

全国学力・学習状況調査に 関わって

—テスト理論の視点からの10年間—

東北大学大学院教育学研究科

柴山 直

教育設計評価講座

(教育測定学＝教育学＋心理学＋統計学)

日時：2016年12月25日(日)13時30分～

場所：法政大学 市ヶ谷キャンパス

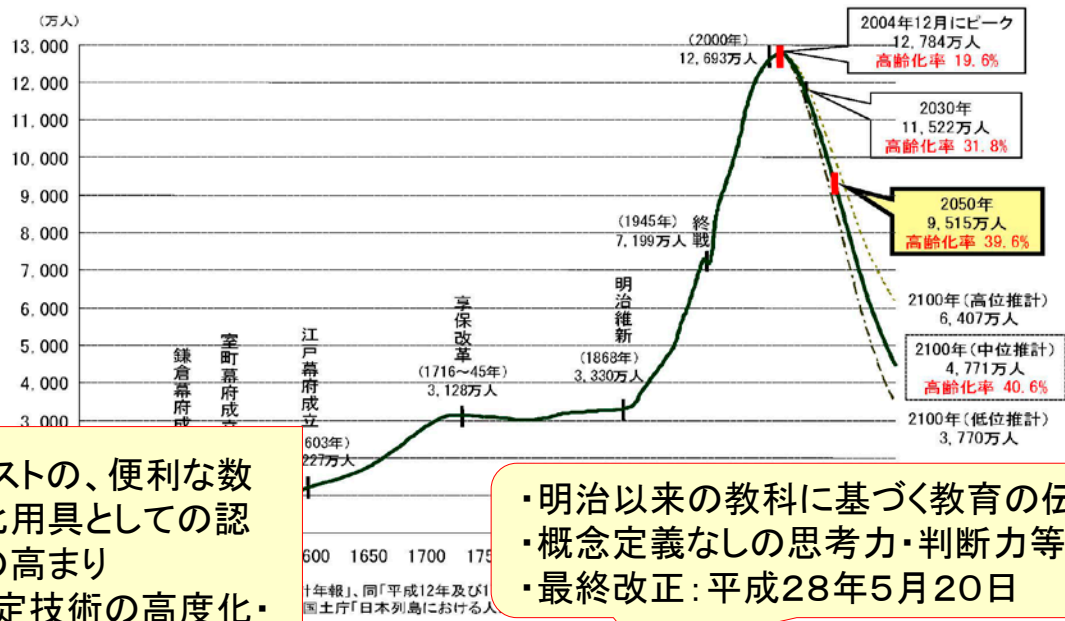
富士見校地内外濠校舎5階 S505教室

学力テスト流行の要因とその影響

(作業仮説: 人口減少 + 財政難 + 職業スキルの抽象化)

【図 I-1】我が国の人口は長期的には急減する局面に

○日本の総人口は、2004年をピークに、今後100年間で100年前(明治時代後半)の水準に戻っていく。この変化は千年単位でみても類を見ない、極めて急激な減少。



人口減少に対しては
人材を量より質で担保:
競争原理(相対評価)



質保証(到達度評価)

財政難に対しては
公共財投入の際の
費用対効果の見積もり:

→ 数値導入の必要性
(「数値で表現できない
ものは管理できない」
という経営学的な信念)

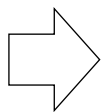
・テストの、便利な数
値化用具としての認
識の高まり
・測定技術の高度化・
複雑化・大規模化

⇒
民間の活用!

・明治以来の教科に基づく教育の伝統は?
・概念定義なしの思考力・判断力等の法定
・最終改正:平成28年5月20日

職業スキルの抽象化:

例) 米国で知的能力を要する
職業に就いていた人々は
1900年 3% → 今日 35%
グローバルに起こっている現象



学校教育法の改定 第四章 小学校 第三十条 2項
前項の場合においては、生涯にわたり学習する基盤が
培われるよう、

- 基礎的な知識及び技能を習得させるとともに、
- これらを活用して課題を解決するために必要な
思考力、判断力、表現力その他の能力をはぐくみ、
- 主体的に学習に取り組む態度を養うことに、
特に意を用いなければならない。

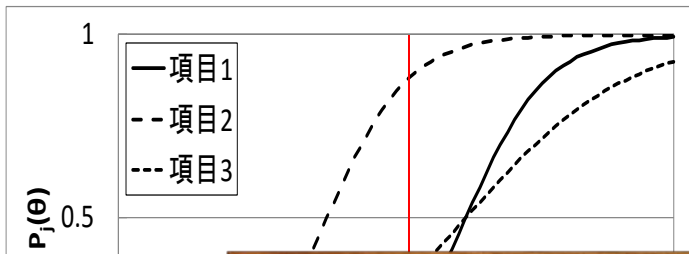
10年間の概略

- 平成16・18年 新潟県全県学力調査(学習状況調査含む)
 - 準備委員会(H15)・仕様設計・分析・報告を担当
- 平成18年 10月16日 日本経済新聞 月曜教育欄 寄稿文
 - A) 国は自らの施策効果の事後チェックのための調査
 - B) 指導のための詳細な調査は地方・学校に任せるべき
 - C) テスト理論にもとづく経年比較調査(標本調査)が必要
 - D) 実施結果を国民にフィードバックできる責任機関の設置
 - E) 大規模試験の設計、実施、分析を担える専門家養成
- 平成19年 全国学力学習状況調査開始(この時点で約40年ぶりの再開)
- 平成21年5月 参加
 - ① 「全国学力・学習状況調査の分析・活用の推進に関する専門家検討会議」
- 以降
 - ② 全国的な学力調査に関する専門家会議(平成23年8月17日～)
 - ③ 全国的な学力調査に関する専門家会議(平成25年7月9日～)
 - ④ 全国的な学力調査に関する専門家会議(平成27年6月24日～)
- 平成28年 経年変化分析調査の本格実施(H25年度はいわば準備)



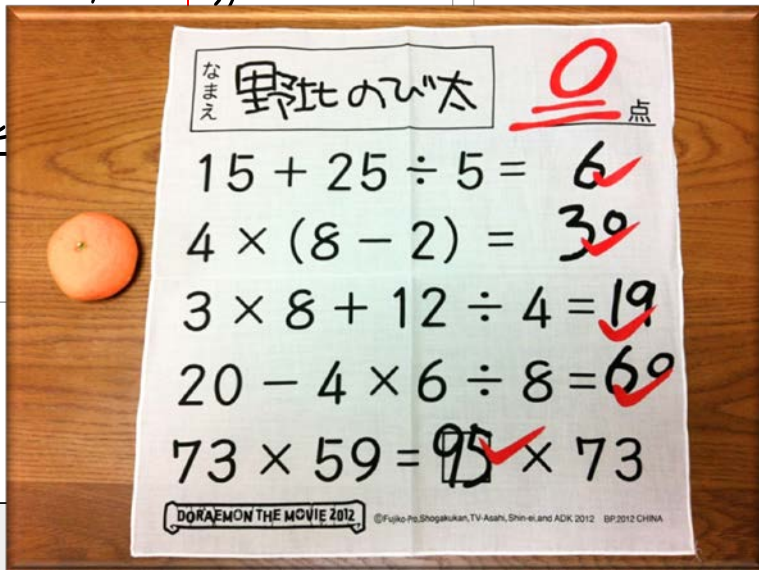
重複テスト
分冊法の
開発調査
研究

良い問題の項目特性曲線のモデル化 (項目反応理論:IRTの考え方)

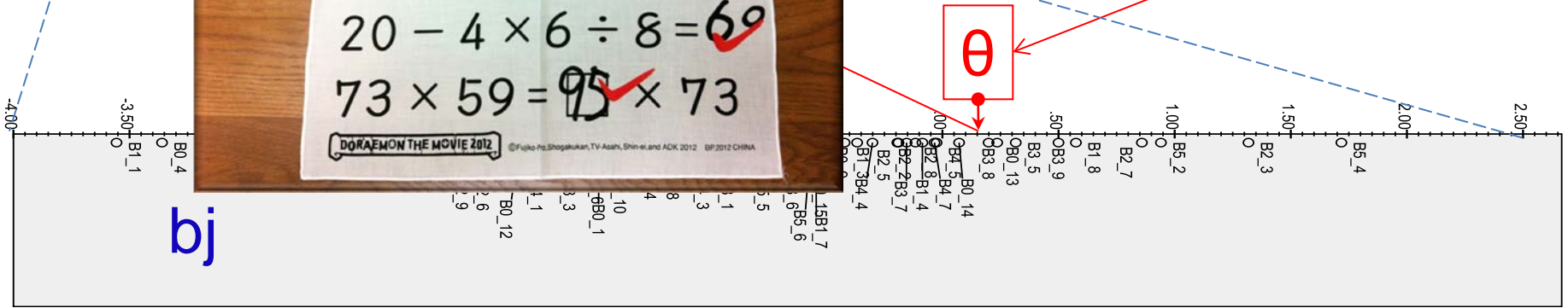


$$P(X_j = 1 | \theta) = \frac{1}{1 + \exp\{-1.7a_j(\theta - b_j)\}}$$

項目jの素点: 正答



θ と項目特性(困難度 b_j)を分離
同じ数直線(尺度)上で表現
み合わせと正誤情報から θ を推定



平成25年度文部科学省委託研究

「東日本大震災への学力の影響～IRT推算値による経年比較分析～」

調査の名称	調査年度（平成）						
	19	20	21	22	23	24	25
文科省委託調査研究 <small>※右の図中、上から調査方法、対象学年、出題範囲、実施月を示す。また、θ, PVはそれぞれ、尺度値、推算値を表す。*印はそれに対応づけされたデータであることを意味する。</small>				新潟市中3 中2まで 10月実施	宮城県中3 中2まで 10月実施	宮城県中3 中2まで 11月実施	宮城県中3 中2まで 10月頃実施
			(θ_{1112})	θ_{1212}	θ_{1322}	θ_{1422}	θ_{1522}
全国学力・学習状況調査 <small>※右の図中、上から調査方法、対象学年、出題範囲、実施月を示す。また、θ, PVはそれぞれ、尺度値、推算値を表す。*印はそれに対応づけされたデータであることを意味する。</small>			θ_{2112}				θ_{2522}
			θ_{2121}				θ_{2521}
	全数 中3 中2まで 4月実施	全数 中3 中2まで 4月実施	全数 中3 中2まで 4月実施	抽出 中3 中2まで 4月実施	東日本大震災 (3.11)	抽出 中3 中2まで 4月実施	全数 中3 中2まで 4月実施
			PV^*_{2121}				PV^*_{2521}

θ_{ghij}

g: 調査別
 g=1 委託調査研究
 g=2 全国学力・学習状況調査

h: 平成年度別
 h=1 平成21年度
 h=2 平成22年度
 h=3 平成23年度
 h=4 平成24年度
 h=5 平成25年度

i: 県別
 i=1 新潟県
 i=2 宮城県

j: 全県・協力校別
 j=1 新潟県
 または
 宮城県の
 全中学校
 j=2 委託調査
 研究への
 協力中学校

等化済みのため同一尺度上で比較可能

対応づけのための係数を求める(H21年度用)

対応づけのための係数を求める(H25年度用)

委託調査研究対象集団の抜き出し

委託調査研究対象集団の抜き出し

宮城県データが共通尺度上で比較可能

等化のためのデータ収集デザイン

(ポイント:どこかに共通する部分がある)

- 等価グループ・デザイン

グループ	テスト X	テスト Y
P	○	
P		○

- 単一グループ・デザイン

グループ	テスト X	テスト Y
P	○	○

- カウンターバランス・デザイン

グループ	テスト X	テスト Y
P	1	2
P	2	1

- アンカーテストを伴う不等価グループ・デザイン (NEAT)

グループ	テスト X	テスト A	テスト Y
P	○	○	
Q		○	○

P,Qはそれぞれ互いに異なる母集団からの標本を表す

偏りをなくすための分冊デザイン (釣合い型不完備ブロックデザイン: BIBD)

	分冊1	分冊2	分冊3	分冊4	分冊5	分冊6	分冊7
位置1	3	4	5	6	7	1	2
位置2	5	6	7	1	2	3	4
位置3	6	7	1	2	3	4	5
位置4	7	1	2	3	4	5	6

- ・どの項目セットも等しく4回使用されている → 使用頻度の効果を相殺
- ・どの項目セットも互いに等しく2回ずつ会合する → 組合せの効果を相殺
- ・どの項目セットも1回ずつ等しく別の位置に配置される → 出現順の効果を相殺

※表内の数字は項目セット(=項目ユニット)

※BIBD=Balanced Incomplete Block Design

データ収集デザインの基本パタン

- 等価グループ・デザイン

グループ	テスト X	テスト Y
P	○	
P		○

- 単一グループ・デザイン

グループ	テスト X	テスト Y
P	○	○

- カウンターバランス・デザイン

グループ	テスト X	テスト Y
P	1	2
P	2	1

- アンカーテストを伴う不等価グループ・デザイン

グループ	テスト X	テスト A	テスト Y
P	○	○	
Q		○	○

計画行列による BIBDの別表現

(フィッシャーの表現・表中の数字は位置)

		項目セット							
		1	2	3	4	5	6	7	
P 母集団	サンプリング	1			1		2	3	4
	分冊	2	4			1		2	3
		3	3	4			1		2
		4	2	3	4			1	
		5		2	3	4			1
		6	1		2	3	4		
		7		1		2	3	4	

(4つの基本パターンが全て含まれている)

BIBD導入の効果

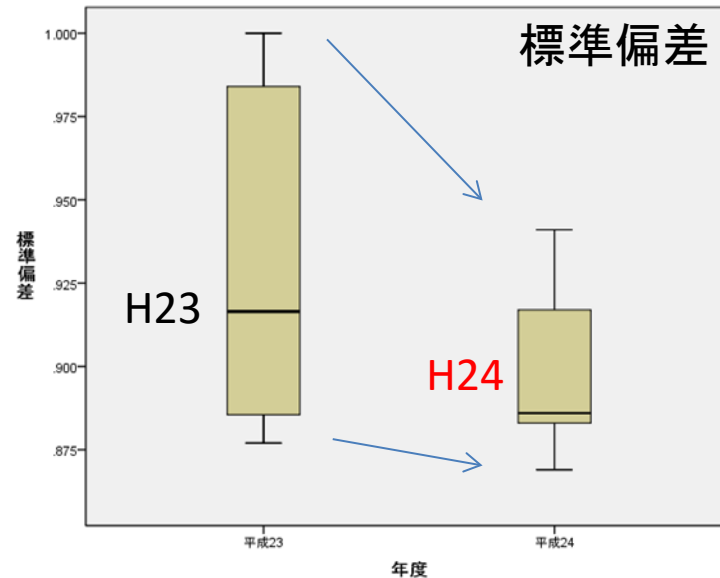
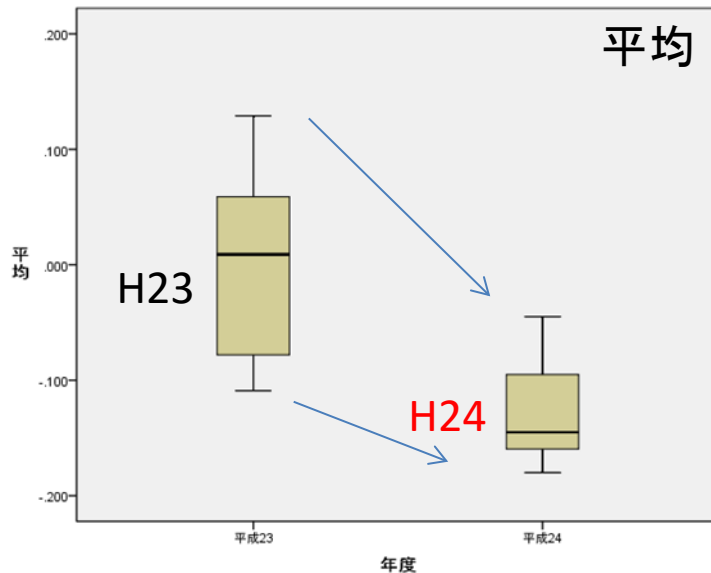
平成23年度 BIBDではない

分冊	S1	S2	S3	S4	S5	S6	S7	S8
人数	256	257	252	251	243	238	233	256
平均	-.048	.035	.065	-.108	-.017	.053	.129	-.109
SD	.892	.997	.941	.885	.971	.886	1.000	.877

平成24年度 BIBD

分冊	S1	S2	S3	S4	S5	S6	S7
人数	368	356	357	356	369	366	367
平均	-.158	-.045	-.047	-.161	-.180	-.143	-.145
SD	.885	.869	.918	.916	.886	.941	.881

分冊ごとの学力特性値 θ の
集団統計量(平均や標準偏差)の
 ばらつきが**H24年度**の方が
均質化されている



• 推算値(Plausible Values:PVs)の必要性

– 項目数が少ないテストの個人スコアのみに基づき
集団の能力分布を推定すると正しい結果が得られない

- 分散の過大評価・過小評価
- 各集団間の平均値差の統計的な判断に支障
- 個人スコアは離散的なため正確なパーセンタイルが得られない

推算値の定義とイメージ

$$h(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)g(\theta)}{\int f(\mathbf{x}|\theta)g(\theta)d\theta}$$

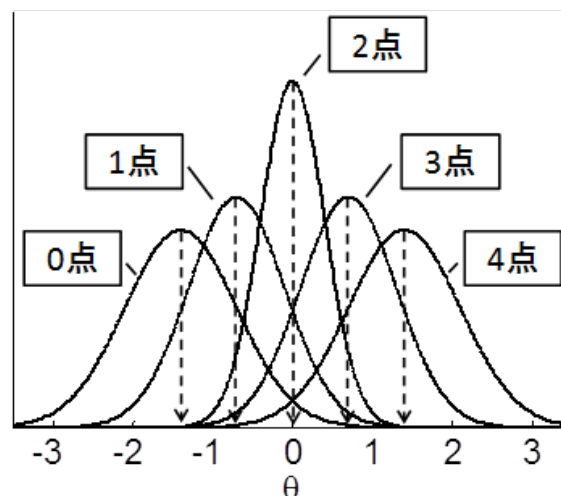
x: 項目反応パターン

θ : 学力特性

$g(\theta)$: 事前分布

$f(\mathbf{x}|\theta)$: 尤度関数

$h(\theta|\mathbf{x})$: 事後分布



各得点の事後分布とEAP推定値
(Wu, 2004, Figureを単純化したもの)

マルチレベル分析

独立変数: 生徒質問紙の項目と学校関係の情報(H21年度学力調査結果含む: 推算値)

従属変数: H25年度学力調査結果: 推算値

－ 個人レベルの変数

- 「学習意欲」変数

- － 「数学の勉強は好きですか」「数学の勉強は大切だと思いますか」等の10項目の第1主成分得点

- 「家庭学習時間」変数

- － 「普段の勉強時間(学習塾や家庭教師含む)」と「休みの日の勉強時間(学習塾や家庭教師含む)」の単純合計を標準化

- 「生活規範」変数

- － 「朝食を毎日食べていますか」「毎日、同じくらいの時刻に起きていますか」「家の人(兄弟姉妹を除く)と普段(月曜から金曜)、夕食を一緒に食べていますか」等の6項目の第1主成分得点

－ 学校レベルの震災の影響を表す変数

- 生徒死者数, 生徒死者数, 生徒行方不明者数, 教職員死者数, 仮設校舎への移転・間借り, の各変数

- 「津波被害」変数

- － スクールバス利用人数, スクールバス概算距離, スクールバス台数, スクールバス概算費用の第1主成分

- 「教員加配」変数

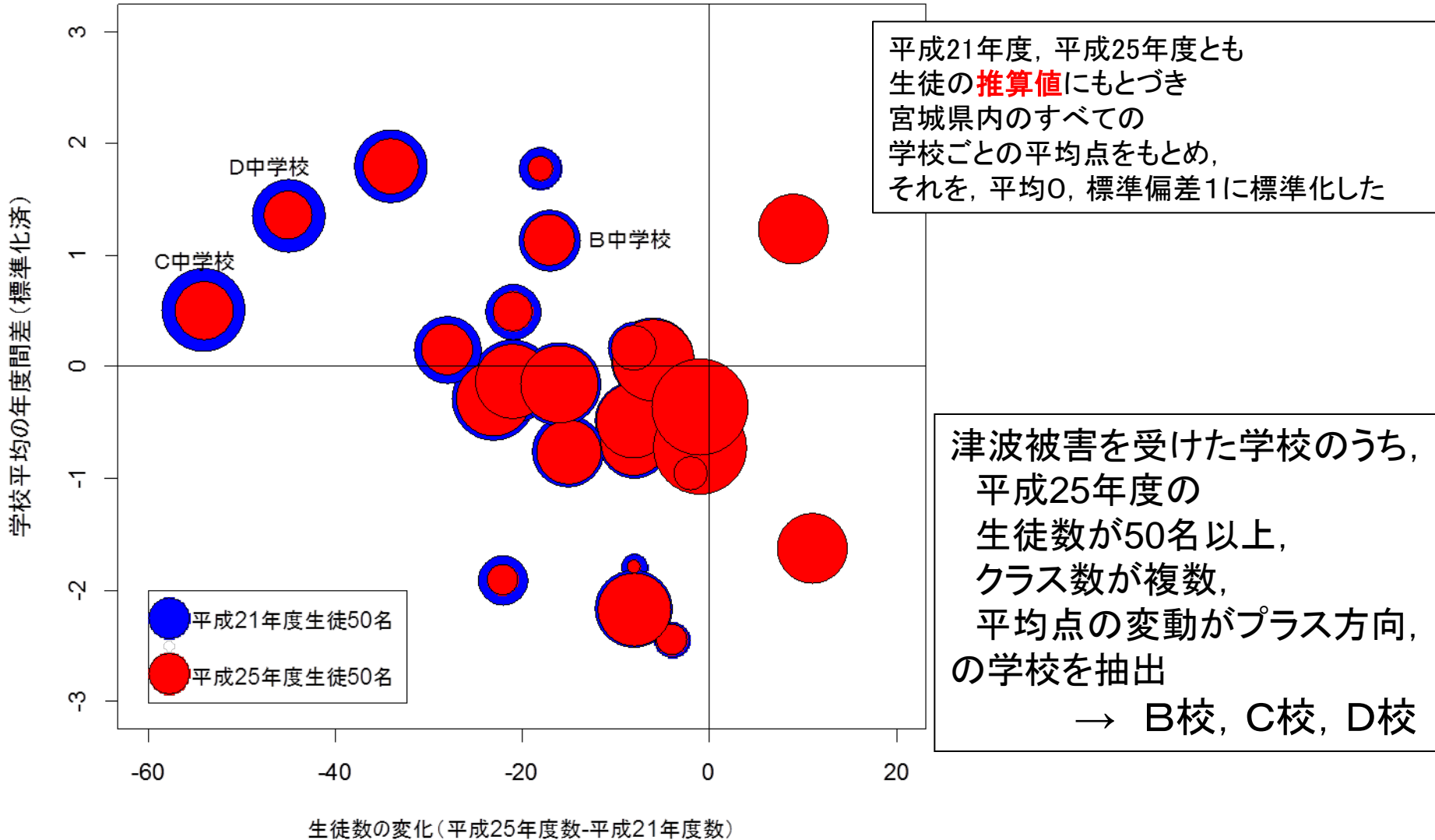
- 「放射性物質拡散」変数

マルチレベル分析の結果

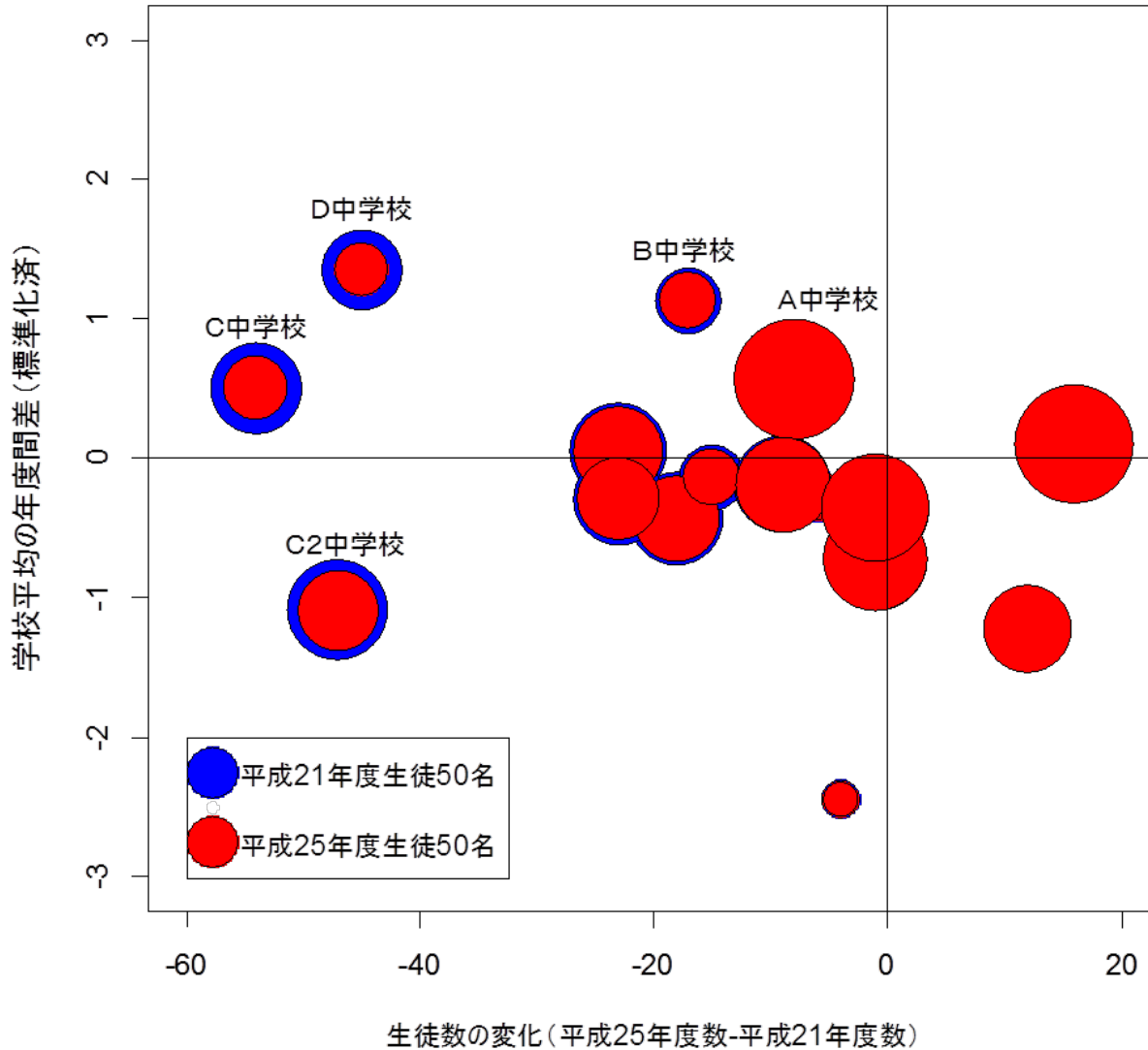
	ヌルモデル	モデル1	モデル2
モデル			
固定パート			
切片	-0.011(.023)	-0.007(.022)	0.030(.019)
PV21			0.539(.057) *
学習意欲		0.292(.008)*	0.291(.008) *
家庭学習時間		0.057(.009)*	0.058(.009) *
生活規範		0.091(.008)*	0.090(.008) *
学習意欲*津波被害			0.017(.007) **
家庭学習時間*津波被害			-0.001(.009)
家庭学習時間*放射線物質拡散			-0.008(.008)
ランダムパート			
σ_{θ}^2	0.824(.004)*	0.671(.003)*	0.671(.003) *
$\sigma_{\mu 0}^2$	0.112(.0011)*	0.099(.010)*	0.069(.007) *
-2対数尤度	272252.442	252295.034	252217.240
AIC	272258.442	252325.034	252255.240
BIC	272287.058	252468.118	252436.479

※表中, PV21はH21年度全国学力調査の推算値にもとづく学校ごとの平均を示す

津波被害を受けた学校の変化図



犠牲となった生徒がいる学校の変化図



生徒の犠牲者がひとりでもいる学校を基準にした場合

津波被害を受けた学校のうち、平成25年度の生徒数が50名以上、クラス数が複数、平均点の変動がプラス方向、の学校を抽出

→ B校, C校, D校
 さらにA校が加わる
 また、C2校は聞き取り調査の中でC校との関係で考察の対象となった学校
 (ただし、C2校への聞き取りは行ってない)

被災小学校
教室
黒板

平成26年3月14日撮影

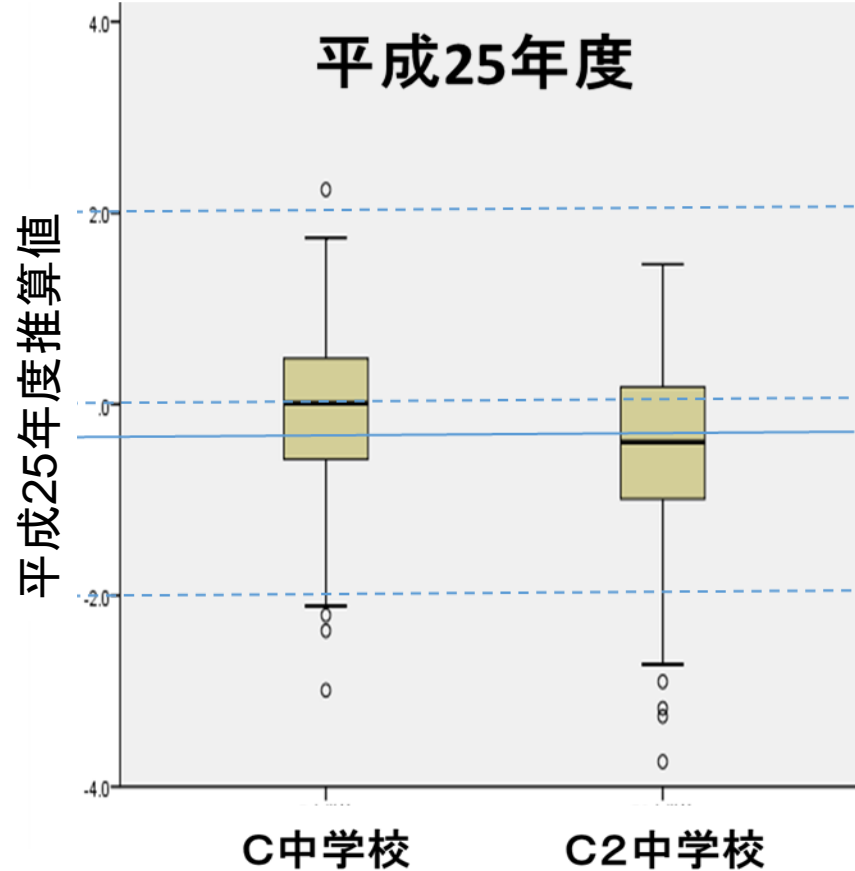
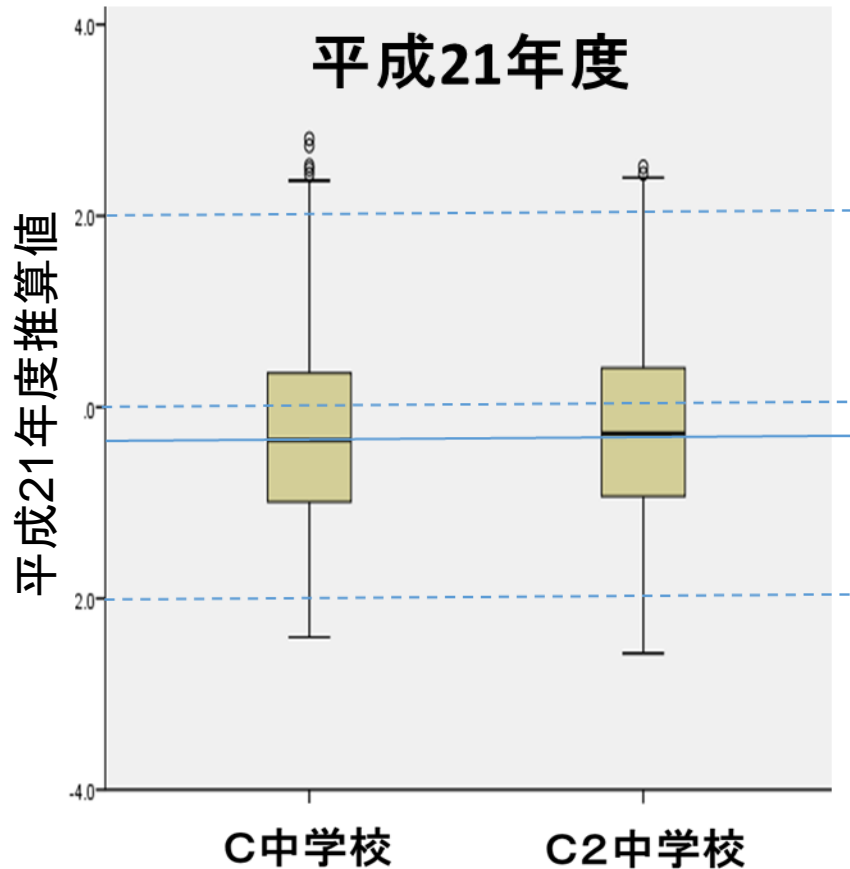
3月14日(月) 日直 ○○ ○○
あ○8日で春休み

東日本太平洋沖地震
平成23年3月11日14時46分

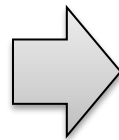


平成25年度文部科学省委託研究
「東日本大震災への学力の影響～IRT推算値による経年比較分析～」

C中学校とC2中学校の変化



平成21年度時点では両校とも生徒の学力分布はほぼ同一



平成25年度時点では両校の学力分布がC校は少し上へ、C2校はその逆方向へ移動している

C中学校およびC教育委員会への 聞き取り調査にもとづく考察

- 震災前はC中学校とほぼ似たような状況にあったC2中学校
 - ・両校とも学力向上ためだけの特別な取り組みはしていない
 - ・震災後は普段の生活を取り戻すのに両校とも精一杯
- C中学校の震災直後の様子
 - ・C中学校は津波被害で使用できず、周辺は誰も住めない
 - ・より内陸の安全なところへ集団移転
 - ・C1中学校に間借り・共存しながら
 - 借りている学校だから一層大切にするという意識
 - 学校として生活規範・学習規範は維持できていた
- C2中学校の状況
 - ・地域は流されたが、校舎には被害なし
 - ・そのため10月頃まで地域住民の避難所になった(500名以上)
 - ・教室に入ってもらうほか手段はなかった
 - ・学校にふさわしくない大人のトラブルを生徒達が直接見てしまう
 - ・生徒達自身の規範意識が混乱した結果
 - ・授業がなかなか成立しない時期あり(→現在は落ち着いている)

「言ってみれば大変地味で目立たない教育統計や心理測定の支えがあって始めて、地に着いた教育のあり方も考えられるのだということを、理解していただければこの展示会も意義あるものになると思いました。」

(池田央先生からいただいたメールから抜粋)

- [fpr 3286]「統計計算の50年」展 SHIBAYAMA Tadashi 2009/03/26 13:16投稿
- 東京理科大学 近代科学資料館
- 「統計計算の50年－統計計算・教育統計の歴史をふりかえる－池田央」展

参考文献・研究紹介等

書籍紹介(テストを技術的な側面から考えるために)

- 日本テスト学会編 (2007) 「テスト・スタンダード～日本のテストの将来に向けて～」 金子書房
 - テストを取り扱う関係者の指針と行動, あるべき姿と規準を提示する
- 池田央監訳(2008)「テスト作成ハンドブック」株式会社教育測定研究所
 - 最新のテスト技術と考え方による公平妥当なテスト作成・実施・利用のすべてについて書かれた本
- 日本テスト学会編(2010) 「見直そう、テストを支える基本の技術と教育」 金子書房
 - テストの仕組みについて、さまざまな疑問にQ&A方式で答える
- 加藤健太郎・山田剛史・川端一光(2014)「Rによる項目反応理論」オーム社
 - 最新のテスト技術体系に関する実践知が統計言語のRで明確に記述されている良書
- 野口裕之、大隅敦子(2014)「テストの基礎理論」 研究社
 - テスティングの最新理論を広く網羅, 特に語学系の検定に役立つ情報も多く便利
- 別府正彦(2015)「新テスト」の学力測定方法を知るIRT入門—基礎知識からテスト開発・分析までの話— 河合出版
 - 実践家を書いたIRTを初めて学ぶときのわかりやすい「教科書」
- 肥田野直 編(1972)「心理学研究法7 テスト I」 東京大学出版会
 - テストの作成から結果の分析までを体系的にまとめている現在でも役立つ本
- 池田央(1973)「心理学研究法8 テスト II」 東京大学出版会
 - テスト理論の名著であり, 現在も研究上の刺激をうけている
- 橋本重治(2001) 続・到達度評価の研究(新装版), 図書文化



研究紹介(文部科学省委託研究関係:Web公開)

1. 平成22年度文部科学省委託研究「全国規模の学力調査における重複テスト分冊法適用の試み」 研究代表者(柴山直)

http://www.mext.go.jp/a_menu/shotou/gakuryoku-chousa/1344297.htm

2. 平成23年度文部科学省委託研究「全国規模の学力調査における重複テスト分冊法の展開可能性について」 研究代表者(柴山直)

http://www.mext.go.jp/b_menu/shingi/chousa/shotou/085/shiryo/attach/1323273.htm

3. 平成24年度文部科学省委託研究「全国規模の学力調査におけるマトリックス・サンプリングにもとづく集団統計量の推定について」 研究代表者(柴山直)

http://www.mext.go.jp/a_menu/shotou/gakuryoku-chousa/1344827.htm

4. 平成25年度文部科学省委託研究「東日本大震災への学力の影響～IRT推算値による経年比較分析～」 研究代表者(柴山直)

http://www.mext.go.jp/a_menu/shotou/gakuryoku-chousa/1361052.htm

内容

PISA, TIMSS等の国際的な大規模学力調査で使われている最新の調査技術を我が国に導入するために行った一連の基礎研究。

平成22年度は基礎技術の導入, 平成23年度は等化技術とリーディング・リテラシー問題の実施と多段階IRTモデルの導入, 平成24年度は釣合い型不完備ブロックデザインおよび推算値の導入を実現した。

さらに, その成果の上にならば, 平成21年度と平成25年度実地の文部科学省全国学力調査の結果を, 対応づけと等化の併用により比較可能とし, その分析結果にもとづき行ったインタビュー調査から, 甚大な津波被害にもかかわらず平均学力を伸ばした学校には, 学校としての一体感や生徒達の規範意識が確立されていたことなどを見いだした。

研究紹介(記述・パフォーマンスアセスメント関係等)

5. 柴山直(2008) 論述式試験の採点システム構築に関する統計科学的研究.[平成17年度～平成19年度科学研究費補助金基盤研究(B)]研究成果報告書,(2008)]研究代表者 柴山直
→ 小論文採点/自動化の実現を試みる
6. 柴山直・前田忠彦(2009) 論述試験の採点デザインと信頼性評価に関する統計的研究.統計数理研究所共同研究レポート244. 統計数理研究所
→ 小論文採点研究の総括・反省
7. 適性試験委員会(2011) JLF叢書 Vol.17 法科大学院統一適性試験テクニカル・レポート2009-2010.[商事法務,(2011)]適性試験委員会編
→ 論理・分析・長文読解・表現力の測定論的研究総括
8. 適性試験委員会(2016) 2016年法科大学院全国統一適性試験 実施報告書 Web公開
→ 実施結果の品質保証の継続的公開例
9. 柴山直(2015)「パフォーマンス・アセスメントのトリレンマ」東北大学大学院教育学研究科教育設計評価講座 研究ノート (未公刊)
10. 井本佳宏・柴山直(2016)「大学入学希望者学力評価テスト(仮称)」の制度設計における限界について.東北大学大学院教育学研究科研究年報, 第64集, 第2号, 181-192.

- ご静聴ありがとうございました