

第5回日本テスト学会記念講演会

全項目が開示されるテスト文化の  
もとでの得点分布の経年比較

—全国テストと自治体テストのリンクング—

石井 秀宗

名古屋大学大学院教育発達科学研究科

2012/3/22 成蹊大学

# 日本の多くの大規模テスト

---

- ▶ 全問題が直後に公開され、毎回新作問題を用いる
- ▶ 入念な試験対策がなされる
- ▶ 同じ科目の問題は全受検者に渡って同一

→

- ▶ 経年変化や発達変化を捉えられる構造になってない
- ▶ 相対順位の変動を学力の経年変化と混同しかねない
- ▶ 受験のための勉強、受験シフトの学習を助長する
- ▶ 巨大なクラスルームテストで、学習領域のごく一部しか評価していない

# 学力の経年比較

---

- ▶ とくにPISA2003以降, 学力の経年変化に注意が向けられている
- ▶ 項目レベルの比較くらいはなされているが, それを学習領域全般の話に拡大解釈するのは無謀  
目標到達率との比較では説得力がない
- ▶ 大規模テストを実施していながらそれを活かさないうえ, PISAの国別順位で議論している

# 先行研究

---

- ▶ 齊田(2003), 吉村他(2005)
- ▶ それぞれ, 一定の実施主体による大規模テストを使用
- ▶ 過去何年分かの問題から項目を抽出してアンカーテストを作成
- ▶ 被験者を集めアンカーテストを実施
- ▶ 項目反応理論を用いて共通尺度下でのパラメタ値を推定し, 経年変化を把握

# 本研究

---

- ▶ 石井・安永(2011)
- ▶ 異なる実施主体による大規模テストを使用(科目は同じ)
- ▶ 全国テストと、それとは異なる年度に実施された自治体テストを利用(アンカーテストは作らない)
- ▶ 被験者を集め自治体テストを実施(被験者は、全国テストも受けている)
- ▶ 得点の対応づけをして得点分布を推定し、経年変化を把握

# テスト実施（データ収集）デザイン

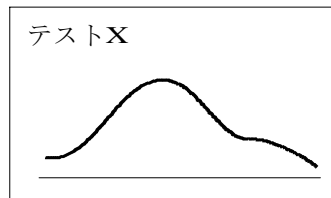
- ▶ 時点T1に、地域Aで、自治体テストXを実施
- ▶ 時点T2に、地域Aと地域Bで、全国テストYを実施
- ▶ 時点T2に、地域Bで、自治体テストXも実施
- ▶ 地域Bでの2つのテストの受検者は同一（同等）集団

	地域A	地域B
時点 $T_1$	テストX	
時点 $T_2$	テストY	テストX   テストY

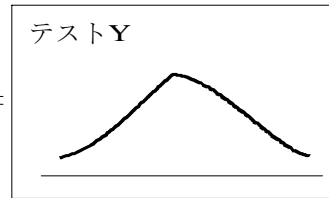
# テスト得点の対応づけ

- ▶ 地域Bで得られた2つのテストの得点分布について、等パーセンタイル法等の手法を用いて得点を対応づけ

時点 $T_2$  地域B

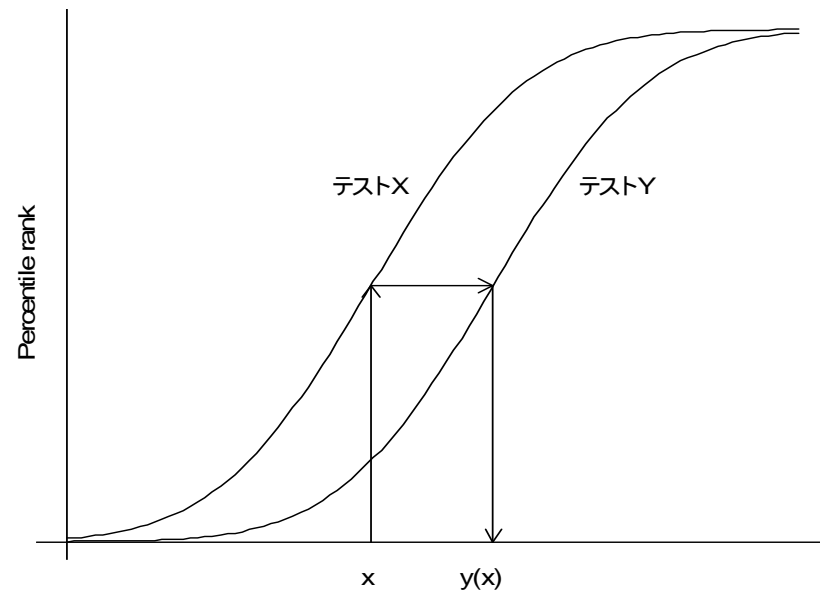


①対応づけ



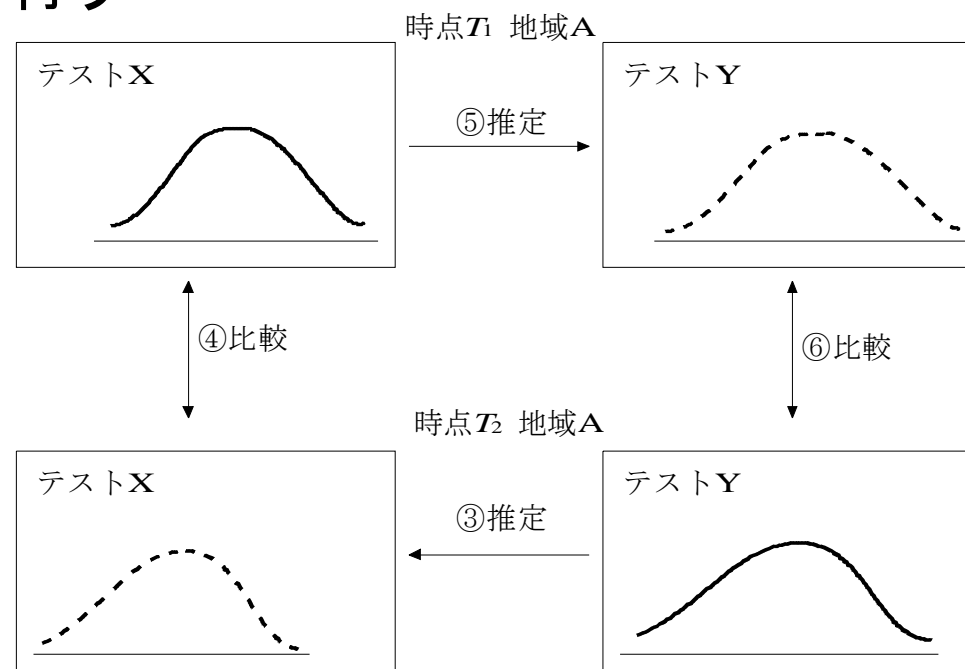
②得点对応表(イメージ)

テストX	テストY
0	3
10	12
20	25
30	34
40	41
50	47
60	59
70	68
80	77
90	85
100	98



# 得点分布の推定と経年比較

- ▶ 対応表を用いて，地域Aにおける，時点T1のテストYの得点分布，時点T2のテストXの得点分布を推定
- ▶ 時点T1とT2の，同じテストの得点分布を比較することにより，経年比較を行う





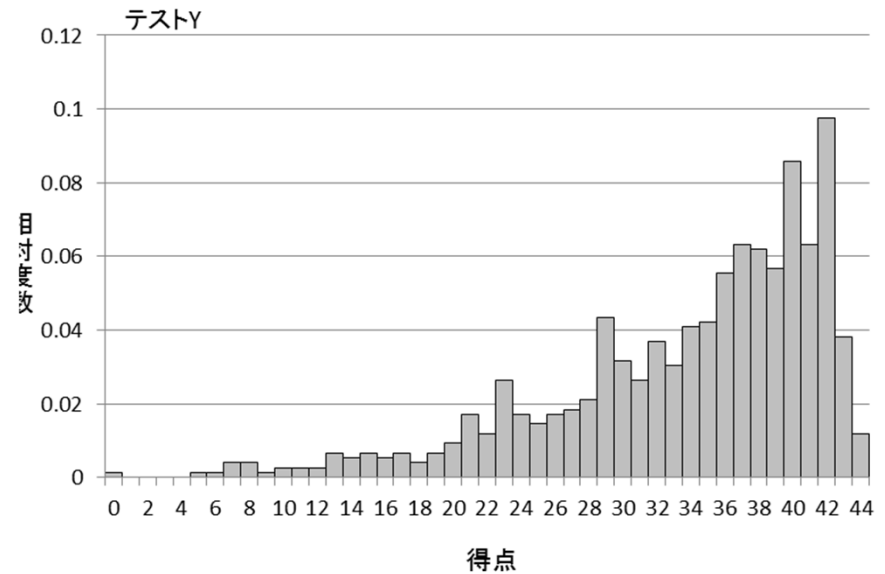
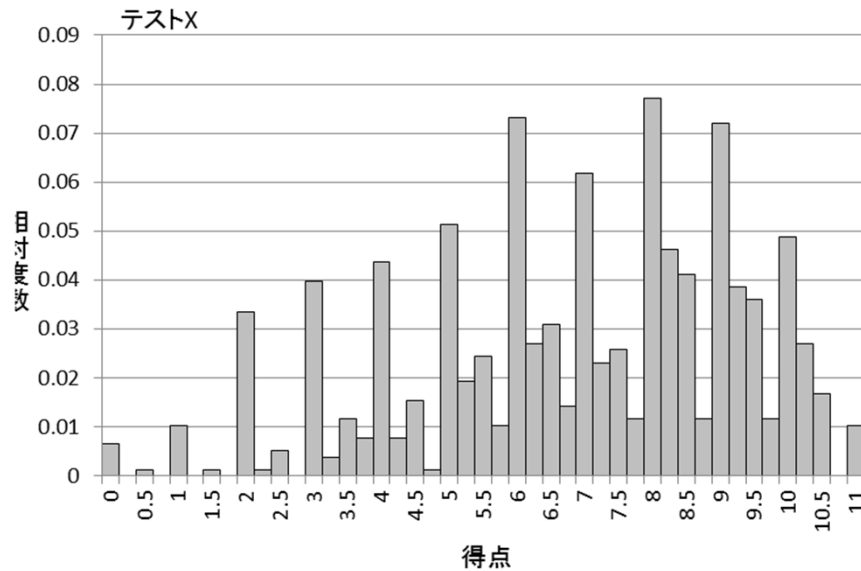
# 適用例

- ▶ テストX: ある自治体で平成18(2006)年度に実施された中学校3年生国語テスト 11問(0~44点)
- ▶ テストY: 平成21(2009)年度全国学力・学習状況調査中学校3年生国語(A+B) 44問(0~44点)

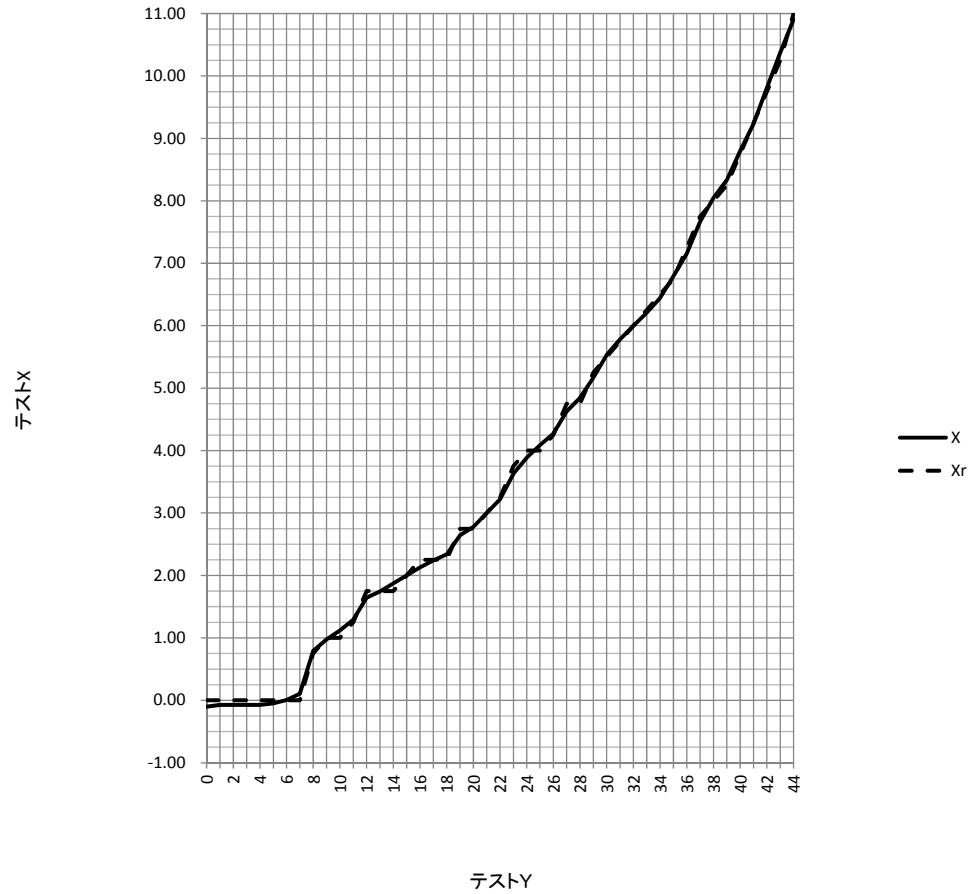
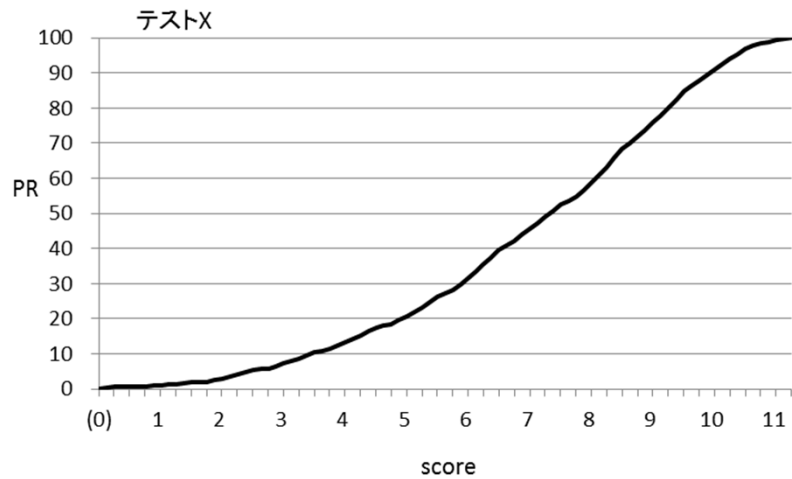
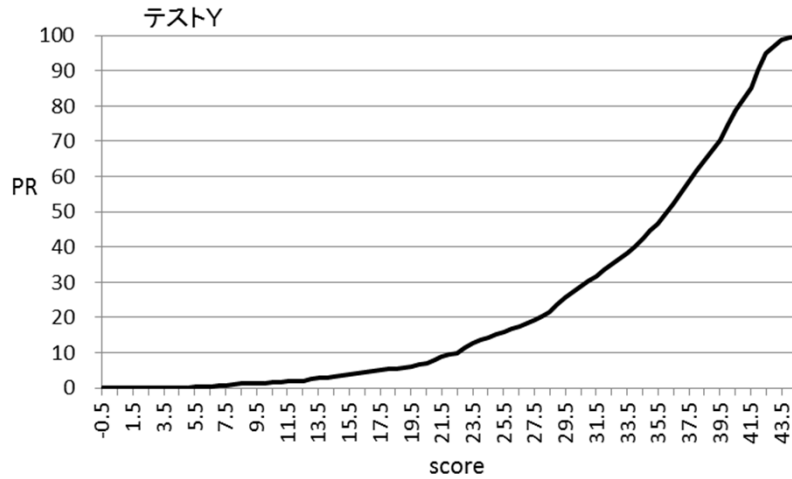
地域		地域A		地域B	
テスト		テストX	テストY	テストX	テストY
時点	T1 H18	3892名			
	T2 H21		19498名	778名	759名

# 地域Bにおける得点分布

	n	Mean	SD	Min	Q1	Q2	Q3	Max
X	778	6.88	2.41	0.00	5.25	7.00	9.00	11.00
Y	759	33.79	8.06	0	29	36	40	44

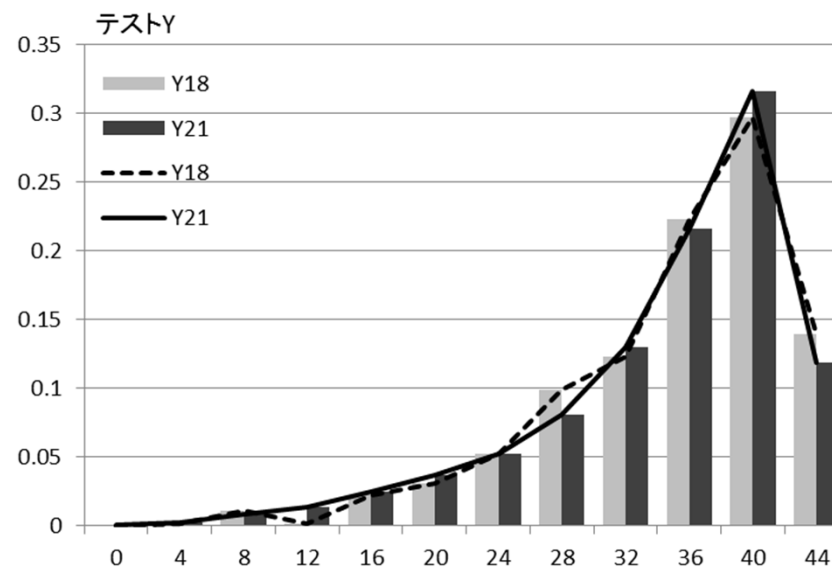
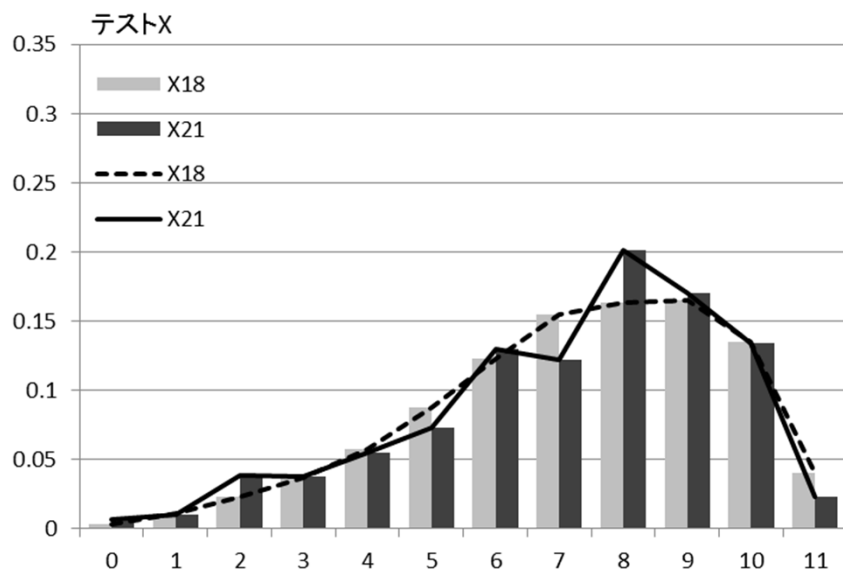


# パーセンタイル関数・対応曲線



# 得点分布の経年比較

	n	Mean	SD	効果量	分散比	Min	Q1	Q2	Q3	Max
X 18	3892	7.27	2.30	-	-	0.00	6.00	7.50	9.00	11.00
X 21	19498	7.16	2.38	-0.05	1.07	-0.10	5.78	7.67	8.80	10.91
X 21r	19498	7.15	2.36	-0.05	1.06	0.00	5.75	7.75	8.75	11.00
Y 18	3892	35.06	7.31	-0.05	1.15	5.9	32.0	36.7	40.4	44.2
Y 18r	3892	35.11	7.26	-0.06	1.16	6	32	37	40	44
Y 21	19498	34.65	7.83	-	-	0	31	37	40	44



# 結果

---

- ▶ 効果量は、いずれもマイナスの値であるが、ほぼゼロで、学力が大きく低下したと言うほどでもない
- ▶ 平均値の非劣性・同等性の検討では、平成18年度に対し平成21年度の平均は、劣らないが同等かそれ以上とは言えない、という結果
- ▶ 分散比は、いずれも1より大きい値であるが、極端に大きいわけでもなく、格差が大きく拡大したと言うほどでもない

# 実現可能性に関して

---

- ▶ 自治体テスト(テストX)を実施する自治体(地域A)は数多い
- ▶ 全国テスト(テストY)は、抽出調査になってもなお、高い割合で実施されている
- ▶ 同一集団であることは、学校単位、地域単位などのまとまりがわかれば確認できる
- ▶ 必ずしも個人のマッチングを必要としないので、2つのテストの受検者コードが異なっても(わからなくても)構わない

# 提案法の問題点

---

- ▶ 地域A, 地域Bの, 2つの地域の協力を得なければならない
- ▶ 個人のマッチングができない場合には, 2つのテストの相関情報等が得られない
- ▶ 学習指導要領が異なる集団間の比較を行うには, 範囲外の内容を除外したテストを構成しなければならず, 冊子全体の内容での比較にはならない
- ▶ 既存のテストをそのまま利用するので, より公的・大掛かりになるほど, 試験対策等の影響が懸念され, 批判されやすい