approach
 - Capture behavioral manifestations beyond naturally-occurring levels
- Item Response Theory approach (Lord, 1980)
 - Model distinct and ordered behavioral reference points

END

- Thanks to all who have helped:
 - Conceptual:
 - Bertram, Kurt, Chris, Paul, Yang
 - Data collection:
 - Scott Leith
 - Assigning Cohen (1994):
 - Lorne

