

論文

項目反応理論による英語能力推移に関する研究の比較

熊谷 龍一 *

*新潟大学全学教育機構

Comparing the research for trends of English proficiencies using Item Response Theory

Ryuichi Kumagai*

* Institute of Undergraduate Programs and Courses, Niigata University

In this paper, we examined three researches which are Saida (2005), Yoshimura (2005) and Kumagai (2007) to discuss for trends of English proficiencies of high school students or students preparing for university examinations. These researches conducted equating, using item response theory.

We discussed about trends of English Proficiencies. Among three researches, we have seen agreement and disagreement. Any research showed that there was a part where an English proficiency had decreased. However, the degrees of the decrease were different.

Keywords : Proficiency of English, Trend, Equating, Item Response Theory (IRT)

キーワード : 英語能力、経年比較、等化、項目反応理論

1. はじめに

2007年度、文部科学省による「全国学力・学習状況調査」が実施され、まもなくその調査結果も公表される見通し(2007年7月現在)である(文部科学省, 2007)。このほかに、経済協力開発機構(OECD)による生徒の学習到達度調査(PISA2003)や、国際教育到達度評価学会(IEA)による国際数学・理科教育動向調査(TIMSS2003)など、様々な形で学力測定調査が実施されている。そのような調査結果の中で注目されるものの一つとして、経年比較、つまり過去と比べて現在の受験者(児童・生徒)の学力がどうなっていたのかが挙げられる。昨今、大きな問題として取り上げられている「学力低下」に関する議論のデータとして、これらの調査結果が引用されることも多い。

このような学力の経年比較を目的の一つとして、米国で大々的に実施されている調査が全学学力調査(The National Assessment of Educational Progress, 以下NAEP)である。NAEPは非営利テスト機関であるEducational Testing Service(以下ETS)により、綿密なテスト計画のもと、実施されているものである(詳細については全米学力調査研究会, 2005を参照されたい)。NAEP調査にお

* 〒950-2181 新潟市西区五十嵐2の町8050番地 新潟大学全学教育機構

Correspondence concerning this article should be sent to: Ryuichi Kumagai, Niigata University, Institute of Undergraduate Programs and Courses, 8050, Ikarashi 2-no-cho, Nishi-ku, Niigata City, Niigata Prefecture, 950-2181, JAPAN
E-mail:ryuichi@ge.niigata-u.ac.jp

ける大きな特徴の一つが、項目反応理論 (Item Response Theory, IRT) と呼ばれるテスト理論を用いた分析が行われていることである。そこでは等化 (Equating) と呼ばれる共通尺度の作成が行われ、異なる版のテストを相互に比較するシステムが構築されている。

我が国において、この等化を用いて能力・学力を測定する研究としては芝・野口 (1982) などが挙げられるが、それ以降この分野に関する研究は活発には進んでこなかった。しかしながら、斉田 (2003)、吉村・荘島・杉野・野澤、他 (2005)、熊谷・山口・小林・別府、他 (2007) といった学力の経年比較に関する研究が報告されるようになってきた。これらの研究には全て IRT が利用されている (木村, 2006)。これら 3 つの研究は奇しくも、高校生もしくは大学受験生における英語能力についての経年比較を扱ったものである。本稿では、これら 3 つの研究について、その結果だけでなく調査方法や分析手法を踏まえて比較することで、高校生および大学受験生の英語学力の推移について検討する。

2. 各研究の概要

2.1 斉田 (2003) の研究

概要 1995 年度から 2002 年度まで実施された茨城県高等学校英語学力テスト (高校 1 年生対象) の等化を行った。テストは聴解、語彙、文法、読解の 4 領域からなり、全項目数は 46 から 50 である。テストの受験者数は 12,234 名から 17,736 名であった。

等化方法 等化の基準となるテスト (1999 年実施テスト) と、それ以外の各年度から項目を抽出した等化用テストを作成し、各テスト 400 名程度が受験をした。等化係数の推定には、能力値による Mean & Sigma 法 (Marco, 1977) を用い、項目反応モデルは 2 パラメタ・ロジスティック・モデルであった。

2.2 吉村・荘島・杉野・野澤、他 (2005) の研究

概要 1990 年から 2004 年に実施されたセンター試験「英語」のテストについて等化を行った。等化対象となる受験者は、センター試験の受験者であり、40 万人から 55 万人程度である。

等化方法 各年度のセンター試験「第 2 問」から 6, 7 項目を抽出した 50 項目からなる等化用テストを 2 セット作成し、大学 1 年生 424 名が受験したデータを使用した。等化係数の推定には、能力値による Mean & Sigma 法のほか、野口 (1983) の方法、Shojima (2003) の方法が用いられたが、それぞれかなり近い推定値を算出しているため、本稿では Mean & Sigma 法による数値のみを比較対象とする。用いられた項目反応モデルは 3 パラメタ・ロジスティック・モデルであった。

2.3 熊谷・山口・小林・別府、他 (2007) の研究

目的 学校法人河合塾において 1995 年度から 2005 年度の 11 年間に年間 4 回実施された英語学力テストのうち、5 月、8 月、12 月に実施された計 33 セットのテストについて等化を行った。テストの内容・構成は、センター試験「英語」とほぼ同様である。各テストの受験者数は 22 万人から 35 万人程度であった。

方法 過去に実施された英語学力テスト、もしくはそれと同質、同量である他の英語学力テスト

(これらをアンカーテストと呼ぶ)を、等化されるテストの前後2週間に実施し、アンカーテストと等化されるテストをどちらも受験した共通受験者を設定した。共通受験者は800名から2万名程度であった。これらの共通受験者に対し、能力推定値による Mean & Sigma 法により等化係数の推定を行った。用いられた項目反応モデルは2パラメタ・ロジスティック・モデルであった。

3. 各研究の結果比較および考察

斉田(2003)、吉村・荘島・杉野・野澤、他(2005)および熊谷・山口・小林・別府、他(2007)では、それぞれ1万人から55万人が受験した大規模な英語学力テストについて、等化による共通尺度化を試みている。その結果は、各テスト受験者集団の能力値における平均値の推移、および標準偏差の比較で示されている。それぞれの結果を以下に示す。

斉田(2003)の結果

図1は斉田(2003)で示された、1995年度から2002年度において英語学力テストを受験した高校1年生の能力平均値をプロットしたものである。この時の各年度における能力値の標準偏差は、どの年度についてもおおよそ1.0(最大値1.00, 最小値0.92)となっていた。

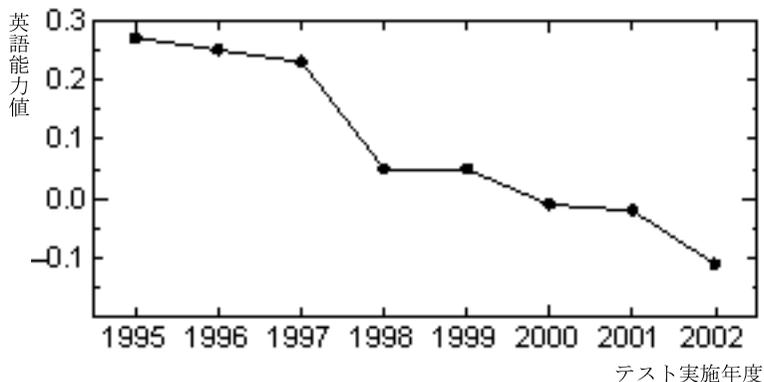


図1 斉田(2003)の結果

吉村・荘島・杉野・野澤、他(2005)の結果

吉村・荘島・杉野・野澤、他(2005)についても、斉田(2003)と同様に、センター試験「英語」受験者の能力値平均を実施年度ごとにプロットしたものを図2に示す。標準偏差については、おおよそ1.0前後であることが報告されており、系統的な経年変化は確認されなかった。

熊谷・山口・小林・別府、他(2007)の結果

上記2つの研究と同様に、熊谷・山口・小林・別府、他(2007)について大学受験生の能力値平均をプロットしたものが図3である。なお熊谷・山口・小林・別府、他(2007)では、年度内に複数実施されたテストについても等化しているため、年度間のほか年度内(5月, 8月, 12月)のデータも示されている。標準偏差については、どの年度・時期においてもほぼ1.0に近い値であった(最大値1.16, 最小値0.84)。

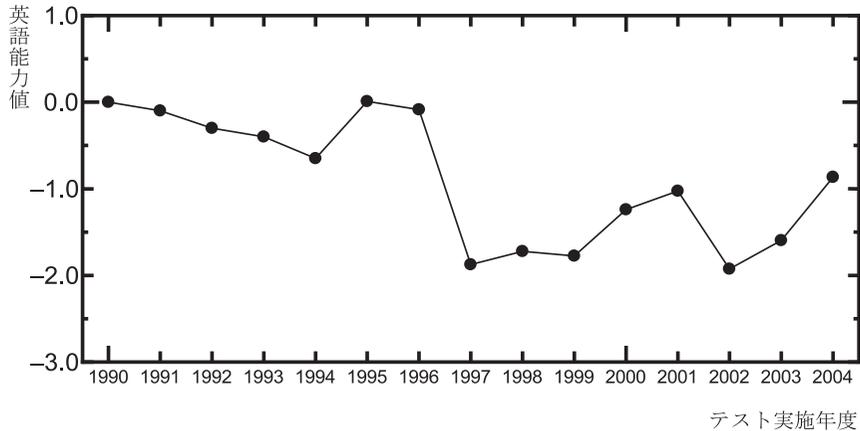


図2 吉村・荘島・杉野・野澤，他（2005）の結果（marco による方法のみ）

また熊谷・山口・小林・別府，他（2007）のデータについては，同一年度内で見ると，複数の実施時期のテストを受験した受験生も相当数存在する．そこで，2005 年度の各時期のテストを全て受験した受験者を抽出し（ $N=104,749$ ），各時期ごとに平均値を算出したところ，5 月が 0.20，8 月が 0.42，12 月が 0.72 であった（図 3 において で表示）．

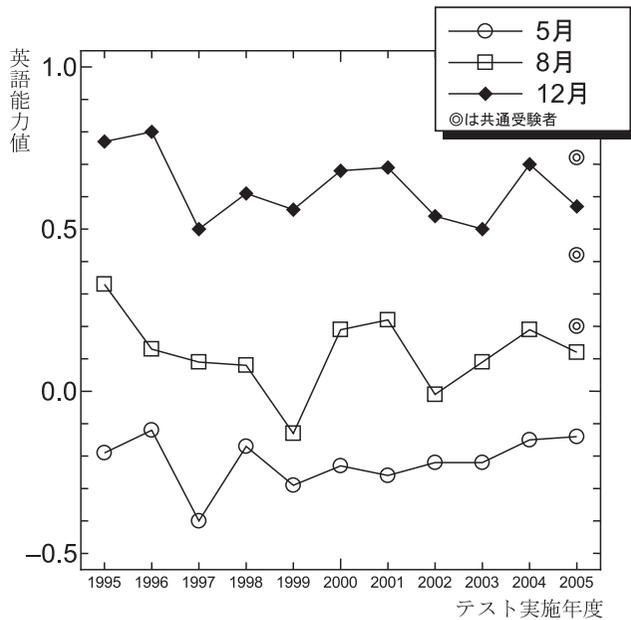


図3 熊谷・山口・小林・別府，他（2007）の結果

結果の比較

3つの研究において，英語能力値の平均値および標準偏差の推移について比較する．比較に際して考慮しておかなければならない点としては，斉田（2003）の研究対象となっているのが茨城県の

高校 1 年生であるのに対して、残りの研究は全国の大学受験生となっていることである。

標準偏差については、全ての研究において各集団の値は 1.0 から大きく変動していないという点で共通していることが確認された（項目反応理論における等化の場合、ある年度の母集団分布について平均値を 0.0、標準偏差を 1.0 と定めることが多い）。この結果から、おおよそ 1990 年から 2005 年にかけて、高校生および大学受験生の英語能力分布の広がりに関しては、系統的に広がっている、もしくは狭まっているということはないことが示唆された。昨今「学力低下」を議論する際の議題として「学力の多様化」が取り上げられることも多いが、量的な広がり、つまり能力分布の散らばりに関しては、そのような傾向は認められなかった。質的な広がりについては、当然、これらのデータのみからは知ることはできない。

次に平均値の推移である。平均値に関しては、3 つの研究間でその数値自体を比較することはできない。それは、各研究内では等化による共通尺度化がなされているものの、研究間で共通尺度が構成されているわけではないからである。しかしながら、各研究での標準偏差がどの時期においてもおおよそ 1.0 であったということから、平均値の変動幅をある程度比較することができる。仮に全ての集団の標準偏差が 1.0 だったとするならば、能力値 0.1 が Z 得点（偏差値）に換算すると 1 に相当する。したがって、例えば斉田（2003）における能力値 1.0 と熊谷・山口・小林・別府、他（2007）における 1.0 ではどちらの能力が高いのかを比較することはできないが、どちらの研究でも能力値が 0.5 減少しているならば、それは母集団に対して同程度の能力低下を示しているということになる（高校 1 年生の母集団と大学受験生の母集団の標準偏差が等しいことが示されているわけではないため、厳密に両者を比較できるわけではない）。

以上を踏まえ能力値平均の推移を見ていくと、研究間での共通点と差異が見えてくる。はじめに斉田（2003）の研究では、能力値平均は年度ごとに低下している。低下の度合いとしては、1995 年から 2002 年にかけて約 0.38 の低下であることが報告されている。また、吉村・荘島・杉野・野澤、他（2005）においては、1990 年から 1996 年までに比べて、1997 年以降の能力が低下していることが報告された。低下の度合いとしては、1996 年から 1997 年にかけて、能力値で約 2.0（先の偏差値換算では 20）という非常に大きな低下が報告されている。

これに対して熊谷・山口・小林・別府、他（2007）では、12 月実施のテストで 1996 年から 1997 年にかけて 0.3 の減少が見られるものの、年度内の推移（例えば 2005 年度において 5 月、8 月、12 月のテストを全て受験した受験者の能力値は、7ヶ月で 0.52 増加）や、8 月実施テストでは、1997 年において顕著な減少が見られなかったことなどから、必ずしも能力値の減少傾向があるわけではないとしている。

さて、吉村・荘島・杉野・野澤、他（2005）と熊谷・山口・小林・別府、他（2007）についてみると、どちらも大学受験生がデータとなっている。吉村・荘島・杉野・野澤、他（2005）のデータは、大学入試センター試験のものであり、テストは毎年 1 月中旬に行われている。また熊谷・山口・小林・別府、他（2007）における 12 月実施部分のデータは、これから大学入試センター試験を受験しようとしている大学受験生のデータであるため、吉村・荘島・杉野・野澤、他（2005）と共通の受験者が多数存在する。そこで、熊谷・山口・小林・別府、他（2007）では、両研究で共通している 1995 年度から 2004 年度の部分を抽出し（図 4）、比較を行っている。図 4 から分かる通り、平均値の変動傾向に関しては両研究で非常に一致していることが分かる。違いは先に述べたとおり、変動幅が吉村・荘島・杉野・野澤、他（2005）では非常に大きく、熊谷・山口・小林・別府、他（2007）での変動幅はむしろ斉田（2003）のものに近い。

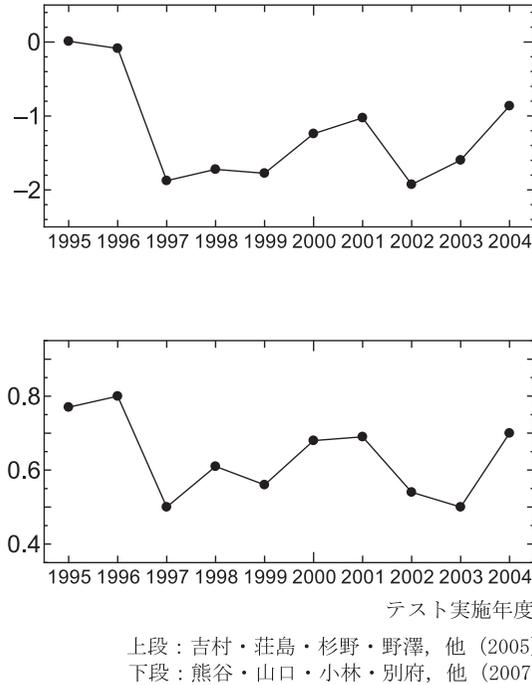


図4 共通受験者による2つの研究結果の比較

吉村・荘島・杉野・野澤，他（2005）では，斉田（2003）や熊谷・山口・小林・別府，他（2007）では示されていない1990年から1994年の情報が示されている．そこで，図4のデータを利用して，吉村・荘島・杉野・野澤，他（2005）から熊谷・山口・小林・別府，他（2007）へと回帰分析を行い，熊谷・山口・小林・別府，他（2007）について1994年以前（12月実施テスト分のみ）を予測したものが図5である．図5から，吉村・荘島・杉野・野澤，他（2005）で示された，1996年以前と1997年以後における能力値の低下が見てとれる．

これら3つの研究で共通しているものとして，ある年度での大幅な能力低下が挙げられる．斉田（2003）では1997年から1998年にかけて，吉村・荘島・杉野・野澤，他（2005）および熊谷・山口・小林・別府，他（2007）では1996年から1997年にかけて，能力値が相対的に大きく低下しているのである．斉田（2003），吉村・荘島・杉野・野澤，他（2005）では，その理由の一つとして，学習指導要領の改訂を挙げている．どちらの研究においても，中学校・もしくは高等学校において学習指導要領改訂後に学習した受験者で，この能力の大幅低下が起きていることを述べており，それは熊谷・山口・小林・別府，他（2007）のデータについても同様のことが言える．

4. 最後に

本稿では，高校生および大学受験生の英語能力の推移を取り扱った3つの研究の比較を行った．このような比較ができたのは，それぞれの研究において等化という手続を利用して，複数年度・時期のテスト結果を相互に比較できるように研究計画がなされたからである．これまで我が国においても膨大な数のテストが実施されてきた．しかしながら，そのテスト結果を相互に比較するという

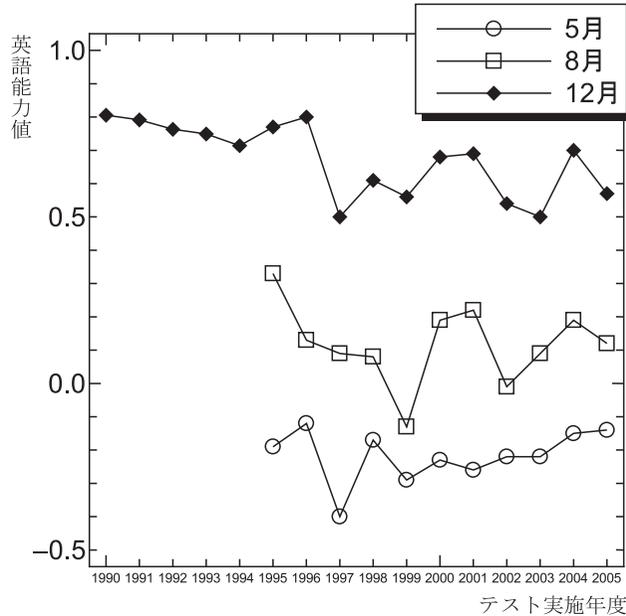


図5 熊谷・山口・小林・別府，他（2007）の結果に対する回帰予測

ことはほとんどなされてこなかった。今回取り上げた3つの研究は、すべて項目反応理論を用いた等化という手続により、そのことを成し遂げたものである。3つの研究はそれぞれが追試研究を目的としたものではないため、たとえば対象学年や、等化計画法など異なる面もいくつか見られる。しかしながら、高校生（もしくは大学受験生）の英語能力を測定することを目的としたテストの経年比較に関して、このように複数の研究者が取り組み、成果を公表することで、英語能力に関して重要な情報が得られたことは間違いない。このような経年比較の研究は、十分に洗練された計画性と大量のデータの蓄積が必要となる。非常に困難ではあるが、英語に限らずその他の教科・能力についても同様の研究がなされることが期待される。

経年比較データを考察する際に、その結果の解釈には十分注意しなければならない。斉田（2003）、吉村・荘島・杉野・野澤，他（2005）、熊谷・山口・小林・別府，他（2007）のそれぞれで注記されているが、「そもそも学力とは何なのか」、「このテストが測定しているものは何なのか」を無視してデータだけの解釈をしてはならない。例えば本稿で取り上げた3つの研究は、どれも「話す能力」などを直接には測定していない。しかしながら、学校現場では当然「話す能力」の学習も行われている。「英語学力」といった場合に、「話す能力」を全く含めないと結論づけることができないのは明白なことである。従って、今回の3つの研究結果だけを見てすぐに「英語学力の低下」と結びつけることはできないのである。既に述べたが、本稿で取り上げた以外にも様々な学力・能力に関するデータを蓄積することで、「学力の経年変化」に関する議論を進めていく必要があるだろう。

参考文献

- Marco, G. L. "Item characteristic curve solutions to three intractable testing problems", *Journal of Educational Measurement* **14** (1977), 139-160.
- Shojima, K. "Linking tests under the continuous response model", *Behaviormetrika*, **30** (2003), 155-171.
- 木村拓也。「何故、日本の学力調査には科学的測定論が根付かなかったのか？ - 戦後日本で実施された全国学力調査の変遷と「テストの専門家」養成能力の実態 - 」, 『日本テスト学会第4回大会発表論文抄録集』(2006), 90-93.
- 熊谷龍一, 山口大輔, 小林万里子, 別府正彦, 脇田貴文, 野口裕之。「大規模英語学力テストにおける年度間・年度内比較」, 『日本テスト学会誌』(2007) **3**, 83-90.
- 齊田智里。「高校入学時の英語能力値の年次推移 - 項目応答理論を用いた県規模英語学力テストの共通尺度化 - 」, 『第15回英検研究助成報告』(2003), 日本英語検定協会, 12-24.
- 芝祐順, 野口裕之。「語彙理解力尺度の研究 - 追跡データによる等化 - 」, 『東京大学教育学部紀要』**22** (1982), 31-42.
- 全米学力調査研究会(編)。『全米学力調査(NAEP)の研究』(2005)。
- 野口裕之。「被験者の推定尺度値を利用した潜在特性尺度の等化方法」, 『教育心理学研究』**31** (1983), 233-238.
- 文部科学省。「全国的な学力調査について」, 最終閲覧日 2007年7月27日, (2007)
http://www.mext.go.jp/a_menu/shotou/gakuryoku-chousa/index.htm .
- 吉村 宰, 荘島宏二郎, 杉野直樹, 野澤健, 清水裕子, 齋藤栄二, 根岸雅史, 岡部純子, サイモンフレイザー。「大学入試センター試験既出問題を利用した共通受験者計画による英語学力の経年変化の調査」, 『日本テスト学会誌』(2005) **1**, 51-58.