流通科学大学論集―人間・社会・自然編―第 27 巻第 2 号，95-104（2015）

＜書評＞

斉田智里　2014.　『英語学力の経年変化に関する研究　―項目応答理論を用いた事後的等化法による共通尺度化』東京：　風間書房

井澤廣行*、平越裕之†

Hiroyuki Izawa, Hiroyuki Hirakoshi

著者の才覚、知恵、及び、献身による偉業であり名著である。その賛辞は、本書業績への以下に挙げる評者視点に基づく。特に、1)が本書独自の意義であり、著者の知恵から生まれた「事後的等化法」は日本におけるテスト研究分野の歴史にその名を残すであろう。

1)　項目応答理論を用いた「事後的等化法」の発案と実践。

2)　「根拠(evidence)」を提示する精緻な研究分析。

3)　「学習指導要領外国語科」、「テスト理論」、「第二言語習得理論」を含む広範な文献調査。

4)　項目応答理論と因子分析を実行する最新ソフトウェアの使用熟練。

本書刊行を評者に知らしめた著者による小論(斉田 [1]、2014)の中に、次の言葉がある。これが本書の核心点である。

> 「エビデンスに基づいた教育効果の検証と教育政策の立案を行うためには､尺度の等化により学力の経年比較が可能となるテスト開発が不可欠です」(p. 18)。

要するに、国際的な学力調査(The OECD Programme for International Student Assessment: PISA)や全米学力調査(The National Assessment of Education Progress)では項目応答理論の適用による尺度等化で以て学力の経年比較が可能である(本書、pp. 17-19)が、日本の学力調査では、「技術的には可能である」としても、その状況にない(p. 96)ということである。なお、村木・斉田 [2] (2008, p. 78)によれば、全米学力調査において、多肢選択形式としての 2 値(0・1)項目群データ、短答形式項目への 2 値処理データ、及び、多値処理される記述式項目群データに対して、それぞれ順に、「3 パラメタ・ロジスティック・モデル」(以降、3PLM)、「2 パラメタ・ロジスティック・モデル」(以降、2PLM)、及び、「一般化部分採点モデル」(村木 [3]、2011, pp. 86-90、参照)が適用されているとのことである。

*流通科学大学サービス産業学部、〒 651-2188　神戸市西区学園西町 3-1
†流通科学大学総合政策学部、〒 651-2188　神戸市西区学園西町 3-1

（2014 年 9 月 1 日受理）
Ⓒ 2015 UMDS Research Association

本書評冒頭で記したように、著者の偉業は項目応答理論(Item Response Theory、以降、IRT)を用いた「事後的等化法」の発案と実践である。渡辺・野口 [4] (1999)は、「等化」を「複数の異なるテスト(尺度)による測定結果を共通に表示できる共通尺度を構成する手続き」(p. 26)と定義している。素点に基づく古典的等化は論理的な矛盾を有しており(池田 [5]、1994, pp. 158-159)、素点上でのその実質的な等化不能性はテスト得点におけるテスト難易度と受験者能力の不分離性に起因している(斉田 [1]、2014, p. 17)。池田 [5] (1994)は次のように述べている。

> 「古典的等化法ではこうした問題を解決することができないが、もし与えられたテストが項目応答モデルに適合するならば、少なくとも理論的には、この問題を解決することができる。なぜなら項目応答モデルによると、項目特性値は用いたサンプルの能力に依存しないし、受験者の能力特性値も用いたテスト項目群に依存しない。項目特性値の分かった 2 つのテストを受けた受験者の能力は同一の共通尺度上に位置づけられ直接比較することができる」(p. 159)。

TOEFL (Test of English as a Foreign Language)を含む国際的大規模テストで用いられている IRT 適用等化法は「係留テスト・デザイン」(「共通項目デザイン」)である(大友 [6]、1996, p. 236)。共通項目が存在しない公的な国内大規模既存テスト群への斉田による本書での IRT 適用「事後的等化法」は、その性質上、「共通受験者デザイン」となる。なお、本書評での斉田 [1] (2014)を除く参照文献すべてにおいて IRT 適用等化法に関する論理と数理が与えられているが、「事後的等化法」への直接的な言及はない。それは、まさしく著者の知恵から生まれたものである。以下に、本書での「事後的等化法」を重要事項として参照する。

2PLM の基本式として次式が与えられている(本書、p. 26)。

$$P_j(\theta) = \{1 + \exp[-Da_j(\theta - b_j)]\}^{-1} \quad (1)$$

θ： 受験者能力パラメタ値

D： 尺度因子であり、D=1.7 のときに θ 全域にわたり正規累積 2PLM との違いが 0.01 以下になることが知られている(村木 [3]、2011, p. 45)。

a_j： 項目 j の識別力パラメタ値

b_j： 項目 j の困難度パラメタ値

$P_j(\theta)$： 能力パラメタ値 θ を持つ受験者の項目 j への正答確率

「項目反応理論では項目の識別力・困難度および被験者の特性尺度値を表わす尺度の原点と単位は線形変換の範囲で任意に定められる」(野口 [7]、1991, p. 55)。従って、*を等化後の尺度の記号として、

$$a^* = a/K \quad (2)$$

$$b^* = Kb+L \quad (3)$$

$$\theta^* = K\theta + L \quad (4)$$

との変換を施しても、$P_j(\theta^*) = P_j(\theta)$となり同一の項目特性曲線が与えられる(本書、p. 27)。なお、

2PLM におけるこの線形変換による尺度共通性は、3PLM 及び $a_j = a = 1$ とする 1 母数モデル(1PLM)においても同様である(豊田 [8]、2002, p. 85)。これは、IRT における項目パラメタ値と能力パラメタ値の両尺度が不定性を持つ(村木 [3]、2011, p. 109)ということであり、「LOGIST5 のように当該被験者集団で推定尺度値の平均が 0.0、標準偏差が 1.0 となるように尺度の原点と単位がとられる場合が多い」(野口 [7]、1991, p. 55)と述べられている。

上に参照した式(2)、(3)、(4)における K と L は等化係数と呼ばれ(渡辺・野口 [4]、1999, p. 28)、斉田による本書でのその推定法は「平均シグマ法(mean-sigma method)」(本書、p. 28)である。平均シグマ法が海外での大規模テストの尺度等化において使用されている一例が野口 [7] (1991, pp. 65-66)により紹介されている。なお、3PLM における当て推量パラメタ値は、そのモデル数式の上で「尺度の変換に対して不変である」(池田 [5]、1994, p. 161)とされて、複数テスト間で比較可能である(本書、p. 83; 豊田 [8]、2002, p. 84、参照)。

等化係数 K と L は、「等化後の尺度における能力推定値を θ^*、等化前の尺度における能力推定値を θ、標準偏差を $s(\cdot)$、平均を $m(\cdot)$とすると、式(5)と(6)から求めることができる」(本書、p. 28)。

$$K = s(\theta^*) / s(\theta) \tag{5}$$

$$L = m(\theta^*) - K \times m(\theta) \tag{6}$$

「平均シグマ法」による K と L のこの推定法は、能力推定値 θ^* と θ の標準化された値が一致するとみなされることにより以下のように算出される(村木 [3]、2011, p. 110、参照)。

$$[\theta^* - m(\theta^*)] / s(\theta^*) = [\theta - m(\theta)] / s(\theta)$$

$$\rightarrow \theta^* = [s(\theta^*) / s(\theta)] \times \theta + \{ m(\theta^*) - [s(\theta^*) / s(\theta)] \times m(\theta) \} = K\theta + L \tag{7}$$

なお、式(7)は能力推定値 θ^* と θ に誤差がないものとして導出されており、「推定精度のよくない被験者の尺度値の影響を受け易い」(野口 [7]、1991, p. 62)と指摘されている。このことに関して、村木 [3] (2011)は次のように述べている。

> 「IRT による等化方法において問題点をあげるなら、その方法のどれもが項目パラメータの推定誤差を考慮に入れていない点であろう。受験者の数が多い場合や項目パラメータの推定誤差が無視できる場合は、等化の結果に誤差の影響はあまりないだろう。項目パラメータの推定誤差と等化の関係のいっそうの研究が望まれる」(p. 113)。

「事後的等化法」は、前述した通り、その性質上「共通受験者デザイン」となる。「事後的等化法の手順」としてその実践法が次のように与えられている(本書、pp. 28-31)。

1) 「一次元性の確認」: IRT 適用の条件として、テスト項目群の高い程度の 1 因子性が必要とされる。著者の確認法は、TESTFACT4.0 を使用しての四分相関係数に基づく因子分析主因子解により出力される第 1 因子負荷量の大きさと固有値の減衰状況の視認である(pp. 55-57)。著者は、第 1 因子負荷量が 0.2 未満の項目を各テストから削除した上で、再度の因子分析主因子解による第 1 固有値と第 2 固有値の比率の大きさと固有値の減衰状況の確

認で以て、項目群 1 因子性の高さを判断している(p. 56)。

2) 「等化の基準となるテストの決定」: 複数の既存テストの中から等化の基準となる 1 つの最も良質なテストを決定する。1)での項目群 1 因子性の高さと削除項目数の少なさ、及び、クロンバックのアルファ係数の大きさが、最も良質なテストの判断材料とされている(p. 59)

3) 「項目パラメタ値の推定」: IRT ソフトウェア BILOG-MG3.0 を使用して、1)での第 1 因子負荷量が 0.2 未満の項目を削除した各テストの項目パラメタ値を推定する。なお、「実データを用いた等化実験」(pp. 31-38)では 2PLM が採用されている。主研究では、「今後のテスト改善の方向性を明らかにする」(p. 79)ことが目的とされて、より精緻な項目分析(pp. 79-89)を可能とする 3PLM が適用されている(p. 58)。1 万人を超える受験者数(表 3.1、p. 43、参照)を持つテストへの 3PLM 適用に支障はないとの著者による判断である(p. 59)。

4) 「等化用テストの作成」: 2)でのその決定された等化基準テストの項目群と「等化をしたいテスト」の項目群から項目数をそれぞれほぼ半々として、既存テストとほぼ同一の項目数・問題領域構成・出題形式から成る「等化用テスト」を作成する。「等化用テスト」の項目群選出に際しては、等化の精度を高めるために、困難度推定値において適当なばらつきがあり、識別力推定値の高い項目群を優先する(p. 30-31)。

5) 「等化用テストの実施」: その作成された「等化用テスト」を既存テストの受験者群母集団からの抽出とみなし得て、項目群パラメタ値の推定精度を高めるために能力値にばらつきのある新たな 400 名以上の受験者群に実施する(p. 31、及び、表 3.9、p. 61、参照)。

6) 「等化係数の算出」: この箇所での著者による記述が評者には理解出来ていない。従って、著者の言葉をそのまま次に引用する。「「等化用テスト」を受検した生徒の能力値は、それぞれのテストの項目パラメタ値を用いて、2 通りに推定される」(p. 31)。この引用文における「それぞれのテストの項目パラメタ値」は、「それぞれの既存テスト項目の項目パラメタ推定値」(p. 34)と書き直されている。更に、「項目パラメタ値を固定した上で、能力値を EAP 法により推定」(p. 91)ともあり、BILOG-MG3.0 の適用により各既存テストにおける項目パラメタ推定出力値を使用して、2 通りの受験者能力推定値を得ると思われる。「2 通りに推定された能力推定値の平均と標準偏差を用いて、平均シグマ法により、式(5)と式(6)から等化係数 K, L を求める」(p. 31)。

7) 「等化後の項目パラメタ値の推定」: 得られた等化係数 K と L の式(2)と式(3)への代入により、等化後の識別力パラメタ値と困難度パラメタ値を算出する。これにより、共通尺度上での「項目プール」の作成が可能となる(p. 31)。

8) 「等化後の能力パラメタ値の推定」: この箇所での著者による記述も評者には理解出来ていない。従って、著者の言葉をそのまま次に引用する。「等化後の項目パラメタ値を用いて、

等化後の能力推定値を求める」(p. 31)。この引用文に関して、「等化後の項目パラメタ値を用いて、各年度テスト受検者の IRT 尺度値(共通尺度の能力推定値)を求めた。IRT 尺度値を求めるにあたっては、BILOG-MG で期待事後(expected a posteriori: EAP)推定値を指定した」(p. 62)とある。これによっても、著者による IRT 関連最新ソフトウェア使用熟練の程が窺われる。

以下に、本書での斉田による「事後的等化法」の実践に基づく「英語学力の経年変化」の研究分析成果を参照する。先ずは、その分析に関係する要素説明を与える。

分析対象テスト群は 1995 年から 2008 年に渡る 14 年間の「茨城県高等学校英語学力テスト A」である。1995 年から 2008 年にかけて、そのテストの受験参加高校数は 80 から 50 へ、又、高校 1 年生受験者数は 17,736 人から 7,791 人へ漸減している(表 3-1、p. 43、参照)。そのテスト項目群は、「リスニング」、「音声」、「語彙」、「会話文完成」、「文法」、「並べ替え」、及び、「読解」から構成されている(表 4-2、p. 84、参照)。項目総数を、1997 年度までは 50、1998 年度以降 46 とされて(p. 42)、4 枝択一選択のマークシート解答方式である(pp. 42-43)。そのテストは、毎年 4 月に茨城県内公立高校(当時ほぼ 111 校)におけるテスト受験希望校にて 1 授業時 45 分間で一斉に実施された(pp. 41-42)。

次に、上記「事後的等化法の手順」に従い、著者の研究分析準備段階でのデータ作成が順次に記述されている。以下がその要旨である。

1) 「一次元性の確認」: 先ず、第 1 因子負荷量が 0.2 未満の項目が各テストから削除された。再度の因子分析主因子解による第 1 固有値と第 2 固有値の比率の大きさと固有値の減衰状況の視認により、14 テストそれぞれの項目群 1 因子性の高さが確認された(p. 56)。
2) 「等化の基準となるテストの決定」: 第 1 固有値と第 2 固有値の比率が 6.94(表 3.7、p. 57、参照)、クロンバックのアルファ係数が 0.88(表 3-8、p. 58、参照)、削除項目数が 6 である 1999 年度テスト A が「等化の基準となるテスト」とされた(p. 59)。
3) 「項目パラメタ値の推定」: BILOG-MG3.0 上での 3PLM の適用により、1)での第 1 因子負荷量が 0.2 未満の項目を削除した各テストの項目パラメタ値が推定された(pp. 58-59)。
4) 「等化用テストの作成」: 2)での「等化の基準となるテスト」とされた 1999 年度テスト A と他年度実施各テスト A から 13 冊の等化用テストが作成された。元のテスト項目数 46 及び問題領域構成・出題形式との同一性・困難度推定値の適度なばらつき・項目識別力推定値の高さが項目選出基準とされた。但し、リスニング問題と読解問題それぞれの不分離制限もあったが故に、等化基準 1999 年度テスト A と 2005 年度までの各年テスト A の項目数比は 28:18 とされた。2006 年度以降の等化用各テストについては、「実施上の理由から」、項目数比は、2006 年度 24:16、2007 年度 18:14、2008 年度 18:14 とされた(pp. 60-61)。
5) 「等化用テストの実施」: 13 冊の等化用テストが、「茨城県高等学校教育研究会英語部」の協力の下、高校単位で実施された。各等化用テストの受験者群能力パラメタ推定値に大

きなばらつきが得られるように、学力水準の幅広い高校群から受験者群が選出されて、1冊あたり受験者数総計 380～550 名へのテスト実施(表 3-9、p. 61、参照)とされた。等化用テストの実施時期については、1995 年度から 2002 年度までの 7 冊が 2002 年 9 月、2003 年度と 2004 年度の 2 冊が 2004 年 6 月、2005 年度の 1 冊が 2005 年 9 月、及び、2006 年度から 2008 年度までの 3 冊が 2008 年 7 月であった(pp. 60-61)。

6) 「等化係数の算出」: 前述したように、この箇所での著者による記述が評者には理解出来ていない。従って、著者の言葉をそのまま次に引用する。「等化用テストの項目は 2 つの既存テストの項目から構成されているので、「等化用テスト」を受けた生徒の能力値は、それぞれの項目パラメタ値を用いて 2 通りに推定される」(pp. 60-61)。「平均シグマ法」による 2 通りの能力推定値分布の各平均値と各標準偏差に基づいて算出された等化係数 K と L の値 13 組が表 3-9(p. 61)に与えられている。なお、等化用テストデータにおいて、20%以上の無回答項目がある受験者は除かれたとあり、13 冊の等化用テストへの有効受験者数合計は 6,050 名であった(p. 61)。

7) 「等化後の項目パラメタ値の推定」: 得られた等化係数 K と L の式(2)と式(3)への代入により、等化後の項目群での識別力パラメタ値と困難度パラメタ値が算出された(pp. 61-62)。

8) 「等化後の能力パラメタ値の推定」: この箇所での著者による記述も評者には理解出来ていない。従って、著者の言葉をそのまま次に引用する。「等化後の項目パラメタ値を用いて、各年度テスト受検者の IRT 尺度値(共通尺度の能力推定値)を求めた。IRT 尺度値を求めるにあたっては、BILOG-MG で期待事後(expected a posteriori: EAP)推定値を指定した」(p. 62)とある。

著者による「英語学力の経年変化」分析成果の最たる 1 点は英語学力漸次低下である。図 3.2「高等学校入学時の英語学力経年変化(平均値)」(p. 64)、及び、図 3.3「高等学校入学時の英語学力経年変化(パーセンタイル)」(p. 65)にそれが端的に示されている。かくして、著者を含む当時の茨城県公立高校教員間に共有されていた高校 1 年生入学時の「英語学力低下懸念」(p. iii)が実証された訳である。1995 年から 2008 年に渡って毎年 4 月に実施された「茨城県高等学校英語学力テスト A」により測定された高校 1 年生の英語学力低下傾向の「エビデンス」(本書で著者により 9 回使用されている言葉：日本語訳「根拠」)を著者が提示したことになる。なお、評者は無知・不詳であったが、2003 年に斉田による「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」が「日本英語検定協会」による「第 15 回英検研究助成報告」として発表されている(p. 116)。

表 3-10「高等学校入学時の英語学力特性値記述統計量(年度別)」(p. 63)、及び、図 3.3「高等学校入学時の英語学力経年変化(パーセンタイル)」(p. 65)のそれぞれに関する著者の言葉を下に引用する。

「本データの開始年度である 1995 年度から 2008 年度まででは、全体で 0.74 の低下である。偏差値換算で 7.4 点も低下していたことになる。2008 年度に偏差値 50 であった成績中位

者(1 万人の受検者中 5,000 番の生徒)が 1995 年度のテストを受けるとすれば偏差値が 42.6 に下がり、順位も約 2,704 番下がって 7,704 番相当の実力になるということである。低下の程度は大きいといえる」(p. 65)。

「平成元年[1988 年]度改訂学習指導要領実施時期[中学校各学年 1 週英語授業時数 4 回(本書、p. 7、参照)]には、成績上位層と中位層に顕著な低下傾向が見られるが、平成 10 年[1998 年]度改訂学習指導要領中学校実施開始[中学校各学年 1 週英語授業時数 3 回(本書、p. 7、参照)]の 2002 年度以降[その指導要領高等学校実施は 2003 年からの年次進行(本書、p. 8、参照)]は、成績中位層と下位層の低下傾向が顕著で、成績上位層と下位層との格差が拡大していることが観察される」(p. 66)。

更に、本書評冒頭提示での 2) 著者の研究分析への精緻を示すものとして、「学校単位」と「継続受検校」での「英語学力の経年変化」が提示されている。前者と後者は、それぞれ、「学力層の高い高等学校の参加が年々減っている可能性」(p. 66)と「受検校の変動要因」(p. 68)を取り除くためである。図 3.4「学校単位での英語学力経年変化(平均値)」(p. 67)において、全体に右肩下がりの傾向が確認される。表 3-11「継続受検校の英語学力特性値記述統計量(年度別)」(p. 68)からも、1995 年度から 2008 年度に渡る英語学力漸次低下傾向における全体平均で 0.64(偏差値換算により 6.4 点)の低下であった。中学校各学年 1 週英語授業時数が 2001 年度までの 4 回であったが、3 回となった 2002 年度(表 1.2、p. 7、参照)から 2003 年度の高校入学時にかけて、特に成績中位層と成績下位層における落ち込みが顕著であった(図 3-6、p. 69、参照」)。著者の言葉を引用すれば、「全体で見ると、[1995 年度から 2008 年度に渡って、]成績上位層と中位層及び成績上位層と下位層との差がやや拡大し、成績中位層と下位層との差はやや縮小する傾向が見られた」(p.70)。

上に参照した表 3-10「高等学校入学時の英語学力特性値記述統計量(年度別)」(p. 63)、及び、表 3-11「継続受検校の英語学力特性値記述統計量(年度別)」(p. 68)における 1995 年度から 2008 年度にかけての「英語学力特性値」のそれぞれにおける平均値低下の値 0.74 と 0.64 の意味が、「項目プールの活用における学力の伸びの可視化」(pp. 91-92)の下に与えられている。それは研究精緻を示すものである。ある県立高等学校 1 年生 319 名が調査対象とされた。2004 年 4 月における「茨城県高等学校英語学力テスト A」における県全体の平均値−0.23(標準偏差 0.83、表 3.10、p. 63、参照)に比して、その 319 名の同テストでの平均値は 0.90 であり、県平均より 1 標準偏差を大きく超える平均値を持つかなり高い学力層であった(p. 91)。2004 年 4 月でのそのテスト A の一斉「受検」に加えて、2004 年 9 月、2005 年 1 月、及び、2005 年 3 月での実力テスト実施の際に、著者による本研究の成果である 569 の共通尺度化項目群(p. 90)から選出された 30 項目程度が各実力テストに含められた。その「項目パラメタ値を固定した上で、」BILOG-MG3.0 を用いて、「能力値を EAP 法により推定」(p. 91)された結果に基づく基本統計量が、表 4.4「高等学校入学後の英語学力特性値の変化(1 年間の追跡調査)」(p. 91)に与えられている。1 年間でのその上昇値平均 0.46 と比較して、「同じ尺度で測定した県全体の高

校入学時の英語学力特性値が 14 年間(1995～2008)で平均 0.74 の低下というのは、1 年間の[かなり高い学力層による]学習でも追いつかないくらいの大きな低下である」(pp. 91-92)と著者は指摘している。

本研究に基づく「考察」として「英語学力の経年変化と学習指導要領実施との関係」が第 3 章第 5 節(pp. 71-78)に与えられている。その結論部において、著者は次のように述べている。

> 「学習指導要領を実施した結果、どのような効果が見られ、どのような点に改善が必要であるかを客観的なエビデンスに基づいて検証し、国民に納得のいく形で示し、検証結果を次の学習指導要領改訂や教育施策実現に活かしていくという教育政策立案の姿勢が強く求められる。なにより大事なのは、次世代の日本を担う子供たちである。まずは国民の重要な学力水準について知り、エビデンスに基づいた教育政策を実施していくために、国際的な学力調査[例えば、TIMSS2003 及び PISA2003、本書、p. 17、参照]に頼るのではなく、自国で信頼ある教育評価システムを作ることが緊急の課題であると思われる」(p. 78)。

以上が、本書評冒頭に挙げた著者による 1)項目応答理論を用いた「事後的等化法」の発案と実践、及び、2)「根拠(evidence)」を提示する精緻な研究分析の真髄である。評者に欠如している「知恵」が著者による英語教育学研究者としての労苦をいとわない実証主義精神により体現されている偉業である。その実証を顕示したいとの著者の真摯な想いと態度が、「事後的等化法」の実践における 13 冊の「等化用テストの実施」に際して、「茨城県高等学校教育研究会英語部」からの実に多大な協力を得さしめたのであろう。3)「学習指導要領外国語科」、「テスト理論」、「第二言語習得理論」を含む広範な文献調査に関しては、国立教育政策研究所による「教育課程実施状況調査」への言及(pp. 12-16)、並びに、「茨城県高等学校英語学力テスト A」の妥当性検討(pp. 40-55)に印象づけられる。著者による先行調査・文献への敬意、及び、自己研究におけるその参照に基づく学術的意義に関する「根拠」提示への徹底的なこだわりが窺われる。4)項目応答理論と因子分析を実行する最新ソフトウェアの使用熟練については、2PLM と 3PLM の適用において必須のソフトウェアと思われる BILOG-MG3.0、TESTFACT4.0、並びに、EasyInfo の著者による使用技量習熟の程が、評者には驚嘆である。まさに、教育学、特に、教育・心理測定を専攻する学徒に大なる学究意欲を喚起する本書内容である。

最後に、「事後的等化法」に関するさらなる議論が以下の点においてなされることを望みたい。これは、著者により、「等化の精度を高めるために」(p. 31、及び、p. 60)との言葉があるように等化用テスト項目選出の上で意識されているが、その精度に関する直接的な言及がなされていないことによる。

- 等化係数 K と L の精度は、等化用テストの項目数によってどのように変化するのか
- 等化用テストの項目数は、実用上どの程度であれば良いと考えられるのか

著者は、14 の異なるテストの尺度等化において 13 種類の等化用テストを作成して実施している。それぞれの等化用テストの項目数は、10 冊において 46、2 冊において 32、1 冊において 40 であり、各等化用テストの受験者数は 380～550 名である(表 3.9、p. 61、参照)。各等化用テスト作成において項目群選出に等化係数を高める工夫がされている(p. 60)とはいえ、10 冊における項目数は 46 であり、

2 通りの受験者能力推定値を得るための項目数比は 28:18 である(表 3.9、p. 61、参照)。2 通りに得られた受験者能力推定値分布における各平均値と各標準偏差に基づく等化係数算出の上で等化後の項目パラメタ値を得る(pp. 61-62)のであるが、IRT の適用と分析における 28 と 18 の項目数は少ないという印象を持つ。46 項目から成る既存テストという制限はもちろん理解できる。然しながら、28 と 18 のこの少ない項目数が IRT による受験者能力推定値の安定性に与える影響、延いては、そこからさらに導出される等化係数に与える影響についての言及が望まれる。一般的に、受験者数が項目数に比べてはるかに多い場合には項目困難度の推定値は安定するが、項目数が少なくなるにつれて受験者能力推定値の真値からのばらつき程度は大きくなる。つまり、能力推定値は真の能力値にノイズを加算したような形となり、能力推定値の分散は真の能力値の分散よりも、ノイズの分散分大きくなるということである。本書の内容が秀逸であるからこそ、「事後的等化法」における「等化用テスト」実施上での項目数と受験者数に関する何らかの議論が欲しい。詳細な理論的分析は困難と思われるが、項目数による等化係数への影響をシミュレーション等による何らかの方法で検討して、「事後的等化法」の適用を望む研究者へのガイドラインとなるような見解が本書の改訂版にあればと考える。この意味で、

- 等化係数 K と L の精度は、等化用テストの項目数によってどのように変化するのか
- 等化用テストの項目数は、実用上どの程度であれば良いと考えられるのか

についての論述がまさしく期待される。

末尾付言として、以下の著者記述 2 点に関する丁寧な説明も改訂版にあればと思う。BILOG-MG3.0 のマニュアルにその推定法に関する記載があるかも知れないが、IRT の実践未経験者にはその論理が分かり難い。「事後的等化法」の実践に不可欠であるが故に、又、才覚のある著者であるからこそ、その論理の説明が望まれる。

1) 「等化用テストの受検者の能力値は、それぞれの既存テスト項目の項目パラメタ推定値を用いて、2 通りに推定できる」(p. 34、及び、その同内容文、p. 61)。
2) 「項目パラメタ値を固定した上で、能力値を EAP 法により推定」(p. 91、及び、その同内容文、p. 62)。

[追記]

名著に値するその研究業績の質の高さにより、将来において、本書の増刷ないしは改訂がなされる可能性を評者は感じる。それを鑑みて、本書における誤字・脱字の類の校正必要箇所を下に指摘する。「A (p.) → B」において「→」が「への変更(が望ましい)」との意味である。著者が本書評を目にされる折に、参考にして下されば幸甚である。

1) 「実態のある英文和訳」(p. 5) → 「実体のある英文和訳」
2) 「習得されていない状況が伺える。」(p. 13) → 「習得されていない状況が窺える。」
3) 「英語学力低下は伺えないが、」(p. 16) → 「→英語学力低下は窺えないが、」

4) 「Organisation」(p. 18) → 「Organization」
5) 「複数のテストを共通の受検者が」(p. 19) → 「後者を複数テストの共通受検者が」
6) 「(2LPM)」(p. 26) → 「(2PLM)」
7) 「同じ学年対象のテストでは、2.5 点、異なる学年対象の」(p. 36)
→ 「同じ 1 年生対象のテストでは、2.5 点、同じ 2、3 年生対象の」
8) 「−1 よりも小さい負の値」(p. 46) → 「−1 と 0 の間にある負の値」
9) 「受検をしていることが伺える。」(p. 50) → 「受検をしていることが窺える。」
10) 「受験者能力値の推定精度」(p. 60) → 「項目パラメタ値の推定精度」
11) 「式(2)〜(4)」(p. 61) → 「式(2)と式(3)」
12) 「20 パーセンタイル」(p. 62) → 「25 パーセンタイル」
13) 「表 3-10」(p. 70) → 「表 3-11」
14) 「成績上位層を中位層の格差」(p. 71) → 「成績上位層と中位層の格差」
15) 「成績上位層と成績中位層及び成績下位層との格差」(p. 72)
→ 「成績上位層と成績下位層、及び、成績中位層と成績下位層の格差」
16) 「会話文完成が最も低く、」(p. 88) → 「音声が最も低く、」
17) 「会話文完成にのみ、中程度の正の」(p. 89)
→ 「会話文完成・音声・並べ替えに中・小程度の正の」
18) 「(表 3-9, p. 61)」(p. 91) → 「(表 3-10, p. 63)」
19) 「能力の育成のためは、」(p. 94) → 「能力の育成のためには、」
20) 「資格を保有していること条件」(p. 99) → 「資格を保有していることを条件」

【参照文献】

1) 斉田智里　2014. 「英語力はどう測るのか。」 『英語教育』第 62 巻、第 11 号、pp. 16-17. 東京：大修館
2) 村木英治・斉田智里　2008. 第 6 章「調査デサインの考え方と方法」、荒井克弘・倉元直樹編著『全国学力調査　日米比較研究』(pp. 66-80) 東京：　金子書房
3) 村木英治　2011. 『項目反応理論』 東京：　朝倉書店
4) 渡辺直登・野口裕之編著　1999. 『組織心理測定論 ― 項目反応理論のフロンティア』 東京：　白桃書房
5) 池田 央　1994. 『現代テスト理論』 東京：　朝倉書店
6) 大友賢二　1996. 『項目応答理論入門』 東京：　大修館書店
7) 野口裕之　1991. 3 章「項目反応理論にもとづくテストの作成法」、芝 祐順編 『項目反応理論 ― 基礎と応用』(pp. 51-86) 東京大学出版会
8) 豊田秀樹　2002. 『項目反応理論 [入門編]』 東京：　朝倉書店