

令和7年度第3回大学等の質保証人材育成セミナー

大学の未来をデザインするPart 2

—指標とロジックモデルを活かした組織が動く計画づくり—

# 指標を設定するだけでなく 妥当性・実用性を 検証するための10の観点

大学改革支援・学位授与機構 研究開発部

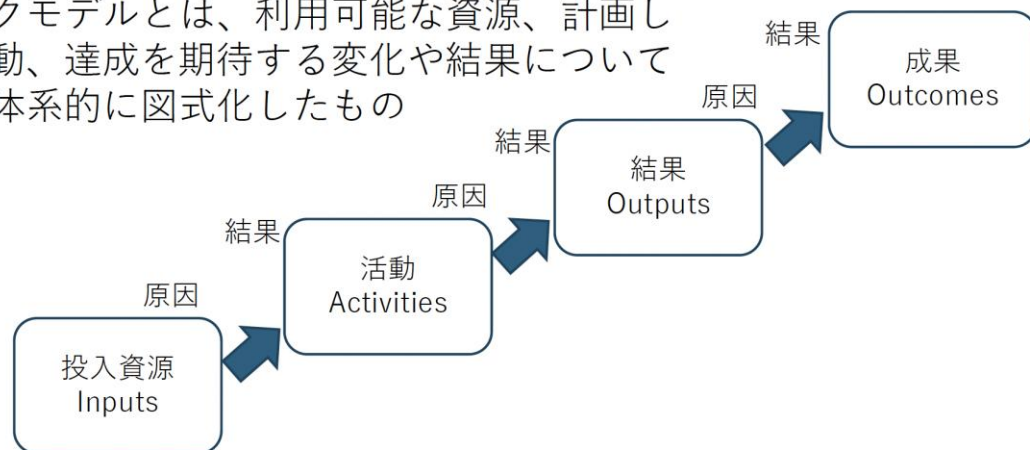
渋井 進

# ロジックモデルって、思考を可視化しただけのような気がする

## 因果関係の連鎖として活動を捉える

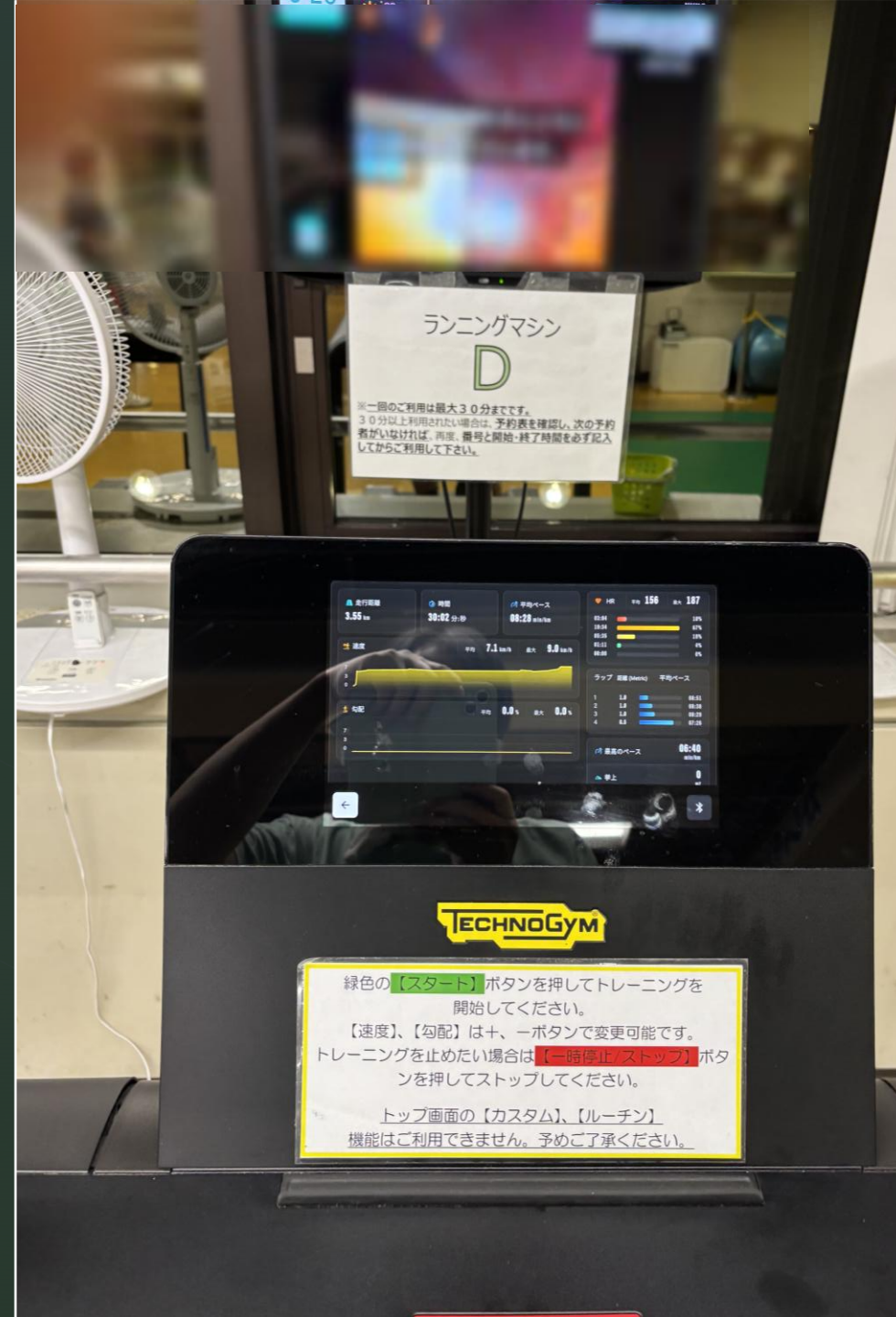
計画は、目標に至るまでの一連の流れが、デザインされていることが重要であり、そのうえで原因と結果の連鎖として捉える必要がある

→ロジックモデルとは、利用可能な資源、計画している活動、達成を期待する変化や結果についての考えを体系的に図式化したもの



- 僕は日々何かを行う際に、どのような結果、成果が得られるかを考えて資源投入している気がする。
- ただ、それがうまく行くかどうかはわからない。
- 例: 勉強、料理、恋愛、etc

# 可視化の効果 は身近に



# 可視化することの価値として

- 構造化と共有可能性
  - 頭の中の考えは人によってバラバラですが、ロジックモデルは共通の枠組みで整理するので、チームや外部に説明しやすくなります。
- 評価設計の基盤
  - 「どの成果を測るべきか」「どの段階で指標を設定するか」を明確にするためのツールになります。これは単なる思考では難しい部分。
- 因果関係の検証
  - 可視化することで「この活動で本当にこの成果につながる？」と批判的に検討できる。頭の中だけだと見落としがち。
- 資源配分や優先順位づけ
  - モデル化することで、どこに力を入れるべきかが可視化される。



前回発表時の導  
入スライドは？

# 問題意識

- 大学は、目的・計画の達成状況や進捗状況について評価を行うが、その判断の根拠となるデータを入手するためには、測定可能な情報から**指標**を設定する必要がある。
- 大学が独自に指標をデザインするが、目的・計画の内容を適切に捉えているものであるか、評価者を十分説得できる内容であるかが気になる。
- せっかくデータを収集・分析しても説得力がないものであったら努力は無駄になってしまう。
- 指標を選ぶにあたって、それが**妥当であるか**、信用できるかを判断するための指針はないだろうか？

# 指標設定が全て？

- 目的の違い
  - 改善へ向けた本質的な体制作りの一環としての指標作りなのか、あるいは(運悪く)回ってきて、評価実施時に(やむなく)収集している指標なのか。
- 効率の問題
  - 指標が評価者等を説得できるものであれば問題ないだろう。
  - ロジックモデルを自己評価書内に書くわけではないし、提出も求められない。
- どうすれば説得力をもつ指標と言えるのか？
  - そこで妥当性！

# 本日の内容

今日のお話と関連する内容が、  
以下の論文にまとめられています。  
ぜひご覧ください。

渋井進, 浅野茂, 橋本貴充, 小野宏, 小野達也, 山崎その, 田中弥生 (2017).「自己評価力向上支援のための評価指標設定に関するチェックリストの開発」,  
大学評価・学位研究 18, 21-36.

- チェックリストの背景にある考え方
  - 妥当性とは？
- チェックリストの説明
  - 妥当性のチェックリスト項目とその事例
  - 実用面のチェックリスト項目とその事例
- チェックリスト使用上の留意点

## ※ 重要なお知らせ

本発表は、信頼できると判断した情報源に基づいて作成されていますが、発表内容を元に実践したことにより生じる直接的または間接的な損失、損害、不利益等に関して、渋井は一切の責任を負いません。最終的な決定は、ご自身の判断と責任において行われますようお願い申し上げます。



## 本日提案のチェックリスト

### 妥当性

- 目的との適合性
- 調査対象・結果への影響
- 信頼性
- 理解可能性
- 包括性・非重複性

### 実用面

- 意思決定者への有用性
- 計測可能性
- 収集の適時性
- データ収集のコスト
- 操作可能性

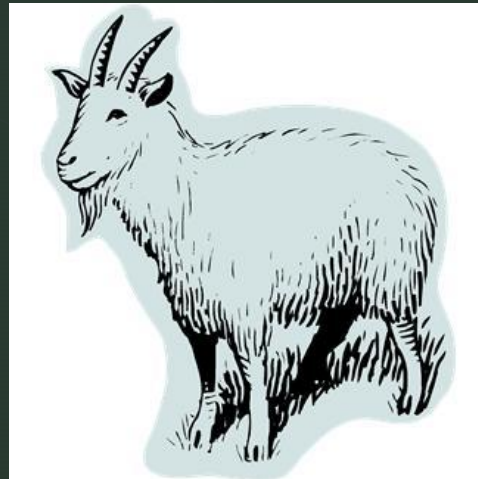
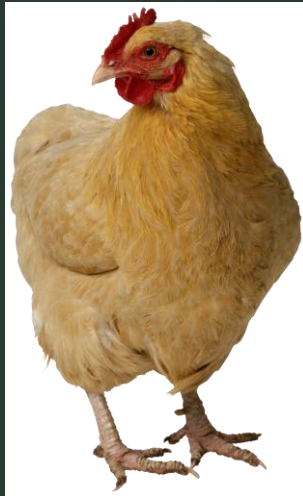
## 妥当性(validity)とは？

- 指標とそのデータが**評価すべき内容を的確に捉えているか。**
- データが得られたとしても、指標が測定したい内容を、適切に捉えていなければ意味が無い。



## 妥当性が問題となる例 (ODAでの指標)

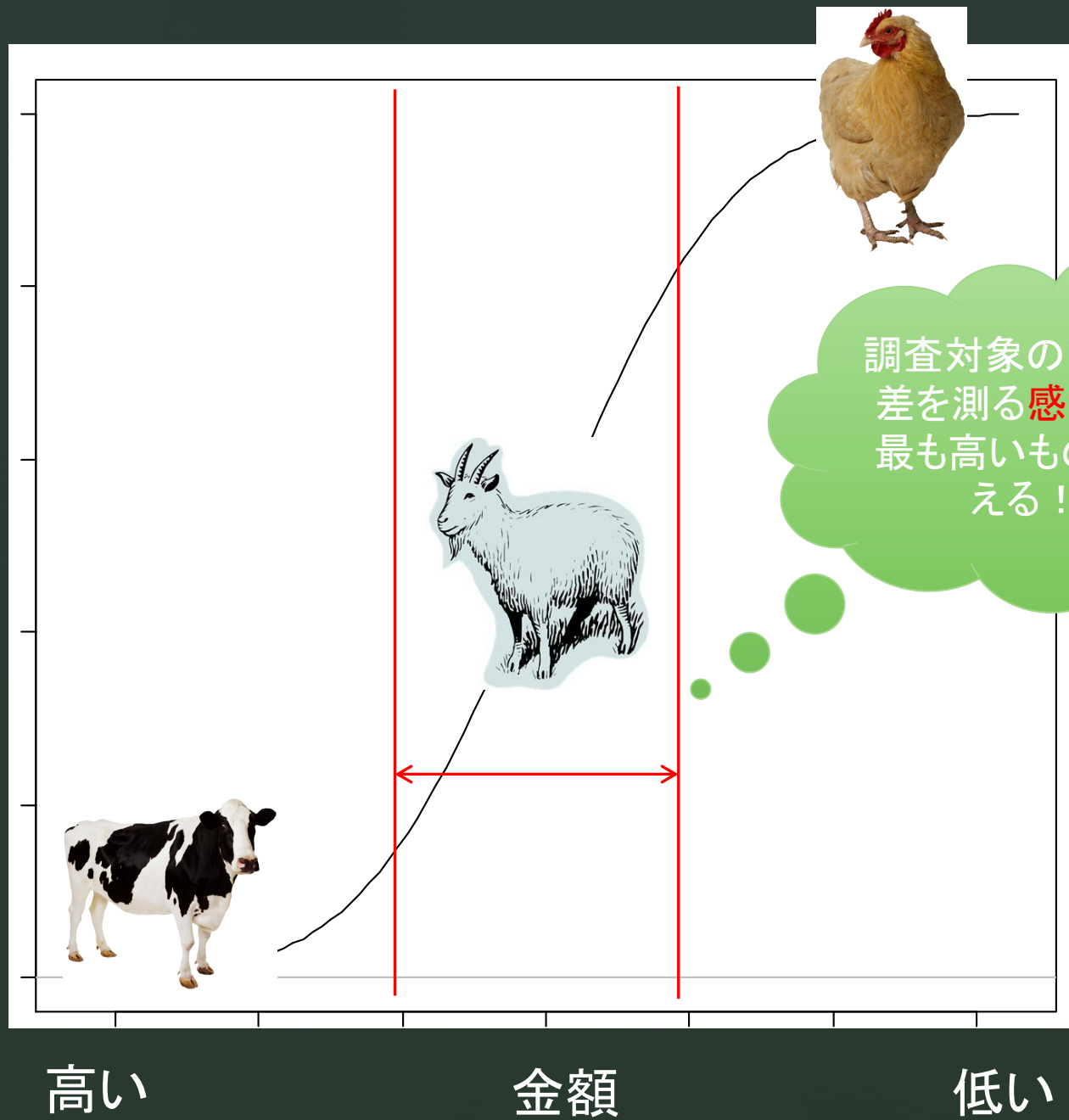
- 途上国の貧困農村の所得の向上度をどのように測るか？
- 持っている家畜の数で考える。
- どの家畜が適切か？



高い

持っている頻度

低い



調査対象の**変化**や  
差を測る**感度**が、  
最も高いものを考  
える！

高い

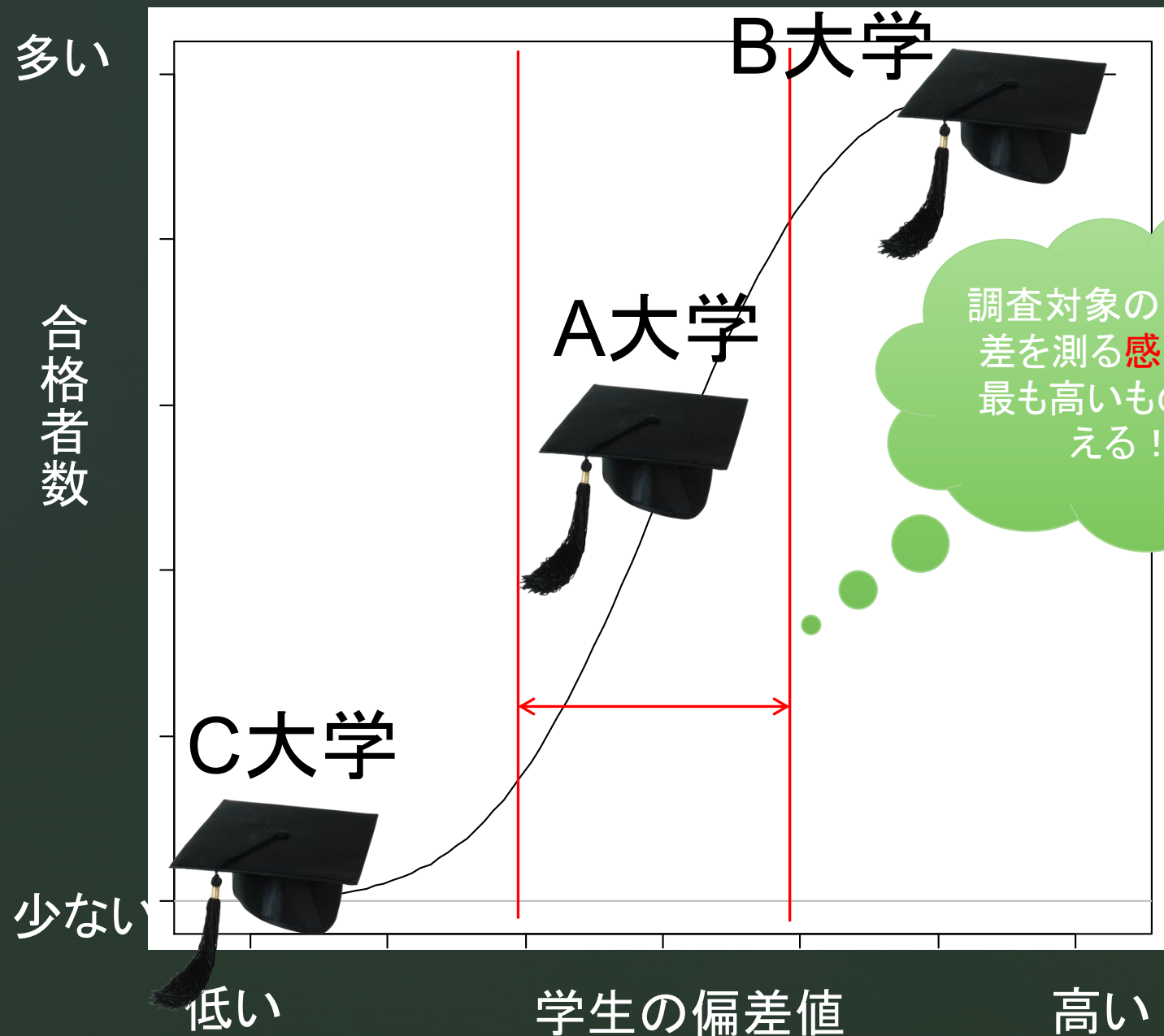
金額

低い

## 妥当性が問題となる例 (英語教育)

- 英語教育の効果を測定するため、英検2級を全学的に実施し、その取得者の数を指標とした。
- 偏差値50のA大学では、取得者の増加が見られ、指標としての機能を果たした。
- ところが、偏差値70のB大学では、全員が取得してしまい意味が無かった。
- また、偏差値30のC大学では、誰も受からずに意味が無かった。





## 以上の例は指標の感度の問題

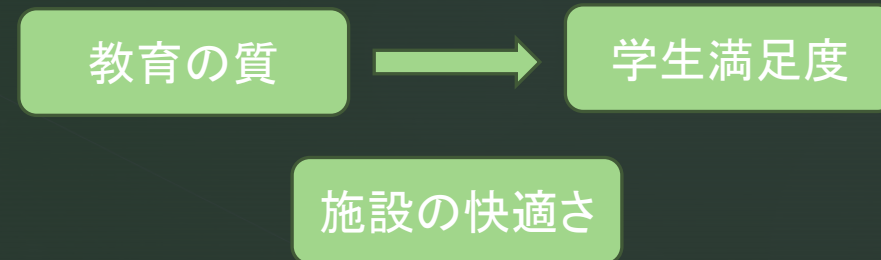
- 妥当性の第2段階というべきなのが評価指標の感度(または感応度)である。
- ロジックモデルの該当ブロックにおける事象の進捗を敏感に反映できるか否かという観点で、多くの場合評価指標の定義の巧拙や各種制約(たとえばデータの入手可能性や測定の難しさ)の程度によって決まる。
- この感度は「測定すべきものを測っているか」という妥当性の一部と考えられる。

## 妥当性研究の流れ

- E. L. Thorndike (1904)『An Introduction to the Theory of Mental and Social Measurements』。心理測定のみならず教育評価や幅広い社会科学の分野において、心理・社会現象を数量化する科学的アプローチを提示する中で、信頼性・妥当性の検討の必要性を示唆。教育測定運動の父
- 1920年代には、Binet, Spearmanなどが、「テストと基準との相関」として議論。測定が目的にあっているか。
- Campbell & Stanley (1963)は、社会科学における実験計画法において、広範な分野で因果推論の方法論を整理。
  - 「内的妥当性」(因果推論の正しさ: 観察された効果が本当に操作した要因によるものか)
  - 「外的妥当性」(一般化可能性: 結果が他の集団・状況に一般化できるか)

## 因果推論とは？（大学評価との関係で）

- 因果推論の目的
  - 「ある要因(X)が結果(Y)を引き起こすか」を検証する
- 反事実の考え方
  - 「XがなかったらYはどうなっていたか？」（統制群）
- 相関と因果の違い
  - 相関:一緒に変化する
  - 因果:XがYを変えるメカニズム
- 大学評価における例
  - 指標:学生満足度
  - 問い:満足度は教育の質を反映しているか？
  - → 実は施設の快適さに依存しているなら因果関係は弱い



相関 ≠ 因果

# 大学評価の指標を例として 内的・外的妥当性を考えてみる

- 内的妥当性との関係
  - 「指標が本当にその大学の教育・研究等の質を反映しているか」という問い。
  - 例えば、学生満足度を教育の質の指標として測定する場合、満足度が教員の教育力や学習成果に結びついているかを検証する必要。誤った因果関係(例:満足度が施設の快適さに強く依存している)なら、内的妥当性が低くなる。
- 外的妥当性との関係
  - 「その指標が他の大学や状況に一般化できるか」という問い。
  - 例えば、ある大学で導入した学習成果を測る指標が、他大学でも同様に機能するか。分野差・地域差・規模差など。
- 大学評価特有の難しさ
  - 実験研究と違い、大学評価は統制条件が取れないため、因果推論が難しく、内的妥当性を高めるには多面的な指標や質的な評価が必要。
  - 外的妥当性は、国際比較や異なる制度環境での適用可能性を考えるとさらに複雑。



## その後の妥当性に 関連した議論の展開

- 政策評価、心理測定、医学等の幅広い分野において、Campbell & Stanley (1963)は引用され、分野ごとの特性を踏まえた妥当性の文脈を考慮しつつ、妥当性概念の検討と細分化がなされてきた。
- 以下では、妥当性概念が整理された3つの例を提示
  - 業績測定型評価における妥当性
    - 田中(2014)は英語圏で幅広く参照される代表的な文献であるHatry(1999, 2006)及びAmmons(1995)における整理を統合して提示。
    - 米国行政学会(American Society for Public Administration; ASPA)が2000年に作成した業績測定型評価のマニュアルも広く利用されている。
  - 心理測定における妥当性
    - 測定されるデータが、その目的にどれだけかなっているか、特にその概念的な適切さの程度をどの程度満たしているか、についての6つの側面(Messick, 1995)

表1 評価指標が満たすべき条件一例1

区分		基 準	説 明
個別指標の基準	1-1妥当性	Validity	計測すべき事象を計測
	1-2信頼性	Reliability	正確に計測
	1-3理解可能性	Understandability	意味が明確で誤解しにくい
	1-4タイムリー性	Timeliness	有用なタイミングで入手可能
	1-5目的との適合性	Relevance to the objectives	目的や成果を適切に反映
	1-6施策の影響の大きさ	Program influence	計測事象への施策の影響度
	1-7計測可能性	Feasibility of collecting data	データ収集が可能
	1-8データの収集費用	Cost of collecting data	費用の大きさ
	1-9操作可能性	Manipulability	計測値の操作可能性が小さい
	1-10意思決定への有用性	Usefulness for decision-making	意思決定に有益な知見を提供
指標群の基準	2-1包括性	Comprehensiveness	重要な側面を漏らさずカバー
	2-2非重複性	Nonredundancy	重複なく異なる側面を計測
	2-3データの収集費用	Cost of collecting data	総費用の大きさ
	2-4反抗的行動への耐性	Resistance to perverse behavior	意図に反する行動を誘発せず

田中（2014）をもとに筆者が作成。

表 2 評価指標が満たすべき条件一例 2

基準	説明
1 Results oriented	アウトカムに焦点
2 Relevant	目的との関係が論理的かつ直接
3 Responsive	パフォーマンス水準の変化を反映
4 Valid	把握すべき情報を把握
5 Reliable	正確でぶれない情報
6 Cost-effective	データ収集費用が過大でない
7 Useful	意思決定者に有益な情報を提供
8 Accessible	定期的に情報が得られる
9 Comparable	時系列比較が可能
10 Compatible	既存の財務・業務システムに適合
11 Clear	様々な立場の人が理解できる
12 Affordable	予算内で運用できる

ASPA（2000）による（筆者訳）。

表3 心理測定における妥当性の側面

側面		説明
内容の側面	Content Aspect	内容が，測りたい内容と関係があるか
実体の側面	Substantive Aspect	反応や回答のプロセスが，理論に合致しているか
構造の側面	Structural Aspect	尺度得点の構造が理論的なものに合致しているか
一般化可能性の側面	Generalizability Aspect	内容を，測りたいことに一般化できるか
外的側面	External Aspect	他の変数との収束的または弁別的な関係
結果の側面	Consequential Aspect	測定結果の解釈が，どのような行動や影響，結果に結びつくか

Messick（1995）をもとに作成。

# 本日の内容

今日のお話と関連する内容が、  
以下の論文にまとめられています。  
ぜひご覧ください。

渋井進, 浅野茂, 橋本貴充, 小野宏, 小野達也, 山崎その, 田中弥生 (2017).「自己評価力向上支援のための評価指標設定に関するチェックリストの開発」,  
大学評価・学位研究 18, 21-36.

- チェックリストの背景にある考え方
  - 妥当性とは？
- チェックリストの説明
  - 妥当性のチェックリスト項目とその事例
  - 実用面のチェックリスト項目とその事例
- チェックリスト使用上の留意点

## ※ 重要なお知らせ

本発表は、信頼できると判断した情報源に基づいて作成されていますが、発表内容を元に実践したことにより生じる直接的または間接的な損失、損害、不利益等に関して、渋井は一切の責任を負いません。最終的な決定は、ご自身の判断と責任において行われますようお願い申し上げます。



区分	基準	説明
妥当性	目的との適合性	指標が、計画の進捗や目指す成果とその進捗を適切に反映しているか。
	調査対象・結果への影響	指標設定の結果、意図しない悪影響を及ぼすものではないか。
	信頼性	誰がいつ測定しても、同じ事象や状態からは同じ測定結果が得られるか。
	理解可能性	指標の意味が、明確でわかりやすく、誤解が生じないか。
	包括性・非重複性	計画の重要な側面が、もらさず指標によってカバーされているか。
		指標間に重複がなく、各指標は異なる側面を計測しているか。
実用面	意思決定者への有用性	指標が、執行部等の意思決定者に対して、有益な知見を提供してくれているか。
	計測可能性	指標となるデータは収集可能か。
	収集の適時性	有用なタイミングで、指標の計測値は入手可能か。
	データ収集のコスト	データを収集するための費用は大きすぎないか。
	操作可能性	指標の計測値は、都合良く操作して変更可能なものではないか。

区分	基準	説明
妥当性	目的との適合性	指標が、計画の進捗や目指す成果とその進捗を適切に反映しているか。
	調査対象・結果への影響	指標設定の結果、意図しない悪影響を及ぼすものではないか。
	信頼性	誰がいつ測定しても、同じ事象や状態からは同じ測定結果が得られるか。
	理解可能性	指標の意味が、明確でわかりやすく、誤解が生じないか。
	包括性・非重複性	計画の重要な側面が、もらさず指標によってカバーされているか。
		指標間に重複がなく、各指標は異なる側面を計測しているか。
実用面	意思決定者への有用性	指標が、執行部等の意思決定者に対して、有益な知見を提供してくれているか。
	計測可能性	指標となるデータは収集可能か。
	収集の適時性	有用なタイミングで、指標の計測値は入手可能か。
	データ収集のコスト	データを収集するための費用は大きすぎないか。
	操作可能性	指標の計測値は、都合良く操作して変更可能なものではないか。

## 目的との適合性、包括性

- 指標が、計画の進捗や**目指す成果**を適切に**反映している**か。
- 指標が**偏ってはいない**か、必要なものは**網羅されている**か？
- 現場で起きていることと、指標の間に乖離が無いのか？

中期目標	中期計画	具体的な取組事項	指標
社会性・倫理観を涵養するボランティア活動を推進する	社会に学ぶ体験教育の強化など、「ボランティア支援センター」を中心とした学生支援体制を拡充する	①ボランティア活動に対する支援基盤の整備 ②ボランティア支援センター運営の充実 ③災害時における支援体制の整備	・ボランティア活動参加率
			・ボランティア関係の授業の受講者数
			・ボランティアセンターの利用者数
			・ボランティア関係の研究の外部資金獲得額
			・インターンシップ・体験学習等の参加者割合
			・インターンシップ・体験学習等の参加学生満足度
			・卒業時アンケート

→現場に近い教員、職員らでチェックする体制  
を作ることによって解決。

区分	基準	説明
妥当性	目的との適合性	指標が、計画の進捗や目指す成果とその進捗を適切に反映しているか。
	調査対象・結果への影響	指標設定の結果、意図しない悪影響を及ぼすものではないか。
	信頼性	誰がいつ測定しても、同じ事象や状態からは同じ測定結果が得られるか。
	理解可能性	指標の意味が、明確でわかりやすく、誤解が生じないか。
	包括性・非重複性	計画の重要な側面が、もらさず指標によってカバーされているか。
		指標間に重複がなく、各指標は異なる側面を計測しているか。
実用面	意思決定者への有用性	指標が、執行部等の意思決定者に対して、有益な知見を提供してくれているか。
	計測可能性	指標となるデータは収集可能か。
	収集の適時性	有用なタイミングで、指標の計測値は入手可能か。
	データ収集のコスト	データを収集するための費用は大きすぎないか。
	操作可能性	指標の計測値は、都合良く操作して変更可能なものではないか。

## 非重複性

- 指標間に**重複がなく、各指標は異なる側面を計測しているか？**
  - 例 「リーダーシップ力」を測る指標をいくつかの学生調査のアンケート項目をもとに作成してみた。ところが、他に作っている「プレゼンテーション力」、「説得力」を示す項目と**同じような項目が並び、その値もほぼ同じ様になっている。**
  - 解決法 構成概念として、他の指標と区別が可能かどうか、ということをチェックすべき。指標は**省力化のため、絞り込む。**



## 他の指標との相関は？

	リーダシッ プカ	プレゼンカ	説得力
Aさん	5	5	5
Bさん	3	3	3
Cさん	2	2	2
Dさん	4	4	3

リーダーシッ  
プカ

プレゼンカ

説得力

それぞれの指標は区別できるのか？

プレゼンカ

リー  
プカ

説得力

区分	基準	説明
妥当性	目的との適合性	指標が、計画の進捗や目指す成果とその進捗を適切に反映しているか。
	調査対象・結果への影響	指標設定の結果、意図しない悪影響を及ぼすものではないか。
	信頼性	誰がいつ測定しても、同じ事象や状態からは同じ測定結果が得られるか。
	理解可能性	指標の意味が、明確でわかりやすく、誤解が生じないか。
	包括性・非重複性	計画の重要な側面が、もらさず指標によってカバーされているか。
		指標間に重複がなく、各指標は異なる側面を計測しているか。
実用面	意思決定者への有用性	指標が、執行部等の意思決定者に対して、有益な知見を提供してくれているか。
	計測可能性	指標となるデータは収集可能か。
	収集の適時性	有用なタイミングで、指標の計測値は入手可能か。
	データ収集のコスト	データを収集するための費用は大きすぎないか。
	操作可能性	指標の計測値は、都合良く操作して変更可能なものではないか。

## 信頼性



- 誰がいつ測定しても、**同じ事象や状態からは同じ測定結果**が得られるか。
- 同一のデータを収集した場合に、同様の値が得られるかどうかという**安定性、一貫性**があるか
- 測定の**精度**について十分であるか

## 信頼性(つづき)

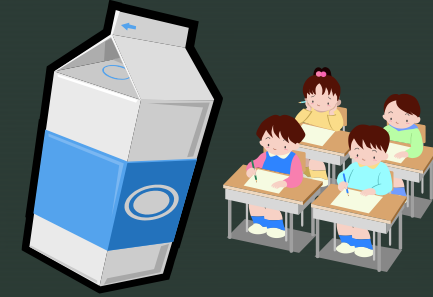
- 部局や担当部署に同一のデータを再度要求しても、収集時に**ばらつき**が生じたり、大きく異なるようなことのない測定法で得られているか。
- ばらつきを減らすため
  - 指標が抽象的な場合には注意！
    - データの定義や、特定の考え方や判断に依存してはいないだろうか。
  - データの入力ミス等も気をつけよう！
    - チェックする体制はとられているだろうか。

## 信頼性が問題となる例

目的番号	目的	評価質問	指標	データ入手手段
2-2	戦略がある	戦略があるか？	共同・受託研究獲得のための 方策の有無	共同・受託研究獲得のため の方策
4-1	アンケート、ヒアリング、などは、ちゃんと信頼できるように定められたものだろうか？			地域連携に係る方針、規定等；教員数；地域連携活動参加教員一覧
4-1-1	ある	あるか？	共同・受託研究の周知状況；共同・受託研究に係る説明会の開催件数、出席率	共同・受託研究の周知のチラシやホームページ；共同・受託研究に係る説明会の開催一覧、アンケート、ヒアリング結果
4-1-1-1	何をもって、どこまでを、関連する「方策」「活動」とするか、皆の合意は取れているか？			アンケート；教員ごとの活動報告書
	になる	になるか？	教員ごとの研究業績リストに占める地域連携課題の比率	教員ごとの活動報告書



区分	基準	説明
妥当性	目的との適合性	指標が、計画の進捗や目指す成果とその進捗を適切に反映しているか。
	調査対象・結果への影響	指標設定の結果、意図しない悪影響を及ぼすものではないか。
	信頼性	誰がいつ測定しても、同じ事象や状態からは同じ測定結果が得られるか。
	理解可能性	指標の意味が、明確でわかりやすく、誤解が生じないか。
	包括性・非重複性	計画の重要な側面が、もらさず指標によってカバーされているか。
		指標間に重複がなく、各指標は異なる側面を計測しているか。
実用面	意思決定者への有用性	指標が、執行部等の意思決定者に対して、有益な知見を提供してくれているか。
	計測可能性	指標となるデータは収集可能か。
	収集の適時性	有用なタイミングで、指標の計測値は入手可能か。
	データ収集のコスト	データを収集するための費用は大きすぎないか。
	操作可能性	指標の計測値は、都合良く操作して変更可能なものではないか。



## 調査対象・結果への影響

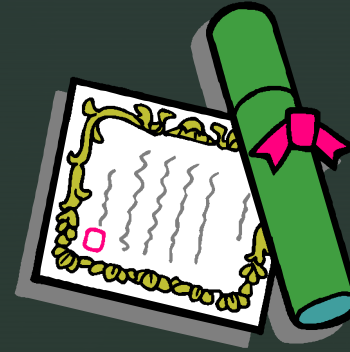
- **評価の意図に反する行動を誘発**するものではないだろうか。
- 例 学生のリサイクルに対する協力度を測定するため、牛乳パックを持参した回数を指標として計測を行った。ところが、牛乳パックを持参するためにパック入りの牛乳を頑張って飲む行動が促進されてしまった。
  - 「園児が牛乳パックをたくさん入れたいため、牛乳を多く飲むようになった」 環境省『分別収集について』参考資料集、環境省、2013年。
- 例 ある都市では、学生の学力向上のため、学校ごとに学力テストの得点を測定することで、教育の効果の指標とすることとした。ところが、一部の学力の低い学生を欠席させる事態が生じた。
  - 「1961年から行われた全国一斉学力テストでは、テストの結果が全国の公立小中学校の教員の勤務評定に結び付けられるようになりました。つまり、子どもたちに高い点を取らせる教員は、人事その他の待遇において優遇されたのです。その結果、次のようなことが実際に行われるようになりました・・・」（東洋経済オンライン, 2018.08.15, 『大阪市長「学力テスト発言」が危険である根拠:日本は過去何度も同じ失敗を繰り返してきた』）

→指標を設定した結果として目的が置き換わってしまわないよう注意！

- **注** 牛乳パックのリサイクルや学力テストが全て失敗していると私が主張するわけではありません。あくまで、1つの例。

区分	基準	説明
妥当性	目的との適合性	指標が、計画の進捗や目指す成果とその進捗を適切に反映しているか。
	調査対象・結果への影響	指標設定の結果、意図しない悪影響を及ぼすものではないか。
	信頼性	誰がいつ測定しても、同じ事象や状態からは同じ測定結果が得られるか。
	理解可能性	指標の意味が、明確でわかりやすく、誤解が生じないか。
	包括性・非重複性	計画の重要な側面が、もらさず指標によってカバーされているか。
		指標間に重複がなく、各指標は異なる側面を計測しているか。
実用面	意思決定者への有用性	指標が、執行部等の意思決定者に対して、有益な知見を提供してくれているか。
	計測可能性	指標となるデータは収集可能か。
	収集の適時性	有用なタイミングで、指標の計測値は入手可能か。
	データ収集のコスト	データを収集するための費用は大きすぎないか。
	操作可能性	指標の計測値は、都合良く操作して変更可能なものではないか。

## 理解可能性



- 指標の意味が、**明確でわかりやすいか、皆に誤解が生じないか。**
  - 例 ある大学(仮想)では若手研究者支援のために、今年書いた論文数によって、研究費を配分する計画を立てた。
  - ところが、様々な分野の研究者から、A学部では査読付き一流雑誌、B学部では査読無し紀要、C学部では学会抄録、異動前のお機関での業績など、様々なものが出てきて、これらを一緒にして合計し、論文数とすることに、不満の声が挙った。
- 指標は誤解のないように、定義を明確に！

# 本日の内容

今日のお話と関連する内容が、  
以下の論文にまとめられています。  
ぜひご覧ください。

渋井進, 浅野茂, 橋本貴充, 小野宏, 小野達也, 山崎その, 田中弥生 (2017).「自己評価力向上支援のための評価指標設定に関するチェックリストの開発」,  
大学評価・学位研究 18, 21-36.

- チェックリストの背景にある考え方
  - 妥当性とは？
- チェックリストの説明
  - 妥当性のチェックリスト項目とその事例
  - 実用面のチェックリスト項目とその事例
- チェックリスト使用上の留意点

## ※ 重要なお知らせ

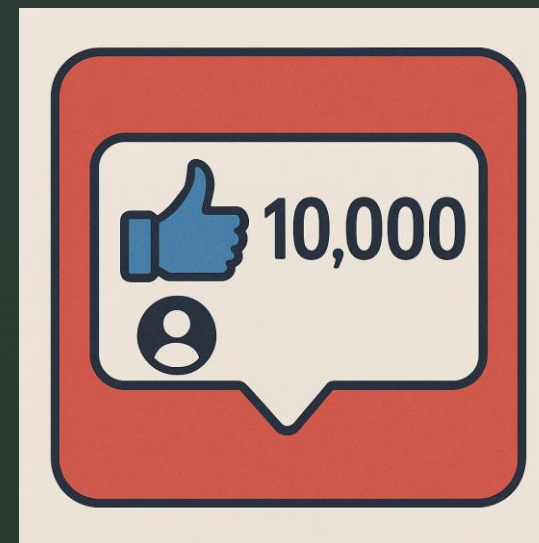
本発表は、信頼できると判断した情報源に基づいて作成されていますが、発表内容を元に実践したことにより生じる直接的または間接的な損失、損害、不利益等に関して、渋井は一切の責任を負いません。最終的な決定は、ご自身の判断と責任において行われますようお願い申し上げます。

区分	基準	説明
妥当性	目的との適合性	指標が、計画の進捗や目指す成果とその進捗を適切に反映しているか。
	調査対象・結果への影響	指標設定の結果、意図しない悪影響を及ぼすものではないか。
	信頼性	誰がいつ測定しても、同じ事象や状態からは同じ測定結果が得られるか。
	理解可能性	指標の意味が、明確でわかりやすく、誤解が生じないか。
	包括性・非重複性	計画の重要な側面が、もらさず指標によってカバーされているか。
		指標間に重複がなく、各指標は異なる側面を計測しているか。
実用面	意思決定者への有用性	指標が、執行部等の意思決定者に対して、有益な知見を提供してくれているか。
	計測可能性	指標となるデータは収集可能か。
	収集の適時性	有用なタイミングで、指標の計測値は入手可能か。
	データ収集のコスト	データを収集するための費用は大きすぎないか。
	操作可能性	指標の計測値は、都合良く操作して変更可能なものではないか。



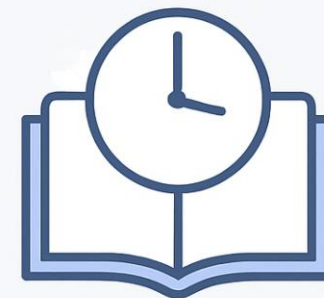
## 意思決定者への有用性

- 指標が、執行部等の意思決定者に対して、有益な知見を提供しているか。
- 理論的に導かれたとしても、担当理事・センター長等の責任者の意思決定に利用可能な指標でないと、実用面では意味を持たないことがある。
- 本観点は評価時のエビデンスだけではなく、実質的な改善との関連で重要。
- 例: ある大学では広報活動の成果の指標として「大学公式SNSのフォロワー数」を指標として測定した。
  - フォロワー数が増えたことは成果の一部と言えるが、今後の広報戦略や志願者数の増加へ向けた方策などの、意思決定に結びつけることは困難。



区分	基準	説明
妥当性	目的との適合性	指標が、計画の進捗や目指す成果とその進捗を適切に反映しているか。
	調査対象・結果への影響	指標設定の結果、意図しない悪影響を及ぼすものではないか。
	信頼性	誰がいつ測定しても、同じ事象や状態からは同じ測定結果が得られるか。
	理解可能性	指標の意味が、明確でわかりやすく、誤解が生じないか。
	包括性・非重複性	計画の重要な側面が、もらさず指標によってカバーされているか。
		指標間に重複がなく、各指標は異なる側面を計測しているか。
実用面	意思決定者への有用性	指標が、執行部等の意思決定者に対して、有益な知見を提供してくれているか。
	計測可能性	指標となるデータは収集可能か。
	収集の適時性	有用なタイミングで、指標の計測値は入手可能か。
	データ収集のコスト	データを収集するための費用は大きすぎないか。
	操作可能性	指標の計測値は、都合良く操作して変更可能なものではないか。

## 計測可能性



- 指標となるデータは収集可能か。
- 理想的には計測したい指標として想定されるものでも、倫理的な側面や現実的に収集不能なものである指標だと、実用的には問題がある。
- 例: 学生を大事にする教育力を重視した大学を示すための指標として、「教員の1コマあたりの授業準備時間」を設定した。
- 自己申告で信頼性が低い(とりあえず多く答えておこう)。可視化されることで、労働条件に関する倫理的問題も浮上の可能性。

区分	基準	説明
妥当性	目的との適合性	指標が、計画の進捗や目指す成果とその進捗を適切に反映しているか。
	調査対象・結果への影響	指標設定の結果、意図しない悪影響を及ぼすものではないか。
	信頼性	誰がいつ測定しても、同じ事象や状態からは同じ測定結果が得られるか。
	理解可能性	指標の意味が、明確でわかりやすく、誤解が生じないか。
	包括性・非重複性	計画の重要な側面が、もらさず指標によってカバーされているか。
		指標間に重複がなく、各指標は異なる側面を計測しているか。
実用面	意思決定者への有用性	指標が、執行部等の意思決定者に対して、有益な知見を提供してくれているか。
	計測可能性	指標となるデータは収集可能か。
	収集の適時性	有用なタイミングで、指標の計測値は入手可能か。
	データ収集のコスト	データを収集するための費用は大きすぎないか。
	操作可能性	指標の計測値は、都合良く操作して変更可能なものではないか。

## 収集の適時性

- 有用なタイミングで、指標の計測値は入手可能か。
- 法人評価や認証評価等の外部の評価時のエビデンスや、補助金申請時の提出データ等、それぞれの必要な時期に入手可能なものであるかは重要。必要な時に入手できない、またはズレがあるデータでは実用性がない。
- 例 授業評価アンケート: 全15回の授業の13回目の終わり頃にアンケートを行なっても、今年の受講者に対して改善はできない。
- データ収集の日常的な体制整備と、調査等によって収集する場合には、その利用時期を考えて事前にスケジュールを立てる必要がある。



区分	基準	説明
妥当性	目的との適合性	指標が、計画の進捗や目指す成果とその進捗を適切に反映しているか。
	調査対象・結果への影響	指標設定の結果、意図しない悪影響を及ぼすものではないか。
	信頼性	誰がいつ測定しても、同じ事象や状態からは同じ測定結果が得られるか。
	理解可能性	指標の意味が、明確でわかりやすく、誤解が生じないか。
	包括性・非重複性	計画の重要な側面が、もらさず指標によってカバーされているか。
		指標間に重複がなく、各指標は異なる側面を計測しているか。
実用面	意思決定者への有用性	指標が、執行部等の意思決定者に対して、有益な知見を提供してくれているか。
	計測可能性	指標となるデータは収集可能か。
	収集の適時性	有用なタイミングで、指標の計測値は入手可能か。
	データ収集のコスト	データを収集するための費用は大きすぎないか。
	操作可能性	指標の計測値は、都合良く操作して変更可能なものではないか。



## データ収集のコスト

- データを収集するための費用は大きすぎないか
- 例: 卒業生、在校生、企業等へのアンケート調査等によるプログラムの効果や学習成果の調査は良く行われているが、大学側や調査対象者の負担を考慮すると、精査して行う必要がある→ 評価疲れの防止へ
- 例: 「授業中の学生の集中度」を客観的に測定するため、全学でカメラを導入して教室に設置し、AIでリアルタイムに覚醒度や表情を分析することとした。
- ⇨ 機器導入費用が高額で、維持費もかかる。「計測可能性」の倫理面にも課題が。
- 現状では、まだ早すぎるかもしれない。

区分	基準	説明
妥当性	目的との適合性	指標が、計画の進捗や目指す成果とその進捗を適切に反映しているか。
	調査対象・結果への影響	指標設定の結果、意図しない悪影響を及ぼすものではないか。
	信頼性	誰がいつ測定しても、同じ事象や状態からは同じ測定結果が得られるか。
	理解可能性	指標の意味が、明確でわかりやすく、誤解が生じないか。
	包括性・非重複性	計画の重要な側面が、もらさず指標によってカバーされているか。
		指標間に重複がなく、各指標は異なる側面を計測しているか。
実用面	意思決定者への有用性	指標が、執行部等の意思決定者に対して、有益な知見を提供してくれているか。
	計測可能性	指標となるデータは収集可能か。
	収集の適時性	有用なタイミングで、指標の計測値は入手可能か。
	データ収集のコスト	データを収集するための費用は大きすぎないか。
	操作可能性	指標の計測値は、都合良く操作して変更可能なものではないか。

## 操作可能性

- 指標の計測値は、都合良く操作して変更可能なものではないか。
- 例：全学的に各部局に調査を行った場合、解釈により都合の良いような数値へと操作することで有利な結論を導きだすことが可能な状況であると、正確な把握ができないという問題が生じることとなり、公平性や正確さの確保という点から配慮すべき。
- 例：いくつかの指標における操作可能性
  - 授業満足度：教員が課題を減らし、成績を甘くして操作
  - 交流協定数：名義だけの交流で数値を稼ぐ
  - 学生出席率：出席確認を緩くして数値を上げる
  - 研究発表件数：内容の質を問わず発表数だけ増やす
  - FD参加率：内容を問わず、何らかのイベントを教員会議の前後に設定

竹橋大学で用いられているいくつかの指標は、以下のチェックリストの観点から問題があると気づくのではないだろうか

- 入試倍率(学部一括)
- 就職率(希望者分母)
- 卒業時アンケート(回収率)
- 授業アンケート(本人通知のみ)
- 自己点検評価(○付けのみ)

区分	基準	説明
妥当性	目的との適合性	指標が、計画の進捗や目指す成果とその進捗を適切に反映しているか。
	調査対象・結果への影響	指標設定の結果、意図しない悪影響を及ぼすものではないか。
	信頼性	誰がいつ測定しても、同じ事象や状態からは同じ測定結果が得られるか。
	理解可能性	指標の意味が、明確でわかりやすく、誤解が生じないか。
	包括性・非重複性	計画の重要な側面が、もらさず指標によってカバーされているか。 指標間に重複がなく、各指標は異なる側面を計測しているか。
実用面	意思決定者への有用性	指標が、執行部等の意思決定者に対して、有益な知見を提供してくれているか。
	計測可能性	指標となるデータは収集可能か。
	収集の適時性	有用なタイミングで、指標の計測値は入手可能か。
	データ収集のコスト	データを収集するための費用は大きすぎないか。
	操作可能性	指標の計測値は、都合良く操作して変更可能なものではないか。

# 本日の内容

今日のお話と関連する内容が、  
以下の論文にまとめられています。  
ぜひご覧ください。

渋井進, 浅野茂, 橋本貴充, 小野宏, 小野達也, 山崎その, 田中弥生 (2017).「自己評価力向上支援のための評価指標設定に関するチェックリストの開発」,  
大学評価・学位研究 18, 21-36.

- チェックリストの背景にある考え方
  - 妥当性とは？
- チェックリストの説明
  - 妥当性のチェックリスト項目とその事例
  - 実用面のチェックリスト項目とその事例
- チェックリスト使用上の留意点

## ※ 重要なお知らせ

本発表は、信頼できると判断した情報源に基づいて作成されていますが、発表内容を元に実践したことにより生じる直接的または間接的な損失、損害、不利益等に関して、渋井は一切の責任を負いません。最終的な決定は、ご自身の判断と責任において行われますようお願い申し上げます。

## ■ チェックリスト活用の留意点

- チェックリストでひとつクリアすれば、次の項目が気になってくるだろう。
- まずは気になったところから活用しよう。最初から全てをカバーしようとしなない。
- チェック項目を用いて確認作業をする際、なぜ、それがOKなのか、NGなのか理由を考える。
- その際、本資料や文献を活用すると理解がより深まるだろう。



# 引用文献

- American Society for Public Administration (2000). Performance Measurement: Concepts and techniques. Washington. D.C.: ASPA Center for Accountability and Performance.
- Ammons, D.N. (1996). Accountability for Performance: Measurement and Monitoring in Local Government. Washington. D.C.: International City/County Management Association.
- Campbell, D.T. & Stanley, J. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), Handbook of research on teaching (pp. 171–246). Chicago, IL: Rand McNally.
- Hatry, H.P. (1999). Performance Measurement: Getting results. Washington, DC: Urban Institute Press.
- Hatry, H.P. (2006). Performance Measurement: Getting results (2nd ed.). Washington, DC: Urban Institute Press.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist, 50(9), 741-749.
- 田中啓 (2014). 自治体評価の戦略 —有効に機能させるための16の原則, 東洋経済新報社.
- Thorndike, E. L. (1904). An Introduction to the Theory of Mental and Social Measurements. New York, NY: Macmillan.