



日本テスト学会第23回大会
研究委員会企画シンポジウム
「学校教育の中のテスト」

これからの学校教育とテスト - 理論と実践のはざまから -

坂本 佑太朗

株式会社リクルートマネジメントソリューションズ
測定技術研究所

*本発表の内容は個人の見解であり、所属機関の見解を代表するものではありません

坂本 佑太朗

株式会社リクルートマネジメントソリューションズ 測定技術研究所

東北大学大学院教育学研究科博士課程修了 博士（教育学）

主な研究業績：

Sakamoto, Y., & Kumagai, R. (2025). A new method for detecting DTF using sum scores. The 90th Annual International Meeting of the Psychometric Society (IMPS).

坂本佑太朗・仲間大輔（2024）．ワーク・エンゲイジメントがパフォーマンスに与える影響——組織内のマネジメントによる調整効果に注目して—— 産業・組織心理学研究，37(2)，173-180.

坂本佑太朗（2024）.学力テストの結果解釈における下位領域の取り扱い —TIMSS2003中学2年生理科データを用いた実証的検証—日本テスト学会誌, 19(1), pp95-109.

坂本佑太朗・肖雨知・中山豊（2023）．プロブスト法を使用した多面観察評価の妥当性検証 日本心理学会第87回大会．

坂本 佑太朗・柴山直（2017）．学力テストの下位領域に関する多次元IRT分析 データ分析の理論と応用，6，pp.31-44．

内容

- 背景
- 関連学術研究の動向
 - 学力研究、測定技術研究
 - テストの活用
- まとめ



内容

- 背景
- 関連学術研究の動向
 - 学力研究、測定技術研究
 - テストの活用
- まとめ



「学校教育の中のテスト」を考えるために、「学校教育」に閉じない視点からそれを捉え直す

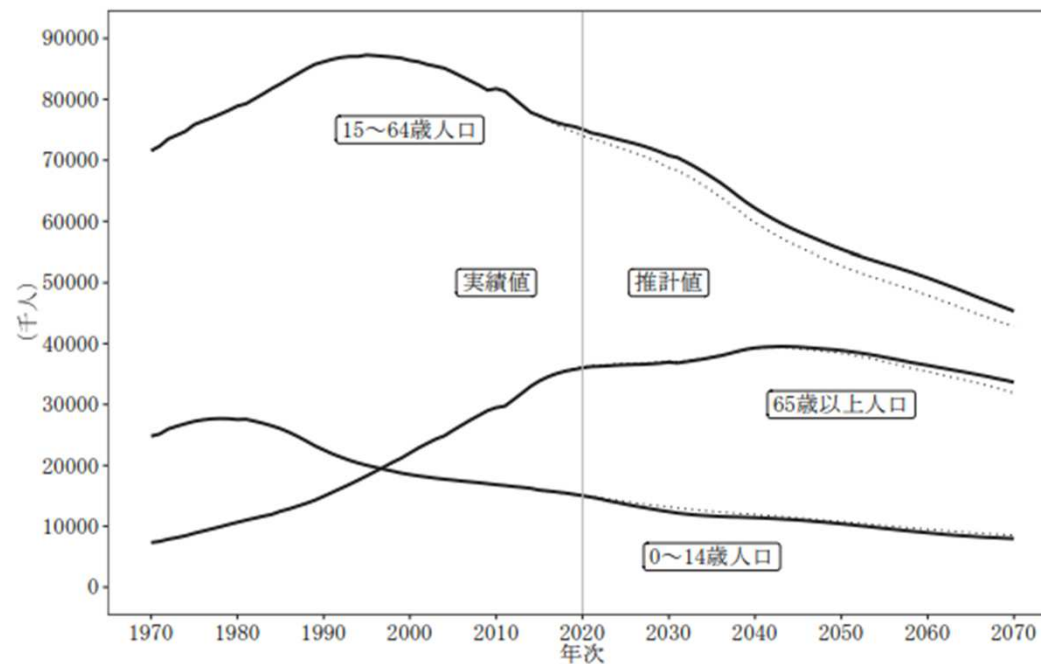
人材不足

リスキリング

非認知能力

人材不足

日本全体として子どもの減少、働く人の減少が著しい。



注) 破線は前回の出生中位推計。国立社会保障・人口問題研究所 (2023) 図 -2-1より抜粋

企業組織における管理職選抜の場面では、「選びたいが選ぶだけ人がいない」といった状況も「選抜」に加えて、目の前の人材「育成」が喫緊の課題に

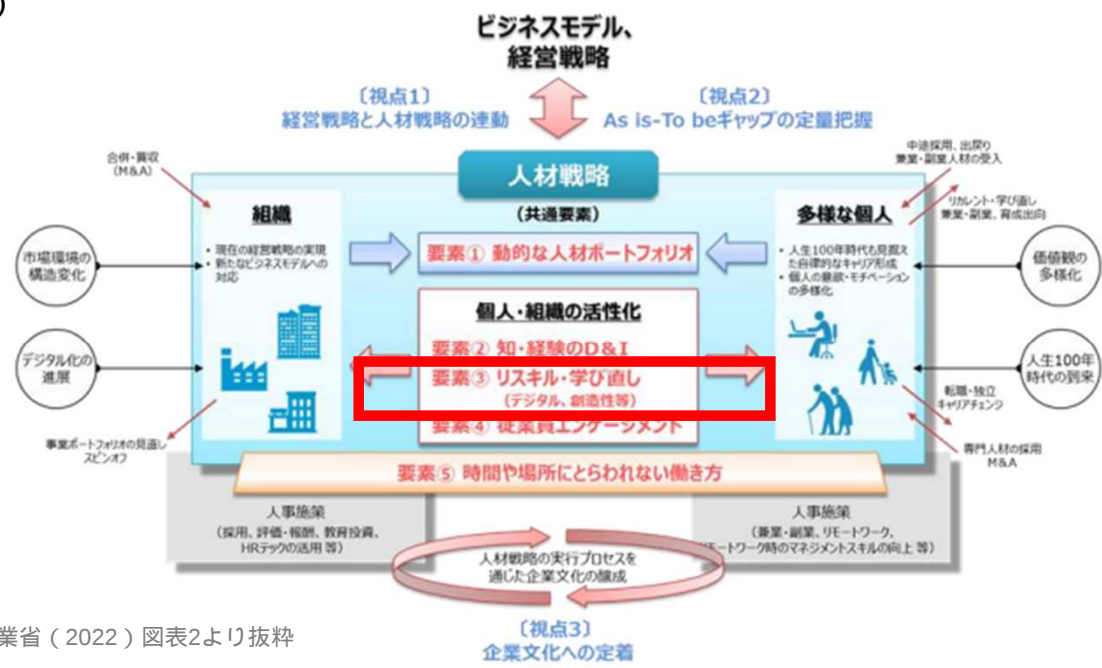
*引用文献

国立社会保障・人口問題研究所 (2023) . 日本の将来推計人口 - 令和5年推計 - .

リスキリング

激しい環境変化の中、DX時代の人材戦略として昨今「リスキリング」が叫ばれている。

*リスキリング：「新しい職業に就くために、あるいは、今の職業で必要とされるスキルの大幅な変化に適応するために、必要なスキルを獲得する / させること（石原, 2021）



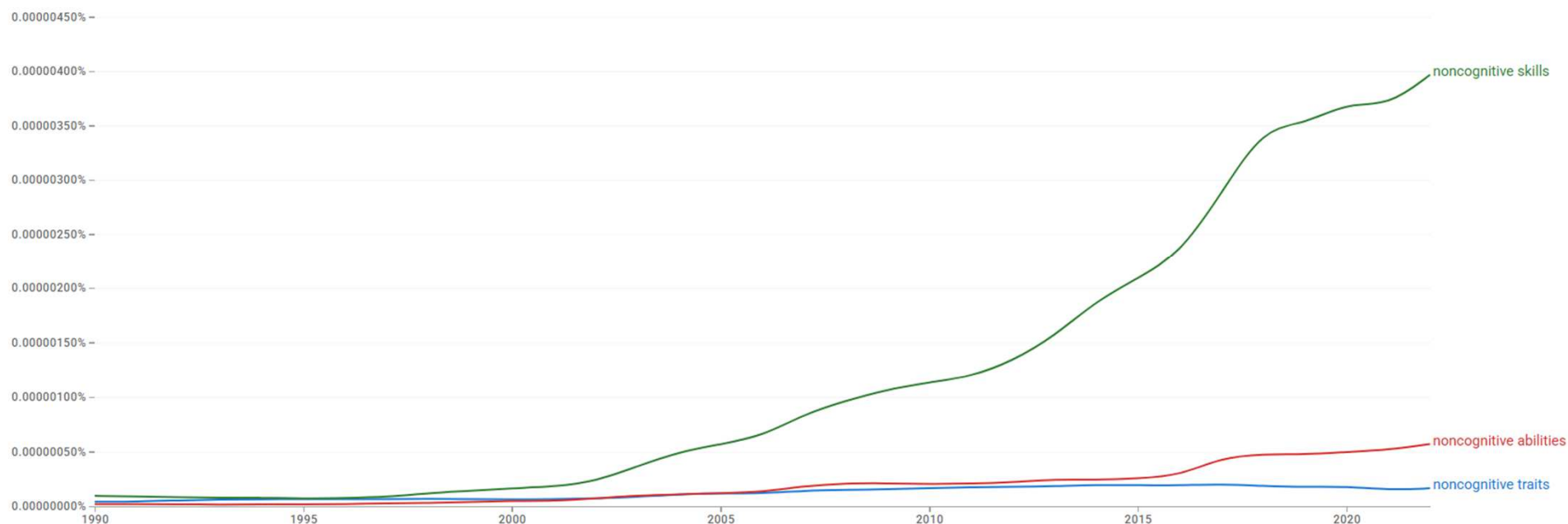
注) 経済産業省 (2022) 図表2より抜粋

「新しい時代の新しい能力」が労働者に要請される時代（勅使河原，2025）
その流れは今後学校教育へも影響してくるだろうと想定される

*引用文献
石原直子 (2021) . リスキリングとはーDX時代の人材戦略と世界の潮流ー第2回 デジタル時代の人材政策に関する検討会 資料2-2 .
経済産業省. (2022). 人的資本経営の実現に向けた検討会報告書 ~人材版伊藤レポート 2.0 ~ . https://www.meti.go.jp/policy/economy/jinteki_shihon/pdf/report2.0.pdf .
勅使河原真衣 (2025) . 学歴社会は誰のため PHP新書 .

非認知能力

学力を中心とした「認知能力」以外の「非認知能力」に注目が集まっている（小塩編, 2021, 2025）



注）小塩編（2021）図1同様の検索を、Google Books Ngram Viewerにて2025年8月現在に実施できる最新年（2022年）までで再検索した結果。

VUCA時代、パンデミックといった時代に非認知能力が着目（小塩編, 2025）

*引用文献

小塩真司（編）（2021）．非認知能力 - 概念・測定と教育の可能性 - 北大路書房

小塩真司（編）（2025）．非認知能力 - 生涯にわたる変化と影響 - 北大路書房

人口減少社会において、さまざまな「能力」の登場

その「能力」の測定・評価と人材育成の必要性
(e.g. 高度化・複雑化する測定対象への挑戦 (日本テスト学会大会実行委員会企画シンポジウム1, 2025))



そのような状況の中で、

私たちテスト関係者はどのように今後の「テスト」を考えていけばいいのか？

特に、学校教育の中のテストはどのようにあるべきなのだろうか？

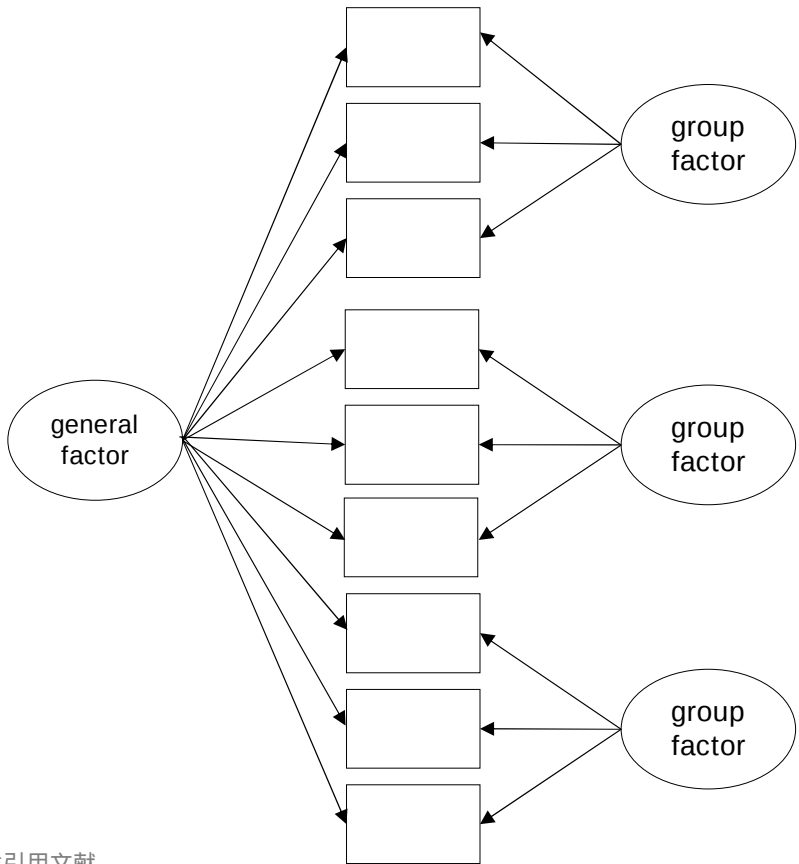
昨今の学力研究・測定技術研究を俯瞰しながら、ヒントを得ていく

内容

- 背景
- 関連学術研究の動向
 - 学力研究、測定技術研究
 - テストの活用
- まとめ



学力をどう捉えられるか？に向き合う研究例。
その視点で、学力テストの下位領域（subscale）に着目した実証研究（坂本, 2024；坂本・柴山, 2017）は、
学力を細分化するには限界があることを指摘（坂本, 2024）



• 坂本（2024）はわが国における
TIMSS2003中学校2年生理科データを用いて検証

- ✓ 下位領域特有の学力に対する規定要因からの影響は限定的
- ✓ 当該テスト結果の解釈では一次元の学力の存在を前提に，それに加えて下位領域ごとの特徴を参考程度に参照することが心理測定学的に支持される

*欧米では下位領域研究の積み重ねがあり（Haberman et al, 2024）、わが国においても引き続きの研究が期待される

*引用文献
Haberman, S., Sinharay, S., Feinberg, R. A., & Wainer, H. (2024). Subscores: A practical guide to their production and consumption . Cambridge University Press.
坂本佑太朗（2024）. 学力テストの結果解釈における下位領域の取り扱い —TIMSS2003中学2年生理科データを用いた実証的検証—日本テスト学会誌, 19(1), pp95-109.
坂本 佑太朗・柴山直（2017）. 学力テストの下位領域に関する多次元IRT分析 データ分析の理論と応用 , 6 , pp.31-44

IMPS2023ではGongjun Xu (University of Michigan) がEarly carrer Award、
IMPS2025ではMatthew Madison(University of Georgia)がInvited Speakerとなるなど、認知診断モデルの研究への着目
国内でも、日本テスト学会2023年大会でも「学力の診断・活用」セッションの設置

Gongjun Xu, University of Michigan
Identifiability of Cognitive Diagnosis Models
JULY 27, 2023 - 3:05PM

2023 Early Career Award

Cognitive Diagnosis Models (CDMs) are popular statistical tools for development in education, psychology, and other social and behavioral sciences. CDMs can be viewed as discrete latent variable models, where the model parameters are restricted by specified diagnostic assumptions. Though widely used, CDMs often suffer from models' discrete nature and complex restricted structure. This talk will introduce results on CDMs by considering both strict and partial identifiability of the model. Identifiability conditions only depend on the design Q-matrix and are easily practical guidelines for designing statistically valid diagnostic tests.

ABOUT THE SPEAKER



Dr. Gongjun Xu is an Associate Professor with a joint appointment in the Department of Statistics and the Department of Mathematics at the University of Michigan. He received his B.S. in Statistics and Technology of China in 2008, and his Ph.D. in Statistics from the University of Michigan in 2013. His research interests include multivariate variable models, cognitive diagnosis models, statistical learning and inference. Dr. Xu is the Chief Editor for Journal of Educational Measurement, Editor for Journal of American Statistical Association, Statistica Sinica, and Journal of the Royal Statistical Society. He has received the CAREER Award (2019), International Outstanding Young Researcher Award (2019), Bernoulli Society New Researcher Award (2020), Presidents of Statistical Societies (COPSS) Emerging Leader Award (2023), and Career Award (2023).

Matthew Madison, University of Georgia
Diagnostic Classification Models: Advancement Through Simplicity
JULY 16, 2025 - 4:45PM - 5:30PM

Diagnostic classification models (DCMs), also known as cognitive diagnosis models (CDMs), are psychometric models designed to classify respondents into interpretable categories. These categories, for example, can be interpreted as proficiency levels (e.g., educational assessment) or behavior diagnoses (e.g., psychological assessment). Despite their potential in many applications where classification is the primary objective, DCMs are not widely used in large scale assessment or research applications. In this talk, I conjecture some reasons why this may be the case. More specifically, I discuss some logistical, methodological, and interpretational challenges, and explore potential solutions to overcome some of these challenges.

ABOUT THE SPEAKER



Matthew Madison is Assistant Professor of Quantitative Methodology at the University of Georgia, where he studies and teaches courses in educational measurement and statistics. His research centers around the advancement of multivariate psychometric models. More specifically, Dr. Madison has focused his research on a class of contemporary item response models called diagnostic classification models (DCMs) and their application in educational contexts. Motivated by his background in mathematics education, his research seeks to use DCMs to improve the utility of educational assessments. Recently, he has investigated longitudinal DCMs as alternative to traditional methods for modeling learning over time. His research has been supported by Institute of Education Sciences and National Science Foundation, and he has published research in journal such as Psychometrika, Journal of Educational Measurement, and Journal of Educational and Behavioral Statistics.

一般研究発表セッション1：学力の診断・活用

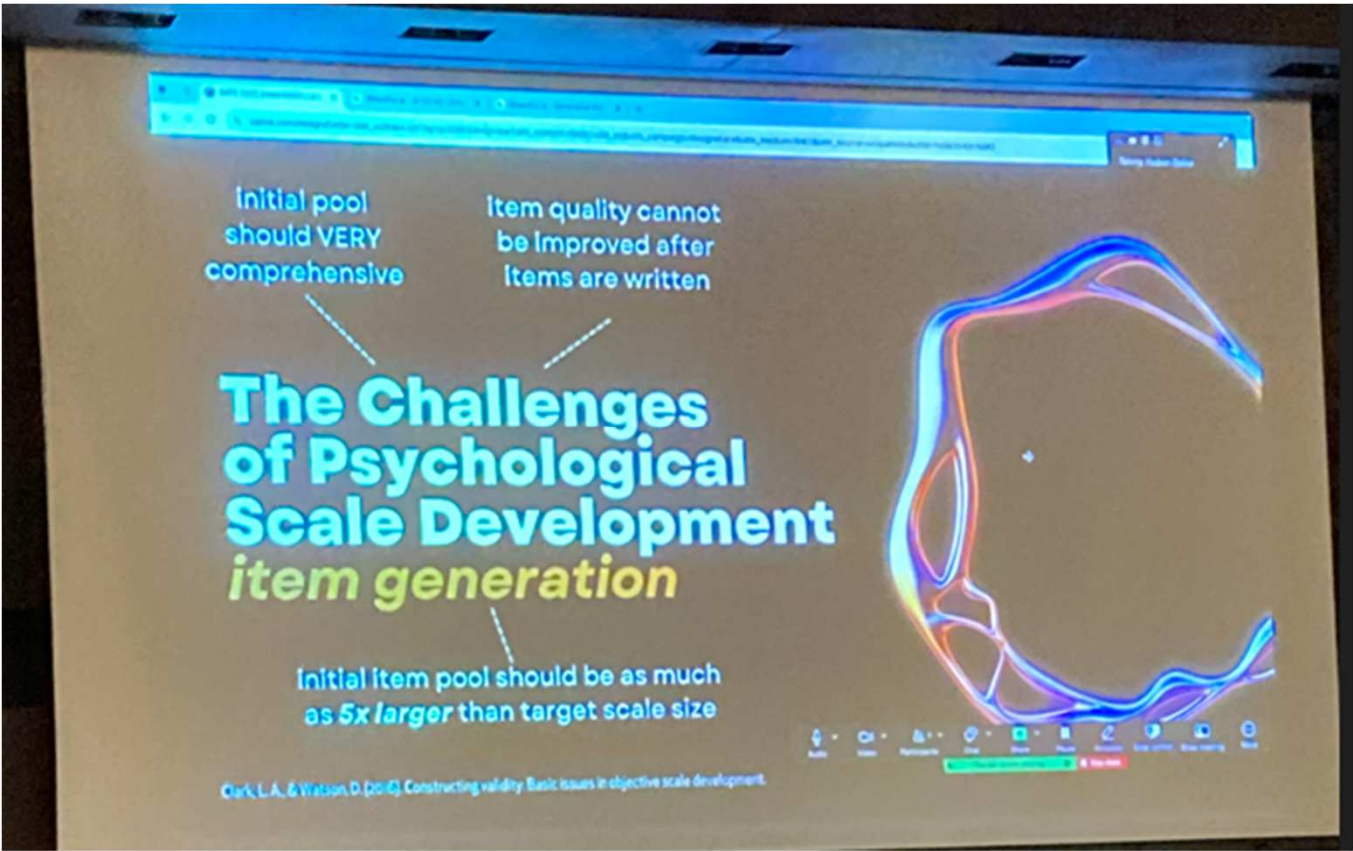
座長 光永 悠彦（名古屋大学）

- 認知診断モデルを用いた深い理解の習得状況の解析—令和3年度全国学力・学習状況調査(数学)を用いた検討—
○植阪 友理（東京大学）
佐宗 駿（東京大学，日本学術振興会）
宇佐美慧（東京大学）
- 認知診断モデルの学校現場での活用と生徒・教師の反応—学び方の改善を目指した定期テストのフィードバックの工夫—
○佐宗 駿（東京大学，日本学術振興会）
柴 里実（東京大学）
- 数学基礎力テストを用いた文科大学生の能力分析
○佐々木 淳（下関市立大学）
- 力学概念理解度調査（FCI）と科学的推論能力テスト（LCTSR）—回答の解釈と解釈から得られる知見—
○庄司 善彦（高エネルギー加速器研究機構）
宗尻 修治（広島大学）
野村 和泉（中部大学）
齊藤 準（帯広畜産大学）
- 教室内テストにおける認知診断モデルを活用したフィードバックの効果
○大橋 亮河（東京学芸大学）
鈴木 雅之（横浜国立大学）

いわゆる形成的評価を目的とした方法論として関心が高まっている状況

*引用文献
<https://www.psychometricsociety.org/keynote-speaker/gongjun-xu-university-michigan>
<https://www.psychometricsociety.org/invited-speaker/matthew-madison-university-georgia>
<https://jartest.jp/meeting/22th/session.html>

海外では、生成AI × Psychometricsとして ‘Generative Psychometrics’セッションが設置@IMPS2025
テスト開発に生成AIを積極的に活用することでコスト削減といった動きが明確に見られる



生成AIの技術的発達、テスト開発はじめテストの在り方を考える上では無視できない

引用: Russell-Lasalandra, L. (2024) AI-GENIE: A simulation study on the fully automatic scale development methodology[Oral]
International Meeting of Psychometric Society (IMPS) 2024, Prague, Czech.

学力の細分化可能性などの学力をどう捉えるか、の問い直し

CDMおよび生成AIはじめとした方法論の発展



今一度、改めてテスト研究者、実務家、テストに関わる人が押さえるべきポイントは何か？

これからの学校教育の中のテストはどうあるべきか？

押さえておきたい心理学的な指摘を2つご紹介

かねてより心理学的立場から、同じ構成概念について多数の測定尺度が存在 = 構成概念の乱立 状態への危惧が指摘（石丸, 2011；南風原, 2011；高本・服部, 2015）

先に見た「背景」にあてはめると・・・

- # リスキリング # 非認知能力 で登場する、新しく見える概念も、
「類似あるいは酷似した構成概念が複数存在する」（高本・服部, 2015） と言える？
- 非認知能力にいたっては、
**昨今「非認知能力」とされる能力（e.g. 表情認知、推論）は
心理学研究の中では「認知能力」として扱われてきた**（武藤他, 2024）

*引用文献

石丸 径一郎（2011）．臨床心理学研究法5 調査研究の方法

南風原 朝和（2011）．量的研究法 下山 晴彦（編）臨床心理学をまなぶ 7 東京大学出版会

武藤世良・篠原郁子・針生 悦子・仲 真紀子・三宮 真智子（2024）．討論—認知と非認知のあいだ——教育・研究・実践の観点から——. 教育心理学年報, 63, 207–225.

高本 真寛・服部 環（2015）．国内の心理尺度作成論文における信頼性係数の利用動向 心理学評論, 58, 220–235.

「能力」の問い直しの動き

「『能力』は周囲との関係次第」（勅使河原, 2022）



- 「『能力』は**環境次第でいくらでも移ろう**もの。個々の身体の中に確固として存在する臓器のようなはっきりしたものではない。**誰となにをどのようにやるか**、これ次第！」
(p36)

*強調は発表者

上記の議論は学校教育段階よりも社会人を対象とした場面が中心
しかし、昨今の非認知能力はじめとしたさまざまな概念が取り込まれる学校教育でのテストを考えていく上では
重要な視点

*引用文献
勅使河原真衣（2022）．「能力」の生きづらさをほぐす　どく社．

このような指摘を踏まえると、

今一度、改めてテスト研究者、実務家、テストに関わる人が認識すべきことは何か？

「テスト」は手段であり、まず第一にテストの目的を明確にしたい

何を、誰が、どう測るのか



何のために、何を、誰が、どう測り、誰にどのように使ってもらいたいのか

改めてテストに関わる関係者が再考したい3つの論点。

- ✓ 「手段」「技術」に関心が集まるあまりに、テスト実施の「目的」やテストの結果をどのように使ってもらいたいのか、の議論が希薄になっていないか？
少子化、人材不足時代に「選抜」の重要度の相対的低下？
- ✓ 当該テストの目的や理想の活用法に照らして、本当にその測定技術は必要なのか？
ムズカシすぎない？
 - 「そもそもIRTが必要ですか？」（熊谷, 2016）
 - 素点の有用性の見直し (Sijtsma, et al., 2024)
- ✓ 高度化・複雑化する測定対象は本当に測る必要はあるのか？
 - 「構成概念の乱立」（南風原, 2011）を自分たちで拍車をかけていないか？

*引用文献

南風原 朝和（2011）. 量的研究法 下山 晴彦（編）臨床心理学をまなぶ 7 東京大学出版会

熊谷龍一（2016）. データにもとづくテスト分析とテスト作成、新しい時代のテストー教師力向上のヒントー日本テスト学会講習会 .

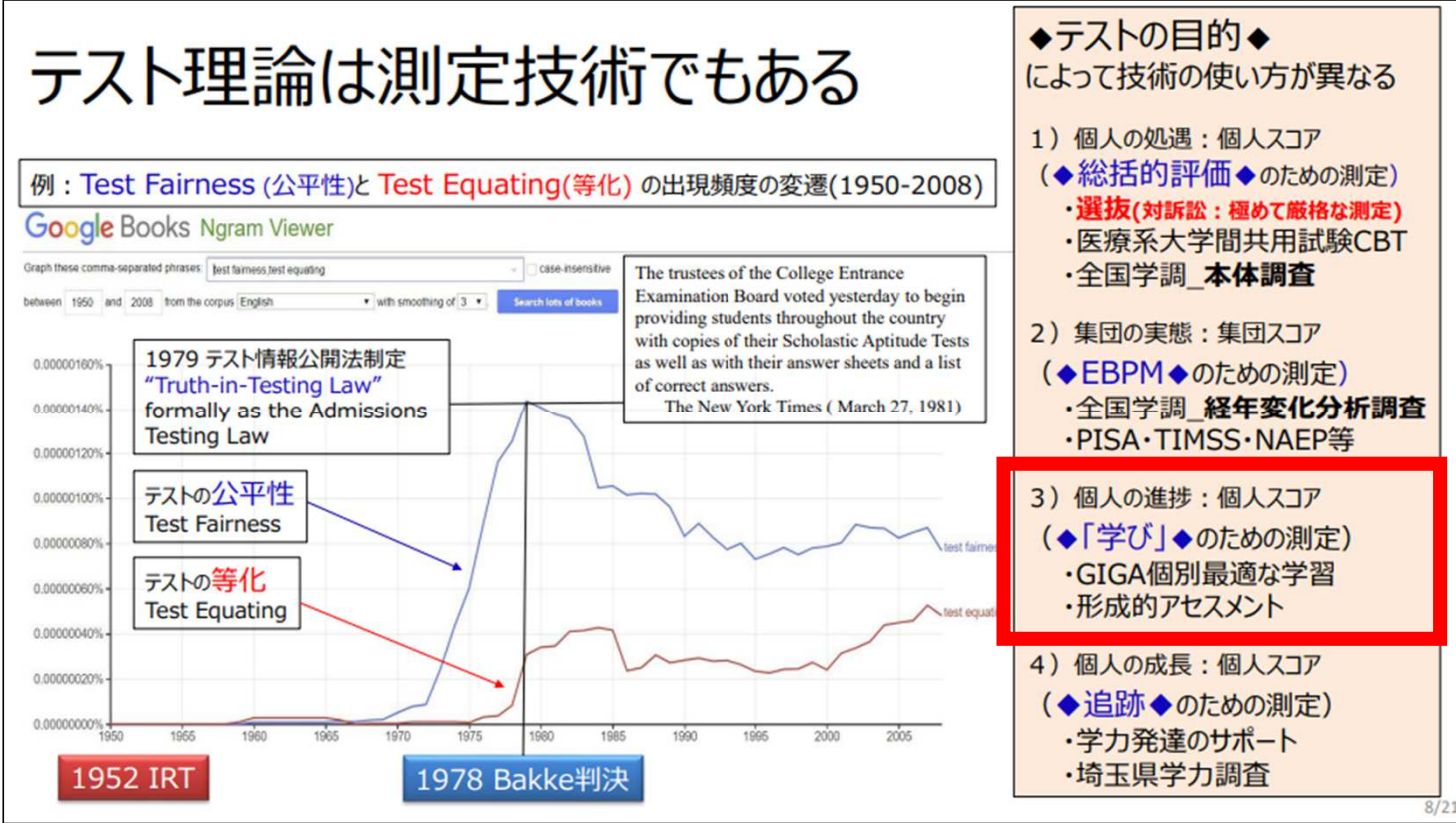
Sijtsma, K., Ellis, J. L., & Borsboom, D. (2024). Recognize the value of the sum score, psychometrics 'greatest accomplishment. Psychometrika, 89(1), 84–117.

<https://doi.org/10.1007/s11336-024-09964-7>

とはいえ、これからの学校教育の中のテストは？

関連研究を下敷きに、2つの方向性を示したい

柴山（2023）からテストの目的別に整理すれば以下の4種類



ここでは、3) 個人の進捗：個人スコア に着目し、昨今の関連議論からの示唆を導いてみる

*引用文献
柴山直（2023）.全国学力・学習状況調査の成果と課題 - 経年変化分析調査を視座として - 基盤研究B「グローバル世界を視野とする学力・非認知能力の効果的学校モデル」
【公開シンポジウム】全国学力・学習状況調査のこれまでとこれから

テスト学ではそれほど注目されないが、テスト結果のフィードバック原則、は関連領域で整理が試みられている。
ただし、その効果は時と場合による。

Kluger and DeNisi (1996) @産業場面

- フィードバックが一律に良い結果をもたらすわけではなく、効果が大きくばらつく
 - フィードバック介入全体としては平均で中程度のプラス効果（効果量 $d = 0.41$ ）
 - 約38%はパフォーマンスがむしろ低下
- Tips
 - フィードバックがタスクそのものに関連すると、パフォーマンスが向上する傾向
 - 「自己」に向きすぎると「自尊感情を守る防御」に走り、パフォーマンス低下のリスク

Wisniewski et al. (2020) @学校場面

- フィードバックは全体として中程度～高い効果（効果量 $d = 0.46$ ）
- 特に「ハイインフォメーションフィードバック」と呼ばれる、タスクだけでなくプロセスや自己調整に関する情報を含むフィードバックが非常に効果的（ $d = 0.99$ ）
- 17%の研究でマイナス効果が報告

*引用文献
Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284. <https://doi.org/10.1037/0033-2909.119.2.254>
Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, 10, Article 3087. <https://doi.org/10.3389/fpsyg.2019.03087>

学校場面におけるテスト結果も用いたフィードバックの原則としては、1) Student中心、2) 関連要素の交互作用 (Panadero & Lipnevich, 2022)

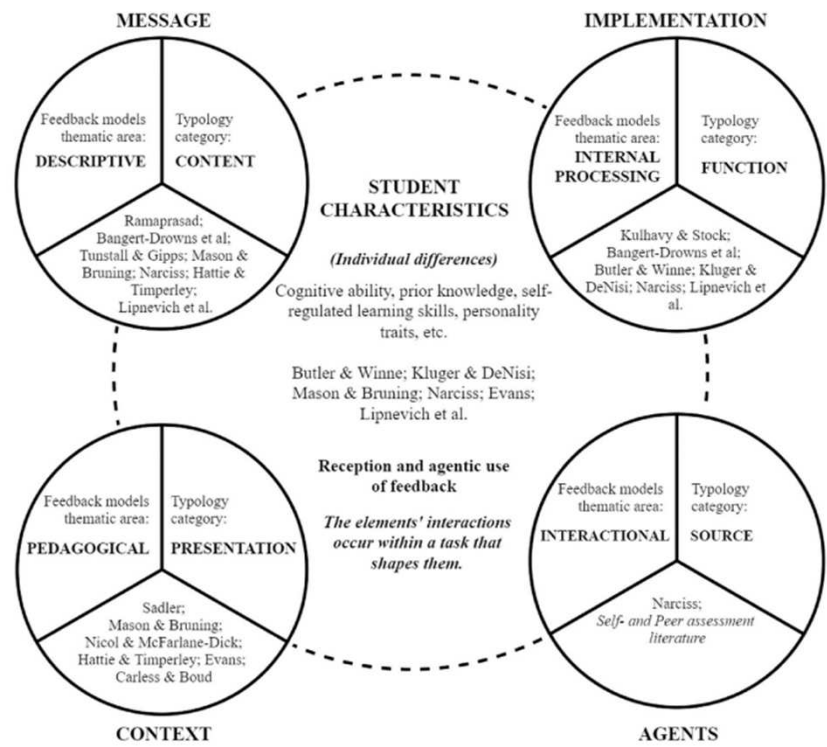


Fig. 5. The MISCA model (Message Implementation Student Context Agents).

どう測るか？だけではなく、測ったものをどのような場面でどう使うか？もセットで議論すべき
教育評価論と教育測定論、量的評価と質的評価の断絶（松下，2025）を今こそ超えていく必要があるのでは？

*引用文献
松下佳代（2025）.測りすぎ時代の学習評価論 勁草書房 .
Panadero, E., & Lipnevich, A. A. (2022). A review of feedback models and typologies: Towards an integrative model of feedback elements. Educational Research Review, 35, 100416.
<https://doi.org/10.1016/j.edurev.2021.100416>

先の方向性 とあわせて重要となるだろう視点は以下の通り

✓ 学習の転移（辰野, 1973）

学習心理学における原則（辰野, 1973）

- 学習材料，教科内容で，同一要素，あるいは類似の要因を考え，転移の起こりやすいように教材を配列すること
- 学習指導のとき，単に知識や技能の指導にとどまらず，一般化の仕方，原理の適用の仕方，研究の方法や態度をも教えること

事例（奈須・岡村, 2023）

- 山口大学教育学部附属山口小学校の事例
- 複数の教科や単元の学習の中で共通する内容、似た内容を意識したり関連付ける指導

✓ 学力の細分化 学習の転移（ 学力の統合）

細分化して捉えたい測点論者

転移を促して「使える」力の育成

*引用文献

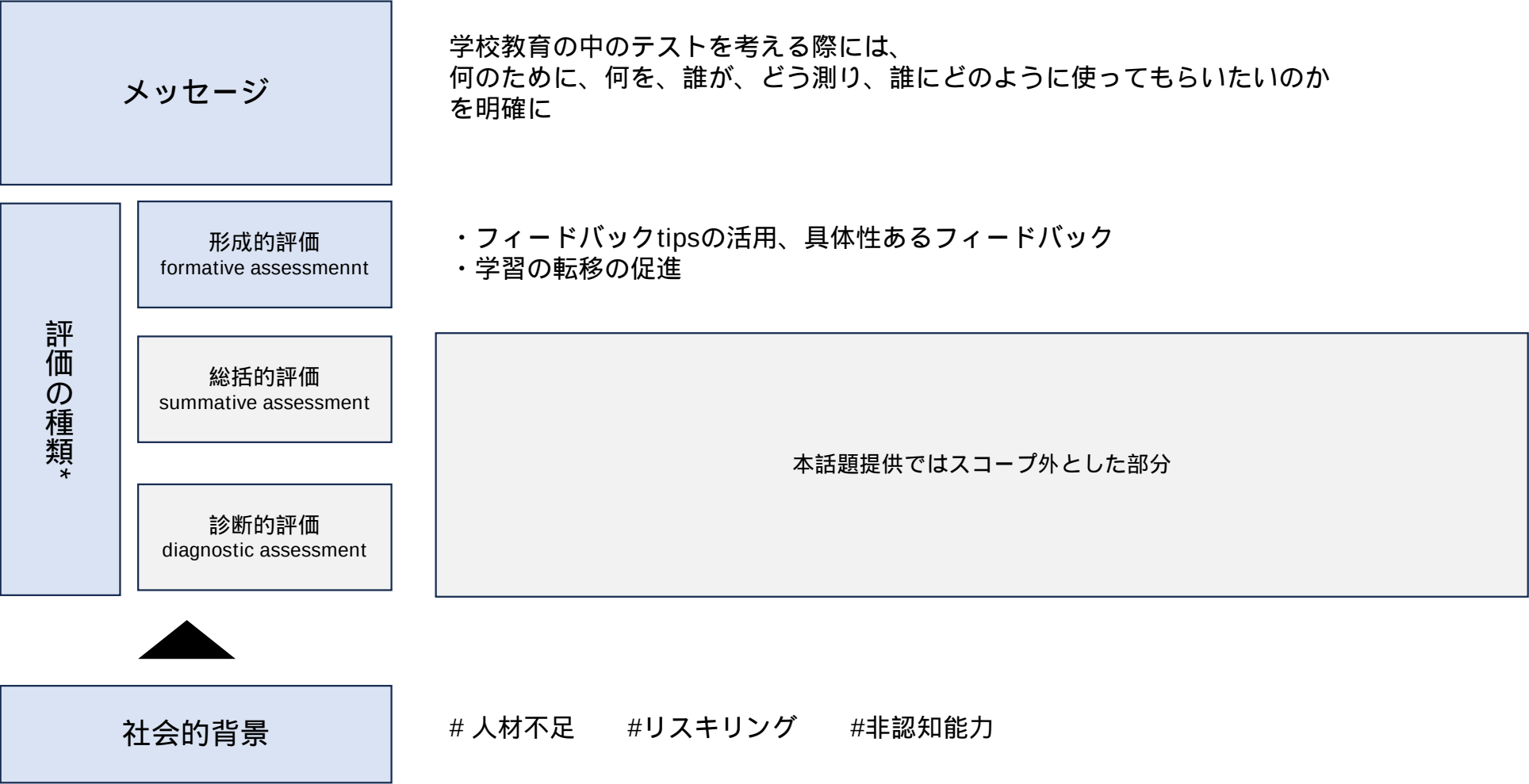
辰野千壽（1973）．学習心理学総説 金子書房

奈須正裕・岡村吉永（編）（2023）．転移する学力 東洋館出版社

内容

- 背景
- 関連学術研究の動向
 - 学力研究、測定技術研究
 - テストの活用
- まとめ





*引用文献
西岡加名恵・石井英真・田中耕治（編）（2015）. 新しい教育評価入門 - 人と育てる評価のために - [増補版] 有斐閣コンパクト

形成的評価における
テスト活用は
具体性を持たせる

- ・測っておしまい、ではなく「次の学びにどうつなげるか」

具体的には、

- ✓ 次の改善に向けたアクションプラン、レコメンドを明示する

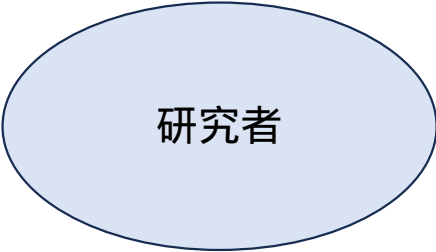
測定技術に振り回されず、
目的を起点に

- ・測ること自体が目的化しないように

「何を、誰が、何のために測るのか」を教育現場と共有し続ける姿勢が必要

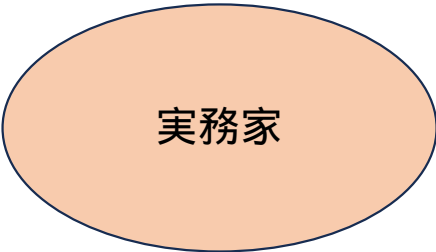
具体的には、

- ✓ 企画構想段階で、「これは本当にテストが必要なんだっけ」を問い直す



研究者

- ✓ 当該技術をどのような目的のどのような場面で活かすことをねらいにするのかの明確化
- ✓ その際の「コスパ」、*「タイプ」*は実務上無視できないことを念頭に
- ✓ 教育測定×教育評価 領域横断での協働
- ✓ 「測りすぎ」(Muller, 2018 松本訳 2019) ない



実務家

- ✓ *「IRT」*、*「CBT」*、*「CDM」*に踊らされない
- ✓ 「テスト」の話のときには、本当に測る必要があるのか？を再考
- ✓ 「測る」場面の精査(多くの組織でテスト祭りになっている、といった声も聞こえる)
- ✓ その上で、テストの目的、意図したい活用、起こしたいこと、実務上外せないこと、は協働先の専門家に明確に伝える