

科学研究費助成事業 研究成果報告書

令和 2 年 6 月 10 日現在

機関番号：32663

研究種目：基盤研究(C) (一般)

研究期間：2016～2019

課題番号：16K02985

研究課題名(和文) 英語読解診断テストの開発と妥当性の評価について

研究課題名(英文) Developing English reading diagnostic test and evaluating its test score validity

研究代表者

熊澤 孝昭 (KUMAZAWA, Takaaki)

東洋大学・経済学部・准教授

研究者番号：20366933

交付決定額(研究期間全体)：(直接経費) 3,100,000円

研究成果の概要(和文)：本研究の成果としては、まず日本人大学生で特にヨーロッパ言語参照枠(CEFR)でのA1またはA2の基礎段階の言語学習者が読解選択式テストの解答時における認知プロセスの解明を試みた点にある。また、CEFRは学習者レベルの大枠を記述したもののだが、より調査対象者に適した読解スキルに於いて「できるか」を自己評価するための質問紙の開発を試みた点である。

研究成果の学術的意義や社会的意義

受験者の中でも特に中級・基礎レベルの対象者はテスト受験中には特に一文一文を正確に読み取ろうとしない傾向にあるということがいえる。また、キーワードを拾い読みするのみで正確な文理解ができないことともいえる。この観点からもリメディアル教育の一環として訳読などを通しての精読の重要性は否定できない。認知診断評価(CDA)の枠組みでのプレースメントテスト開発ならびに認知診断モデルの有用性である。CDAの一環としてカリキュラム内での学修到達度目標の習熟度を測定するプレースメントテストを開発し、さらに一つの項目で複数の認知スキルを測定することができるCDMから得られるフィードバックは有益的な情報となりうる。

研究成果の概要(英文)：One of the major outcomes of this research is that this research made an attempt to find out basic learners' (A1 and A2 as reference to CEFR) cognitive processing while taking a diagnostic reading test. The other major outcome is that while CEFR is used to describe overall levels, this research made an attempt to develop a self-evaluation can-do survey that is designed for specific targeted learners' reading achievement.

研究分野：応用言語学

キーワード：言語テスト理論 英語教育学

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

近年、英語外部テストを含む受験者数が多い大規模なテスト (large-scale) は得点のみではなく、受験者の英語力を記述することが重要視されており、英語力の指針としてヨーロッパ言語共通参照枠 (Common European Framework of Reference for Languages) や Can-Do リスト (can-do statement) が用いられている。また、成績評価をするための学修度合いを評価する総括的評価 (summative assessment) ではなく、テスト結果からフィードバックを与え、より学習者の学習や指導を向上させるための形式的評価 (formative assessment) の研究が注目されている。それ以外にもテスト結果をもとに学習の方向性を判断するための学習志向的評価 (learning oriented assessment [Jones, & Saville, 2016; Purpura, 2004]) や社会的文化論 (socio-cultural theory) を基盤とし、より評価および指導を統合するダイナミック評価 (dynamic assessment [Antón, 2009]) などが注目されている。つまり、英語力の度合いを評価するのみではなく、評価をいかに学習に結びつけるかが言語テストの分野での喫緊の課題といえる。その一環として特に注目を浴びているのがよりきめ細かいスキルの到達度合いを診断する診断的評価 (diagnostic assessment) についての研究 (e.g., Alderson, 2005; Alderson, Haapakangas, Huhta, Nieminen, & Ullakanoja, 2015) で、その中でも特に注目されているのが認知診断評価 (Cognitive Diagnostic Assessment [CDA]) である。

Kumazawa (2006) では、教科書にある語彙から出題した語彙診断テストを分析し、熊澤 (2010) ではプレースメントテストながら学修到達目標に準拠した診断テストを分析した。さらに熊澤 (2013) では Kane (2006) が考案した論証的妥当性の枠組み (argument-based approach to validity) を用いて開発したプレースメントテスト得点の解釈および使用の妥当性を評価した。本稿では、そのプレースメント中にあった項目を部分的に抜粋したプレースメント用項目と、さらに新たにより細粒な読解スキルの到達度度合いを診断するための診断テスト用項目からなるテストの得点解釈の妥当性について論証することを主な研究目標とする。

2. 研究の目的

従来は相対的な順位付けでの得点のみであったが、近年ではより診断情報をフィードバックすることによって学習改善に役立てるまでテストの役割は拡張している。そのためのテスト理論である学習志向的評価、ダイナミック評価などが提案されているが、その中でも特に注目されているのがより診断情報をテストから抽出できる CDA である。しかしながら、国内ではまず言語テストの分野で CDA についての実証研究が発表されていなく、さらに診断情報をより得るため、また CDM により適した診断テストを設計し、そのテスト得点解釈の妥当性を評価する研究が進んでいないのが現状である。本研究では、読解時における認知スキルの習熟度合いについての診断情報を引き出すためのテストを開発し、クラス分けする目的で実施したその診断テストの得点解釈の妥当性を評価する。よって、本研究の研究課題を以下の通り設定し、その妥当性を検証する。なお、この研究課題 (RQ) は Kane (2006) の枠組みでの解釈的論証に該当する。

・得点化の推論として、クラス分け目的で実施された読解診断テスト項目に対しての受験者の解答は観測 (observation) され、適切に得点化されることによって観測得点 (observed score) とする (RQ1)

・一般化の推論として、観測得点は一貫して受験者の得点を推定でき、また一貫してレ

ベル分けできることによって一般化され測定領域得点 (universe score) とする (RQ2) ・外挿の推論として, 測定領域得点は受験者が振り分けられたレベルでの履修者としての授業パフォーマンスを十分に説明できると外挿することで目標得点 (target score) とする (RQ3)

3. 研究の方法

本研究の対象者は, 私立大学社会系学部の 2017 年度新入生 334 名である。英語力に関しては, ヨーロッパ言語共通参照枠では, 大半が基礎段階の英語使用者である A1 または A2 に該当するであろう。クラス分けに用いられたテストは二つのテストから構成されている。テスト 1 は文法, 語彙, 読解セクションからなるプレイスメントテストである ($k = 20, 20, 5$)。テスト 2 である英語読解総合的診断テストは 4 択 30 問が出題され, 三種類の項目形式, ならびに 7 種の問題文を用いた。職員の協力により, 教室確保からテスト問題およびマークシートの準備を含む事前準備がされた。テスト実施日は 2017 年度オリエンテーション期間中で, 事前説明を含めて 90 分が当てられた。

4. 研究成果

表 1 はテスト 1, テスト 2, 全体のそれぞれ記述統計と信頼度指数を示す。分析対象であるテスト 2 ののみについて言及する。信頼性 係数は中度 ($\alpha = .63$) であり, クラス分けの決定に用いるのには若干低い値となった。標準測定誤差 (standard error of measurement [SEM]) は 2.43 となり, 64% の確率で同一受験者が画一的条件で同一テストを受験した場合, ± 2.43 点の範囲に収まるということになる。例えば, 受験者が 14 点であれば, 確率的におおよそ 12 点から 14 点の範囲に収まるということになる。ファイ・ラムダ信頼度指数は分割点を 61% に設定した場合, .62 となった。つまり, 62% の一貫性をもって上級群と中級・基礎群を区別していると解釈できる。同様に, 32% に分割点が設定された場合も同程度の指数となった。なお, この指数の特徴として, 分割点と平均点の距離によって指数が変動する。よって, 同一データをもとに分割点を 100% に設定した場合, 指数は .97 となる。信頼度区間 (confidence interval [CI]) は標準測定誤差と同様に解釈でき, 真の得点 (true score) は 9% の範囲に確率的に収まるということになる。

表 1

記述統計と信頼度指数

テスト	<i>M</i>	<i>SD</i>	<i>k</i>	Min	Max	Skewness	Kurtosis	SEM	(.61)	CI	(.32)	CI	
テスト 1	22.81	7.74	45	6	43	0.42	-0.58	.86	1.55	.53	0.07	.86	0.07
テスト 2	13.95	4.00	30	4	25	0.15	-0.15	.63	2.43	.62	0.09	.62	0.09
合計	36.75	10.79	75	17	67	0.47	-0.35	.87	1.44	.78	0.06	.89	0.06

古典的テスト理論では, プレイスメントテストのような相対評価用テスト項目であれば, 正答率は 30% から 70% の範囲に収まるのが目安として適度 (Brown, 2005) で, この範囲外の項目は 11 項目 (項目 1, 2, 3, 4, 7, 10, 14, 19, 21, 26, 27) であった。一方, 診断テストのような絶対評価用テスト項目であれば, 事前テストとして実施したのであれば, 未習熟であるということを示す正答率が低いほうがよい (Brown, 2005)。

正答率が.70以上で高すぎる項目は5項目(項目1, 2, 3, 4, 19)であった。項目弁別力(r_{pbi})は.20以上が適正(Brown, 2005)だが, 7項目(項目2, 7, 10, 12, 22, 24, 27)が基準値以下である。

RUMモデル(reparameterized unified model [Roussos, DiBello, Stout, Hartz, Henson, & Templin, 2007])では, a パラメータは項目困難度推定値となるが, Q 行列中にある項目に指定された認知スキルを習熟した場合の項目困難度推定値となる。つまり, 項目1であれば, スキャニングの認知スキルを習熟していれば, 99%の確率で正解できるということになる。目安としては.60以下を下回ると指定された認知スキルを習熟しても正解できない難易度が高すぎる項目となる(Jang, 2009)。この基準を満たしていない項目は13項目(項目7, 10, 12, 13, 17, 18, 21, 22, 23, 24, 26, 27, 30)ある。 r パラメータは指定された認知スキルの習熟がいかにかに習熟者・未習熟者を識別するかで, IRTでは項目識別力となる。目安としては, .90以上であると指定された認知スキルの習熟は習熟者・未習熟者を識別していないこととなり(Jang, 2009), その基準を満たしていないのは4項目(項目1, 2, 7, 10)ある。この4項目は古典的テスト理論での項目分析結果でも機能していないとの結果であった。 c パラメータが高くなると, Q 行列中に指定されていない認知スキル以外のスキルが正答に関係しているということになる(Jang, 2009)。 c パラメータが高い傾向にある項目は, a パラメータが低いまたは r パラメータが高くなる。例えば, 項目10の a , r^*1 , r^*4 のパラメータはそれぞれ.31, .74, .91となった。この項目は文中にあるpurchase(JACET8000で1979レベル)という語彙の受容的意味を知りつつ, さらにその文を読解できないと正解できない項目となっている。また, その文中の代名詞がなにを指しているのか, 理解できないと正答確率が上がらないような項目で, まずこのような認知スキルは他の項目では測定していないため c パラメータが上昇した, さらに複数のスキルを要するため難易度が高くなったと推測する。観測値およびMCMCにて推定された正答率で.50以上の乖離がみられた項目はなく, 適合度合いはよい結果となった。

参考文献

- Alderson, C. (2005). Diagnosing foreign language proficiency: The interface between learning and assessment. New York, NY: Continuum.
- Alderson, C. Haapakangas, E., Huhta, A., Nieminen, L. & Ullakanoja, R. (2015). The diagnosis of reading in a second or foreign language. New York: Routledge.
- Antón, M. (2009). Dynamic assessment. In G. Fulcher & F. Davidson (Eds.), The Routledge handbook of language testing (pp. 120-134). Abingdon, England: Routledge.
- Brown, J. D. (2005). Testing in language programs: A comprehensive guide to English language assessment. New York: McGraw-Hill College Press.
- Jang, E. E. (2009). Cognitive diagnostic assessment of L2 reading comprehension ability: Validity arguments for Fusion Model application to LanguEdge assessment. *Language Testing*, 26(1), 031-73.
<https://doi.org/10.1177/0265532208097336>
- Jones, N., & Saville, N. (2016). Learning oriented assessment: A systemic approach (Studies in Language Testing Vol. 45). Cambridge: UCLES & Cambridge

University Press.

Kane, M. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed.), (pp. 17-64). Washington, DC: National Council on Measurement in Education and the American Council on Education.

Purpura, J. (2004). *Assessing grammar*. Cambridge: Cambridge University Press.

Roussos, L. A., DiBello, L. V., Stout, W., Hartz, S. M., Henson, R. A., & Templin, J. (2007). The fusion model skills diagnosis system. In J. Leighton, & M. Gierl (Eds.). *Cognitive Diagnostic Assessment for Education: Theory and Applications* (pp. 275-318). Cambridge: Cambridge University.

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 0件/うち国際共著 0件/うちオープンアクセス 0件）

1. 著者名 熊澤孝昭	4. 巻 29
2. 論文標題 学内開発英語ブレイスメントテスト得点の信頼性について	5. 発行年 2018年
3. 雑誌名 関東学院教養論集	6. 最初と最後の頁 1-7
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計1件（うち招待講演 0件/うち国際学会 0件）

1. 発表者名 Takaaki KUMAZAWA
2. 発表標題 Using a cognitive diagnostic model to examine remedial Japanese university students' grammatical knowledge
3. 学会等名 Asian Association for Language Assessment
4. 発表年 2016年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	佐藤 敬典 (Sato Takanori) (60758506)	上智大学・言語教育研究センター・助教 (32621)	
研究分担者	N Timothy (Newfields Timothy) (40246161)	東洋大学・経済学部・教授 (32663)	