World Health Organization

Introduction to Rasch Measurement

Technical Training WHO

Carolina Fellinghauer, PhD, Dpl.Stat.
fellinghauerc@who.int

What is measurement?

Main ideas of measurement

What are the main ideas of measurement?

What is measurement?

Main ideas of measurement

What are the main ideas of measurement?

– **Objects** have **properties** that can be thought in terms of more or less, larger or smaller, stronger or weaker – example?

What is measurement?

Main ideas of measurement

What are the main ideas of measurement?

- **Objects** have **properties** that can be thought in terms of more or less, larger or smaller, stronger or weaker – example?
- This **property** can be measured through its **manifestation** – example?

What is measurement?

Main ideas of measurement

What are the main ideas of measurement?

- **Objects** have **properties** that can be thought in terms of more or less, larger or smaller, stronger or weaker – example?
- This **property** can be measured through its **manifestation** – example?
- This **manifestation** can be mapped onto a **scale** – example?

What is measurement?

Main ideas of measurement

What are the main ideas of measurement?

- **Objects** have **properties** that can be thought in terms of more or less, larger or smaller, stronger or weaker – example?
- This **property** can be measured through its **manifestation** – example?
- This **manifestation** can be mapped onto a **scale** – example?
- **Measurement** can have some **error** involved, and may not be perfectly precise – example?

Main ideas of measurement

Scale types with their properties according to Stanley Smith Stevens					
		Nominal scale	Ordinal scale	Interval scale	Ratio scale
Logical/ math operations	× ÷	✗	✗	✗	✓
	+ −	✗	✗	✓	✓
	< >	✗	✓	✓	✓
	= ≠	✓	✓	✓	✓

Main ideas of measurement

Scale examples

What are some examples of each type of scale?

Main ideas of measurement

Scale examples

What are some examples of each type of scale?

Scale examples

What are some examples of each type of scale?

Main ideas of measurement

Scale examples

What are some examples of each type of scale?

Scale examples

What are some examples of each type of scale?

– Nominal scale: 0=Apples, 1=Pears, 2=Bananas

Main ideas of measurement

Scale examples

What are some examples of each type of scale?

Scale examples

What are some examples of each type of scale?

– Nominal scale: 0=Apples, 1=Pears, 2=Bananas
– Ordinal scale: 0=Never, 1=Sometimes, 2=Often

Main ideas of measurement

Scale examples

What are some examples of each type of scale?

Scale examples

What are some examples of each type of scale?

– Nominal scale: 0=Apples, 1=Pears, 2=Bananas
– Ordinal scale: 0=Never, 1=Sometimes, 2=Often
– Interval scale: 0=0°C, 1=1°C, 2=2°C

Main ideas of measurement

Scale examples

What are some examples of each type of scale?

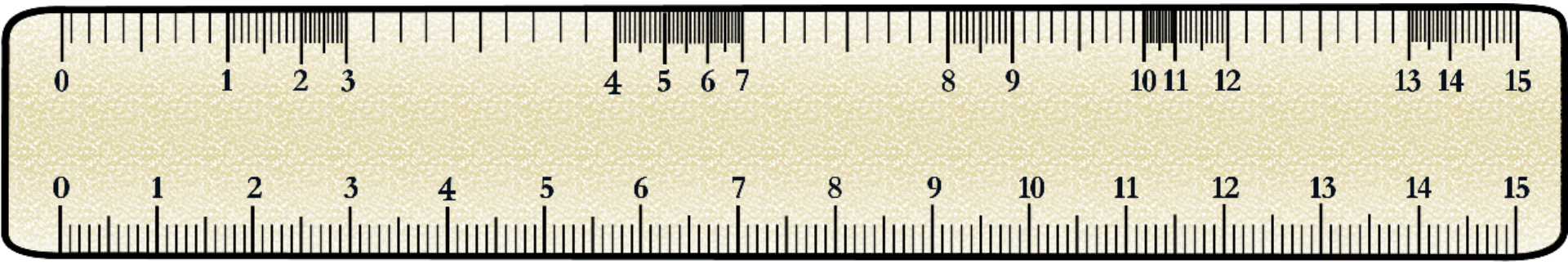
Scale examples

What are some examples of each type of scale?

- Nominal scale: 0=Apples, 1=Pears, 2=Bananas
- Ordinal scale: 0=Never, 1=Sometimes, 2=Often
- Interval scale: 0=0°C, 1=1°C, 2=2°C
- Ratio scale: 0=0cm, 1=1cm, 2=2cm

Our goal

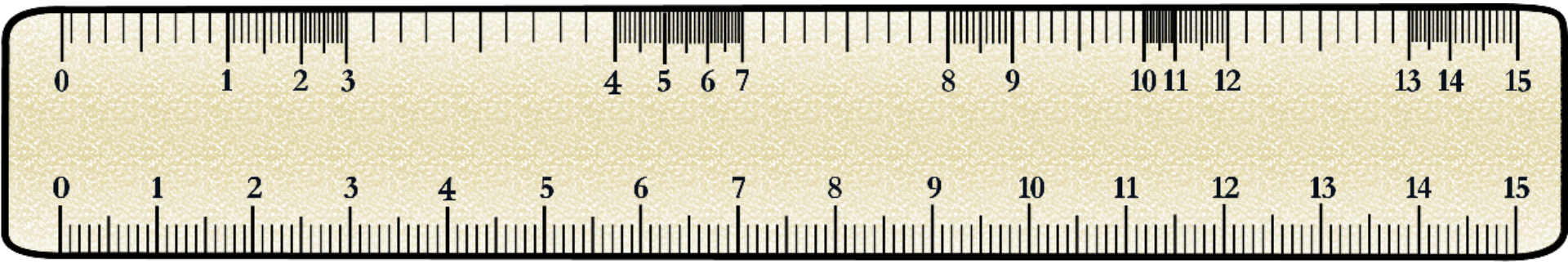
Going from ordinal…



…to interval

Our goal

Going from ordinal…

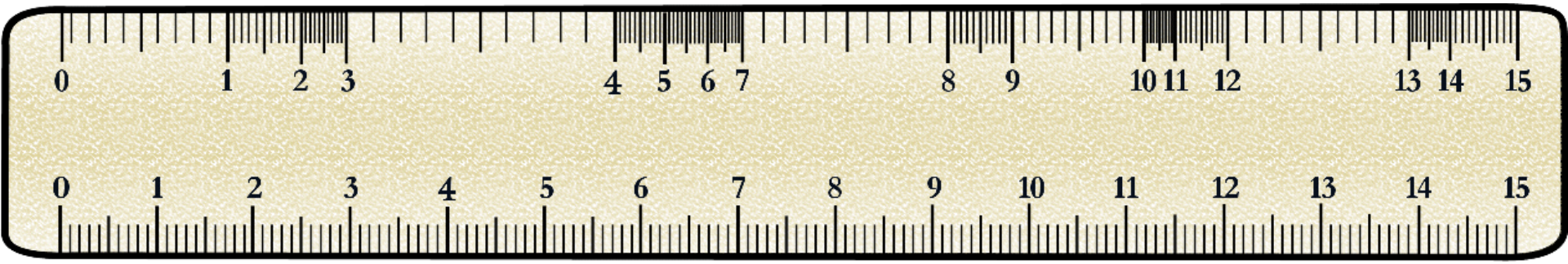


…to interval

How?

Our goal

Going from ordinal…



…to interval

with psychometrics

Psychometrics

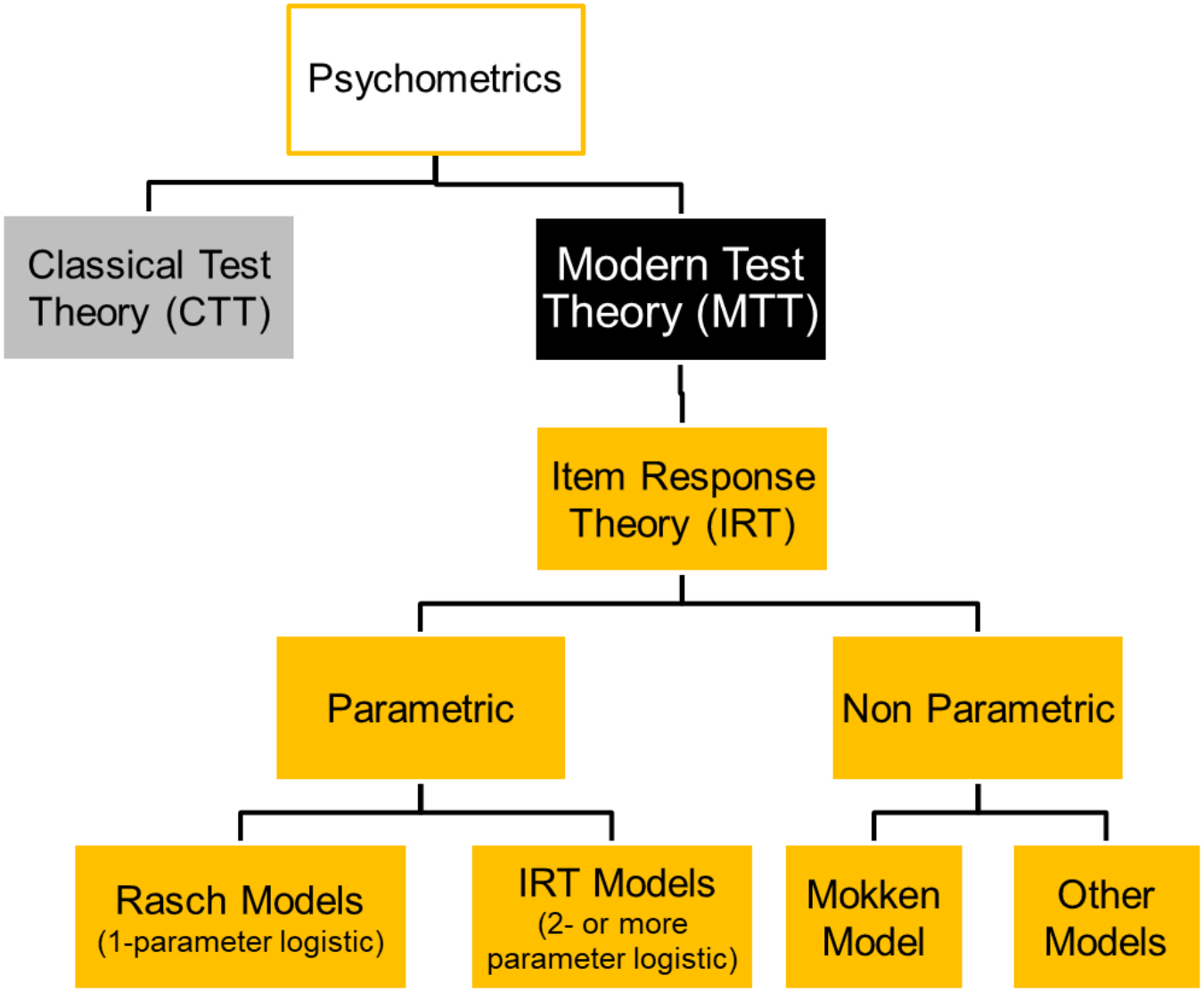
Psychometrics is understood as the field of study concerned with the theory and technique of psychological measurement including the measurement of knowledge, abilities, attitudes, and personality traits. The field is primarily concerned with the study of differences between individuals.

It involves two major research tasks:

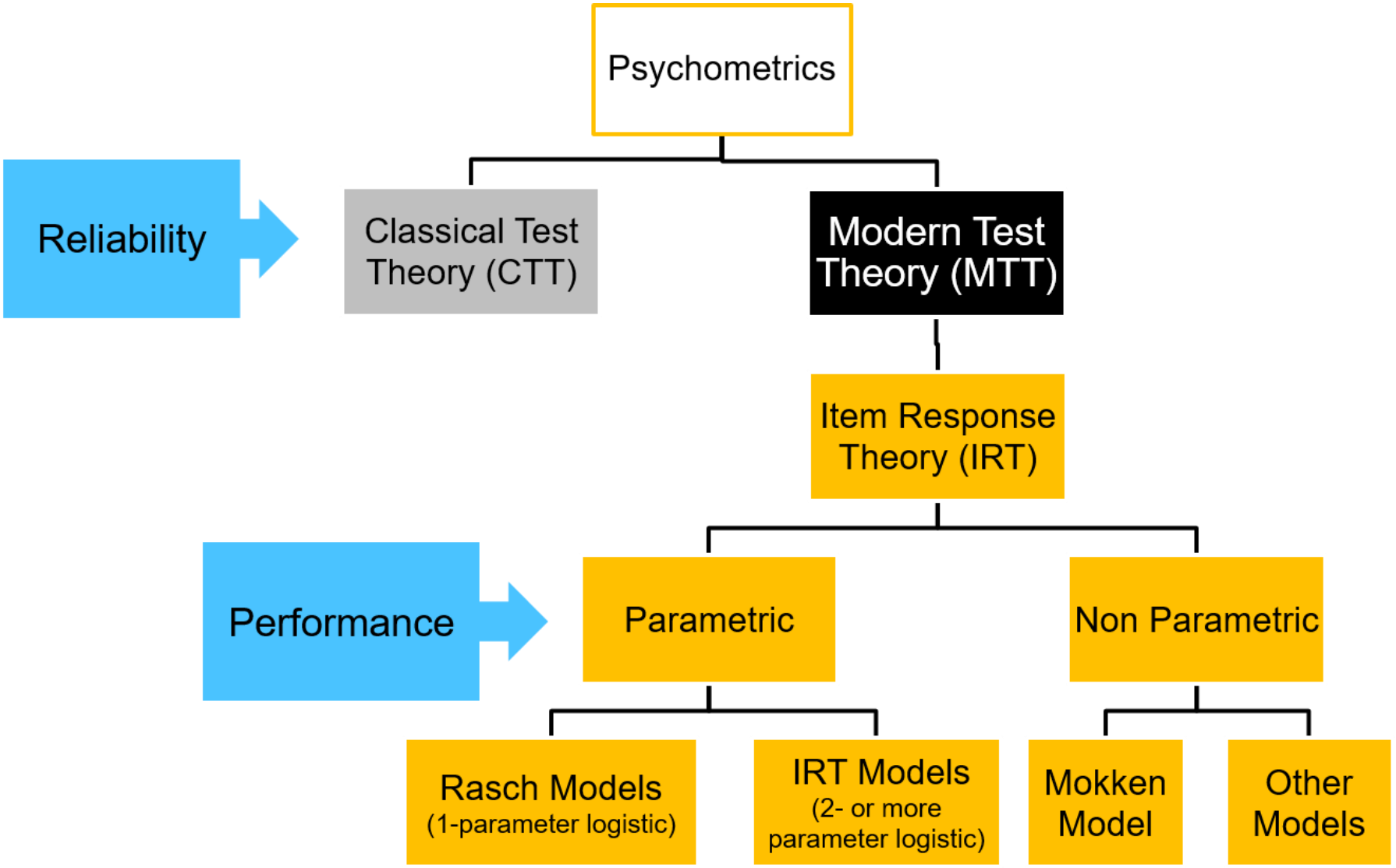
(i) the **construction** of instruments and procedures for measurement; and

(ii) the **development and refinement of theoretical approaches** to measurement

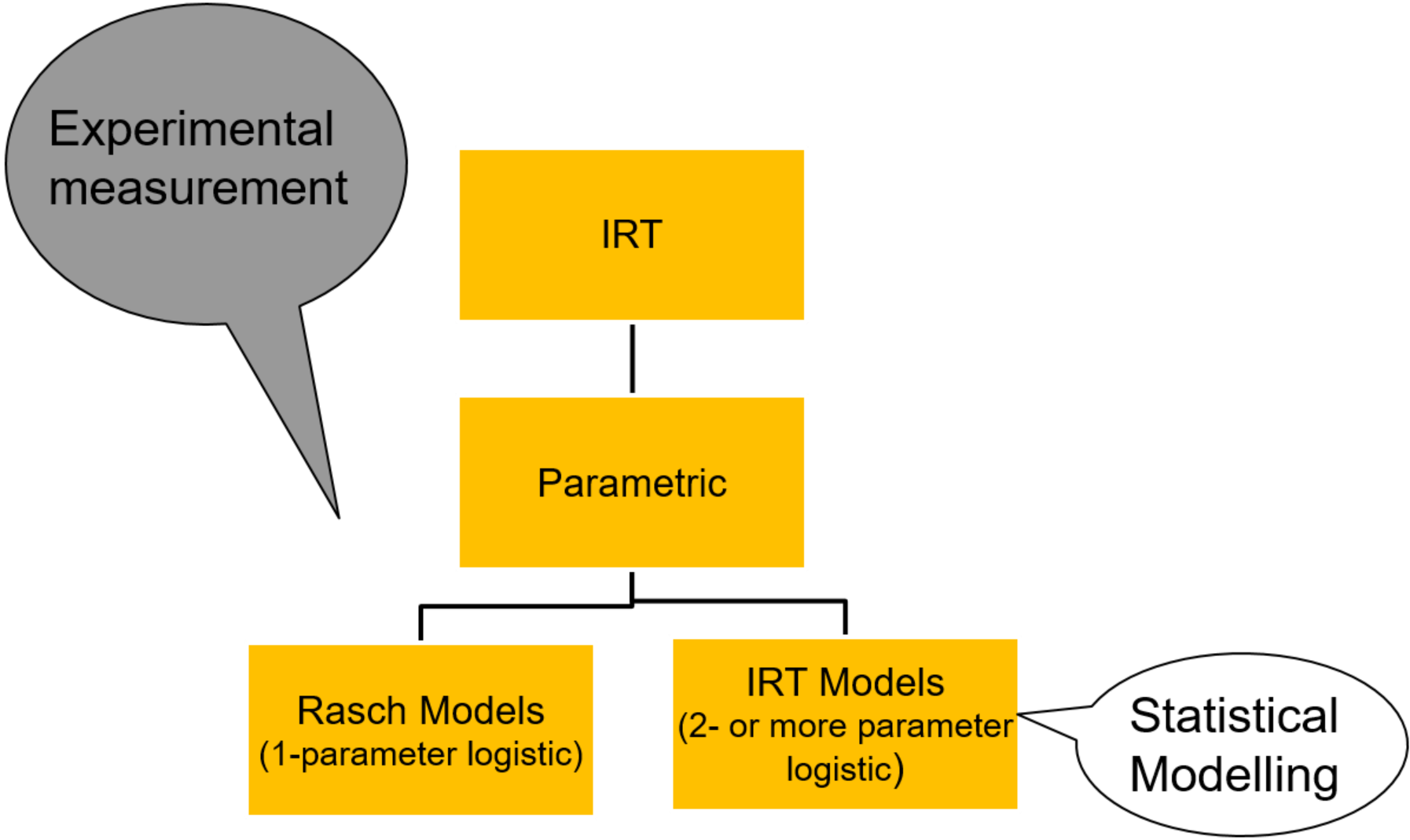
Psychometrics



Psychometrics



Two Competing Paradigms



Rasch vs IRT[1]

The two Paradigms

The Traditional Paradigm

Chosing one model over another because it accounts better for the data. The data is given.

The model has to fit the data = model fitting

The Rasch Paradigm

The data has to fit the model.

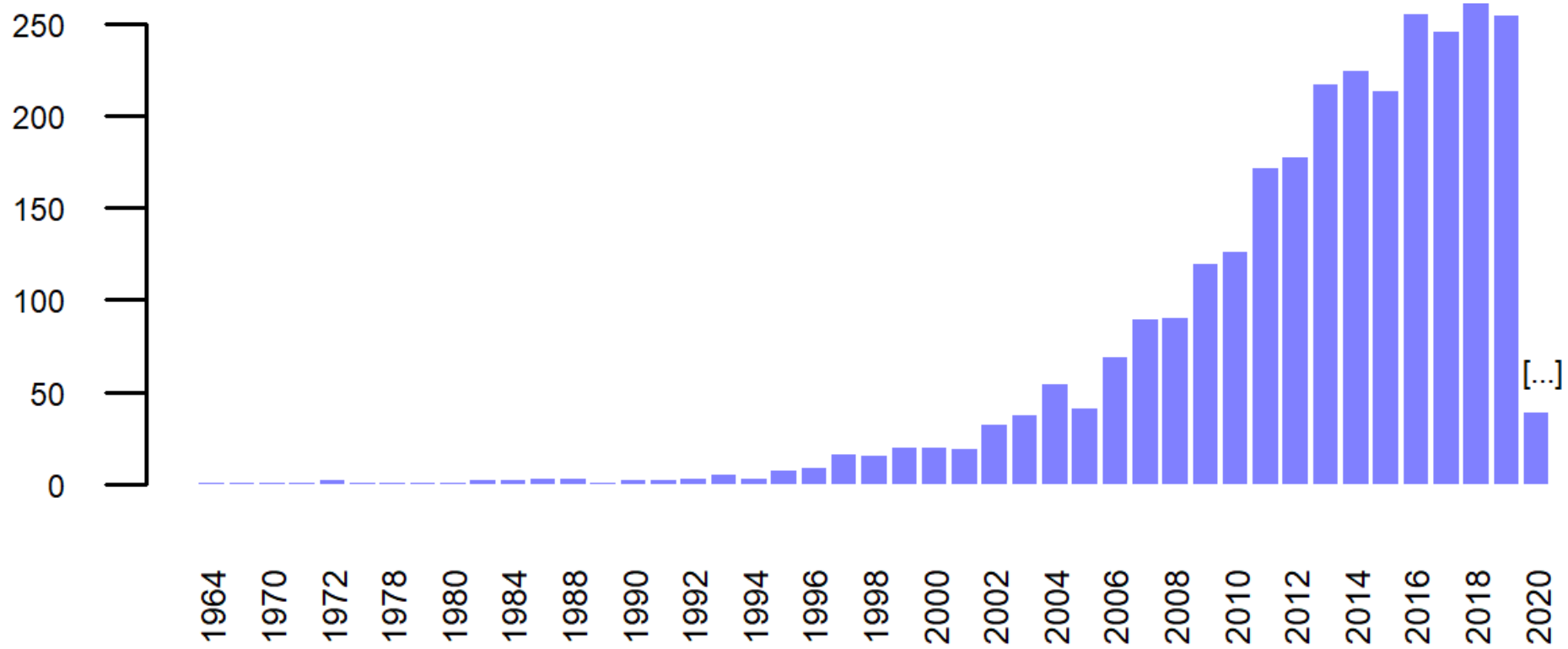
Controversy

IRT is not measurement, it is modelling. Only Rasch analysis stands for objective measurement.

Rasch Analysis: What is that?

- This statistical method is named after the Danish Mathematician George Rasch who developed it[2]
- Statistical method from the field of modern/probabilistic test theory.
- Rasch analysis is mainly used to determine metric properties of questionnaires.
- It is widely used and especially in the fields of educational, psychological, and health assessment.
- Once fit to the Rasch model is established a score is considered interval scaled and usable for measurement and parametric analyis.

Rasch-Publications per Year (N = 2855)



Pubmed search on Feb 2020:
https://pubmed.ncbi.nlm.nih.gov/
(Rasch Analysis) AND (Psychometry OR Test Theory OR Methods)

Before the Rasch model : L.L. Thurstone (1887 - 1955)

Universal characteristics of measurement

- A scale must **transcend** the group measured (1928)
- The measurement of any object or entity describes **only one attribute** of the object measured (1931).
- A unit of measurement is always a process of some kind which **can be repeated** without modification in the different parts of the measurement continuum.
- Thurstone's **Law of Comparative Judgement** makes the link to modern approach for social and psychological measurement.

Before the Rasch model : L.L. Thurstone (1887 - 1955)

Universal characteristics of measurement

- A scale must **transcend** the group measured (1928)
- The measurement of any object or entity describes **only one attribute** of the object measured (1931).
- A unit of measurement is always a process of some kind which **can be repeated** without modification in the different parts of the measurement continuum.
- Thurstone's **Law of Comparative Judgement** makes the link to modern approach for social and psychological measurement.

A person has to order a number of stimuli (with regard to a property). Repeated comparative judgements will be more efficient in sorting it out than absolute ratings on all stimuli at once. He applied this approach to the measurement of psychological values.

Check: https://www.nomoremarking.com/demo2#

Before the Rasch model : N.R. Campbell (1880 - 1949)[3]

- Order alone will not allow a property to be measured.
- The first condition for measurement - the number assigned represent a **transitive and asymmetrical** relationship (R) (e.g. heavier as)

a) Transitive: If A has it to B and B has it to C, then A has it to C

b) Symmetrical: If A has it to B then B has it to A.

Additivity (whole is the sum of the parts) is the hallmark of measurement.

Before the Rasch model : N.R. Campbell (1880 - 1949)[3]

- Order alone will not allow a property to be measured.
- The first condition for measurement - the number assigned represent a **transitive and asymmetrical** relationship (R) (e.g. heavier as)

a) Transitive: If A has it to B and B has it to C, then A has it to C

b) Symmetrical: If A has it to B then B has it to A.

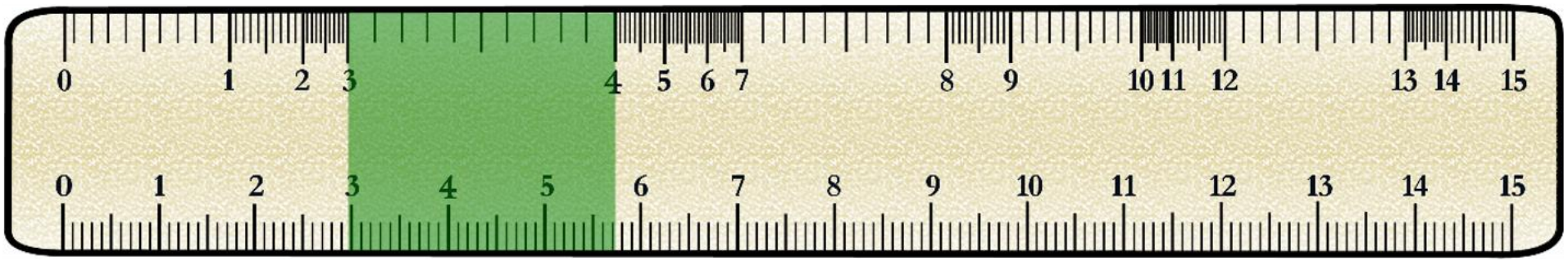
Additivity (whole is the sum of the parts) is the hallmark of measurement.

The way to construct additivity for psychological measurement is to design an operation which answers the question : “If person A has more ability than person B , then how much ‘ability’ must be added to B to make the performance of B appear the same as the performance of A?”

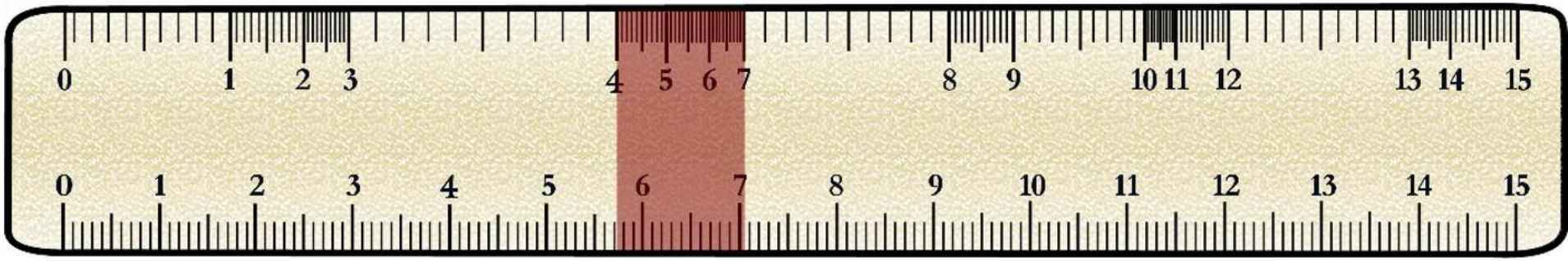
Check :
https://archive.org/details/physicselements00campuoft/page/266

Measurement invariance

Patient One



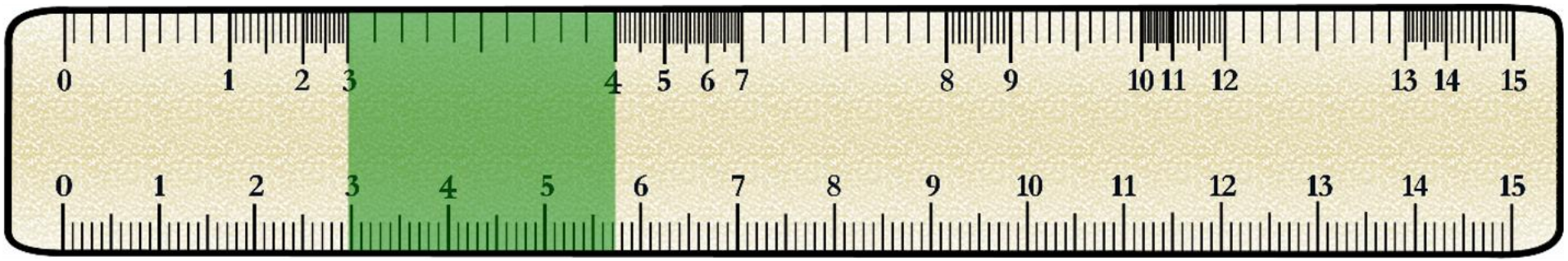
Patient Two



Measurement invariance

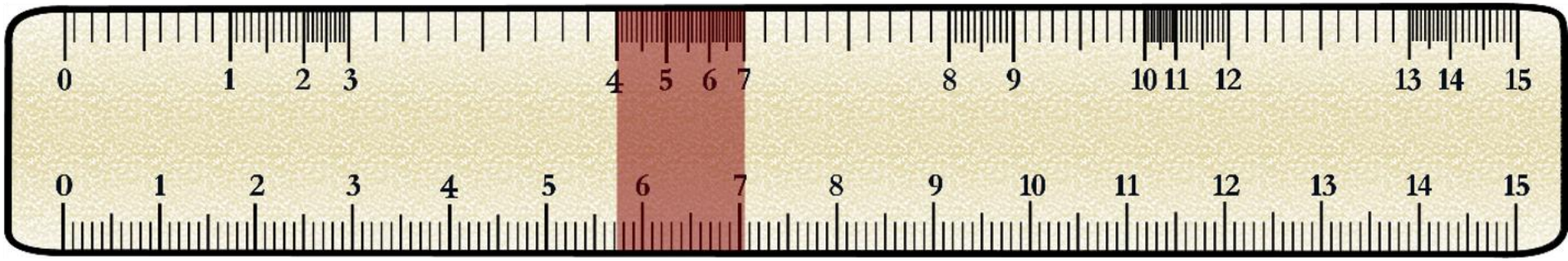
Patient One

+ 1



Patient Two

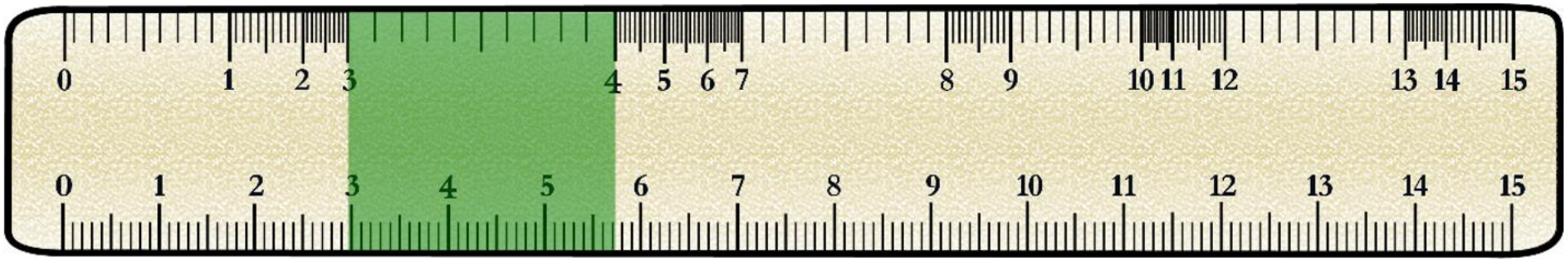
+3



Measurement invariance

Patient One

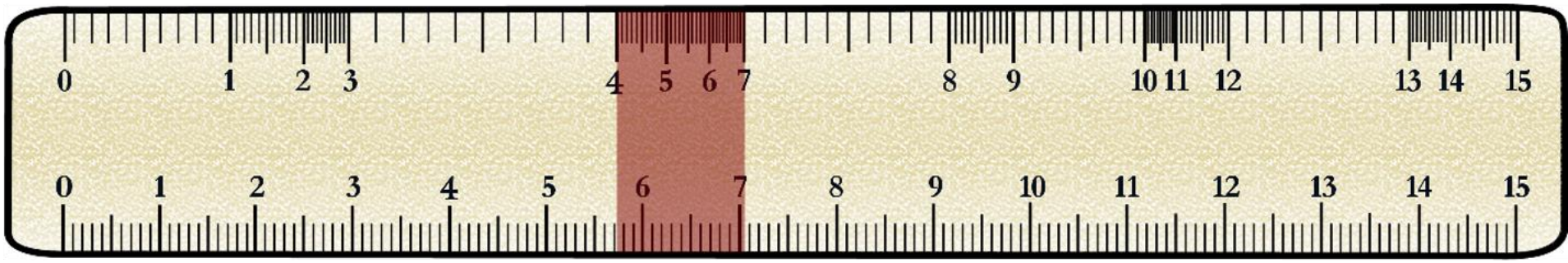
+ 1



+ 2.7

Patient Two

+3



1.2

Misinference from Ordinal Scales[4]

"Ordinal scales of measurement
do not support the mathematical operations
needed to calculate means and standard
deviations.

Why is ordinal not good enough?

Pretend you are the CEO of an orange juice company

5000 oranges vs. 750 kg of oranges

Measuring versus counting

1. Do you have difficulty engaging in vigorous activities, such as playing football or doing sport?
2. Do you have difficulty going outside of your home?
3. Do you have difficulty moving around inside your home?

Not all questions are equally difficult!

Before the Rasch Model: S.S. Stevens (1906-1973)

SCIENCE

Vol. 103, No. 2684 Friday, June 7, 1946

On the Theory of Scales of Measurement

S. S. Stevens

Director, Psycho-Acoustic Laboratory, Harvard University

FOR SEVEN YEARS A COMMITTEE of the British Association for the Advancement of Science debated the problem of measurement. Appointed in 1932 to represent Section A (Mathematical and Physical Sciences) and Section J (Psychology), the committee was instructed to consider and report upon the possibility of "quantitative estimates of sensory events"—meaning simply: Is it possible to measure human sensation? Deliberation led only to disagreement, mainly about what is meant by the term measurement. An interim report in 1938 found one member complaining that his colleagues "came out by that same door as they went in," and in order to have another try at agreement, the committee begged to be continued for another year.

For its final report (1940) the committee chose a common bone for its contentions, directing its arguments at a concrete example of a sensory scale. This was the Sone scale of loudness (S. S. Stevens and H. Davis. *Hearing.* New York: Wiley, 1938), which purports to measure the subjective magnitude of an auditory sensation against a scale having the formal properties of other basic scales, such as those used to measure length and weight. Again the 19 members of

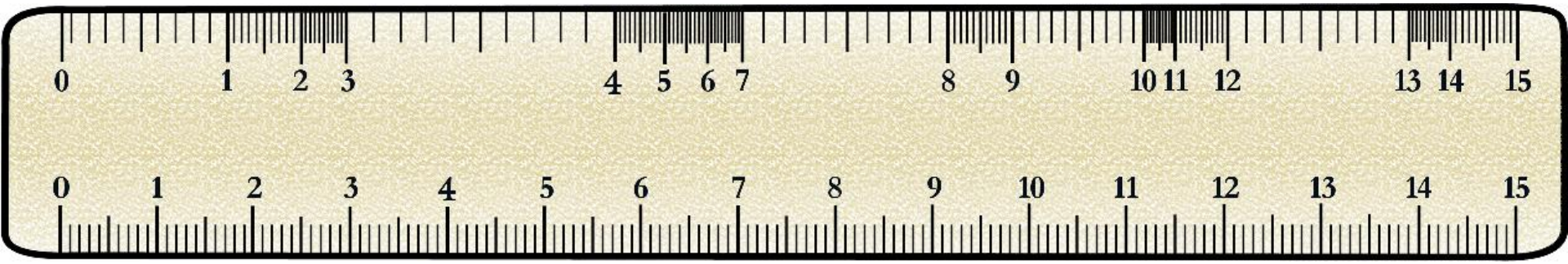
by the formal (mathematical) properties of the scales. Furthermore—and this is of great concern to several of the sciences—the statistical manipulations that can legitimately be applied to empirical data depend upon the type of scale against which the data are ordered.

A Classification of Scales of Measurement

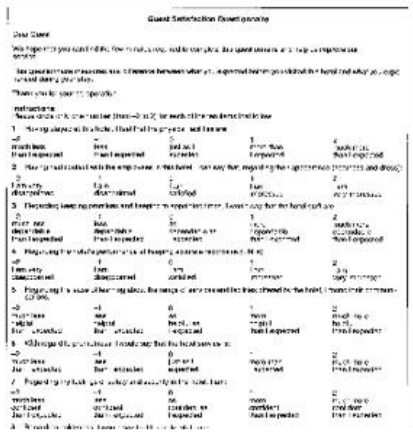
Paraphrasing N. R. Campbell (Final Report, p. 340), we may say that measurement, in the broadest sense, is defined as the assignment of numerals to objects or events according to rules. The fact that numerals can be assigned under different rules leads to different kinds of scales and different kinds of measurement. The problem then becomes that of making explicit (a) the various rules for the assignment of numerals, (b) the mathematical properties (or group structure) of the resulting scales, and (c) the statistical operations applicable to measurements made with each type of scale.

Scales are possible in the first place only because there is a certain isomorphism between what we can do with the aspects of objects and the properties of the numeral series. In dealing with the aspects of objects we invoke empirical operations for determin-

How to Make the Transition From the Top to the Bottom of the Ruler?



Rasch Analysis



$$p_i(\theta) = \frac{e^{(\theta-\delta_i)}}{1+e^{(\theta-\delta_i)}}$$

Georg Rasch (1901-1980)

Important Concepts

Measurement entities cannot be measured but properties of entities.

A researcher does not measure a person but attributes of the person.

In social sciences these properties are referred to as **latent** *[…]*

$$property \approx trait \approx construct \approx attribute \approx variable$$

where these terms have similar connotation:

– **Properties** can be measured through their manifestation.

Important Concepts

Assessment is a systematic collection of a set of **observations** that arise from the manifestations of some property.

– **Observations** are produced from

$$instrument \approx test \approx scale \approx questionnaire \approx set\ of\ items \approx a\ measure \approx assessment\ tool...$$

Important Concepts

Observations from an assessment are referred to as **scored responses**

– **Dichotomous scoring**: 2 response-levels
e.g. *agree/disagree*, *correct/incorrect*, *problem/no problem*

– **Polytomous scoring**: more than two response-levels
e.g. *strongly agree*, *agree*, *disagree*, and *strongly disagree*.

Recommended litterature

Rasch:

- T.G. Bond: Applying the Rasch Model: Fundamental Measurement in the Human Sciences. Lawrence Erlbaum Associates, Mahwah New Jersey 2007.
- D. Andrich and I. Marais: A Course in Rasch Measurement Theory: Measuring in the Educational, Social, and Health Sciences. Springer Nature, Singapore 2019.

R:

- Contributed documentation with introductions to R : https://cran.r-project.org/

Recommended litterature

Rasch & R:

- An overview to packages for psychometric analysis are found on https://cran.r-project.org/web/views/Psychometrics.html
- C. Strobl: Das Rasch-Modell - Eine verständliche Einführung für Studium und Praxis. 2. erweiterte Auflage. Rainer Hampp Verlag, Muenchen 2012.
- Overview to psychometric methods including Rasch in R: https://epubdev.wu.ac.at/4010/1/resrepIRThandbook.pdf

Notes

1. Andrich D. (2004) The Rasch Model: A Characteristic of Incompatible Paradigms? Medical Care 42 (1) suppl

2. G. Rasch: Probabilistic models for some intelligence and attainment tests. Danish Institute for Educational Research, Copenhagen 1960

3. N.R. Campbell (1920). Physics: the Elements. London Cambridge University Press.

4. Merbitz C, Morris J, Grip JC. Arch Phys Med Rehabil 1989: 70:308-312