平成２７・２８年度

広域科学教科教育学研究経費報告書

アクティブ・ラーニング型理科授業とその評価法の系統的研究

研究代表者　　新田　英雄

平成 29 年 3 月

東京学芸大学

研究組織

（◎は研究代表者）

新田英雄◎	東京学芸大学	教授	自然系教育講座
植松晴子	東京学芸大学	准教授	自然系教育講座
市原光太郎	附属高等学校	教諭	
宮崎達朗	附属世田谷中学校	教諭	
堀井孝彦	附属世田谷小学校	教諭	
西村塁太	連合学校教育学研究科（神奈川県立逗葉高等学校）	課程所属学生	自然系教育講座

上記の他，本学大学院修士課程学生であった金森大和（H27 年度修士課程修了），後藤敬佑（H28 年度同修了）の研究協力を得て実施した。

平成２７・２８年度広域科学教科教育学研究経費
「アクティブ・ラーニング型理科授業とその評価法の系統的研究」研究報告書

目次

１　研究の概要

2014 年に中央教育審議会から『新しい時代にふさわしい高大接続の実現に向けた高等学校教育、大学教育、大学入学者選抜の一体的改革について』が公表されてから，アクティブ・ラーニングが急速に注目されるようになった。その影響は高校・大学教育にとどまらず，義務教育にも及んでおり，具体的な授業方法の開発が喫緊の課題となっている。アクティブ・ラーニングが普及していくためには，従来の伝統的な講義の授業効果に比べてアクティブ・ラーニングの教育効果の方が高いことを，定量的評価に基づいて明確に示す必要がある。しかし，そのような評価法の研究は進んでいない。

一方，海外，特に米国では自然科学，特に物理学分野において，アクティブ・ラーニングの導入とその評価法が発達している。その背景として，学生・生徒は，自然科学を学ぶ前から自然現象に対して経験に基づいた強固な素朴概念を有しており，これらが正しい科学概念に置き換わるためには，学生・生徒自らが能動的に科学概念を現象に適用していく学習が必要であることが明らかになったことがある。

そこで本研究では，米国のアクティブ・ラーニングとその評価法を参照しつつ，日本の生徒の発達段階を踏まえ，小中高大を見通した一貫性のある知識構造の形成過程を重視し，素朴概念を系統的に克服させることに主眼を置いたアクティブ・ラーニング型理科授業および定量的評価方法を，実践的に研究開発することを目的とする。

本研究報告書では，まず，海外の科学教育研究の動向について述べる。次に，科学的な授業効果の評価法を確立する原動力となった Force Concept Inventory（FCI）を小中高に実施し，科学概念の形成における男女差の変化に注目した分析結果を示す。その後，高等学校物理において実施したピア・インストラクション（アクティブ・ラーニング型授業のひとつ）の中で，具体的に生徒がどのような議論を行っているのかの詳細な分析について述べる。さらに，調査テストの分析および授業ゲインの新たな指標を与えるために今後重要となる現代テスト理論に基づいた FCI の分析について述べる。最後に，本論のまとめを行う。

２　海外におけるアクティブ・ラーニング型科学教育研究の動向

日本の初等中等教育においては，「仮説実験授業」や「到達目標・課題解決方式（玉田方式)」といったアクティブ・ラーニング型の理科教育が半世紀以上も前から実践されてきている。しかしながら，高等教育においてはアクティブ・ラーニングを自然科学教育に取り入れる試みは始まったばかりである。

一方，米国では，物理，化学，工学，生物学，地球科学，天文学といった，理学・工学の各専攻分野に固有な教授 - 学習過程を対象とする研究領域が成立している。これらを総称して「専門分野に基盤をおいた教育研究」(Discipline-Based Education Research：DBER) という[1]。DBER に属する各研究領域は，背景となる教育学，心理学等の知見や研究手法においては共通部分を多く持つが，立脚している学問分野の専門性に直結した領域固有性を有している。

物理教育研究（Physics Education Research，PER）を例にとって具体的に説明するならば，PER とは，教育心理学，認知心理学等で確立された知見と実験手法，データ取得手法などに基盤をおき，それを物理教育固有の問題に適用するという研究手法をとる。なお，データを取得する際の条件制御や誤差の評価，仮説の検定といった統計学的データ処理も，研究に必須の要素である。

1980 年代に McDermott のグループは，発達心理学者 Piaget が開発した課題の手法を学生の物理概念理解の調査に応用し，多くの学生が速度・加速度といった基礎的な概念を理解していないままであることを定量的に示した。なお，同時期には，M. McCloskey の素朴インペタス理論や，J. Clement の MIF（motion implies a force）誤概念といった，科学教育研究として広く知られる成果も発表されている。これらの研究によって，多くの学生が物理を暗記科目としてとらえており，一貫性のある物理の理解に至っていないことが明らかにされた。特に力学教育の研究においては，学生が正しい物理概念とは異なる強固な素朴概念を持っており，力学の授業が終わった後もそれら素朴概念が解消されずにそのまま保持されてしまうことが明らかにされた。これは教員側として予想しない驚くべき結果であった。

例えば，真上に投げ上げられた，上昇中の物体に加わっている力は何かと学生に尋ねると，「重力と，投げ上げたときに手から受けた力」と答える学生が予想を超えて多い。これは上記 McCloskey のインペタス素朴概念，MIF 誤概念の典型例である。しかし，このように答える学生でも，試験では自由落下の試験問題に正しく解答する場合が多い。附属高校のある生徒は，投げ上げの問題に回答する際に，思った通りに回答しなさいと言った教師に対し，

「私は，空中に投げ上げられた物体には重力しかはたらいていないというのが物理の正解なのを知っています。でも，本当に自分の考えを答えてよいならば，手からもらった力もはたらいているはずだと答えます。」

と告白している。この生徒は，物理の授業で教えられた力学と，実生活の経験に根差した概念で構成された「力学」像という 2 重の力学像を形成してしまったのである。そして，前者

は非常に特殊な状況でしか成り立たないのであって，現実の世界は後者が支配しているのだと考えてしまっている。上記の 2 重構造を解消するためには，後者が前者によって再解釈できること，つまり経験によって形成された運動に関する知識と概念が，正しい物理法則によって一貫性のある形で解釈できることを本人が理解し，その合理性に納得する必要がある。このような学習は受け身な学習態度では決して進まない。自ら積極的に物理法則を現象に適用していくアクティブ・ラーニング[1]が欠かせない理由はここにある。

学生の持つ素朴概念や物理学習における問題点が明らかになっていくと，今度は，素朴概念の保持率を精度よく測定する方法の開発が課題となった。この方向の努力の結晶として誕生したのが，Hestenes らによる「力の概念調査」FCI（Force Concept Inventory）である[2]。Hestenes らは，第 1 段階の物理教育研究で見出された知見に基づき，学生の持つ素朴概念と正しい物理概念とを識別する多肢選択方式の調査紙を開発した。FCI の真価は，その素朴概念識別力にある。FCI の誤答選択肢は，膨大なインタビューと記述式調査に裏付けられた素朴概念で構成されており，学生がどのような素朴概念を持っているかを定量的に調査するのに極めて効果的である。なお，誤答選択肢が素朴概念ごとに分類されたご概念分類表が発表されている。本報告書では，FCI とその誤概念分類表を後で用いる。

FCI の開発によって，同一の基準により学生の理解度や素朴概念の保持率が測定できるようになった。また，事前・事後調査に利用することによって，授業効果を同一の基準で測定することが可能となった。このことは非常に大きな意義を有している。なぜならば，「本当にアクティブ・ラーニングを取り入れた授業は，伝統的な講義よりも効果的なのだろうか」という素朴な疑問に答えられなければ，アクティブ・ラーニング型授業の普及は困難だからである。

1998 年に R. Hake によって発表された論文は，その意味で画期的であった[3]。Hake は，公立高校から米国トップの研究大学までの，FCI を用いた事前・事後調査データ[2]を，数年間かけて集めた。データは，62 の授業（総学生数 N=6542 人）に達した。事前調査と事後調査の平均得点率(%) をそれぞれ$\langle S_{\mathrm{pre}}\rangle$および$\langle S_{\mathrm{post}}\rangle$と置くと，その授業の平均ゲイン$\langle G\rangle$は次の式で定義される。

$$\langle G\rangle = \langle S_{\mathrm{post}}\rangle - \langle S_{\mathrm{pre}}\rangle$$

平均ゲインは使用した調査紙で測定できる概念の理解が授業によってどれだけ増加したかを表しており，授業効果を表す指標である。しかし，様々な校種間での授業効果を比較するとなると，$\langle S_{\mathrm{pre}}\rangle$の違いによって授業による伸びの余地に差ができてしまうため，$\langle G\rangle$はそのままでは比較に適さない。そこで，事前調査の平均得点率が$\langle S_{\mathrm{pre}}\rangle$である授業の，伸びの余地（$\langle G\rangle$の最大値）

---

[1]認知心理学，教育心理学の基本的な知見であるが，学習には，知識や情報を選択し，構成し，統合する過程が必須である。この過程を含む学習を「アクティブ・ラーニング」と呼ぶならば，すべての学習は何らかの形でアクティブ・ラーニングでなければ成立しない。

[2] ここで事前・事後調査とは，授業効果を測定するために，同一の調査紙を用いて 1 学期間の授業の開始前と開始後に調査を行うことをいう。

$$\langle G_{\max} \rangle = 100 - \langle S_{\mathrm{pre}} \rangle$$

によって規格化されたゲイン（規格化平均ゲイン）

$$g = \frac{\langle G \rangle}{\langle G_{\max} \rangle} = \frac{\langle S_{\mathrm{post}} \rangle - \langle S_{\mathrm{pre}} \rangle}{100 - \langle S_{\mathrm{pre}} \rangle}$$

を Hake は導入し，授業効果を比較した。

データとして得られた各授業に対してgを求め，それらの平均値を授業形式ごとに求めると，伝統的な講義が 0.23，能動学習型が 0.48 となり，2 倍以上も異なることを Hake は見出した。この結果は大きな衝撃をもたらし，米国の多くの大学でアクティブ・ラーニングが取り入れられる契機となった。また，Hake の論文以降，規格化ゲインを用いて授業効果を表すことが定着した。

最近の PER で取り組まれている難題のひとつに，ジェンダーギャップ問題がある。物理についての関心や学力に男女差があることは古くから気づかれていたが，物理教育研究が進むにつれて，その実態が定量的に示されるようになった。Madsen らは，FCI 等の概念調査を用いた米英の調査結果をレビューし，平均 13%の得点差（ジェンダー・ギャップ値という）があることを見出した[4]。なお，日本でも同様のジェンダーギャップ値が見出されている[5]。Madsen らはジェンダーギャップ値が生じる原因を詳細かつ多角的に分析したが，その要因は見出せていない。

また，物理の学習によって得られるものに対する期待や学習観はなかなか向上せず，物理の授業後にはさらに低下する傾向にあることがわかっている。この傾向はアクティブ・ラーニング型の授業でも同様である。アクティブ・ラーニングによって物理の理解度を高めることはできても，授業が終わった後に物理嫌いを増やしてしまうようでは片手落ちである。この問題を解消するための努力も研究ベースで多方面から行われ，いくつかの成功例も報告されているが，まだ明確な結論を得るには至っていない。

参考文献

[1] S. R. Singer et al. (eds.): *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering* (The National Academies Press, 2012) 同出版社のウェブサイトから無料でダウンロードできる。

[2] D. Hestenes, M. Wells and G. Swackhamer: Phys. Teacher **30** (1992) 141.

[3] R. Hake: Am. J. Phys. **66** (1998) 64.

[4] A. Madsen, S. B. McKagan and E. C. Sayre: Phys. Rev. ST Phys. Educ. Res. **9** (2013) 020121.

[5] 新田英雄・植松晴子・森口真靖: 大学の物理教育 **20S** (2014) S53 -S56.

３　小中高を通じたFCI調査の分析：ジェンダー差の観点から

理工系の大学・学部に進学する高校女子生徒の割合は，男子生徒よりも低く，特に物理を専攻する女子学生は他の自然科学分野と比較して少ないことはよく知られている。また，英米の多くの大学で実施されたFCI (Force Concept Inventory)，FMCE (Force and Motion Conceptual Evaluation)等の力学概念理解度調査において，男子学生の成績が女子学生の成績を有意に上回ることがMadsenらのメタ分析によって示された[1]。Madsenらはこの男女差（ジェンダー差）を

$$G = \langle \text{男子学生の正答率\%} \rangle - \langle \text{女子学生の正答率\%} \rangle$$

で定義されるジェンダーギャップ値Gで表し，Gの平均値が約13%であることを報告している。ここに $\langle \cdots \rangle$ は被験者集団の平均値を表す。なお，日本でもほぼ同程度のジェンダーギャップ値が報告されている[2]。このようにジェンダーギャップが存在することは多くの国で共通する傾向であるが，その原因は明確ではない。Madsenらは，様々な原因が複合的に作用しているのではないかと述べている。

一方，ジェンダーギャップを解消あるいは減少させるにはどのような授業法が効果的かの研究もおこなわれてきている。Lorenzoらは生徒・学生のアクティブ・ラーニングを取り入れた相互作用型授業によってジェンダーギャップが解消されると報告しているが，逆にジェンダーギャップが増大したという結果もあり，結論は得られていない[1)]。ただし，アクティブ・ラーニングを行うことは，従来の講義型授業と比較して，より概念理解度調査の成績を向上させるという結果は変わらない。

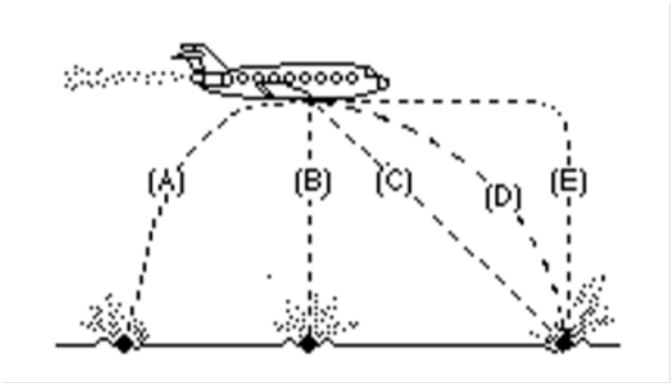
以上のようにジェンダーギャップ問題は，海外においても未解決の研究課題であるが，国内では研究例そのものが少ない。そこで，本学と附属高校で実施してきたFCIデータを調べてみたところ，英米の大学と同様に，10%程度のジェンダー差が見出された[2]。特に，附属高校においても必修2単位物理・物理基礎の授業開始前に10%以上のジェンダー差が生じていることは注目に値する。すなわち，高等学校から物理としての学習が開始される以前から，ジェンダー差は生じているのである。このジェンダー差は，学習歴のいつ頃から生じるのであろうか。

上記を調べるため，附属世田谷小学校，附属世田谷中学校において，小学6年生，中学校2年生（力学分野履修前），3年生（同履修後），高校2年生（物理基礎履修前および後）を対象にFCIを系統的に実施した。その結果，小学生で7%，力学分野履修前の中学2年生で8%，同履修直後の3年生で6%程度のジェンダー差があることがわかった[3]。

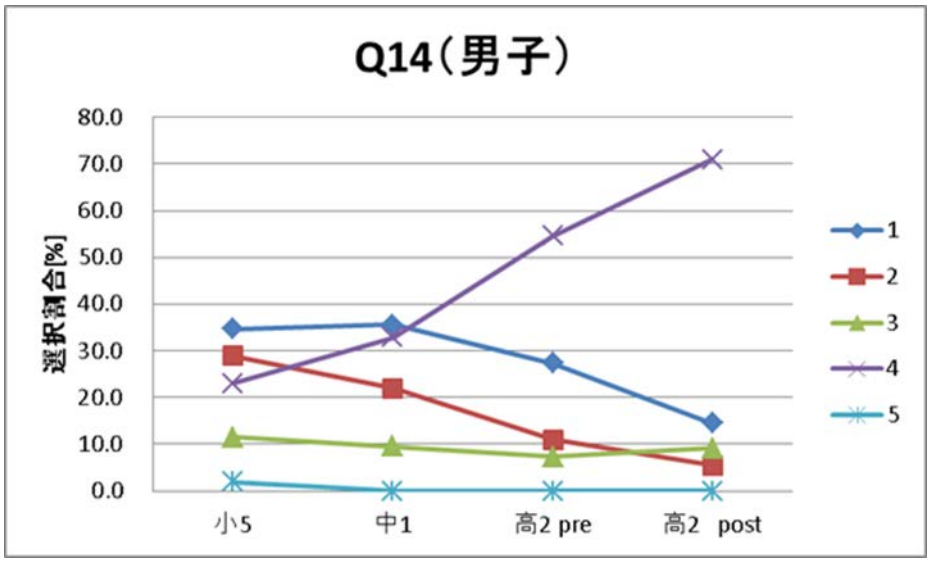
興味深いのは中学3年生で，履修直後でのFCIは履修の影響が強く反映され，作用反作用などの中学校理科で扱っている内容に関係する設問ではジェンダー差が縮まる傾向にある。ところが，高校2年生でのプレテストでは再びジェンダー差が広がっている。これは物

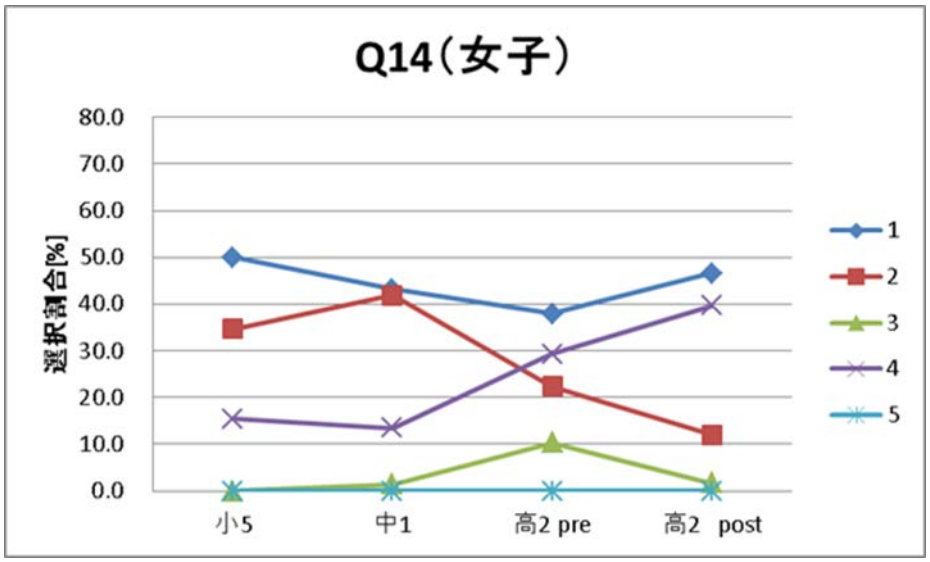
理概念の定着に問題があることを示唆する。なお，当然ながら本研究の母集団は各学年で異なっているが，附属学校ということから教育内容，児童・生徒は比較的均質であると思われる。

以下に，特にジェンダー差の大きかった設問について，具体的な結果を示す。下図は，FCI の設問 14（地上から眺めた時の飛行機から落ちたボールの軌道を回答する）について，横軸に学年，縦軸に各選択肢の回答率をとった男女別グラフである。正答は 4 であるが，女子は選択肢 1（後方に落下）が高 2 のポストテスト段階においても正答を上回っており，その数は 50%近い。真下に落下する選択肢 2 を回答する女子も高 2 段階では男子より 2 倍ほど残っている。この結果は高 2 における平均的なジェンダー差が 10%程度であることを考えると，本設問で問うている物理概念に対する学習の効果が女子において低い，ということにはならないであろう。スポーツをはじめとした経験に原因があるのか，知覚に関連するのか，可能性はさまざまであるが，特定するためにはさらに慎重にジェンダー差の研究を発展させる必要がある。



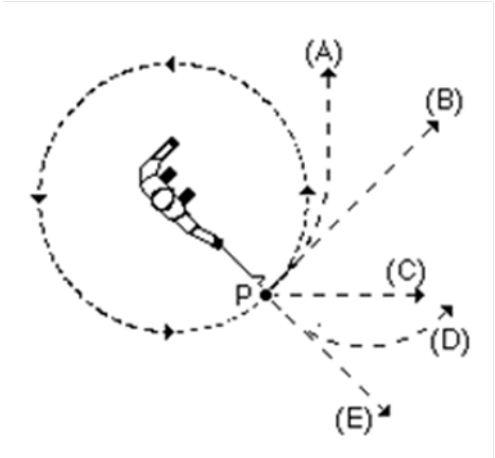
図：設問 14 の選択肢。



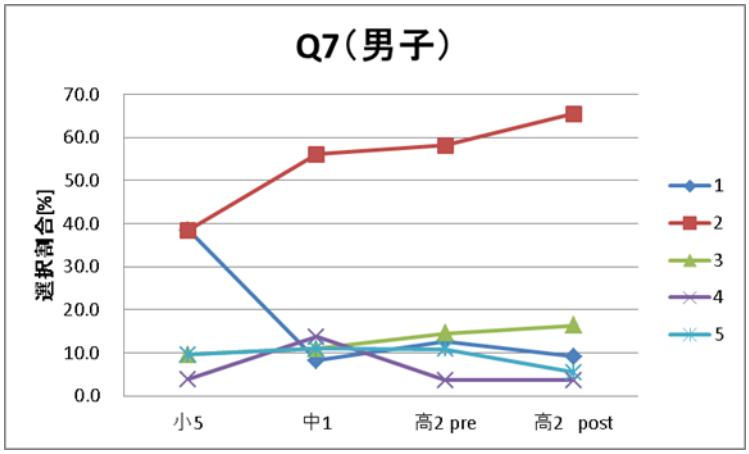


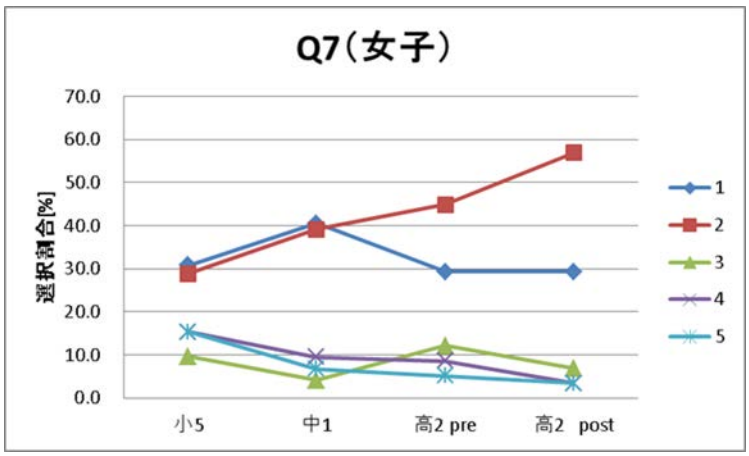
図：FCI 設問 14 の回答分布。グラフの 1～5 は設問の選択肢 A～E に対応する。

下図は FCI 設問 7 の図で，砲丸投げのロープが P 点で切れた時の砲丸の軌道を問う問題である。ひもが切れた位置から慣性の法則にしたがって直進する軌道を表す選択肢 2 が正答であるが，女子の場合，高 2 段階でも選択肢 1 が 30%程度選択されている。これは，円軌道を描くための拘束が途絶えた後も円運動していた効果が残るという，広い意味での「インペタス誤概念」を有する女子生徒が多いことを示している。

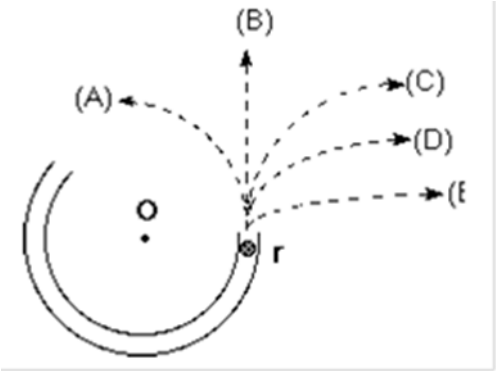


図：設問 7。

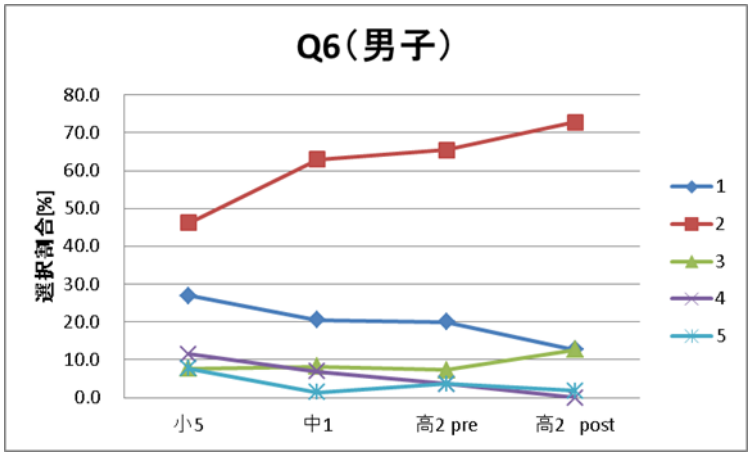


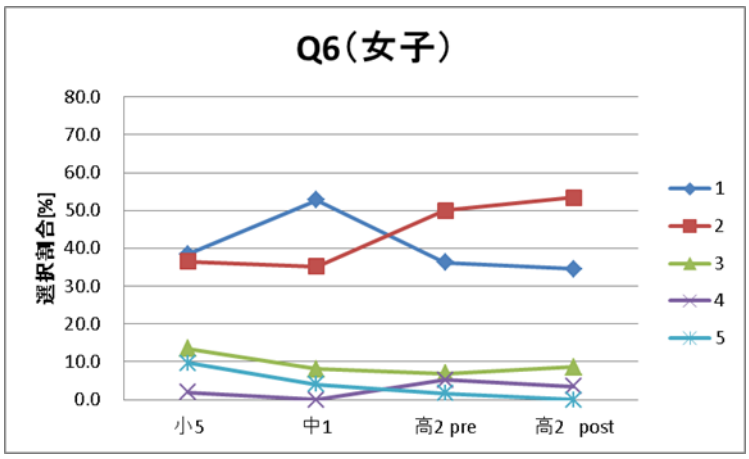


下図は設問 6 の結果を表す。物理的には設問 7 と類似のものであるが，正答率とその変化や，誤答選択肢 1 の割合は設問 6 と 7 で若干異なっている。選択肢 1 を回答する生徒は男女ともに若干多くなっている。



図：設問 6 の図。円形のチューブを進む球の軌道を上から見たところ。





他の設問を含めた意味で，小中高を通じた FCI 実施によるジェンダー差調査で分かったことをまとめる。まず，FCI で分類される素朴概念において，ジェンダーギャップは物理未習段階から生じていることがあげられる。このことは，本論で例示した設問の男女別グラフからも見て取れる。また，小学校では見られなかったジェンダーギャップが中学校段階で生じているケースもある。

ジェンダー差をアクティブ・ラーニング型授業で解消する方向に持っていけるかを調べるため，附属高校で実施している PI 型授業において，ジェンダー差が特に大きい誤概念に対してアクティブ・ラーニングを行うことにより変化が生じるかを 2015 年度に実践的に調べた。具体例としては，運動の法則を扱った後，手で加えた力が運動物体に蓄えられるというイメージは間違いであり，運動の法則に矛盾することを，簡易ホバークラフトを用いた速度の合成の PI と簡単な生徒実験を通じて理解させることを試みた。その結果，当該の FCI 設問８および 13 において，下表のように，男女ともに当該の誤概念が減少することがわかった。しかしながら，他の誤概念においては，PI を用いたアクティブ・ラーニング型授業においては，男女ともに誤概念を減少させることに効果があり，ジェンダー差を縮める方向に向かわせることがわかった。しかしながら，ジェンダーギャップ値は FCI ポストテストで

やや減少したものの，依然として 10%近い値となっている。

	性別	プレ平均正答率(%)	ポスト平均正答率(%)	規格化ゲイン
2014 年				
設問 8	女子	50.0	58.6	0.17
	男子	50.9	56.4	0.11
設問 13	女子	8.6	44.8	0.40
	男子	32.7	70.9	0.57
2015 年				
設問 8	女子	47.2	91.7	0.84
	男子	65.7	94.3	0.83
設問 13	女子	13.9	61.1	0.55
	男子	37.1	60.0	0.36

表：FCI の設問におけるジェンダー差の変化

教育基本法第四条には「すべて国民は，ひとしく，その能力に応じた教育を受ける機会を与えられなければならず，人種，信条，性別，社会的身分，経済的地位又は門地によって，教育上差別されない.」と記されている。法的に，ジェンダー差が生じるような教材を用いて授業を行ってはならないことが規定されているのである。しかしながら，物理の教材には，野球をはじめとしたスポーツ，自動車をはじめとした機械など，男子児童・生徒の方が興味を持ちやすい題材が使われている場合がある。物理教育には長い歴史があり，少なくとも 20 世紀半ばまでは男性を中心として行われてきた。そこで使われてきた教材の中には，教師側が気づきにくいジェンダーバイアスが潜んでいる可能性がある。今後，ジェンダー差に焦点を当てた教材研究，授業研究の充実が望まれる。

参考文献

[1] A. Madsen, S. B. McKagan and E. C. Sayre: Phys. Rev. ST Phys. Educ. Res. **9** (2013) 020121.

[2] 新田英雄・植松晴子・森口真靖: 大学の物理教育 **20S** (2014) S53 -S56.

[3] 本研究のデータ解析には，金森大和氏の協力を得た。

４　ピア・インストラクション型授業のプロトコル分析

これまでも述べてきたが，生徒同士の議論や，実験などを取り入れたアクティブ・ラーニングの中で，ピア・インストラクション(PI)型授業は，比較的導入しやすい授業法の一つである。生徒同士の議論を取り入れた授業を成功させるための鍵は，議論の内容をどれだけ充実させることができるかであるが，充実した議論をさせるためには，生徒がどのような議論をするのかを教師がある程度予測できなければならない。

しかしながら，海外における研究では，PI における生徒の発言は，教師の予想を超えた内容が多く，予測することは困難であるということが示されている。日本においてはそのような研究がなされてきていない。

そこで本研究では，PI の議論を文章化することで詳細にプロトコル分析し，生徒の知識構造の変化，グループ内の相互作用がどのように機能しているのかを探った[1]。なお，プロトコル分析を行う中で，生徒の発言をより詳細に分析するために，Sweller らによる認知負荷理論(Cognitive Load Theory (CLT))を援用して発言を分類した。さらに，議論全体の流れを踏まえ，相互作用の度合いなどに応じて議論の分類も行い，議論パターンの比較を行った。

プロトコル分析した授業実践は，東京学芸大学附属高等学校の平成 27 年度第２学年 1 クラスと，平成 28 年度第２学年 3 クラスで行った。授業実践校では実験を中心とした探求的な授業カリキュラムが構成されており，それにできるだけ沿う形で実験前の講義や実験後の分析において PI を行った。各グループ（10 グループある）の机上に本研究経費で購入したミーティングレコーダー(KING JIM MR 360)を置き，PI における生徒同士の議論を録画・録音し，得られたデータをもとにプロトコル分析を行った。

プロトコル分析は，授業実践時に録画・録音したビデオデータの中から，ある程度議論が長く続いているグループを抽出して生徒の発言の文字起こしを行い，発言の時系列で並ぶように表でまとめて行った。生徒の発言をさらに詳細に分析するために，認知負荷理論を援用し，認知負荷に対応する発言があると仮定し，Intrinsic Remark (IR), Extraneous Remark (ER), Germane Remark (GR)の 3 種類に発言を分類した。それぞれの発言の特徴を下表に示す。

表：生徒の発言の種類と特徴

発言	特徴
IR	問題解決に必要不可欠な発言
ER	問題解決には不必要な，誤った考えなどの発言
GR	問題解決だけでなく，学習をさらに促進するような発言

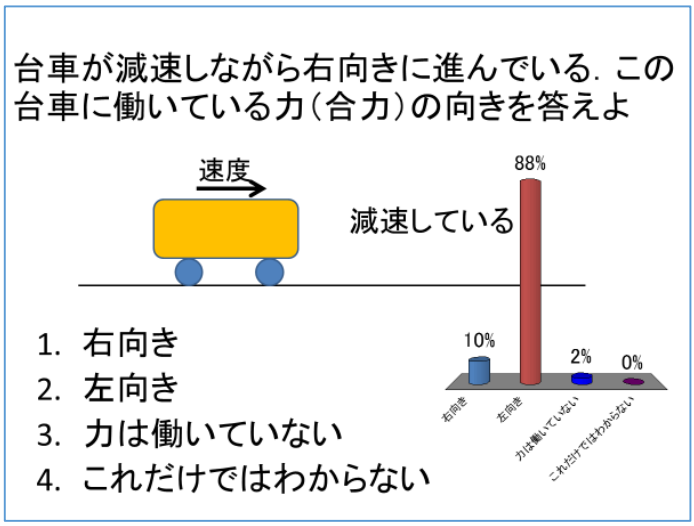
ただし，表におけるそれぞれの発言の分類は，議論の流れや生徒の様子などから判断されるものであり，グループ内の相互作用の様子を考察する際に定性的に分析するためのひとつの目安にすぎないということを注意しておく。

プロトコル分析を行ったデータに対して，議論全体の流れや，生徒の納得の様子などから判断し，議論パターンを以下の表のように 5 つに分類した。

表： 議論の分類

議論パターン	分類基準
PI 成立型	生徒同士の相互作用があり，問題点がほぼ解決している議論
リーダー主導型	一人が意見を述べ周りがそれに同意して終了している議論
PI 未解決型	相互作用はあるが，議論中にでた疑問点が解決しなかった議論
分裂型	相互作用をしているとは言えない議論
PI 失敗型	相互作用が間違った方向に進んでしまった議論

次に，プロトコル分析の一例を示す。



図：プロトコル例の PI 問題

下のプロトコルは「PI 成立型」に分類した議論である。表の列について説明する。通番は発話の順序，発言者欄の「４－A5」といった数字は，学級 4，5 班，生徒 A，発言は PI における生徒の発言，備考欄は発言内容への注釈，分類の意図は「生徒の発言の種類と特徴」を分類する根拠などを表す。右端欄の棒状の部分は，生徒が理解しているかどうかを示しており，桃色は理解した状態，グレーは理解に至っていない状態，白は判断できないものを表す。青色に塗られたプロトコルは Extraneous remark（ER），茶色は Intrinsic remark（IR），桃色は理解した瞬間を表すと思われる発話，抹茶色は Germane remark（GR）を表す。議論前は 2 名が選択肢１（進行方向に力がはたらくというインペタス誤概念に対応する選択肢）を選んでいるが，議論後に，全員正答（2 番）に到達している。

通番	発言者	発言	備考	分類の意図	A	B	C	D	E
1	4-A5	②			2	2	1	1	
2	4-B5	②							
3	4-D5	あたし①							
4	4-C5	まっておれも①なんだけど							
5	4-D5	減速してるけど…でも結局右に動いているから…		誤った考え					
6	4-B5	わたしも最初そう思ったんだけど…							
7	4-D5	うん，だよね，難しい…							
8	4-B5	なんか力を加えないと，ずっとそのまま等速で，動いてるっていう風に考えたら，こっち(左)にかかってるからなのかなーみたいな．		必要な考え					
9	4-C5	あぁ…そういうことか							
10	4-B5	力を加えてないと，加速度…加速してないって考えたら，こう…減速してるってことはこっち(左)かなーみたいな		必要な考え					
11☆	4-A5	例えば右向きに力かけたらさ，加速するじゃん？手で右に押したら．							
12	4-D5	うんうん							
13	4-B5	あー							
14	4-C5	え，でも左にかけたら左に向かって加速するんじゃないの？	力をかけると加速する場合だけ考えている						
15	4-A5	それは減速							
16★	4-D5	あー…！そっか！							
17	4-C5	それが左方向に進んじゃうってことはないの?							
18	4-A5	たとえば台車がこう…手でパンと離して，ちょっとずつ力こっち(左)に押してもさ，進むけど，止まりはしないじゃん？でも減速はしてるじゃん？							
19	4-C5	あー…	納得はしているが，イメージと違う様子						
20	4-B5	チョンチョンって感じでしょ？							
21	4-A5	そうそう							
22	4-A5	例えば運動方程式から見てさ，加速度負だから，力も負の左向き		補足説明					
23	4-B5	運動方程式から何だっけ？							
24	4-A5	ここがいまさ，加速度がマイナスじゃん，で，ここ質量が正じゃん，こっちプラスで，こっちマイナスじゃん，で，加速度マイナスってさ，右側を正ってとったときに，マイナスにいくってことじゃん，だから力も，マイナスの向き							
25	4-B5	うんうん…ここがマイナスプラスで，マイナスだからってこと？							
26	4-A5	そう．だから…台車の動きを数直線で見ると，こっちプラスでこっちマイナスにすると，加速度がこっち(負)向いてるじゃん？だから，力も左向き			2	2	2	2	

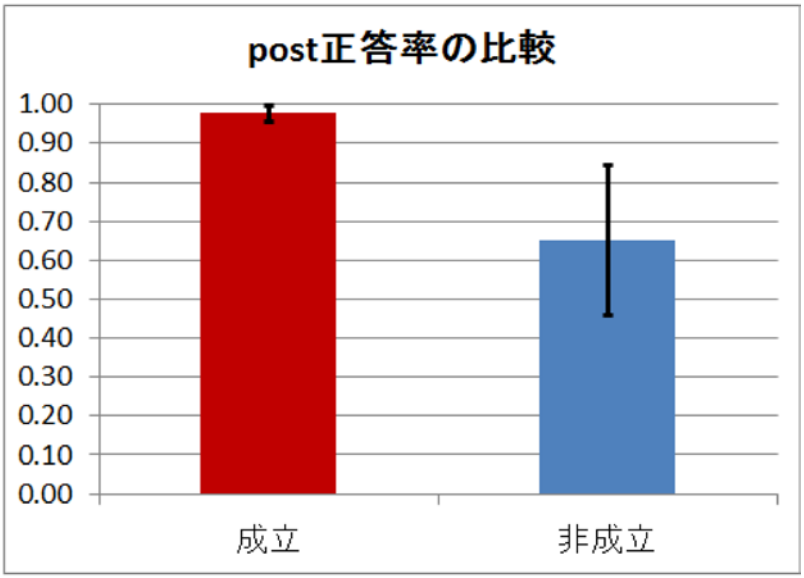
次のプロトコルは PI 不成立型を表す。ER に分類される考え方が多く，進行方向に力が働いているというインペタス誤概念に基づく考え方で議論している生徒が見られる。IR に分類した発言「わかってることは…減速していると，速度の向きと，加速度の向きは逆らしい」が見られるが，時間内に議論は収束せず，結果としてインペタス誤概念に基づく誤答１が議論後に増えている。

通番	発言者	発言	備考	分類の意図	A	B	C	D	E
1	4-D2	何にした？			1	2	2	1	
2	4-B2	②							
3	4-C2	②							
4	4-A2	①							
5	4-D2	私も①だわ							
6	4-A2	えーっとまって…							
7	4-D2	どうして2なの？							
8	4-A2	あー2かも…2かもしれない…							
9	4-B2	まぁ減速してるってことはなんか合力的に言えば逆向きにいかないと遅くならないなっていう感じ		誤った考え					
10	4-C2	逆向きに…うん							
11	4-D2	あー…							
12	4-C2	なんで右向きなの？							
13	4-B2	進んでるってことはなんかこっちに力が働いてるんだろうけど，でも減速してるってことはこっち(左向き)の力がちょっとくらい大きくならないと減速しないなって思って…	進んでいる向きに力が必要だと思っている	誤った考え					
14	4-D2	こっち(左向き)の力が大きかったらこっち(左)行っちゃうんじゃないかなって思った		誤った考え					
15	4-B2	…あ…どうなんだろうね…あ，そっか…どうなんだろ，わからない(笑)							
16	4-B2	でも完璧に打ち消し合うわけじゃない…かな？どっちの向きにも働いているんだけど…		誤った考え					
17	4-C2	こっち(右)に進む力を超えたらこっち(左)に進むけど，超えない場合は減速するんじゃない？		誤った考え					
18	4-D2	あー…でも超えないと右になるんじゃない？							
19	4-B2	右になる？							
20	4-C2	右になんなくない？右に進む力より小っちゃい力が左に働いてたら……あー…							
21	4-B2	でも減速してるから右かなぁ…えー…わかんない		誤った考え					
22	4-D2	A2君どう思う？							
23	4-A2	えっと…わかってることは…減速していると，速度の向きと，加速度の向きは逆らしい．		必要な考え					
24	4-D2	…え…？どういうこと？加速度の向きが逆だから…	加速度の向きと合力の向きの関係を理解していない	混乱					
25	4-B2	合力の向きを…							
26	4-C2	合力だから…							
27	4-A2	そうそうだから…合力だから，わかんないねって話							
28	4-B2	(笑)							
29	4-D2	(笑)			2	1	1	1	

なお，本研究では，のべ 55 グループ分の議論を分析した。下表はそれらがどのタイプに分類されたかを表す。

議論型		議論数
PI 成立型		23
リーダー主導型		3
非成立型	(PI 未解決型)	13
	(分裂型)	6
	(PI 失敗型)	10

次のグラフは，議論タイプによる議論後正答率の違いを表す。誤差は平均値の標準誤差である。当然ながら，PI 非成立型は成立型よりも正答率が低く，また，分散が大きくなっている。



プロトコル分析によって分かったことは以下の通りである。

まず，一般に，正答率は理解した生徒の数より高くなっていることがあげられる。これは，まだ明確な理解には至っていないが「なんとなく」正答が分かっている生徒や，あてずっぽうで回答する生徒がいることを考えれば当然であるが，PI の議論後正答率を見て生徒の理解度とすることは，生徒の理解度の過大評価になることを改めて警告している。拡大解釈するならば，定期試験の結果を生徒・学生の理解度に対応していると解釈することに同様の危険性があるといえる。

次に，議論によって，生徒ひとりでは考えることができなかった新たな視点や，考え方に気づくような場面が多く見られたことがあげられる。これはまさにアクティブ・ラーニングの目指すところであり，多様性を活用した対話的な学びが実際に成立していることがわかる。一方では，議論中に多くの典型的誤概念に基づく考え方が多く見られており，その誤概念にひきつけられて議論が誤概念選択肢に収束することもみられた。これは，教師が議論後に的確な介入しなければ，誤概念が議論によってさらに強固になっていく危険性をはらんでいることを示唆する。

また，リーダー主導型の議論は，一人の生徒の意見だけで議論が終了し，生徒間の相互作用が生じていない様子がみられる。この場合，グループにミニ教師が誕生し，結局は講義型の授業と類似の学習になってしまっているといえる。すなわち，リーダー以外の生徒は受身の状態になってしまい，アクティブ・ラーニング型授業の目指す多様性を生かした協同的な学びが成立していない。もちろん，教師が質問に来た生徒にするように，グルー

プ内の生徒が抱える疑問にリーダーが答えることにより，理解が深まっていくメリットはある。

PI のプロトコル分析により，対話的な学びとしてのアクティブ・ラーニングの典型的な様相を詳細に示すことができたと考えられる。生徒の対話をさせることにより，主体的な学びが行われることは間違いないが，それが効率よく深い学びに結びつくためには，議論の後に教師の的確な介入が必要になることを最後に指摘しておく。もちろん，生徒・学生の実態に合っており，議論のメリットが最大限に生かせる教材を研究ベースで作成することが極めて重要であることは言うまでもない。

参考文献

[1] 本研究の詳細については，後藤敬祐：ピア・インストラクションにおける生徒間相互作用の分析（平成 28 年度東京学芸大学大学院教育学研究科修士論文）を参照のこと。

５．現代テスト理論による概念調査問題の再評価

第 1 章，第 2 章などで述べたように，力の概念調査 FCI は力学範囲における生徒の概念理解の指標であり，授業の効果を表す指標となっている。力学以外の分野でも同様の考え方によって作成された概念調査問題が存在する。それらは世界共通で用いられており，国際比較をしたりするのに活用されている。一方，同一の問題をテストとして使い続けることや，Hake の規格化学習ゲインが間隔尺度になっていないことなど，テスト解析の面では従来の古典テスト理論による分析と評価に早晩限界が来ることが予想される。

そこで本研究では，主に今後の分析・調査法としての活用を想定し，現代テスト理論を用いた FCI の解析を試みた。以下ではまず古典テスト理論と現代テスト理論を概観し，その後，具体的に FCI を現代テスト理論によって分析した結果を示す。

５．１　古典テスト理論（CTT)の概要

$i = 1, 2, \cdots N$を被験者，$j = 1,2, \cdots, J$を問題項目とする。古典テスト理論（Classical Test Theory, CTT）では，ある被験者集団に対して実施したテストの観測得点$\boldsymbol{x_i}$が

$$\boldsymbol{x}_i = t_i + \boldsymbol{e}_i$$

で表される確率変数であると考える。ここにt_iは真の得点と呼ばれる量で，構成概念であり，直接の観測量ではない。$\boldsymbol{e}_i$は誤差と呼ばれ，その平均値$\langle \boldsymbol{e}_i \rangle$は

$$\langle \boldsymbol{e}_i \rangle = 0$$

を満たすと仮定する。したがって，

$$\langle \boldsymbol{x}_i \rangle = t_i$$

が成立する。ただし，ここでいう平均⟨　⟩は被験者が同条件でテストを無限回繰り返した（繰り返しの度に前回テストを受けたことは忘れる）場合に得られる値であり，理想化された量である。（または，同じ真の得点を有する被験者を無限に用意して一斉に試験を受けさせた場合の平均と言ってもいい。）

古典テスト理論では，主に次の 3 つの量でテストを評価する。

- 項目困難度（item difficulty）
- 項目識別力(item discrimination)
- 信頼度（reliability）

以下，これら 3 つの量について簡単に説明する。

項目困難度はテストを構成する項目（設問）の難しさを表す量であり，項目の平均得点率によって一般的に表される。例として，各項目の得点が 1 点の単純化されたテストを考える。被験者iの項目jに対する得点をu_{ij}とすると，

$$u_{ij} = \begin{cases} 1 \ (\text{correct}) \\ 0 \ \ (\text{incorrect}) \end{cases}$$

とおける。このとき，項目jの通過率すなわち項目困難度は

$$p_j = \frac{\sum_{i=1}^{N} u_{ij}}{N}$$

で表される。要するに，設問ごとの平均点（に比例した量）が項目困難度である。

項目識別力は，テストの各項目（設問）が，どれだけ被験者の能力を識別できるかを表す量である。項目識別力は，入学試験を典型とする，被験者を能力別に順位付けすることが目的のテストにおいて特に重要な意味を持つが，一般にテストは被験者を識別することができなければ実施した意味が失われる。古典テスト理論では，項目識別力は各項目の得点と総得点との相関係数で定義される。

$$\rho_j(u_{ij}, x_i) = \sum_{i=1}^{N} \frac{(u_{ij} - \langle u_i \rangle)(x_i - \langle x \rangle)}{\sigma_i \sigma_x}$$

ここにσ_iとσ_xはそれぞれ項目jと全体の偏差，$\langle u_i \rangle$は項目jの平均得点，x_iは被験者iの得点，$\langle x \rangle$は被験者全体の平均得点を表す。

信頼度はテストがどれだけ被験者の真の得点を反映しているかを表す量で

$$\frac{\text{真の得点の分散}}{\text{実際の得点の分散}} = \frac{\sigma_t^2}{\sigma_x^2}$$

によって定義される。しかしながら，実際には真の得点の分散は測定によって得られないので，近似によって信頼度を表現する必要がある。近似にはいろいろあるが，

$$\alpha = \frac{J}{J-1}\left(1 - \frac{\sum_{j=1}^{J} \sigma_j^2}{\sigma_x^2}\right)$$

で表される Cronbach の α 係数が有名である。

なお，テストで測定したいのは真の得点である。しかしながら，真の得点がそのテストで測定しようとしている構成概念が必ずしも真の得点で表されているとは限らない。構成概念がテストによって測定されていることを評価することを，妥当性評価という。妥当性評価は，そのテストを実施した結果からは行えず，被験者に対して問題を回答した意図を記述式アンケート調査や面接調査で確認することで行う必要がある。

５．２　現代テスト理論の概要

古典テスト理論は，テストの分析結果が被験者集団の特性に依存することになる。このことは，異なるテストの結果の比較や，異なる被検者集団の比較を困難にする。例えば，同一の授業の期末テストの結果を異なる年度で比較するにしても，一般にテストの内容は毎年異なるし，受講する学生も毎年異なる。この場合，同じ 60 点という成績がどのような意味を持つのかを客観化することは困難である。

異なるテストの結果の比較や，異なる被検者集団の比較を可能とするためには，テストの難易度の評価と被験者の能力の評価とを独立させる必要がある。このような要求をかなえるのが現代テスト理論である。本論では現代テスト理論の代名詞ともいえる項目応答理論（Item Response Theory，IRT）について述べる[1]。ただし，IRT の理論的側面には触れない。

項目困難度だけをパラメタとする 1 母数（1 パラメタ）ロジスティック IRT（Rasch モデル [3]）では，潜在特性値（latent trait）（能力値 ability, proficiency などと呼ばれることもある）がθ，項目困難度（item difficulty parameter）がδ_jの項目j（FCI の場合は$j = 1,2,\cdots,30$）に正答する確率$p_j(\theta)$は

$$p_j(\theta) = \frac{e^{D(\theta-\delta_j)}}{1+e^{D(\theta-\delta_j)}} = \frac{1}{1+e^{-D(\theta-\delta_j)}}$$

で表される。Dは尺度因子であり，$D = 1.7$のとき，正規累積モデルによる値とよく一致することが知られている。

2 母数の IRT では，項目識別度（item discrimination parameter）a_jがパラメタとして追加される。次の式で表される。

$$p_j(\theta) = \frac{1}{1+e^{-Da_j(\theta-\delta_j)}}$$

3 母数 IRT は，当て推量母数（guessing parameter）c_jが加わる。次の式で表される。

$$p_j(\theta) = c_j + \frac{1-c_j}{1+e^{-Da_j(\theta-\delta_j)}}$$

本分析では，フリーの統計解析ソフトウェアである R を用いている。パッケージとしては，1 母数 IRT は eRm と ltm，2 母数以上の IRT には ltm を用いている。ltm は，被験者の 2 値（正解は 1，不正解は 0）データから ICC の母数と潜在特性の値を推定する。母数及び特性値の推定には周辺最尤推定法を用いている。なお，eRm は同時最尤推定法を用いている。

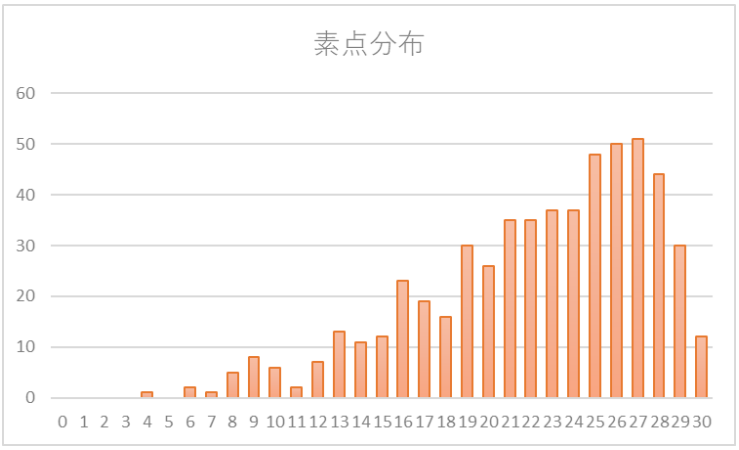
データは本学 2 年生対象の物理学演習（植松担当）のプレテスト被験者 561 名（複数年度分）を用いている。本分析は logistic IRT モデル [4]による分析の試行が目的であり，データの詳細な質は問題としていない。 また，IRT は原理的には項目困難度等の問題のパラメタと被験者能力値パラメタとが分離されるので，複数年度の被験者を集めたデータの解析にも一定の意味がある。

下の図は，被験者全体の素点分布を表す。高得点側に分布が偏っているので，IRT 分析に適しているとは言えない。

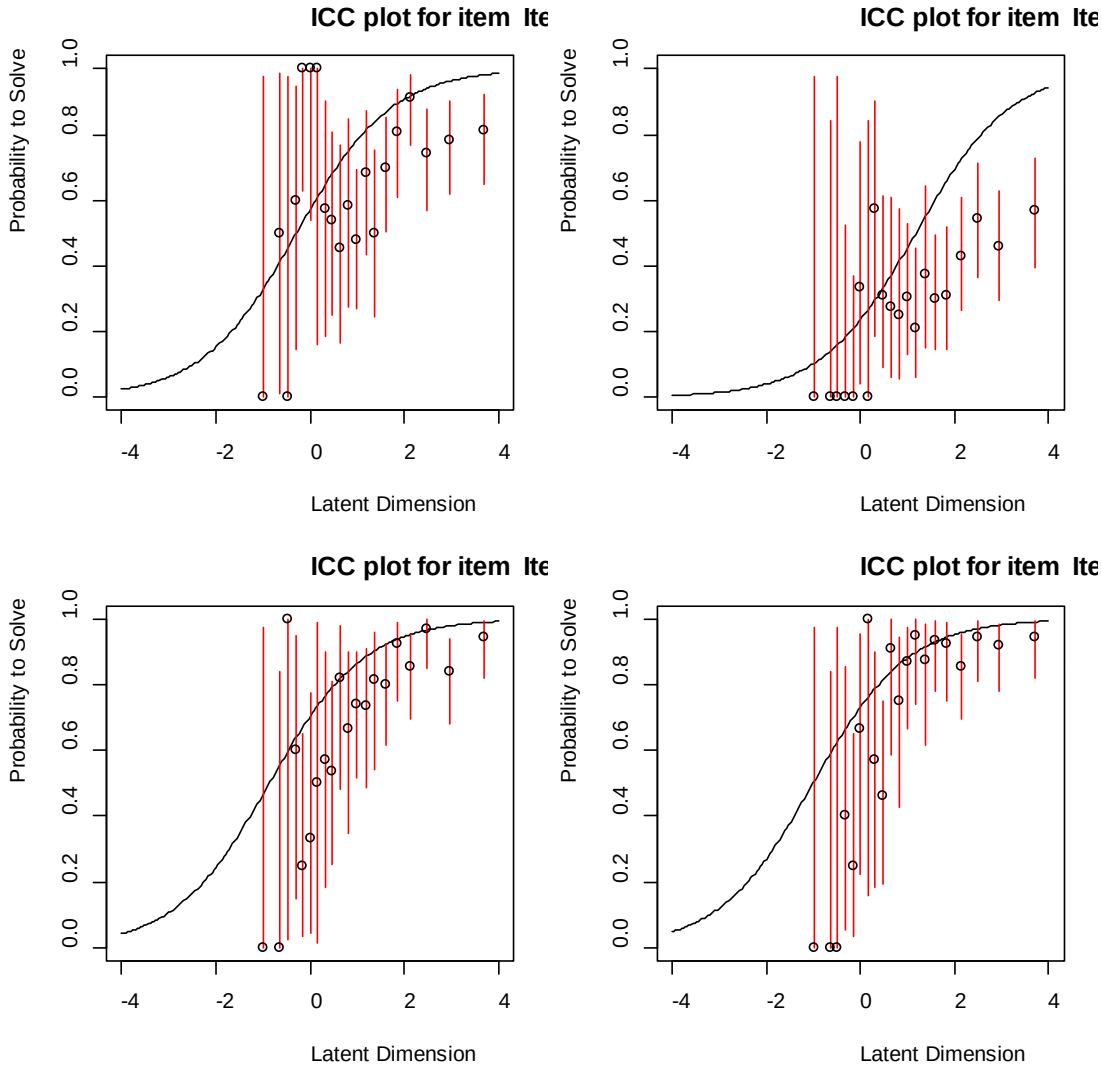
---

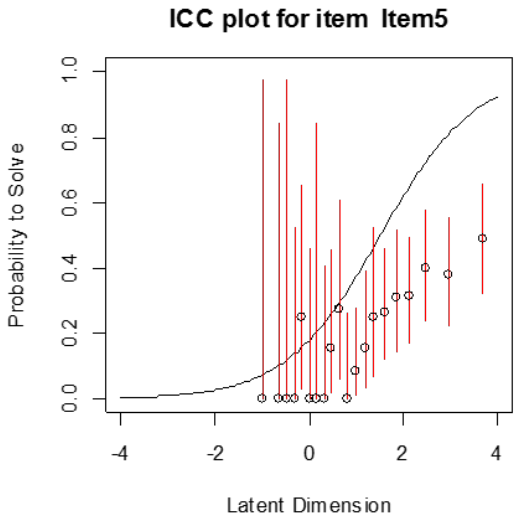
[3] Rasch モデルは，IRT と異なる論理によって展開された数学的な確率モデルであるが，数式的には 1 母数 logistic IRT モデル（以下 logistic 略）と等価である。そこで，IRT モデルによるデータ分析の試行の報告を目的とするこの資料では，以下 Rasch モデルという言葉は原則用いないことにする。

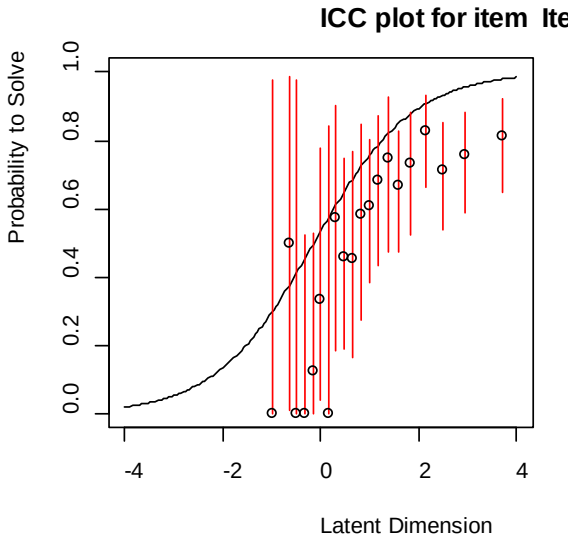
[4] 本論では logistic モデルしか扱わないので，以下 logisitic を省略し，単に IRT モデルと呼ぶ。

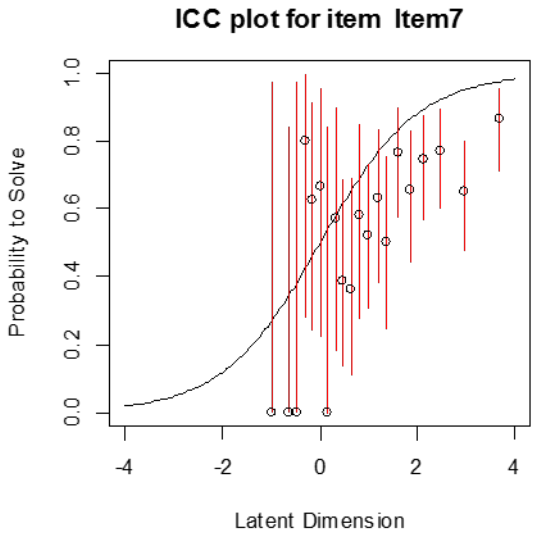


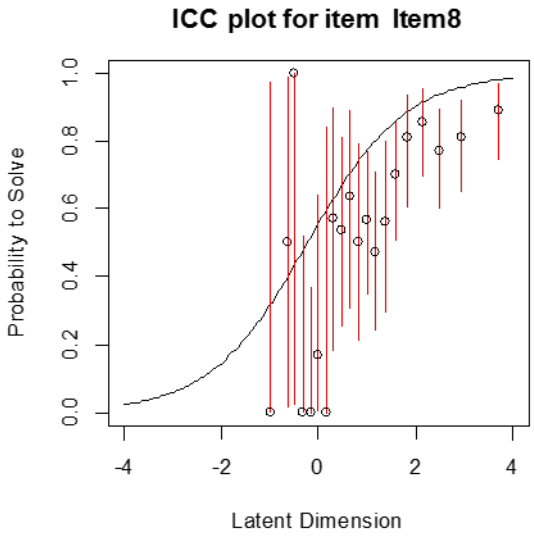
次に，eRm によって求められた 1 母数 ICC（Item Characteristic Curves）のグラフを示す。FCI30 問すべての ICC を示している。

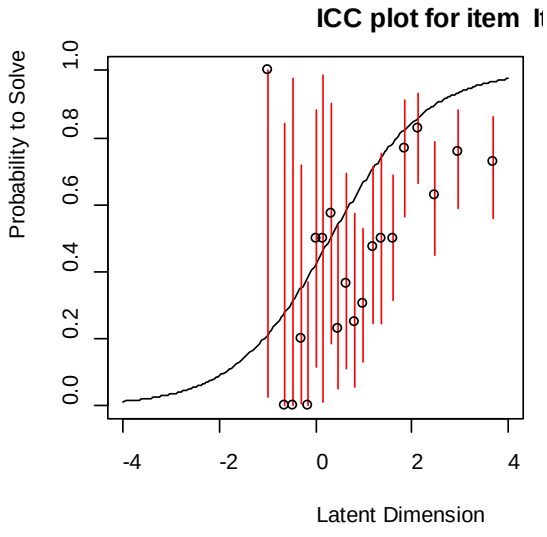


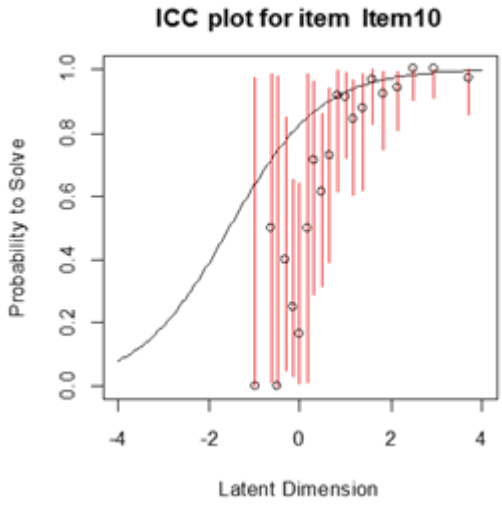

ICC plot for item Item5
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Latent Dimension

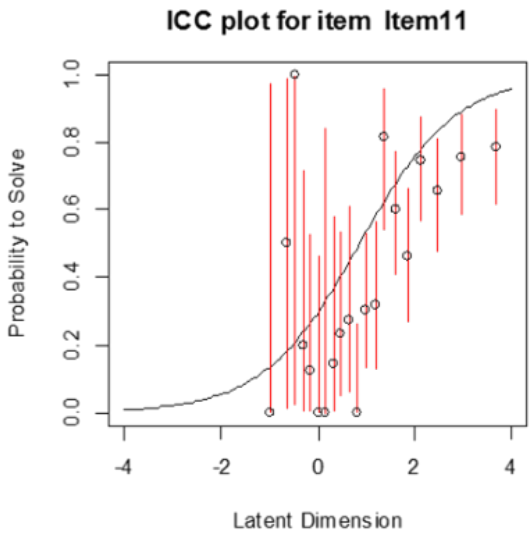

ICC plot for item Ite
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Latent Dimension

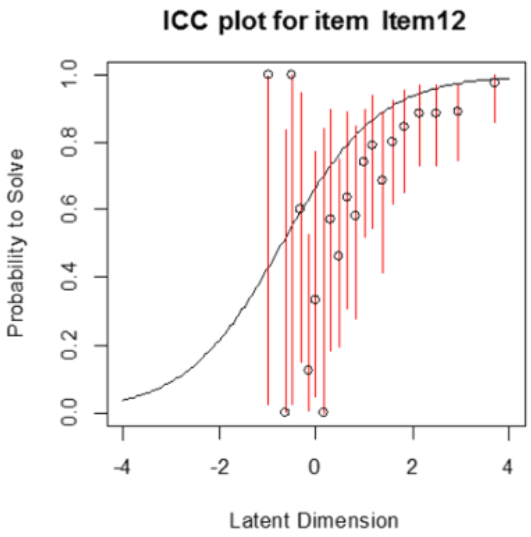

ICC plot for item Item7
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Latent Dimension

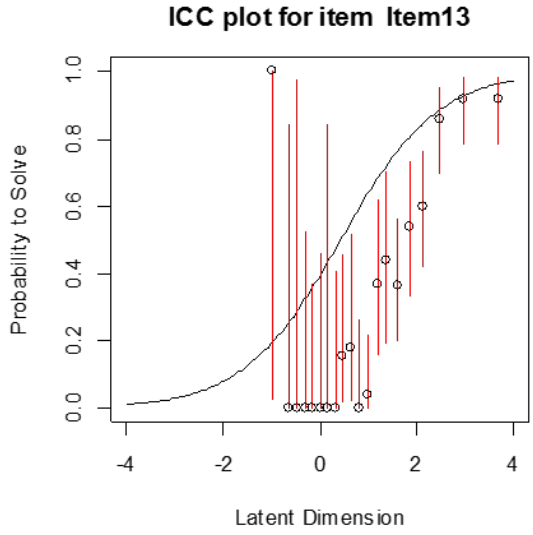

ICC plot for item Item8
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Latent Dimension

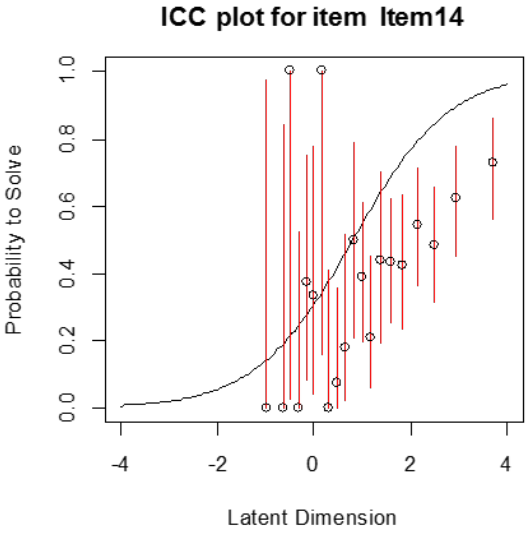

ICC plot for item I
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Latent Dimension

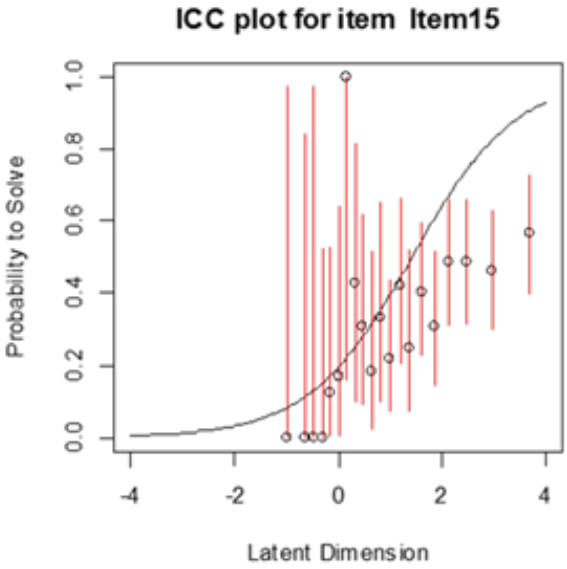

ICC plot for item Item10
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Latent Dimension

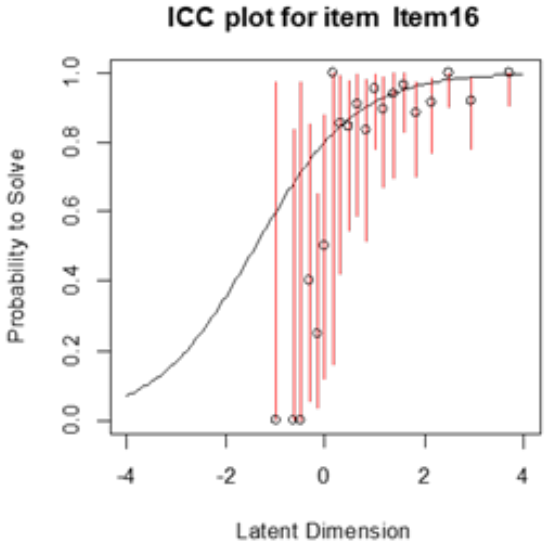

ICC plot for item Item11
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Latent Dimension

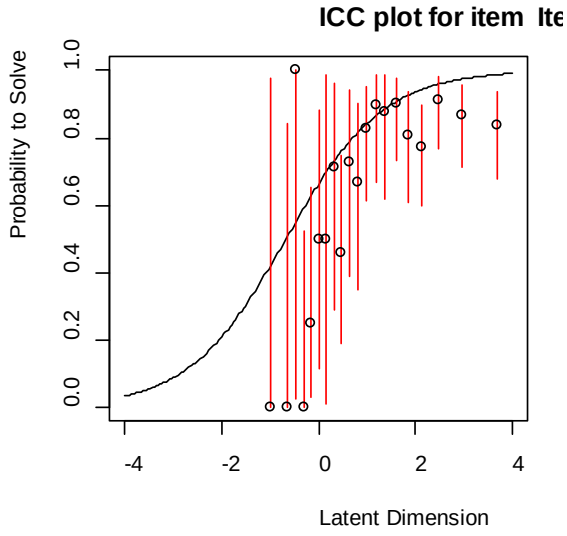

ICC plot for item Item12
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Latent Dimension

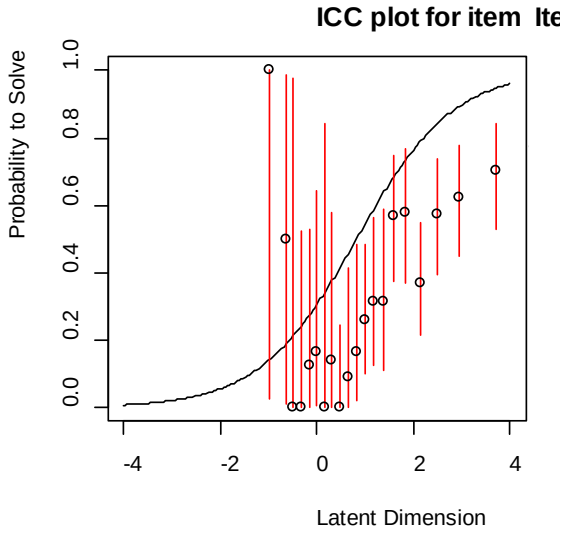

ICC plot for item Item13
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Latent Dimension

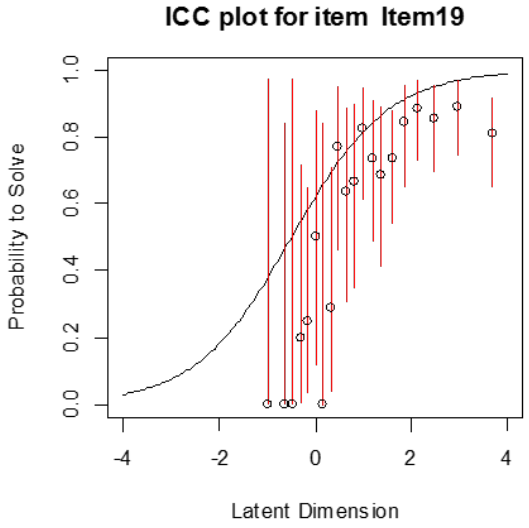

ICC plot for item Item14
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Latent Dimension

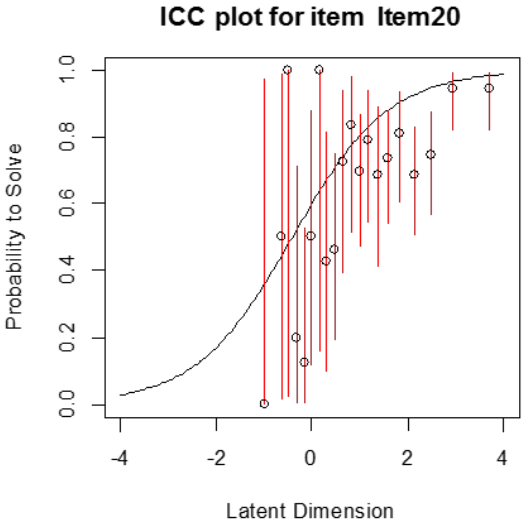

ICC plot for item Item15
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Latent Dimension

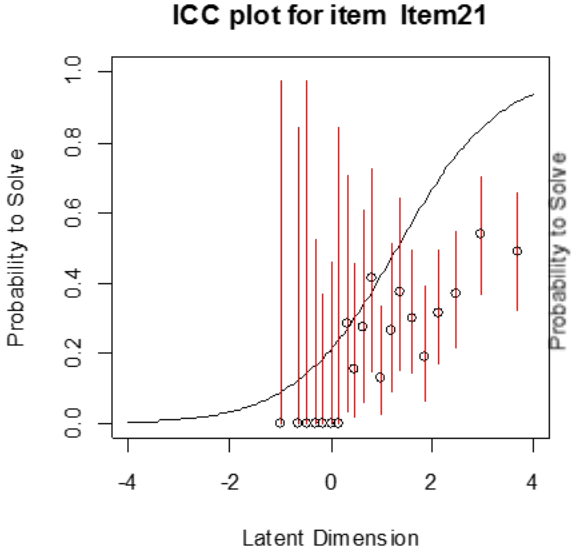

ICC plot for item Item16
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Latent Dimension

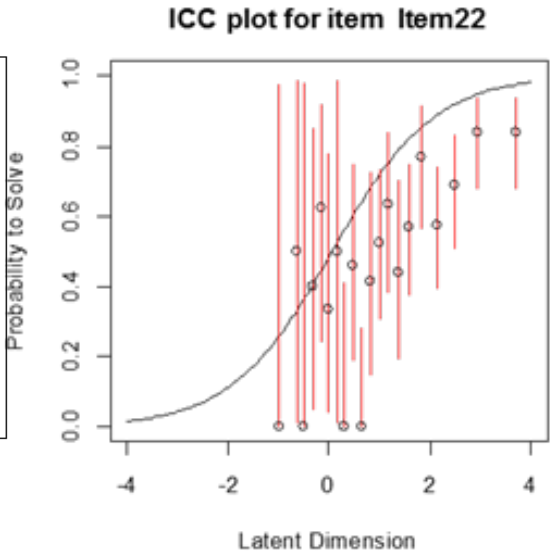

ICC plot for item Ite
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Latent Dimension

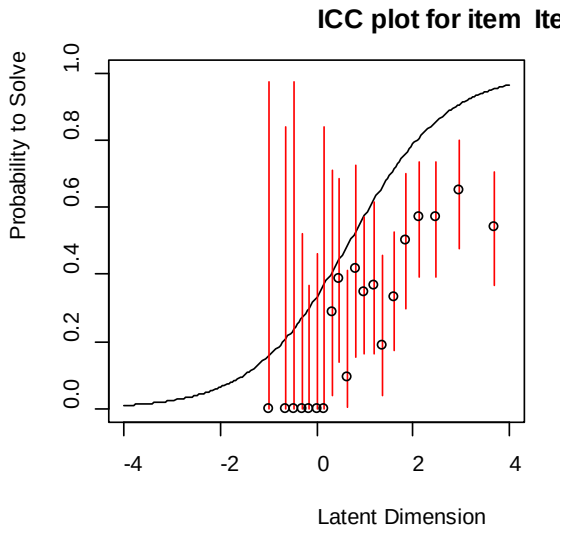

ICC plot for item Ite
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Latent Dimension

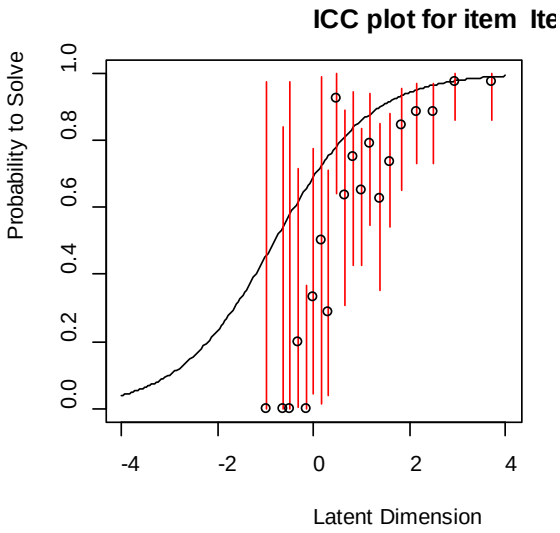

ICC plot for item Item19
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Latent Dimension

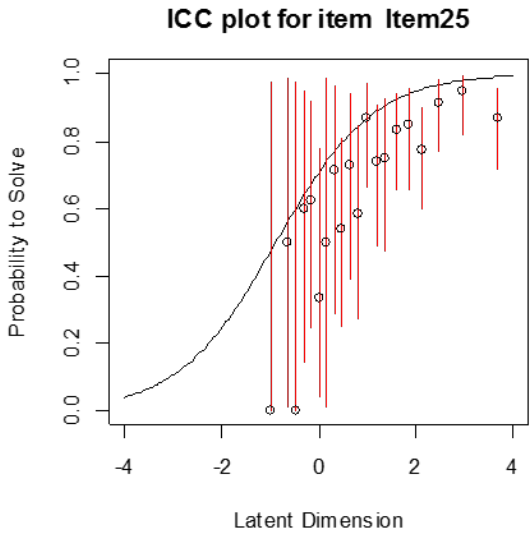

ICC plot for item Item20
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Latent Dimension

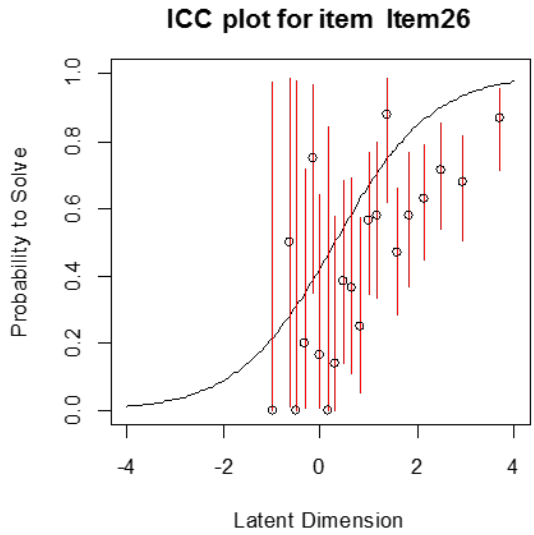

ICC plot for item Item21
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Latent Dimension

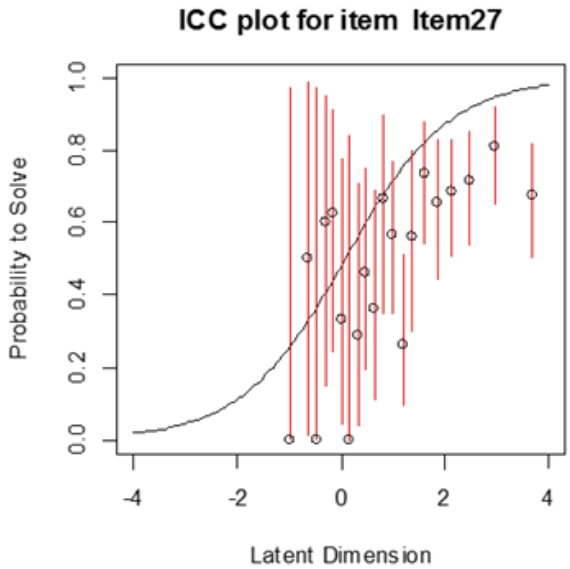

ICC plot for item Item22
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Latent Dimension

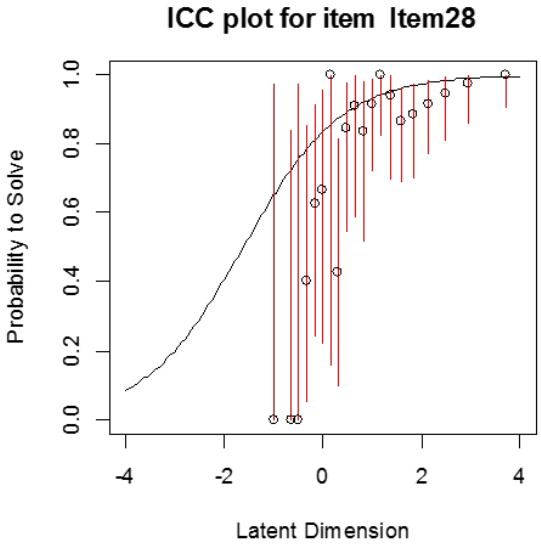

ICC plot for item Ite
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Latent Dimension

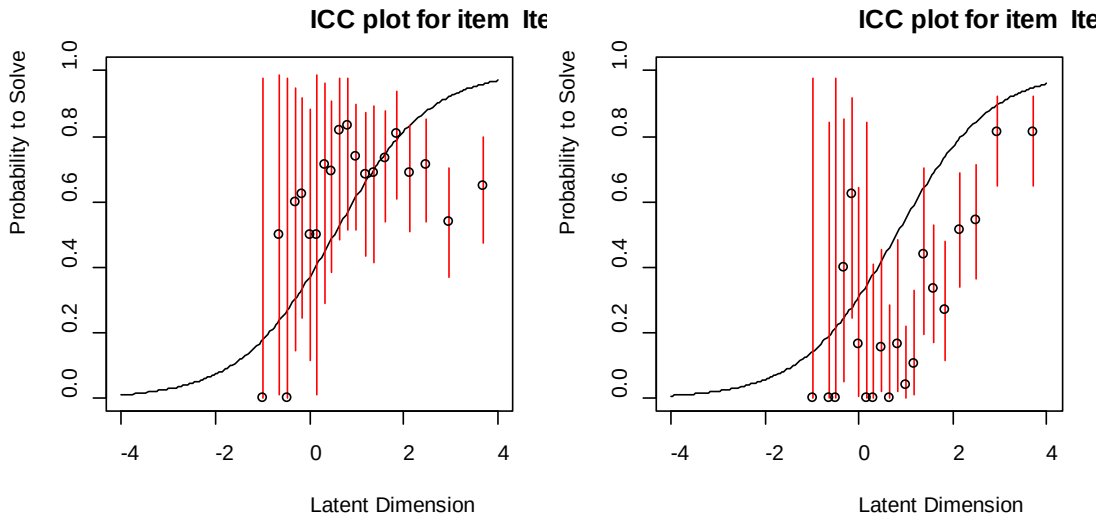

ICC plot for item Ite
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Latent Dimension


ICC plot for item Item25
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Latent Dimension

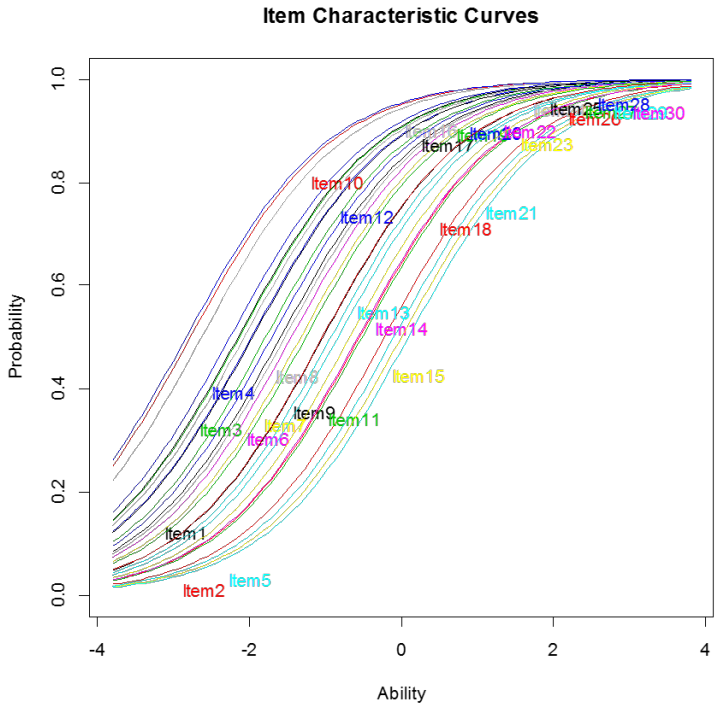

ICC plot for item Item26
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Latent Dimension


ICC plot for item Item27
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Latent Dimension

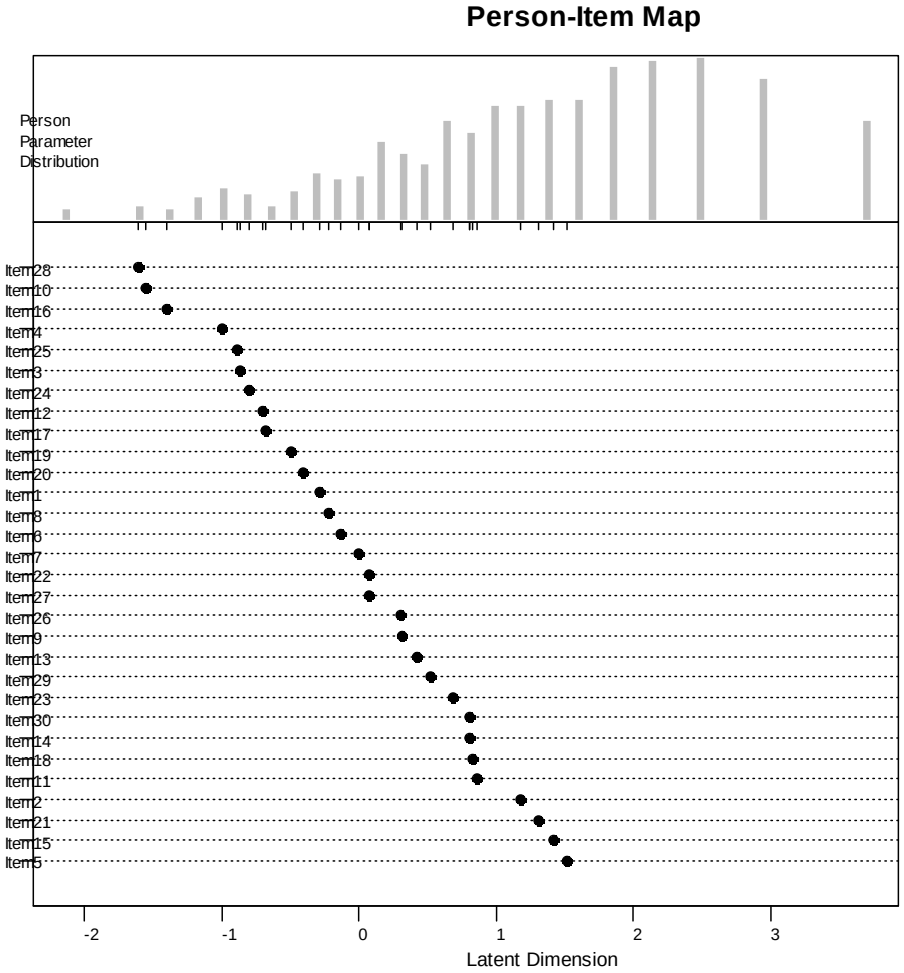

ICC plot for item Item28
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Latent Dimension



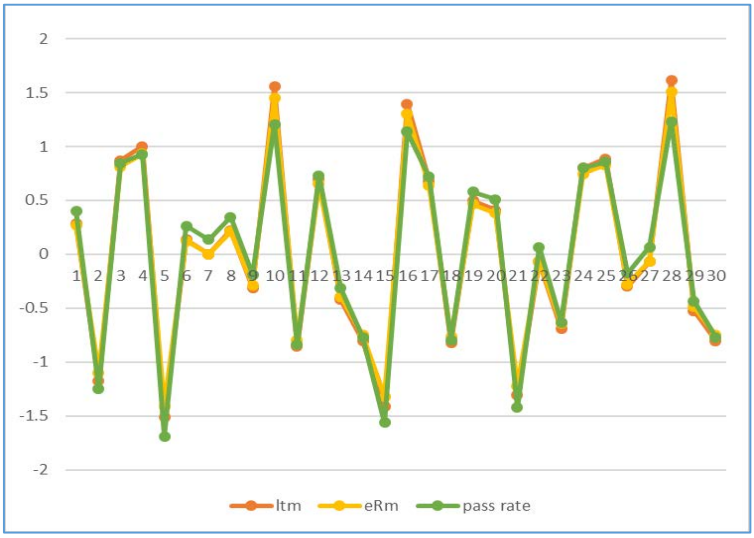
下図に，被験者の能力値の平均が 0 となるようにスケールした場合の，すべての設問の ICC を一緒に描いたグラフを示す。Rasch モデルは 1 母数なので，ICC は項目困難度順に並べただけである。能力値（ここでは ability と表示されている）が 0 で最も難しい項目の正答確率が 0.5 程度になっていることから，被験者集団にとって問題群が易しすぎることがわかる。FCI の利用から離れ，IRT の本来の目的であるテストづくりの観点のみから眺めると，この被験者集団の能力値を正確に測定するためにはは，もう少し難しい問題を入れるべきだあることが示唆される。



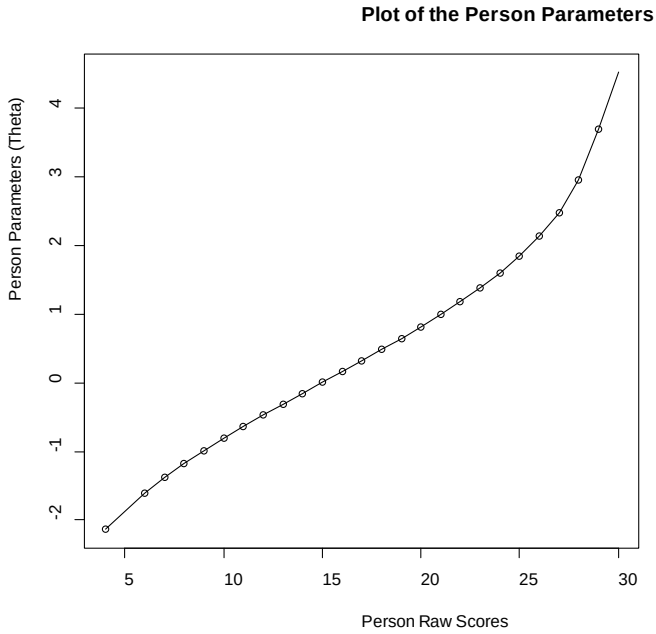
次に示すのは eRm による潜在特性値分布と項目困難度（Person-Item マップ）である。図の上側の棒グラフは特性値別の被験者分布，下側は易しい順に並べた各項目の項目困難度を表す。この図からも被検者特性値分布に対してテスト項目が易しい方に偏っていることがわかる。



次に，古典テスト理論との比較を行う。下図は，古典的な通過率（正答率）を項目容易度（困難度にマイナスをつけた値）と比較したものである。通過率は適当に比率を変換し，項目容易度と比較しやすくしてある。Rasch モデルにおいて最尤法で推定された項目容易度は，通過率そのものとそれほど違わないことがわかる。logit を用いれば，さらにフィットすることが予想される。なお，下図には ltm と eRm の双方で求めた項目困難度（容易度）を掲載してあるが，両者はほぼ一致している。すなわち，推定法が異なっても推定値に事実上の差がないことがわかる。

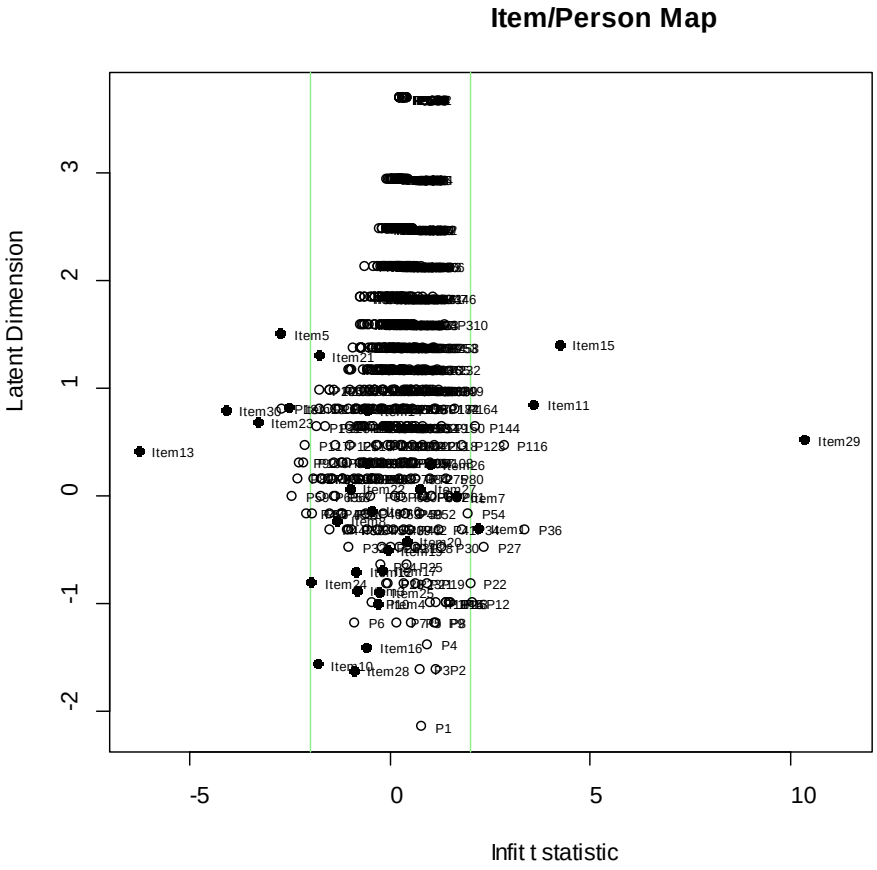


次の図は，潜在特性値と素点との対応関係を表す。ここでは潜在特性値が Person parameter と表示されている。person raw scores が FCI の素点である。素点と特性値は 1 対 1 対応をするが，対応関係は非線形である。



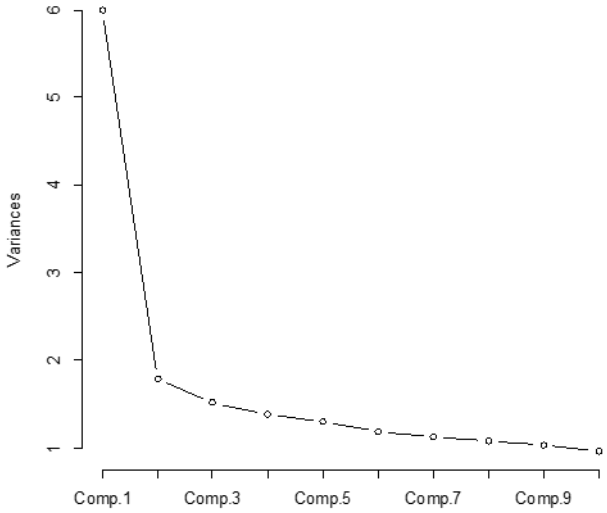
次に，データの 1 パラメタ ICC への適合度について述べる。これはインフィット（infit, information-weighted mean-square-fit statistics の略）という量を用いてチェックする。不正確な表現だが，インフィットは推定値に対する実際の得点の分散のようなものである。下図で，緑色の線の内側にあれば適合度の良いデータとされる。Item29 は不適合（アンダ

ーフィット）の度合いが大きい。Item15, Item11 もテスト理論の観点からはテスト項目からの削除対象となるものといえる。緑線の左側（オーバーフィット）のデータは有害ではないが，なくてもよい項目とされる。テスト全体の傾向と合いすぎているから省略可能なのである。もちろん，FCI は能力値を求めるのが目的ではなく，生徒・学生の持つ素朴概念を詳細に調べるための調査紙であるから，それらの問題が不要ということにはならない。むしろ，オーバーフィットしている項目は，被験者の（FCI で測定可能なものという意味での）ニュートン力学理解度を象徴する設問であるという意味付けができるといえる。



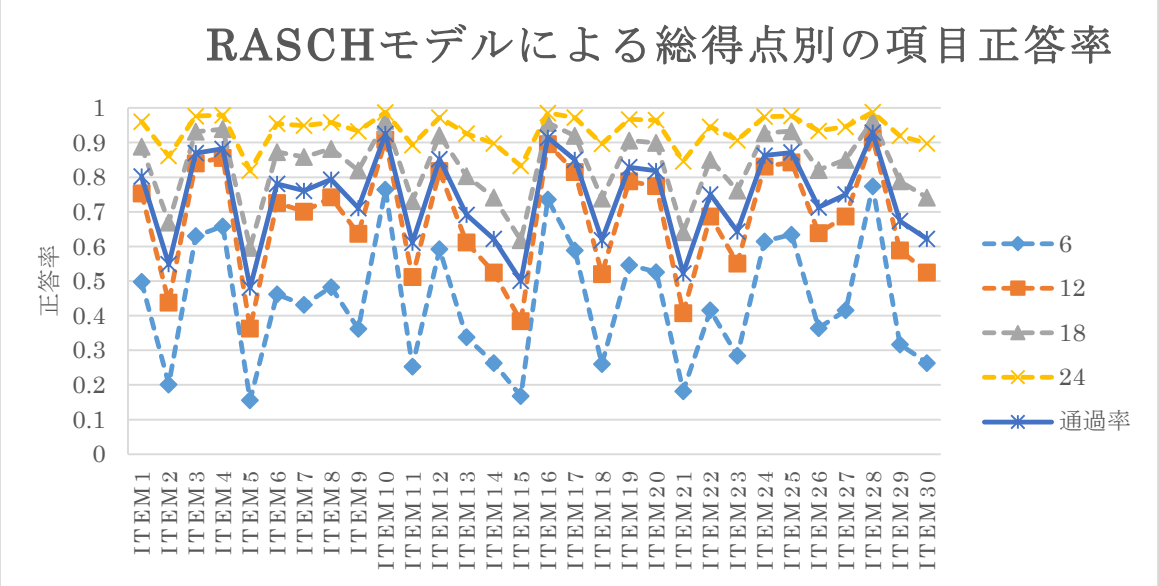
ところで，IRT 分析を行うためには，データの 1 次元性が前提条件となる。つまり，あるテストが被験者のある能力値パラメタを測定しているというためには，そのテストの結果の解釈が多次元的，つまり複数のパラメタを必要とするものであってはならない。

FCI は Newton 力学の概念理解を測定する調査紙であり，被験者に「Newtonian 度」とでも呼ぶべき 1 次元の潜在特性を仮定している。これは因子分析でいう構成概念（construct）の 1 次元性を仮定していることに対応する。そこで，R のパッケージ fa によって scree plot を描き，データの 1 次元性を目視で調べたところ，1 次元性が確認された（下図）。



最後に，ICC の応用として，総得点ごとの項目別正答確率について述べる。

IRT で$p_j(\theta)$が求まれば，様々な能力値ごとの項目正答確率が示せる。下の図は，正答数が 6 問，12 問，18 問，24 問の被験者がそれぞれ各項目に正答できる確率を表す。1 母数なので項目ごとの比率がそれほど変化することはないが（それでも$p_j(\theta)$は非線形なので比率は一定ではない）2 母数以上でフィットできればより情報量の多いデータとなる可能性がある。



上図のようなグラフによって，例えば，クラスの FCI 平均値から，特定の問題の正答率が推定できる。したがって，授業計画を策定する際に，プレテストの結果を利用して，FCI

の項目ごと綿密に時間配分を考えるといった授業改善に応用することができると考えられる。

最後に，プレテストとポストテストによって求めた被験者の能力値の差を用いれば，従来の Hake の規格化ゲインを援用した個人ゲインを用いることなく，個人のゲインを定義することができることを強調しておく。これを Rasch ゲインと呼ぶならば，Rasch の個人ゲインは，その値をクラスで平均すればそのまま平均 Rasch ゲインを与えることになる。それに対し，Hake の個人ゲインを平均してもクラス平均の Hake ゲインは得られない。このことは，個人データを用いた他の調査との相関を求めたりする際に障害となっていた。現代テスト理論を用いることにより，さらに授業の定量的な分析と評価の精度が高まることが期待される。

参考文献

[1] 豊田秀樹：項目反応理論（入門編）[第 2 版]（朝倉書店，2002）.

６．まとめ

平成 24 年度に中央教育審議会大学分科会大学教育部会の審議まとめ『予測困難な時代において生涯学び続け主体的に考える力を育成する大学へ』が公表された（中央教育審議会，2012)。その中に，「学士課程教育の質的転換」を求める記述があり，「求められる質の高い学士課程教育とは，教員と学生とが意思疎通を図りつつ、学生同士が切磋琢磨し、相互に刺激を与えながら知的に成長する課題解決型の能動的学修（アクティブ・ラーニング）によって、学生の思考力や表現力を引き出し、その知性を鍛える双方向の講義、演習、実験、実習や実技等の授業を中心とした教育である」と記されてある。この指摘は，教員が講義ノートや教科書を用いて一方通行的に学生に向かって解説する伝統的な授業スタイルからの脱却という，大きな変化を求めたものであった。

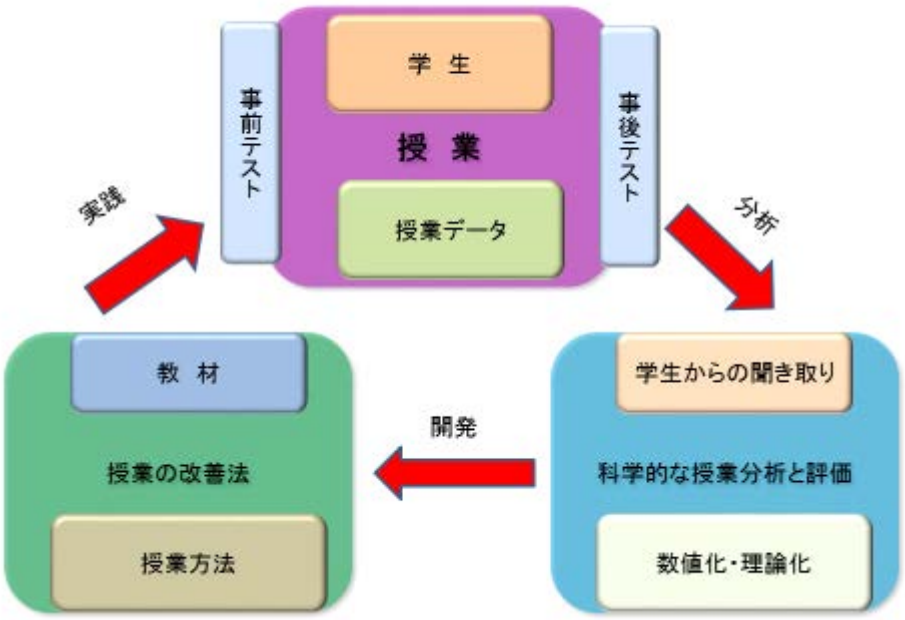
かつての大学の講義とは，「自分がいかにわかっていないかを知る場所」にすぎず，ノートの行間や知識・理解のギャップを自学で埋めていくことが要求された。このような自学の態度は研究者には不可欠のものだが，学習法の視点から見れば，学生に自立したアクティブな学習者になることを要求していたことになる。大学教育の重要な要素として研究者の養成があった頃には，上記のような，ついて来られる者だけを相手にした授業形態もそれなりに機能していたといえる。研究者になるためには，受動的な態度で学習を進めるのではなく，自立したアクティブな学習をすることが不可欠だからである。

しかし，時代は大きく変わり，大学の様相は一変した。かつての高校進学率よりも高い進学率で，高校生は大学へと進学していく。大衆化した現代の大学では，学生にどれだけの学問や技能を身につけさせたかを問われるようになった。特に，現代のように高度に進歩した科学技術が社会の基盤となっている時代においては，どのような職業に携わる者であろうとも，「科学的な見方・考え方」を身に着けていなければ社会全体の存続を危うくする。自然科学の基礎教育によって，すべての学生に対して「科学的な見方・考え方」を育成することができるか否かが，その国の命運を決定づける重要な要素になるであろう。

また，初等中等教育においても，次期学習指導要領の改訂の方向性の中で「主体的･対話的で深い学び（「アクティブ・ ラーニング」）の視点からの学習過程の改善」が示されて以来，アクティブ・ラーニングへの取り組みが活発化している。

このようにアクティブ・ラーニングが注目され，様々な授業法が提案され試されている一方，その効果を示すための科学的な分析評価方法の確立への取り組みは乏しいと言わざるを得ない。

教員養成系大学である本学は，実践的な授業開発とその分析・評価法の開発研究において最先端を歩む義務があると言える。そのためには，次の図のような分析・開発・実践のループによって授業の分析・評価そして改善に取り組むのが理想的である。


学 生
授 業
授業データ
事前テスト
事後テスト
実践
分析
教 材
授業の改善法
授業方法
開発
学生からの聞き取り
科学的な授業分析と評価
数値化・理論化

図： 授業開発のループ

平成２７・２８年度　広域科学教科教育学研究経費報告書

アクティブ・ラーニング型理科授業とその評価法の系統的研究

研究代表者　　新田　英雄

平成 29 年 3 月

東京学芸大学