

氏名	望月 昭彦		
学位の種類	博士（言語学）		
学位記番号	博乙第2817号		
学位授与年月日	平成29年3月24日		
学位授与の要件	学位規則第4条第2項該当		
審査研究科	人文社会科学研究科		
学位論文題目	Comparison Among the Appropriate Word Scoring Cloze Test, the Exact Word Scoring Cloze Test, the C-Test and the Multiple-Choice Cloze Test (適語法クローズ、正語法クローズ、Cテスト、選択式クローズテストの比較)		
主査	筑波大学 教授	久保田 章	
副査	筑波大学 教授	磐崎 弘貞	
副査	筑波大学 教授	Ed.D. (教育学)	平井 明代
副査	大東文化大学大学院 外国語学研究科教授	Ph.D. (言語学)	静 哲人

論文の要旨

言語テスト論、あるいは言語教育評価論においては、Taylor (1953)が開発したクローズテスト及びそれに関連する事項が長年に渡って大きな研究テーマのひとつとなってきた。特に Taylor 以後に合理的クローズテスト、選択式クローズテスト、Cテストなどの様々な修正版クローズテストが開発されたこともあり、それぞれのテストの妥当性や信頼性の分析に加え、様々な観点から複数のテスト間の比較分析が実施されてきた。この方面の研究は国内外で盛んであったが、テストに用いられる文章テキスト自体や受験者の熟達度など、様々な調査条件の違いから、研究結果が異なるのが実情である。

以上の問題意識に基づき、本研究では代表的な4つのタイプのクローズテスト（正語法クローズ (CLex)、適語法クローズ (CLap)、多肢選択式クローズ (MC)、Cテスト (CT)) を取り上げ、(1) どのテストの妥当性が最も高いか、(2) 各テストにおける受験者の誤答傾向にどのような特徴があるか、という2つの大きな研究テーマを設定し、それぞれ受験者の熟達度との関係を考察しながら分析検証することが目的とされている。さらに第1のテーマに関しては、(1) 受験者の熟達度と各クローズテストの成績との関係、(2) 各クローズテストの文章テキスト中の位置の違い（冒頭、中盤、終盤）と受験者の成績との関係、第2のテーマについては、(3) MCにおける錯乱肢と受験者の熟達度との関係、(4) 他のクローズテストにおける受験者の誤りの傾向と受験者の熟達度との関係という、合計で4つの課題について考察している。

論文は、第1章（導入）、第2章（先行研究）、第3章（研究1）、第4章（研究2）、第5章（結論）で構成され、研究1が上記の第1の研究テーマ、研究2が第2のテーマにそれぞれ対応している。また、本研究では、Chapelle and Abraham(1990)を参考にして同一の文章テキストを用い、CLex / CLap (30項目)、MC (30項目)、CT (80項目)を各テストの順序が異なるように配列した3つのテスト様式、Form1（冒頭：MC－中盤：

CT-終盤：CLex/CLap)、Form2 (CT-CLex/CLap-MC)、Form3 (CLex/CLap-MC-CT) を独自に作成して調査に用いている。参加者（受験者）は総数で 770 名（最終的に 578 名）の大学 1 年生で、参加者は英語熟達度試験の他に、約 190 名ずつ 3 群に分かれて、それぞれ Form1、2、3 を受験した。

研究 1 では、まず Messick (1965, 1966) の構成概念妥当性検証枠組に基づき、テストの内容、構造、一般化可能性、外的側面の 4 つの側面について 4 種のクローズテストの比較検証が実施された。内容的側面については、文章の難易度に関する 1 指標、構造的側面については、ラッシュモデルへの適合度に関する 1 指標、一般化可能性については、期待信頼性アルファ係数などの 4 指標、外的側面については、構造方程式モデリングのパス係数など 3 指標、計 9 個の統計的指標を用いて分析が行われた。さらにそれらの分析結果に基づいて各指標の数値を順位化し、最終的に各順位を合計して 4 つのクローズテストの総合的な順位付けを行っている。その結果、CLap が最も妥当性が高く、僅差で CLex が次点、第 3 位、第 4 位はそれぞれ CT、MC であることがわかった。

次に、先行研究ではクローズテストが文章テキスト中のどの部分を用いて作成されているかについて不統一であった点を踏まえ、4 種のクローズテストの文章中の位置の違い（冒頭、中盤、終盤）がクローズテストの成績に与える影響について調査が行われた。用いられた文章テキストの難易度を Flesch-Kincaid の読み易さの指標を用いて分析したところ、言語的には冒頭が最も難しく、反対に終盤が最も易しいという結果となったが、各テストの項目困難度は、文章中の位置によって変わることはなく、いずれの場合も MC が最も易しく、CLex が最も難しいことが確認された。

研究 2 では、はじめに上記の目的(3)に対応した分析が行われた。MC では通例 4 肢選択の形式が用いられることが多いが、従来正答以外の 3 つの錯乱肢は、テスト作成者の経験や過去の受験者の誤答などを基にして言わば帰納的に作成されることが専ら行われてきた。本研究では、錯乱肢を演繹的に 3 種類の範疇（(1)文法的に可、意味的に不可、(2)文法的に不可、意味的に可、(3)文法的に可、意味的に反対の概念）に分類して作成した上で、4 選択肢について項目困難度、語彙レベルなどの観点で分析を行っている。その結果、熟達度と削除タイプの相関はほとんどないこと、熟達度上位群は、Form によって意味的に反対の概念を選択する傾向があること、下位群は、語彙知識及び文法知識が不十分であるために 3 つの錯乱肢を無作為に選ぶ傾向があることなどが明らかとなった。

さらに本研究では、MC の各項目に正解するために必要な「文脈理解の程度」という課題を設定し、Bachman(1985)で提唱された「項目の削除タイプ：(1)節の中、(2)節を超えて文の中、(3)文を超えてテキストの中、(4)テキストの外」という概念を援用して、各 Form に対する分析を行った。その結果、熟達度上位群では各 Form に共通する削除タイプはなかったが、中位群では削除タイプ(1)において意味的に反対の錯乱肢を選ぶ傾向が見られ、下位群では同様に削除タイプ(1)で文法的に誤りのある錯乱肢を選ぶ傾向が観察された。

続いて MC 以外の 3 種のクローズテストにおいて、MC の場合と同様に 3 つの誤りの範疇を中心に、「各クローズテストは受験者の熟達度によってどのような誤りの傾向を示すか」など 8 つのリサーチクエスチョンを立て、誤り分析を実施して受験者の誤答の傾向と熟達度の関係について記述している。この部分の研究については、受験者全員のデータを詳細に分析するのは事実上難しいため、各 Form 毎に熟達度の上位群、中位群、下位群からそれぞれ 20 人ずつ、計 60 人を無作為に抽出し、3 つの Form で総計 180 名の解答状況を誤りの種類や頻度などの観点から精察した。その結果、上位群、中位群、下位群共に、文脈条件が「節内」の場合に意味的な誤りが多いこと、さらに下位群では、「節を超えて文内」の場合に文法的誤りが有意に多く発生することなどがわかった。

以上の調査に基づき、第 5 章では、言語テスト・言語教育におけるクローズテストの教育的意義や示唆、実施容易性等について議論している。

審査の要旨

1 批評

本研究については、まず先行研究の方法論上の主要な課題の克服を試みている点を評価できる。第一に、受験者の熟達度の不均衡を解消するため、約 600 名の参加者を確保して調査している。当該分野でこのような大規模な研究が行われた例は過去にない。第二に、異なるクローズテストの比較検証を行う際には、(1)「受験者を同一にしてテストごとに異なる文章テキストを用いる」か、(2)「同一の文章テキストにして受験者ごとに異なるクローズテスト形式を用いる」か、のいずれかが一般的であったが、いずれの場合も方法論的な課題がないわけではなかった。一方、(3)「同じ受験者が同一の文章テキストを用いながら異なる形式のテストを受験する」場合は、文章中のどの位置にどのような形式のクローズテストが配置されるかによって、結果に影響が出る可能性があった。本研究では(3)の方法を改良し、同一の文章テキストを用いながら、3つのクローズテストの配列順序を入れ替えた3種類のテスト様式(Form)を作成・実施することで、1回のテストだけでできるだけ客観的なデータが収集できるよう工夫している。

さらに研究1では、構成概念妥当性検証枠組の複数の側面から、最終的に9個の統計的指標を用いて詳細な分析を行っている。「信号雑音比」のような新しい指標を含め、これほど多くの数量的な観点から行われた研究は他にない。とりわけ、「期待信頼性」に関する議論で、各テストが十分な信頼性レベルを得るために必要なテスト項目数の目安を提示している点は、テスト開発の観点からも重要な成果である。

MCについては選択肢ないし錯乱肢の作成方法が重要であるが、研究2では、まずMCの錯乱肢の作成原理を事前に決定し、それを基に錯乱肢を3種類に範疇化するなど創意が見られる。さらにクローズテストの削除対象となる語についてBachman(1985)の4つの削除タイプと錯乱肢のタイプ、学習者の熟達度の関係について検証を試みており、研究方法の独自性と的確性を評価できる。

一方将来的な検討課題として、少なくとも以下の点を指摘できる。まず本論文のように、同一の文章テキスト内に異なる種類のクローズテストを配置してテストの比較を行う場合、削除対象となる項目はすべてのテストで同一であることが理想であろうが、本論文では、CLex / ClapとMCが30項目であるのに対し、CTが80項目であったこともあり、文章中の各テストの区切りが曖昧となったり、テスト間で削除項目の不一致が生じている。CTの項目数は先行研究に基づいて決定されたものではあるが、項目数の違いによる効果については検証の余地がある。次に、研究2の誤り分析において、著者は事前に基準を決めて誤りを3つに範疇化している。これ自体は新しい試みではあるが、実際のところMCの錯乱肢や他のクローズテストの誤答のすべてがその範疇に分類可能なのか微妙な項目がある。また、MCの各範疇に属する錯乱肢の数もやや不均衡であるなど、将来的にはいっそう適切な範疇化を求める必要がある。以上に加えて、表の活字などに表記の不統一が散見される点が惜まれる。

2 最終試験

平成29年2月9日、人文社会科学研究科学学位論文審査委員会において、審査委員全員出席のもと、本論文について著者に説明を求めた後、関連事項について質疑応答を行った。なお、学力の確認は、著者が「人文社会科学研究科論文審査等実施細則」第10条(3)に該当することから免除し、審議の結果、審査委員全員一致で合格と判定された。

3 結論

上記の論文審査ならびに最終試験の結果に基づき、著者は博士(言語学)の学位を受けるに十分な資格を有するものと認める。