

# 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

東北大学大学院教育学研究科  
柴山 直

平成25年10月21日 於 文部科学省  
「全国的な学力調査に関する専門家会議」

## 大規模学力調査を設計する際の 技術的なポイント

- A) 対象集団は全体なのか標本なのか
- B) 調査領域は全体なのか部分なのか
- C) 学力変動をモニターするのか否か
- D) 実施は単体なのか相関も必要か



何を優先するかは調査の目的による

# 技術開発の際に想定した状況

限られた予算と時間の中で、もっとも効率よく  
全国レベルの教育施策に必要な情報を集める



- 標本調査方式を採用すること,
- **調査領域全体**を対象とすること,
- **経年変化**がモニターできること,
- 関連する外的情報との**相関分析**ができること

学力の測定モデルとして**項目反応理論(IRT)**を採用

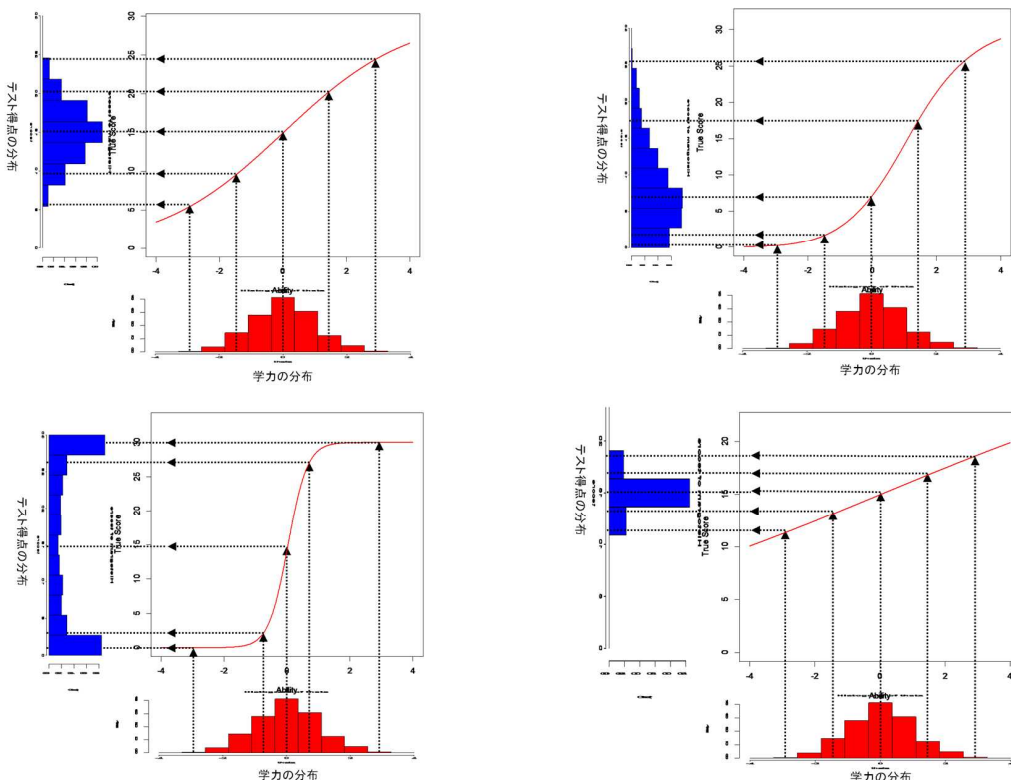


- 項目特性(困難度)と学力特性値( $\theta$ )が分離できる
- 項目を個別に管理できる

3

## 項目特性(困難度)と学力特性値( $\theta$ )の分離

(平成22年度報告書から再掲)



# 平成22年度の成果概略

成果 IRTベースの重複テスト分冊法の導入(算数・数学)が可能であることを示した。

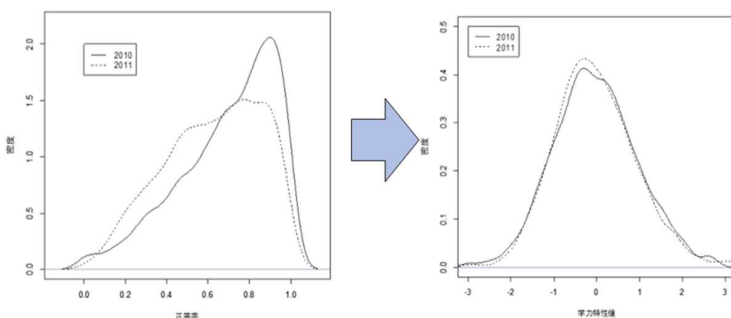
分冊番号	項目セットの番号 (項目数)					
	1	項目セット 1 (8)	項目セット 2 (8)			
2	共通セット (16)		項目セット 2 (8)	項目セット 3 (8)		
3				項目セット 3 (8)	項目セット 4 (8)	
4					項目セット 4 (8)	項目セット 5 (8)
5						項目セット 5 (8)

- 項目を個別管理できる性質を利用
- 従来の一冊子一斉実施方式によるよりも幅広い調査領域をカバーできる
- 問題は非公開

5

# 平成23年度の成果概略

## 成果1 経年比較を実現



テスト得点(正答数)      尺度値(学力特性値:  $\theta$ )

- 学習指導要領の改訂にともなう項目の一部入れ替え
- テストの難易度が上昇

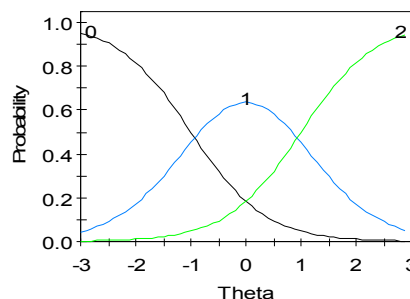


- 項目の困難度と学力特性の分離
- 項目の個別管理



- 尺度の等化(equating)

## 成果2 国語の記述式問題にも対応した



6

# 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 (平成24年度の成果概略)

1. 集団統計量のバイアスの最小化のため、**釣合い型不完備ブロックデザイン(BIBD)**を分冊を組み上げる際のデザインとして導入した。
2. 数学は3年間、国語は2年間の**経年変化を等化法によって把握**することができた。
3. 学力の形成要因などの背景情報と学力とのより精度の高い、また多様な分析が可能となる**推算値**を導入した。
4. 問題非公開の中での参加児童生徒への指導に役立つ**個別フィードバック**の方法を国語においても開発した。
5. IRT分析のためのツールとして**EasyEstimation**を導入した。

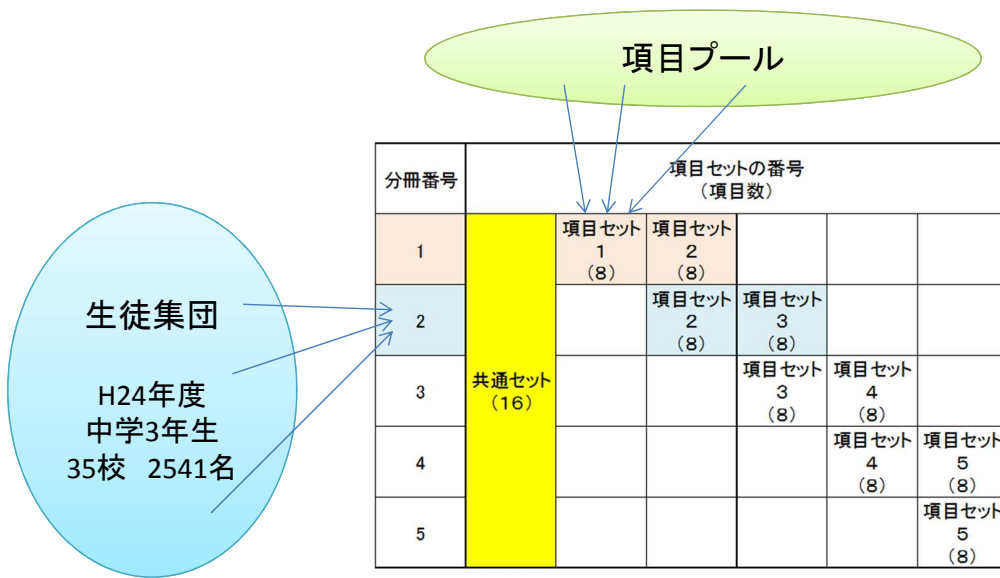
7

## 平成24年度の実施内容概略

- 対象学年 **中学3年生**
- 対象人数 **2541名**
- 実施校 協力自治体の全中学校 **35校**
- 実施時期 平成24年10月1日(月)～10月12日(金)
- 実施時間 上記の期間内の2授業時間  
+ 質問紙調査のための15分程度  
(説明や配布、回収の時間を含めて1授業時間45分)
- 出題教科 数学・国語・質問紙
- 出題範囲 現行の学習指導要領の内容構成にあわせる  
国語は**PISA型のリーディング・リテラシー問題**中心  
質問紙については学力の形成要因等
- 分冊数 数学・国語とも**7分冊**／質問票1部
- 解答数 中学校ともに一人当たり30問程度 問題冊子は回収
- 解答方式 **記述式**を含む記入式
- 個票返却 平成23年12月初旬

8

# (アイテム・)マトリックス・サンプリングのイメージ



! 単に項目セットを組み上げただけでは偏りが生じる  
(項目の出現回数・組合せパターン等々)

9

## 1. 釣り合い型不完備ブロックの導入 (Balanced Incomplete Block Design : BIBD)

- どの項目セットも分冊全体を通して配置される回数が等しくなるようにする.
- 項目セットの組み合わせパターンがたがいに同じ頻度で出現するようにする.
- 冊子の中での項目セットの配置位置に偏りがないようにする.

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

項目セット3

項目セット5

本調査研究(数学)で採用したBIB(7,4,2)デザイン

数学:1人1校時で3.5校時分

国語:1人1校時で2.3校時分

(PISA:1人2時間で6.5時間分)

10

# BIBD導入の効果

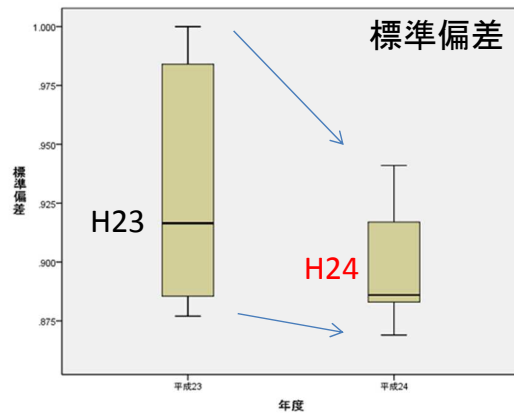
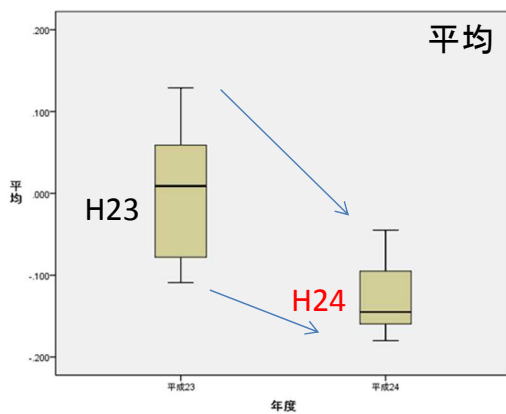
## 平成23年度 BIBDではない

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

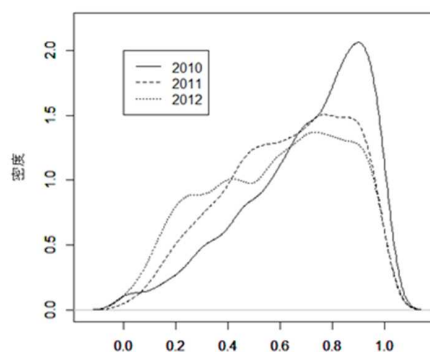
分冊ごとの学力特性値 $\theta$ の  
**集団統計量**(平均や標準偏差)の  
 ばらつきが**H24年度の方が**  
**均質化**されている



11

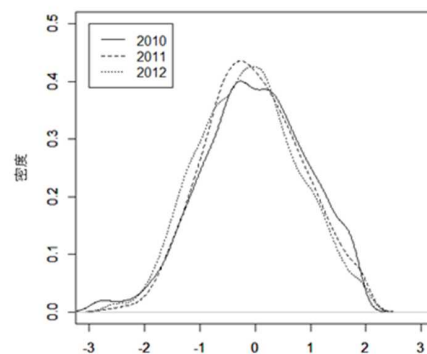
## 2. 尺度等化と経年比較分析(数学)

### 正答数得点による比較



	H22	H23	H24
平均	0.696	0.624	0.588
標準偏差	0.234	0.235	0.258

### 学力特性値による比較



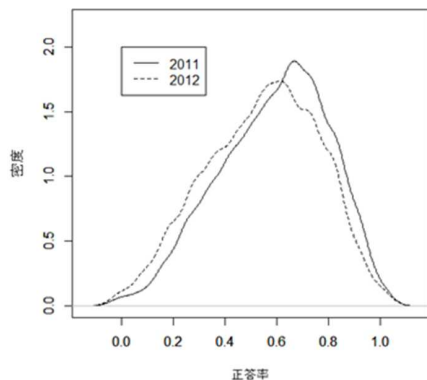
	H22	H23	H24
平均	0.000	-0.019	-0.126
標準偏差	0.945	0.879	0.900

- ・指導要領改訂等により毎年数項目を入れ替えている
- ・項目数は毎年異なる
- ・テストの難易度は毎年高くなっている
- ・H22とH24は10月実施, H23は11月実施

12

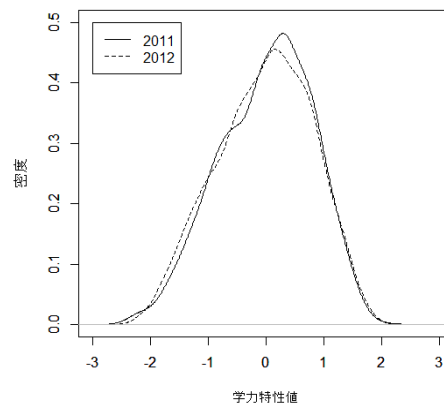
# 尺度等化と経年変化分析(国語)

## 正答数得点による比較



	H23	H24
平均	0.590	0.543
標準偏差	0.209	0.217

## 学力特性値による比較



	H23	H24
平均	0.000	-0.030
標準偏差	0.809	0.826

- ・分冊ごとの信頼性係数 数学 0.92前後 国語 0.44~0.67
- ・国語の信頼性が数学に比べて低い理由 大問形式への依存度が強い
- ・国語の問題は報告書に掲載して公開 項目セット1, 2が要検討

13

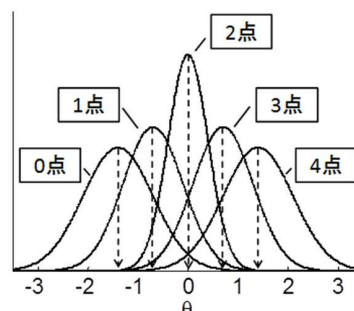
## 3. 推算値の導入

- 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

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

x: 項目反応パターン  
 θ: 学力特性  
 g(θ): 事前分布  
 f(x|θ): 尤度関数  
 h(θ|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)

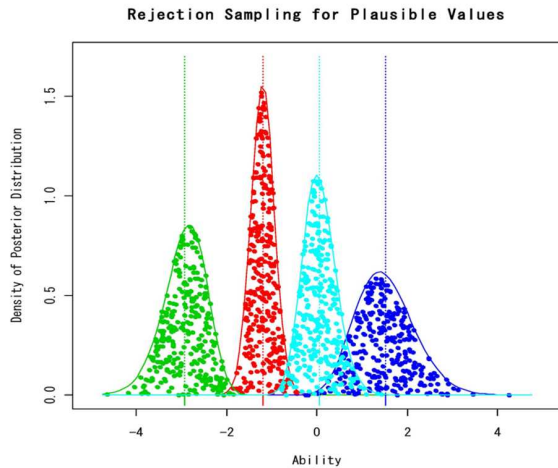
14

# 推算値を求める計算アルゴリズム(R)の公開

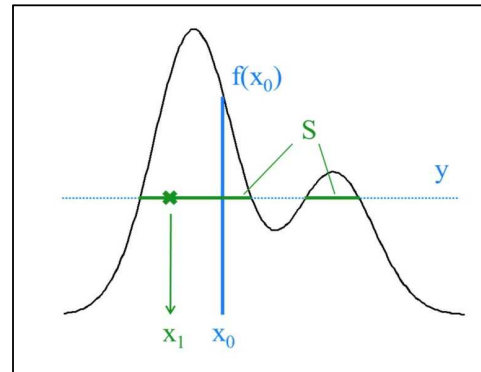
フォン・ノイマンの棄却法  
(R: 柴山)



スライス・サンプリング法による  
クロスチェック (MATLAB: 佐藤)



柴山(2011)より引用

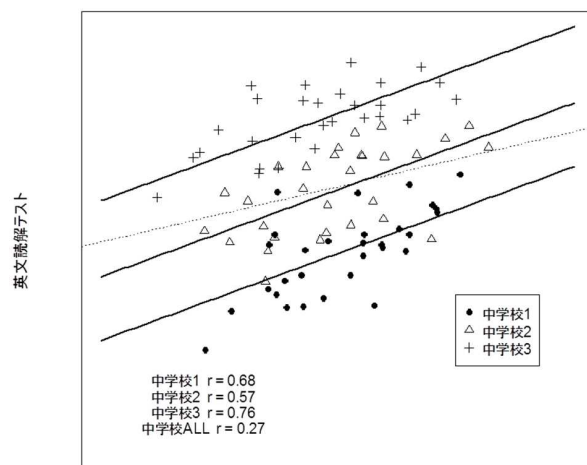
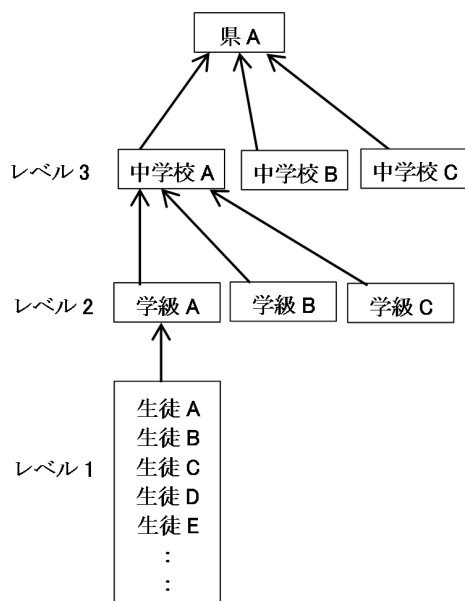


柴山・佐藤・宮田・中野(2013)より引用

報告書: 付録2にRスクリプトは掲載

EasyEstimationにもオプションとして付加予定

## 階層化されたデータとマルチレベル分析 (説明のための仮想例)



英単語テスト



# 推算値を使ったマルチレベル分析の例(第9章)

## 推算値を利用する場合のデータ構造

生徒	推算値	質問項目				
		1	...	j	...	m
生徒 1	PV1_1	X11 <sub>1</sub>	...	X1j <sub>1</sub>	...	X1m <sub>1</sub>
	PV2_1	X11 <sub>1</sub>	...	X1j <sub>1</sub>	...	X1m <sub>1</sub>
	PV3_1	X11 <sub>1</sub>	...	X1j <sub>1</sub>	...	X1m <sub>1</sub>
	PV4_1	X11 <sub>1</sub>	...	X1j <sub>1</sub>	...	X1m <sub>1</sub>
	PV5_1	X11 <sub>1</sub>	...	X1j <sub>1</sub>	...	X1m <sub>1</sub>
}						
生徒 i	PV1_i	Xi1 <sub>i</sub>	...	Xij <sub>i</sub>	...	Xim <sub>i</sub>
	PV2_i	Xi1 <sub>i</sub>	...	Xij <sub>i</sub>	...	Xim <sub>i</sub>
	PV3_i	Xi1 <sub>i</sub>	...	Xij <sub>i</sub>	...	Xim <sub>i</sub>
	PV4_i	Xi1 <sub>i</sub>	...	Xij <sub>i</sub>	...	Xim <sub>i</sub>
	PV5_i	Xi1 <sub>i</sub>	...	Xij <sub>i</sub>	...	Xim <sub>i</sub>
}						
生徒 n	PV1_n	Xn1 <sub>n</sub>	...	Xnj <sub>n</sub>	...	Xnm <sub>n</sub>
	PV2_n	Xn1 <sub>n</sub>	...	Xnj <sub>n</sub>	...	Xnm <sub>n</sub>
	PV3_n	Xn1 <sub>n</sub>	...	Xnj <sub>n</sub>	...	Xnm <sub>n</sub>
	PV4_n	Xn1 <sub>n</sub>	...	Xnj <sub>n</sub>	...	Xnm <sub>n</sub>
	PV5_n	Xn1 <sub>n</sub>	...	Xnj <sub>n</sub>	...	Xnm <sub>n</sub>

## マルチレベル分析の結果

固定効果	ヌルモデル		モデル1	
	推定値	標準誤差	推定値	標準誤差
切片	-0.17***	0.04	-1.73***	0.05
将来の進学希望			0.19***	0.00
知的好奇心			0.02***	0.00
外発的動機付け			0.02***	0.00
震災ダミー			0.12	0.08
変量効果				
$\sigma^2_{u_{0j}}$	0.04		0.03	
$\sigma^2_{e_{ij}}$	0.74		0.57	

### モデル1

レベル1 (生徒レベル) 数学PV<sub>ij</sub> =  $\beta_{0j}$  +  $\beta_1$  将来の進学希望<sub>ij</sub>  
 +  $\beta_2$  知的好奇心<sub>ij</sub>  
 +  $\beta_3$  外発的動機づけ<sub>ij</sub> +  $\beta_4$  震災ダミー<sub>ij</sub>  
 + e<sub>ij</sub>

レベル2 (学校レベル)  $\beta_{0j} = \beta_0 + u_{0j}$

## 4. 問題非公開の場合における指導のための個別フィードバックの方法(国語)

国語調査結果シート							
学校番号	学校名		組	出席番号	性別		
学年	領域	目標	指導事項の内容	読みの種類	平均正答率	解いた問題の正誤	解かなかった問題の推定正答率
第一学年	「書くこと」	目的や意図に応じ、日常生活にかかわることなどについて、構成を考えたの順に書く能力を身に付けさせること。進んで文章を書いて考えをまとめようとする態度を育てる。	(イ) 集めた材料を分類するなどして整理することともに、段落の役割を考慮して文章を構成すること。	記述	29%	-	43%
			記述	33%	×	-	
			記述	71%	○	-	
			記述	53%	-	71%	
			記述	24%	-	33%	
			(フ) 伝えたい事実や事柄について、自分の考えや気持ちを根拠を明確にして書くこと。	記述	29%	-	43%
	「読むこと」	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けさせること。読書を通してものの見方や考え方を広げようとする態度を育てる。	(イ) 文章の中心的な部分と付加的な部分、事実と意見などを読み分け、目的や必要に応じて要約したり要旨をとらえたりすること。	選択	59%	○	-
			選択	46%	×	-	
			選択	62%	×	-	
			選択	32%	-	44%	
			選択	58%	-	74%	
			(フ) 場面や登場人物などの描写に注意して読み、内容の理解に役立てること。	選択	79%	-	90%
選択	23%	-	27%				
選択	43%	-	53%				
選択	66%	○	-				
選択	62%	○	-				
選択	55%	-	63%				
選択	82%	-	96%				

# リーディング・リテラシー(RL)と従来の国語科 における「読解力」の違い(第5章)

## リーディング・リテラシー

- 読まれるテキスト
  - 文章形態をとる「連続型テキスト」
  - 図や表, 公告, 証明書等々文章形態とは言いがたい「非連続型テキスト」
- 書くことと読むこととの関係
  - 長く記述する解答が求められる
  - 自分が理解したことを表現する
  - 異質な文化背景の相手の存在
- テキストの内容への態度
  - 自分の先行知識・価値観などから評価・判断させる

## 従来の国語科における「読解力」

- 読まれるテキスト
  - 文学的文章
  - 説明的文章が2大ジャンル
- 「書くこと」と「読むこと」
  - カリキュラム上は別
  - 比較的等質の家庭環境・文化環境をもった生徒達が対象
- テキストの内容への態度
  - 自分の読み方について述べる  
ときは、テキストの中に根拠を見いだす指導

- ! 個別フィードバックのためのシートではRL問題を国語の指導要領(現行)に対応させた
- ! 数学の活用能力とリーディング・リテラシーには中程度の相関がある(第6章)

19

## 5. IRT分析プログラムEasyEstimation(第10章)

- EasyEstimationシリーズ(熊谷2009,2012)
  - <http://irtanalysis.main.jp/>
- 特徴
  - 研究目的に限り無料で利用できる国産のフリーソフトウェア
  - マウス操作のみで分析可能なGUIにより, テスト分析の入門者においても容易に分析を実行できる.
  - 多母集団分析, 一部の項目母数固定による分析, 項目特性曲線やテスト情報量曲線の描画など, 実用上必要な分析オプションを用意
  - 計算結果は他のソフトウェアとのクロスチェック済み
- シリーズ構成
  - EasyEstimation: 2値型データ用
  - EasyEstGRM: 順序付多値型データ用  
~たとえば数学の証明問題・国語の記述問題の採点結果
  - EasyNominal: 名義尺度データ用
  - EasyDIF: DIF(特異項目機能)分析用
  - 推算値(PVs)計算機能追加予定

20