

妥当性概念の展開

日本学術振興会・東京工業大学
村山 航

発表の構成

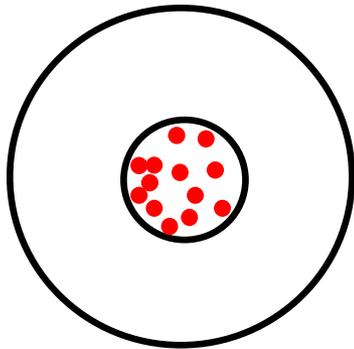
- 歴史的変遷
- 近年における捉え方
- さらに考えを発展させるために

発表の構成

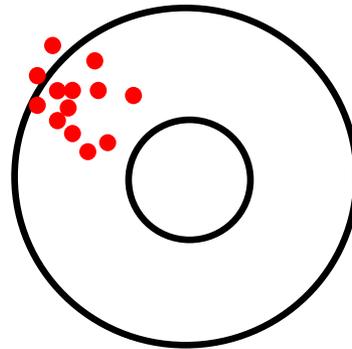
- 歴史的変遷
- 近年における捉え方
- さらに考えを発展させるために

妥当性 (validity) とは何か

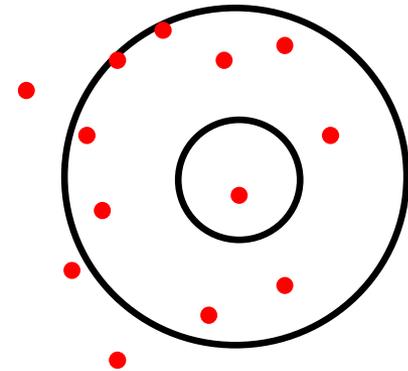
- 測定したいものが測定できているのか
 - “A test is valid if it measures what it purports to measure” (Kelley, 1927)
- 信頼性 (reliability) との違い
 - ダーツのアナロジー



信頼性: 大
妥当性: 大



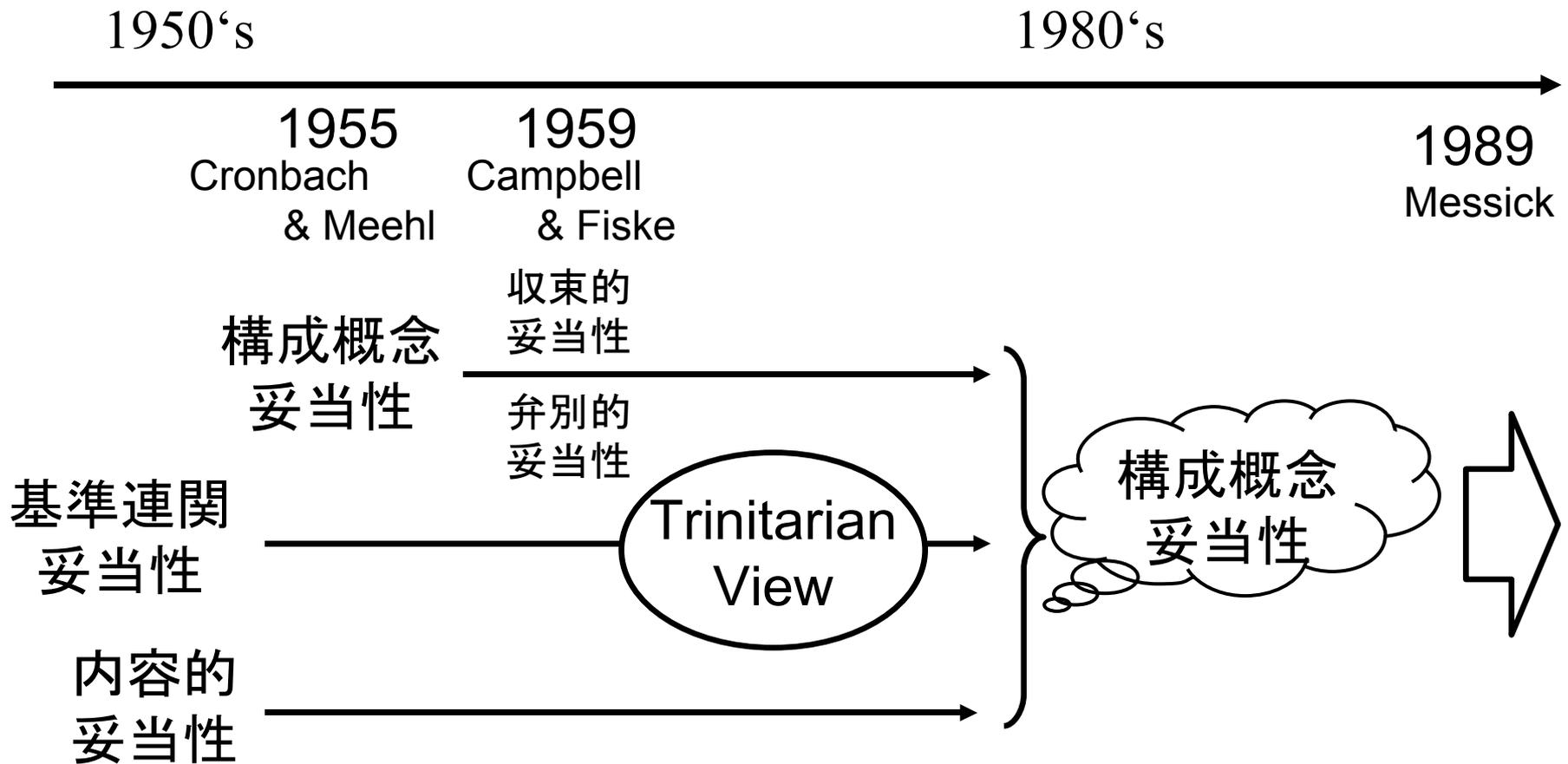
信頼性: 大
妥当性: 小



信頼性: 小
妥当性: 小

※ 信頼性が低くて妥当性が高いものは想定しにくい

妥当性概念の歴史的変遷



操作主義 論理実証主義 反証主義

実用論

1950年代まで

- 背景：操作主義 (operationalism)

- 測定の方法(尺度)自体が概念の定義である
- 尺度が何を測定しているかは考えない



- 基準関連妥当性 (criterion-referenced validity) の登場

- 尺度がその概念を反映している外的基準と相関するか

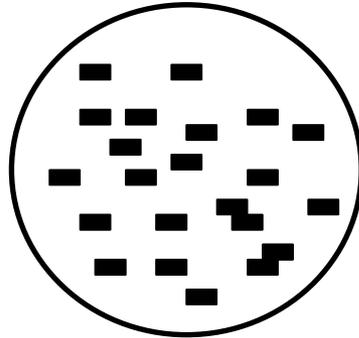


- 予測的妥当性・併存的妥当性の2タイプ
- 尺度が測定しているのは外的基準そのもの (Anastasi, 1950)

項目内容に対する視点

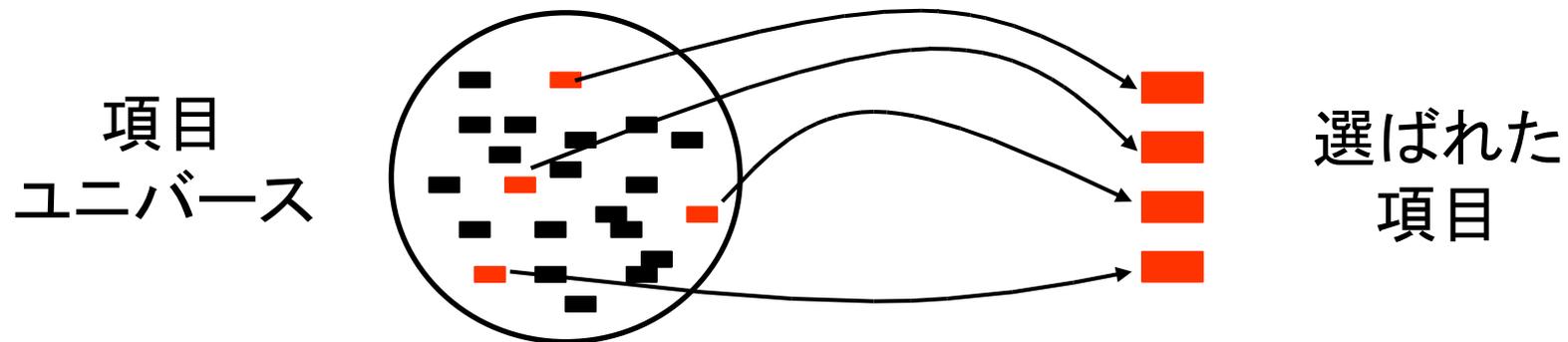
- 内容的妥当性 (content validity; e.g., Rulon, 1946)
 - 問題や質問の内容が測定したい領域を反映しているか
 - 領域の範囲内から選ばれているか
 - 領域から偏りなく選ばれているか

項目
ユニバース



項目内容に対する視点

- 内容的妥当性 (content validity; e.g., Rulon, 1946)
 - 問題や質問の内容が測定したい領域を反映しているか
 - 領域の範囲内から選ばれているか
 - 領域から偏りなく選ばれているか



- 専門家のチェックによる検討
 - e.g., 分数の割り算能力を測定するテストの開発
 - 問題点: 主観的になりやすい

Cronbach & Meehl のブレイクスルー

■ 背景：論理実証主義

- 現象の背後に一般的な法則(理論)を想定する
- 理論から得られる命題を実証的に検討



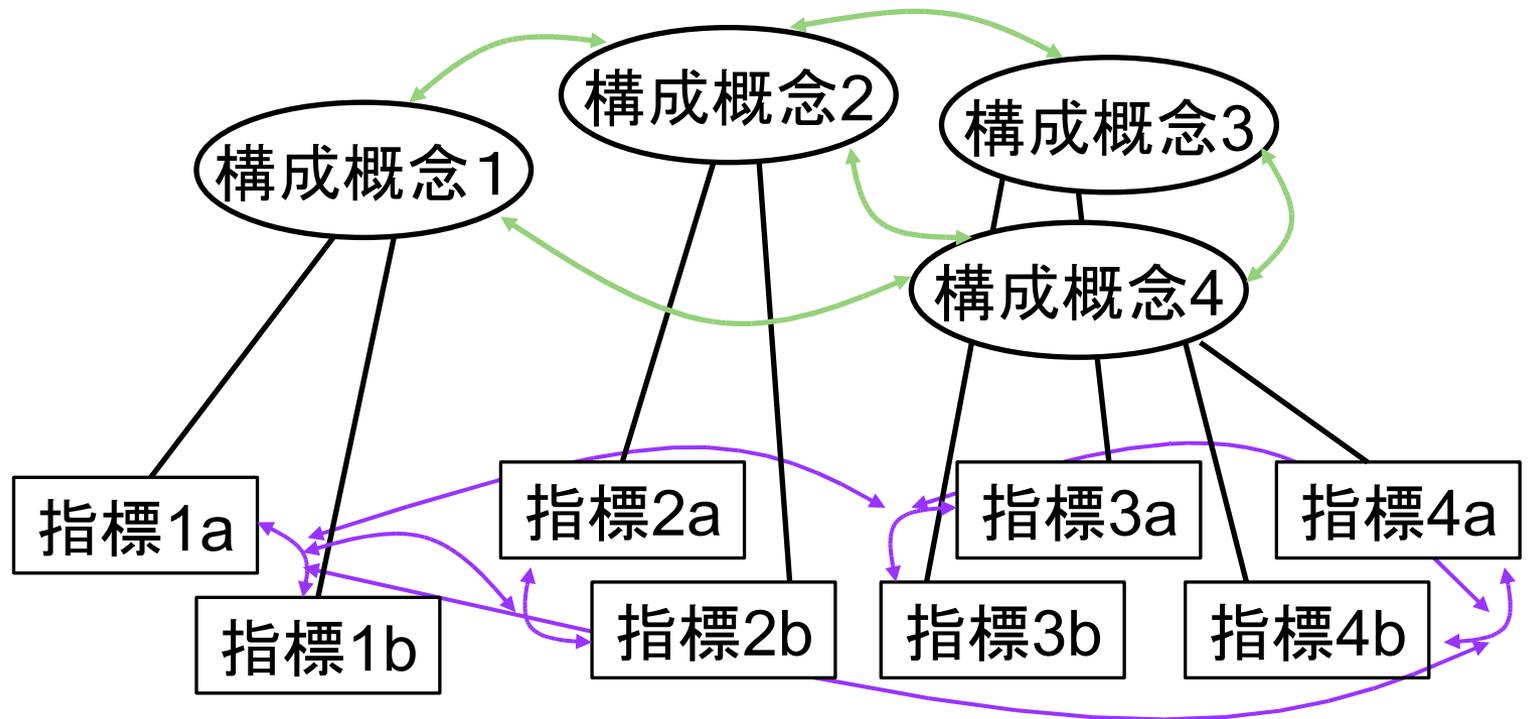
■ Cronbach & Meehl (1955)

- 構成概念妥当性 (construct validity) の重要性を主張
- 尺度は理論的・仮説的な構成概念を測定している
- 構成概念は他の構成概念との関係で定義される
⇒ 法則定立ネットワーク

法則定立ネットワーク (nomological network)

理論の
世界

現実の
世界



- Cronbach & Meehl流の構成概念妥当性とは、データによってこのネットワークの理論を検証すること

収束的妥当性と弁別的妥当性

- 構成概念妥当性の2つの検証方法
 - 収束的妥当性 (convergent validity): 理論的に関連の強い構成概念を測定する指標との相関が高い
 - 弁別的妥当性 (discriminant validity): 理論的に関連の弱い構成概念を測定する指標との相関が低い
- Campbell & Fiske (1959) の提唱
 - 多特性多方法行列(後述)による検証を主張するが、他の場面にも当てはまる概念

構成概念妥当性の特徴

- 妥当性を仮説検証の繰り返しプロセスと考える
 - 仮説(理論): 1つのデータで検証されるものではない
 - あるデータを説明できる仮説は必ず複数存在する
 - 対立仮説を棄却して仮説の確証度を高めることが大切

反証主義の影響

データ

「尺度得点X」と「失敗時の皮膚電気反応」に0.5の相関

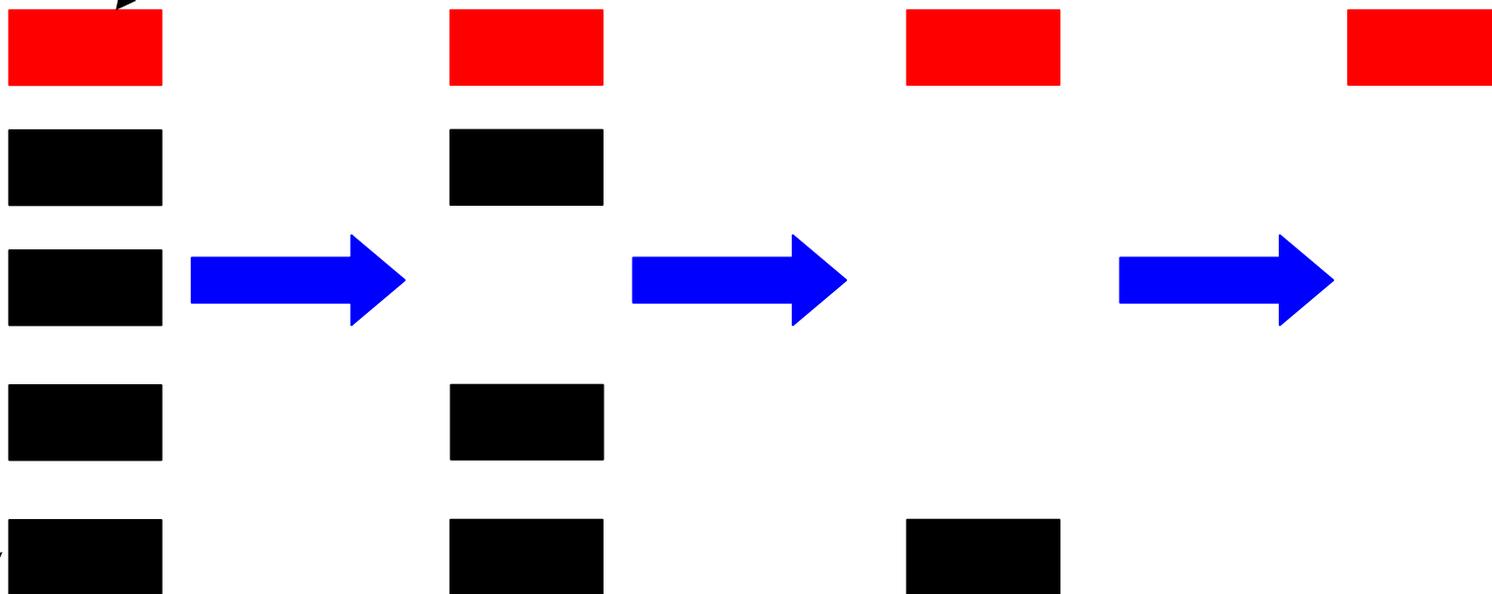
「尺度得点X」と「身体の震え」に0.4の相関

仮説

「X」は「不安」を測定している

「X」は「成功欲~~求~~」を測定している

支持したい仮説



対立仮説

〇〇尺度との
相関大

△△尺度との
相関小

××尺度との
相関大

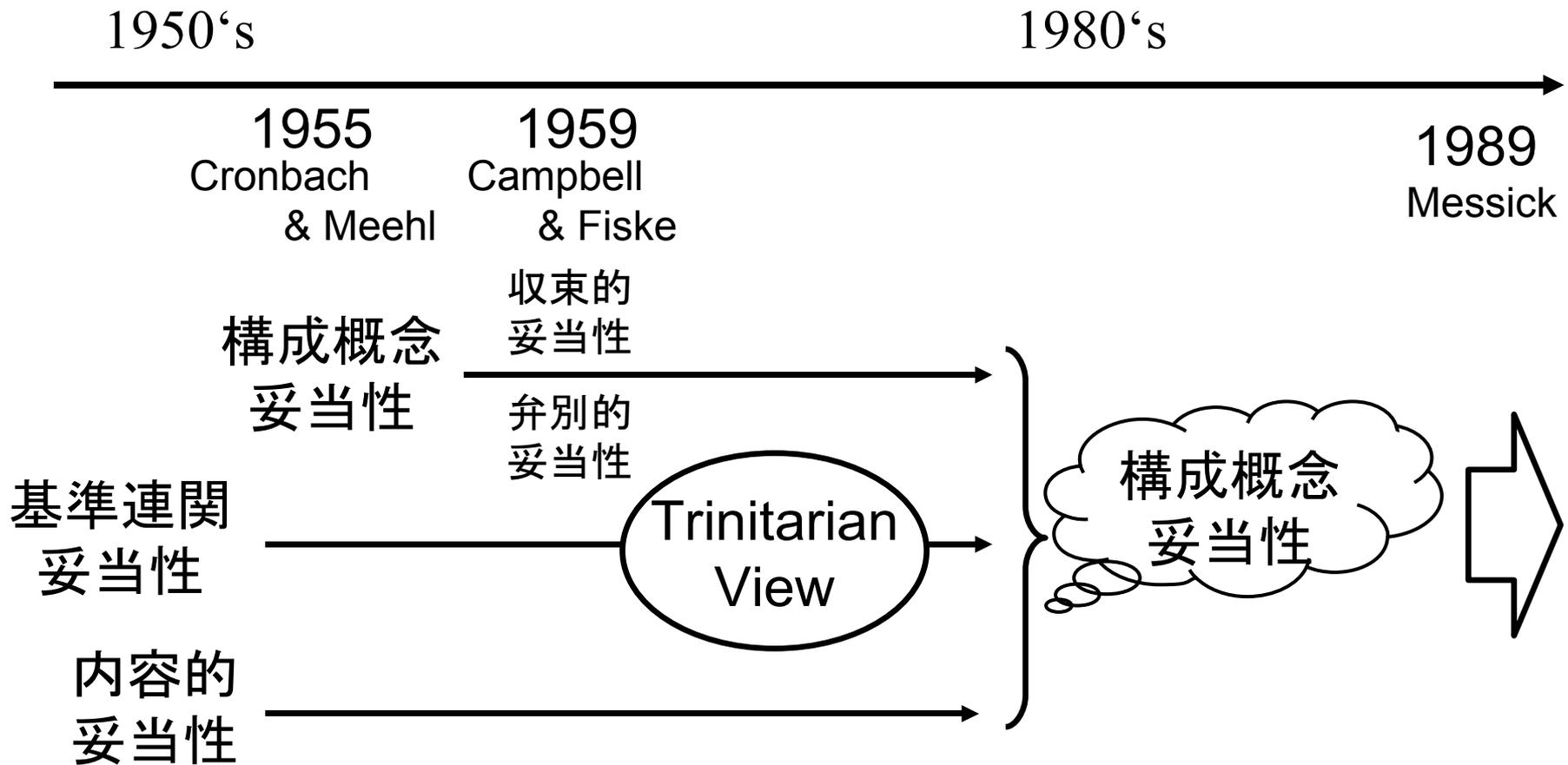
Trinitarian Viewの時代

- APA et al.(1954, 1966, 1974)のテストスタンダード
 - 基準関連妥当性, 内容的妥当性, 構成概念妥当性を, 妥当性の3つのタイプとして記述
- その結果, 3つの妥当性の関係について詳細な考察が行われず
 - 3つの妥当性を, ただ形式的に Stamp Collectingすればよいという風潮 (Landy, 1986)



Trinitarian View

妥当性概念の歴史的変遷



操作主義 論理実証主義 反証主義

実用論

発表の構成

- 歴史的変遷
- 近年における捉え方
- さらに考えを発展させるために

教育心理学ハンドブック



日本教育心理学会 編

心理尺度の作り方

村上宣寛 著

北大路書房

心理学研究法入門

調査・実験から実践まで

南風原朝和 市川伸一 下山晴彦 編

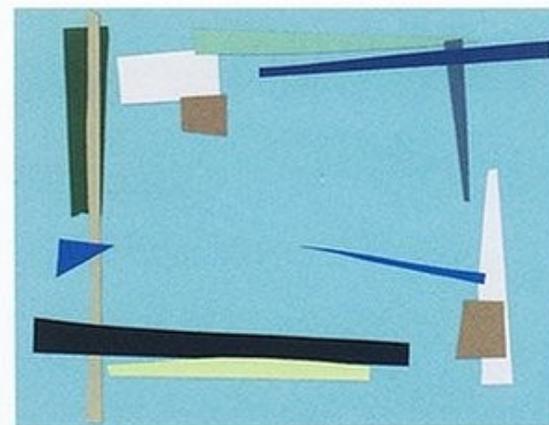
東京大学出版会

心理学の新しいかたち3

心理学 研究法の 新しいかたち

吉田寿夫 著

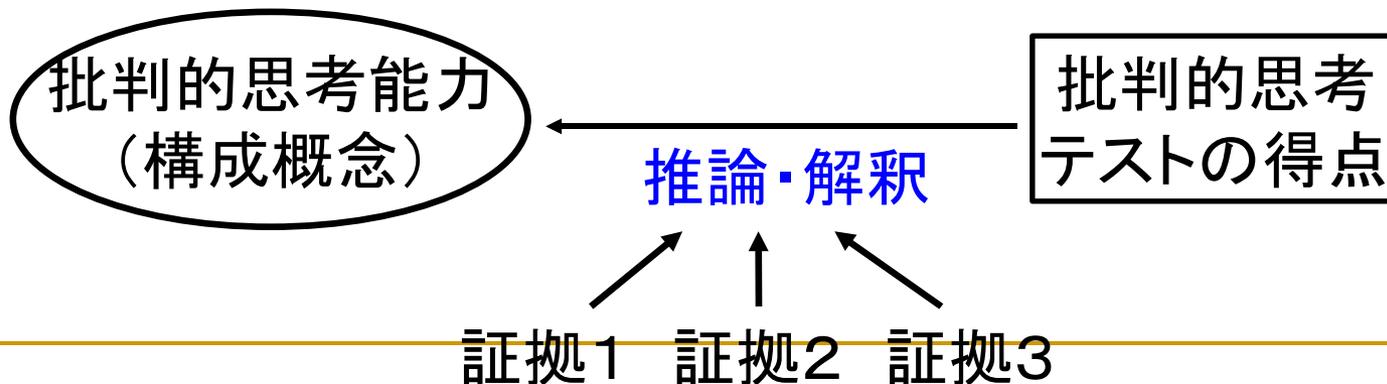
シリーズ企画・編集 下山晴彦



誠信書房

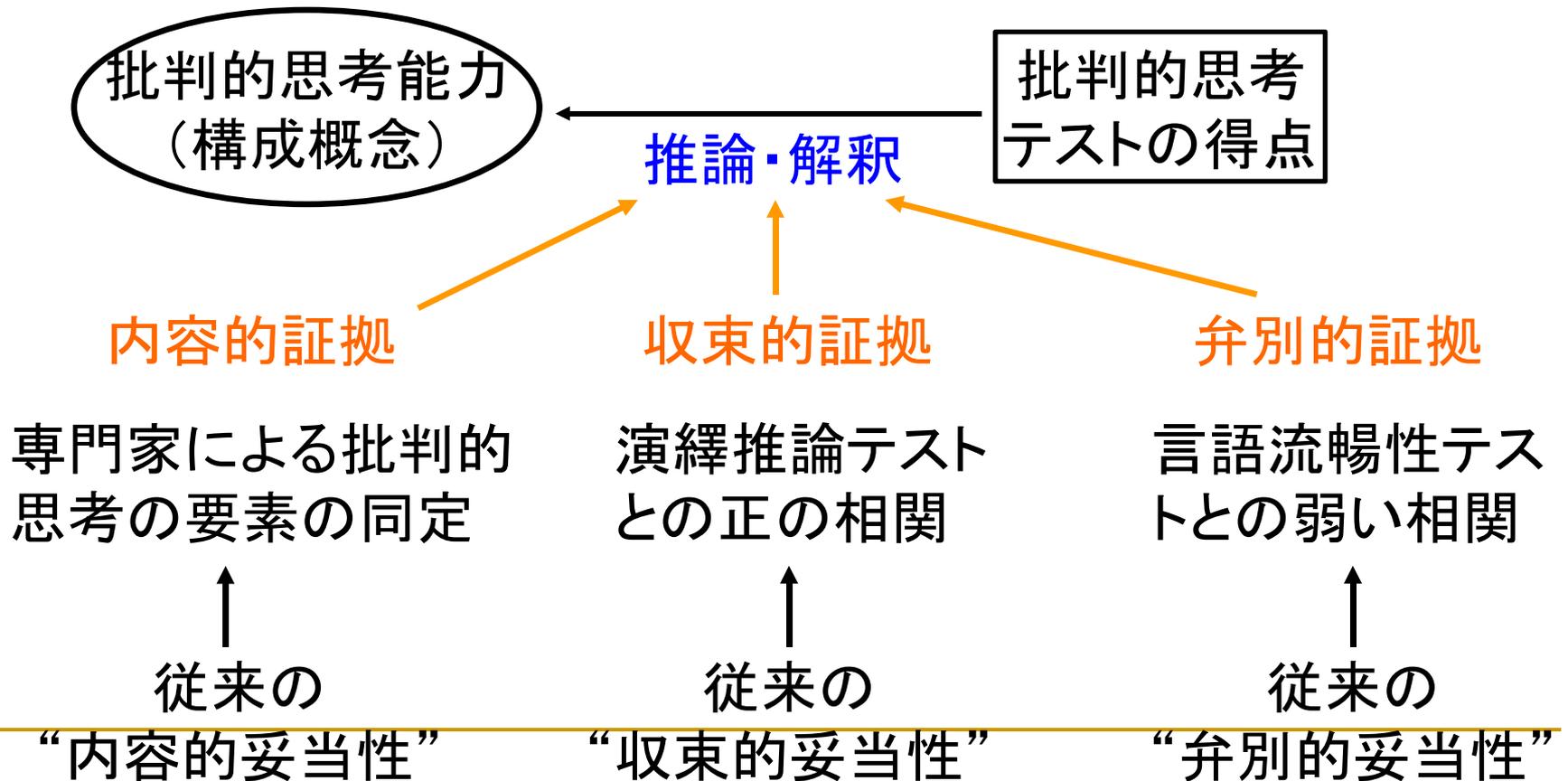
「構成概念妥当性」による統合

- 構成概念妥当性は妥当性の下位概念でなく，“妥当性そのもの”：妥当性は単一の概念 (unitary concept)
- 構成概念妥当性とは (Messick, 1989)
 - テスト得点に基づいて構成概念に対する推論・解釈をするとき、その推論・解釈を支える証拠の適切性に対する統合的な評価（=テスト得点の解釈の適切性）



「〇〇妥当性」は何だったのか？

- 構成概念妥当性を検証するための証拠・方法のタイプ
 - 妥当性 (validity) と妥当化 (validation) の区別



Messick(1995)の妥当化に関する6つの基準

- 内容的側面：専門家による判断など
- 本質的側面：プロセスの分析など
- 構造的側面：因子分析など
- 一般化側面：信頼性など
- 外的側面：相関パターンなど
- 結果的側面：社会的影響の分析など

e.g. パフォーマンスアセスメント

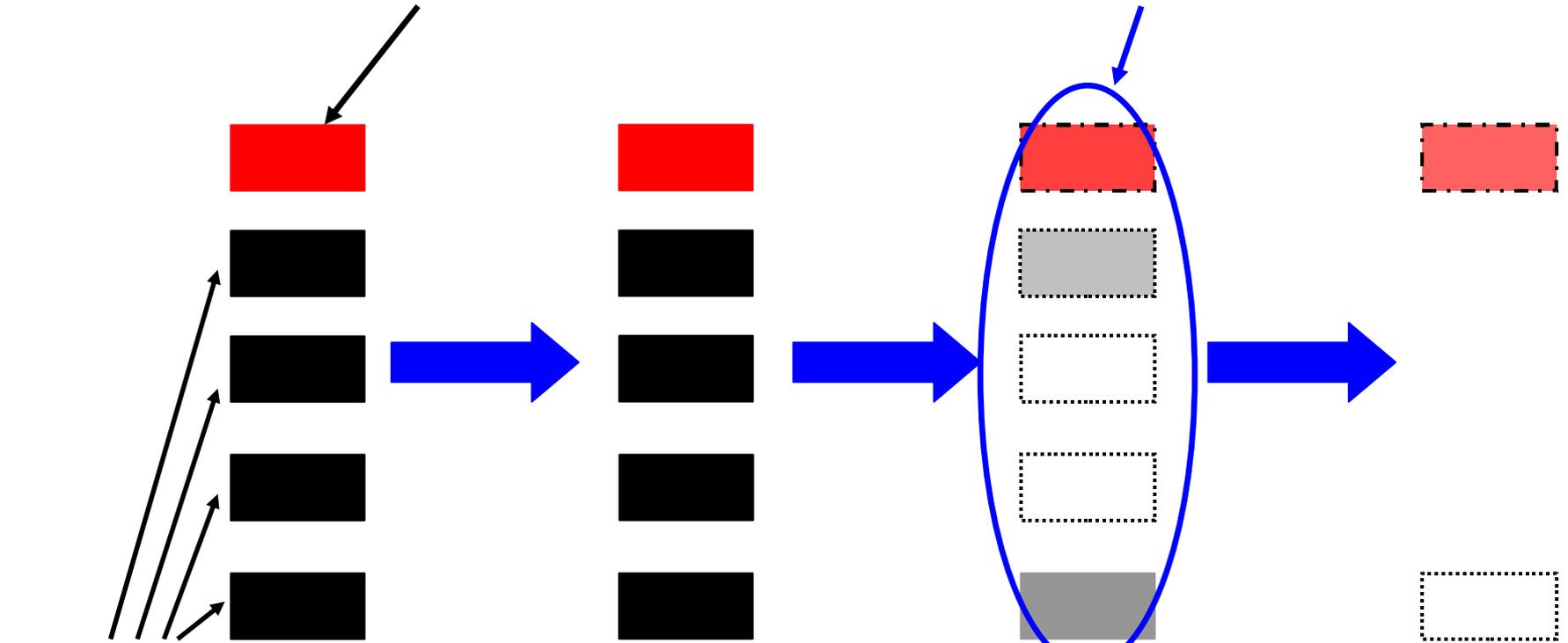
(Linn et al., 1991; Moss, 1992; 村山, 2006など)

基本はやはり「繰り返しの仮説検証」

- ただし、その仮説検証の範囲は、もはや法則定立ネットワークだけに留まらない
 - さまざまな基準からの多面的検討が必須
 - 妥当性はテストの属性ではない：目的・文脈依存性
- また、仮説検証も厳密な論理実証主義で行うわけではない：**实用主義的な論証アプローチ** (Kane, 1992)
 - 仮説は単一の証拠で完全に棄却されるわけではない：妥当性のある・なしではなく「どの程度あるのか」
 - 証拠には強い前提を支える証拠と弱い前提を支える証拠があり、**全体的な妥当性は最も弱い前提を支える証拠に規定される**
 - この証拠を補強することが効果的な妥当性検証

支持したい仮説

仮説が完全に確証/棄却
されるわけではない



対立仮説

信頼性係数
が△△

内容的な
吟味の結果

××尺度との
相関大

収束・弁別的証拠だけでない

弱い前提を支える証拠

発表の構成

- 歴史的変遷
- 近年における捉え方
- さらに考えを発展させるために

さらに考えを発展させるための3つの視点

- 「人間のモデル」を考える
- 尺度得点を算出することの難しさを考える
- 数量化の方法を考える

「人間のモデル」を考える

- 解答/回答者は人間
- 項目に答えるときの「人間のモデル」を知ることなしに妥当性のあるテストの作成は無理！

人間は積極的に解釈を作りあげてしまう

Please describe your mood right now

作成者の
意図

ニュートラル 1 2 3 4 5 6 7 8
少し嬉しい

1
not happy

2

3

4

5

6

7

happy

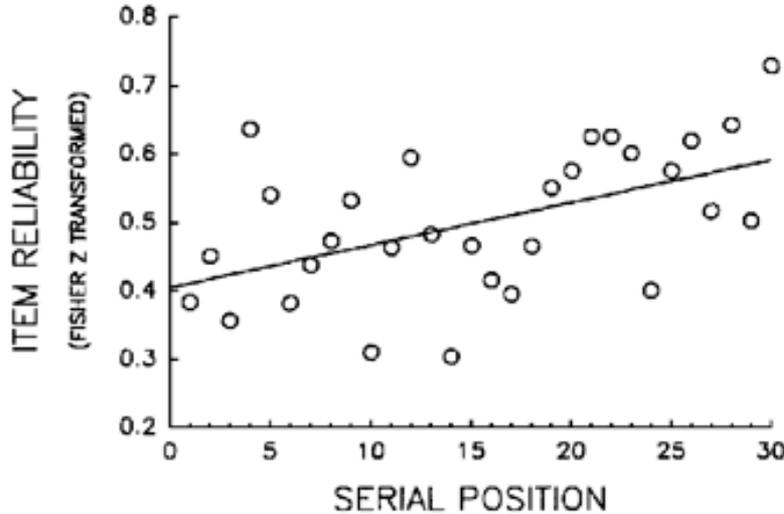
回答者の
読み取り

やや“悲しい”

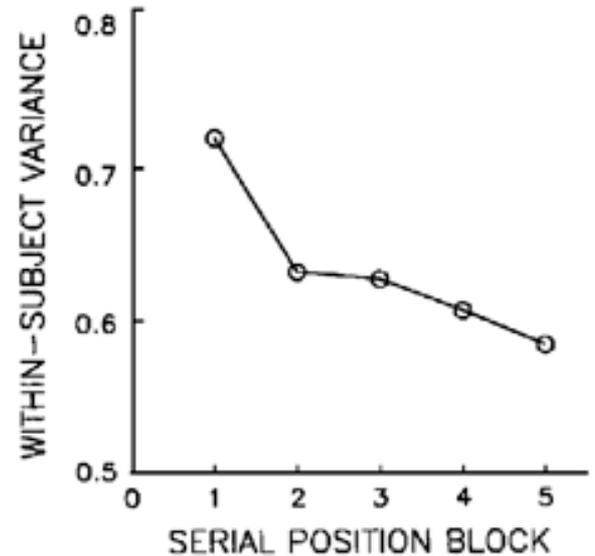
ニュートラル

単極尺度が
双極尺度に！

項目-全体間相関



被験者内項目分散



回答者は徐々に一貫した回答をするようになる！



回答者は「何が測定されているか」を積極的に解釈し、トップダウン的に回答



相関のある誤差の蓄積

(Drolet, 2001)

内的一貫性の増大

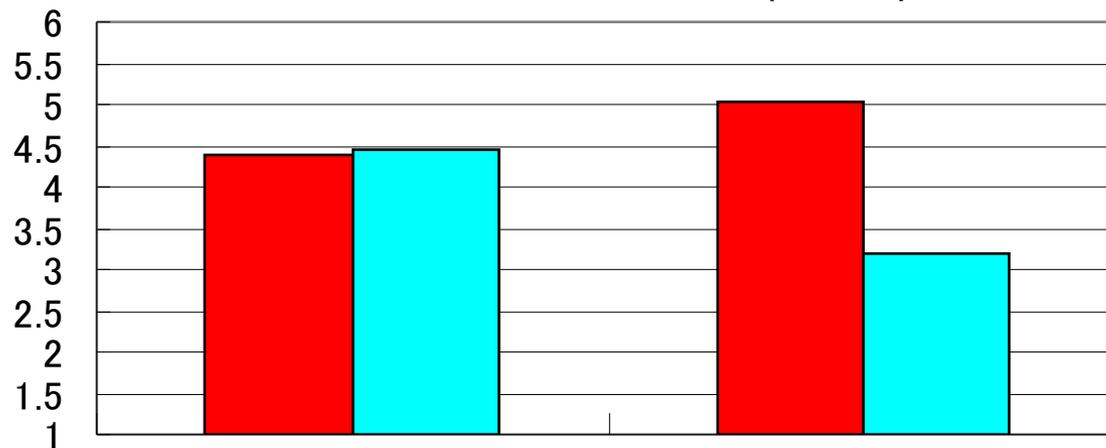


妥当性の低下？

人間はテストにあわせて方略を変える

学習方略使用の程度

村山 (2004)



空所補充クラス

記述クラス

■ 意味理解方略

■ 暗記方略

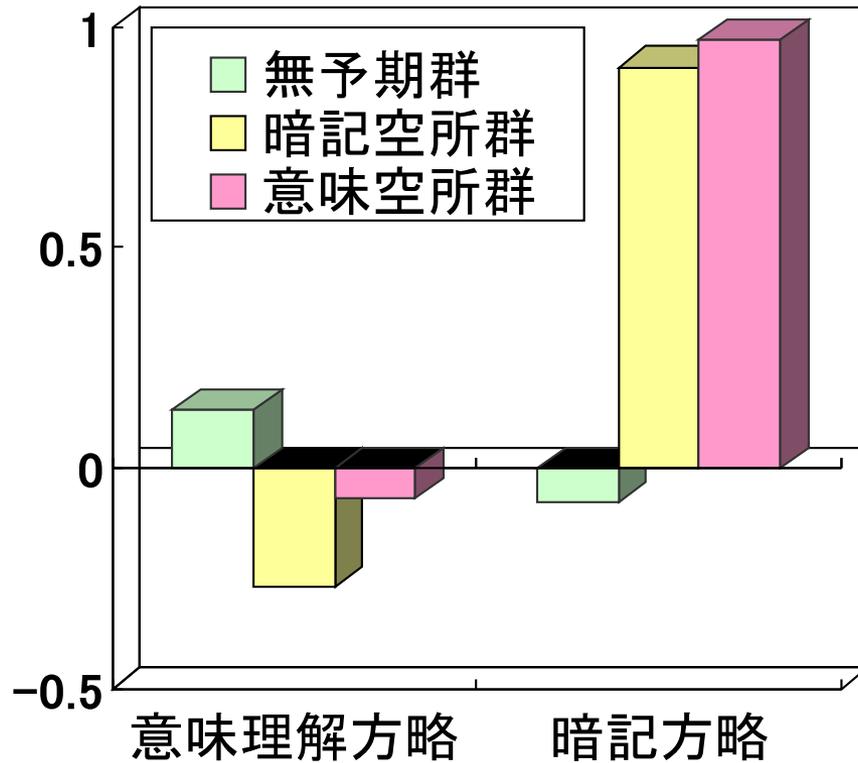
妥当性の増大？
(Powers, 1985)

妥当性の低下？
(Mehrens & Kaminski, 1985)

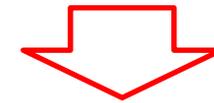
人はテストにあわせて柔軟に方略を調整

妥当性はテストと人との相互作用で変化する

■ さらに…



解答者の学習行動は、テストの表面的な「見え方」に引きずられる



「人間のモデル」を考えると、表面的妥当性も重要！

村山 (2005)

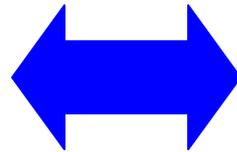
■ 他にも…

- 内省能力の限界 (Nisbett & Wilson, 1977; 吉田, 2002)
 - 反応バイアスの存在
 - 選択肢を呈示する文脈の効果 (Sudman et al. 1996)
 - 問題文の表現の効果 (Hudson, 1983)
-

尺度得点を算出することの難しさを考える

■ 項目作成のときの2つのベクトル

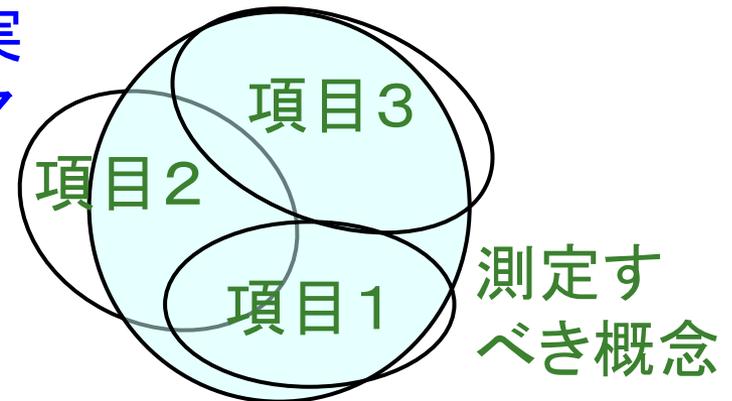
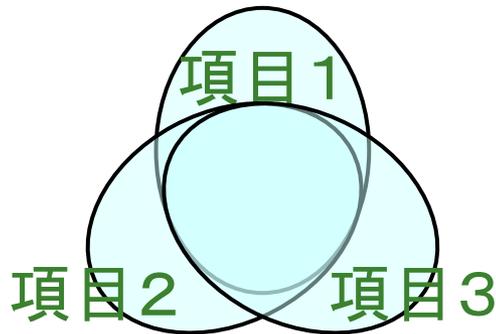
できるだけ共通性の
高い項目を作成する



できるだけ領域の代表
性が高い(幅広い)項
目を作成する

ジレンマ!

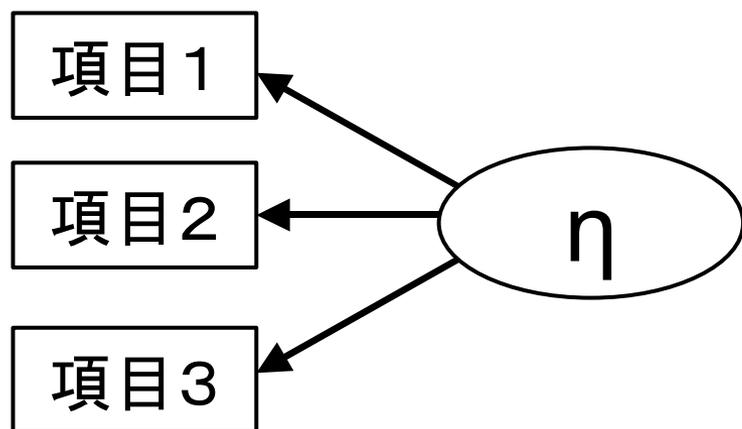
帯域幅と忠実
度のジレンマ



項目作成のときには、このジレンマと戦いつつ、
両方を満たすような尺度を作成する必要性

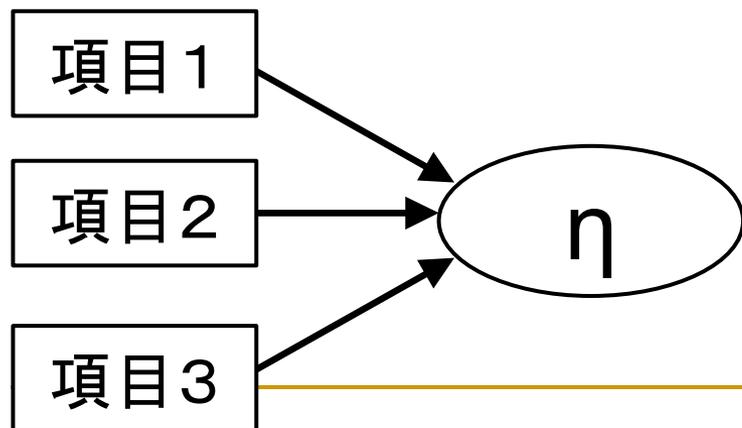
- しかし、これらを満たす「いい尺度」ができたとしても、「尺度得点の算出方法(モデル)」によって切り捨てられる部分が出てくる

Bollen & Lennox (1991)



- 結果指標モデル

- 通常使われるモデル
- η は全項目の共通成分
- 項目の独自因子は誤差に



- 原因指標モデル

- 項目の独自部分も η に寄与
- 高い共通性(項目間相関)は η の解釈を困難に(多重共線性)

- 普通に足し合わせる方法 (parceling?)
 - スケーリングの問題
 - 他変数との相関は項目内の相関関係に依存する
 - ⇒ 内的一貫性と基準連関妥当性のジレンマ

$$r_{XY} = \frac{\sum_{i=1}^k r_{iY}}{\sqrt{k + k(k-1)\bar{r}_{ii}}}$$

\bar{r}_{ii} が大 $\Rightarrow r_{XY}$ の低下
 \bar{r}_{ii} が小 $\Rightarrow r_{XY}$ の増大

内的一貫性

基準連関
妥当性

Y : 外的基準

X : k 個の尺度項目の合計点

\bar{r}_{ii} : すべての尺度項目間の相関の平均

数量化の方法を考える

多特性多方法行列 (Multitrait-Multimethod Matrix, MTMM行列)

	方法1(自己報告)			方法2(他者評定)		
	A	B	C	A	B	C
<u>方法1</u>						
特性A	(.82)					
特性B	.13	(.80)				
特性C	.24	.23	(.43)			
<u>方法2</u>						
特性A	.65	.14	.10	(.28)		
特性B	.06	.73	.16	.27	(.38)	
特性C	.01	.08	.69	.19	.37	(.42)

信頼性

収束的妥当性

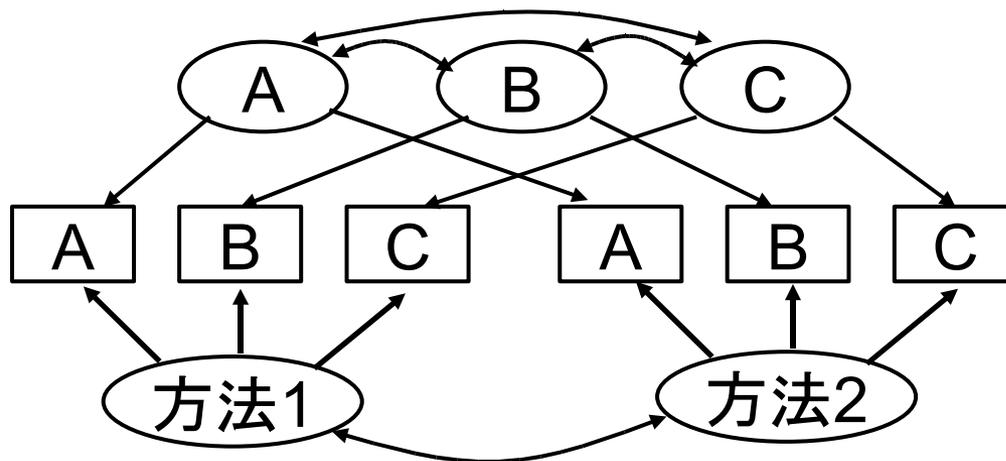
三角の枠が弁別的妥当性

- 収束的・弁別的妥当性を統合的に評価できるが…

- 方法・特性の分散を定量的に評価できない
- 方法・特性の共分散や交互作用を評価できない

- 加法モデル (Jöreskog, 1974)

$$x = \mu + A_T \xi_T + A_M \xi_M + e$$



- 直積モデル (Browne, 1984)

$$x = \mu + (A_M \otimes A_T)(\xi_M \otimes \xi_T) + (e_M \otimes e_T)$$

■ 構成概念妥当性の数量化

- Westen & Rosenthal(2003): Quantifying construct validity
cf. pattern matching法 (Trochim, 1985)

予測値

実測値

Diagnoses involved in prediction	Westen and Rosenthal's predicted correlations	Lambda weights	Obtained z correlations
Histrionic	.60	7	.62
Borderline	.30	4	.56
Dependent	.10	2	.20
Antisocial	.00	1	-.06
Narcissistic	.00	1	.10
Paranoid	-.10	0	-.04
Obsessive-compulsive	-.40	-3	-.23
Avoidant	-.50	-4	-.20
Schizoid	-.50	-4	-.15
Schizotypal	-.50	-4	-.02
$r_{alerting-cv}$.90
$r_{contrast-cv}$.72*

$r_{alerting-CV}$

予測値と実測
値との相関

$r_{contrast-CV}$

対比の考えを
用い、絶対値を
考慮した相関

-
- 村上先生の方法 ⇒ これから詳細に報告
 - 注意点
 - あくまで妥当性検証の1つのステップ。これだけで妥当性が保障されるわけでは決していない (Smith, 2003)
-

The End of Presentation

Thank you!

Murayama Kou

質問がありましたら
murakou@orion.ocn.ne.jp
までお願いします

補足 : Borsboom et al. (2004)

- Messickを中心とした近年の構成概念妥当性の考え方を真っ向から否定
- 妥当性の定義：“測りたいものが測れているか”
- その基準：構成概念から指標への“因果”関係
 - この因果のプロセスを明らかにすることこそがすべて



絶対的な存在
(他の変数間の相関
関係に依存しない)

- 具体例：工夫速算問題の妥当性検証 (村山・市川, 2006)
 - 工夫速算とは：普通に前から計算をしたり筆算をしたりしても解けるが，ちょっとした工夫をすることでより速く・正確な計算が可能になるテスト。

$$\text{例} : 42 \times 8 \div 7 =$$

- 他指標との相関で妥当性を検討するのが困難
 - そもそも問題を見ただけでも「工夫をすると早く正確に解ける」という因果の流れが明確
 - さらに，“工夫をした人が早く正確に解ける”という因果関係を，インタビューを用いてより直接的に検証
 - 加えて，“工夫速算スキルを教えると点数が上がる”という因果関係も検証