

9月8日(土) 研究発表4 第1室(9314)

分野: Others (測定, 質問紙)

リカート・スケールにおける選択肢表現の違いによる反応の比較

A Comparison of Scale Descriptors in the Likert Scale: Frequency vs. Trueness

水本 篤 (関西大学大学院)

竹内 理 (関西大学)

1. 研究の背景と目的

第二言語研究において、学習者の個人差を測定する道具として、質問紙調査は最も多く使用される方法の1つである (Dörnyei, 2003)。特に、学習方略の研究では Oxford (1990) の SILL (Strategy Inventory for Language Learning) に代表されるような質問紙が、開発・実施が容易で、量的にデータを分析することができるという利点から広く用いられている。質問紙調査において最もスタンダードな方法になっているのは、Likert (1932) によって開発された、リカード・スケールであり、回答者がある項目に対して自分の態度・意見を「1. まったくそう思わない」から「5. とてもそう思う」の選択肢から選ぶというものである (Dörnyei, 2003)。リカート・スケールの質問紙のみを研究の測定道具として用いた場合の問題点がいくつかの研究 (Gu, Wen, & Wu, 1995; Ikeda & Takeuchi, 2000; Takeuchi & Wakamoto, 2001) では報告されていることから、近年、Hiromori (2006) や竹内 (2003) の研究に見られるように、質問紙の回答から得られる量的データだけではなく、質的アプローチが用いられるようになってきている。しかし、多くの研究が、利用の容易さから質問紙のみを学習者の反応を測定する唯一の道具として用いているのが現状であるといえる。

Dörnyei (2003, 2005) は Oxford の SILL を例に挙げて、5件法のリカート・スケールの質問紙の項目が特定の方略使用の頻度 (frequency of behavior) を尋ねている場合には、項目得点を合計したときに、背後にある特性を反映しない結果となると批判している。また同時に、選択肢の表現が「1. まったく行わない (never or almost never)」から、「5. ほぼいつも行う (almost or almost always)」となっているような、頻度を尋ねる形式の質問紙は、因子分析などでも一貫性のない結果を生み出す原因となるので用いるべきではないと主張しており、「1. まったくあてはまらない (not at all true of me)」から、「5. とてもよくあてはまる (very true of me)」というような適合度を尋ねる形式にして、下位尺度を構成する項目を合計して尺度得点を算出したときに累積 (cumulative) になるようなものであるべきだとしている。この主張では、項目の問題と選択肢の表現の問題が同時に議論されているため、どちらの影響により質問紙の妥当性に問題が生じるのかはわからないが、学習者の内的な態度や意見を測定するための道具として質問紙を使用するためには、それぞれの問題に対して、信頼性、妥当性の検証が不可欠になってくるといえる。そこで、本研究ではこれらの問題のうち、選択肢の表現の問題を取り上げ、同じ質問項目に対して、選択肢の表現を頻度と適合度に変えて2度実施することによって、回答者の反応の違いが出るのかを実証的に示すことを目的とした。

2. 研究の方法

2.1 参加者

関西の5つの私立大学で、2006年12月から2007年1月にかけて質問紙調査を実施した。

頻度と適合度の質問紙を同じ参加者に2度実施したため、欠席者、または欠損値があるものを除外したところ、最終的に参加者数は408名となった。学習者のレベルは多様になるように配慮し(TOEICスコアで最小90点, 最大830点), 偏った習熟度の影響が質問紙の回答に反映されないようにした(TOEICスコア, $n=281, M=397.91, SD=126.0$)。

2.2 手順

質問紙は Hironori (2004) で開発された学習方略尺度の認知方略(10項目)とメタ認知方略(10項目)に関するものを使用し, 質問項目と選択肢の表現を今回の研究目的に合うように書き換えた。質問紙調査の実施にあたって, 順番の効果を相殺するため, 半数の学習者($n=189$)は頻度の選択肢の質問紙を, 残りの半数($n=219$)は適合度の選択肢の質問紙を先に回答させカウンターバランスを取った。実施の間隔は早い場合には2週間, ほとんどが1ヶ月の間隔を空けた。間には冬季の休暇(2週間程度)が入っており, 英語習熟度(もしくは方略の使用)がその間に変化するとは考えにくいいため, 実施にはふさわしい間隔であると考えられた。また, 1回目と2回目の間にはフィードバックを与えることはないように注意が払われた。以下は実際に使用された項目と選択肢の表現の具体的な例である。

【例】新しく習った英単語に似た意味の日本語の単語を考える

(頻度)	(適合度)
1. まったくしていない	1. まったくあてはまらない
2. あまりしていない	2. あまりあてはまらない
3. ときどきしている	3. いくらかあてはまる
4. よくしている	4. よくあてはまる
5. ほぼいつもしている	5. とてもよくあてはまる

2.3 分析

実施した質問紙の回答を認知方略とメタ認知方略のそれぞれに対して, 主成分分析とラッシュ・モデルを適応し, 好ましくない項目を削除し, 尺度の一元性を確認した。その結果, 認知方略は6項目(頻度 $\alpha=.70$, 適合度 $\alpha=.69$), メタ認知方略は8項目(頻度 $\alpha=.82$, 適合度 $\alpha=.78$)を用い, それぞれの項目合計得点を項目数で割った尺度得点を使って分析が進められた。

まず, 相関分析と t 検定でそれぞれの選択肢の表現で算出される尺度によって結果に違いがあるのかを比較した。また, 検証的因子分析を行い, どちらの選択肢の表現の方がモデルにあてはまっているかを適合度指標で比較した。

質問紙は本来, 順序尺度のデータであるが, 5件法以上であれば間隔尺度として扱い, 因子分析などを行うことは可能であるとされる(鎌原, 1998; 前田, 2004)。そこで, 頻度と適合度のどちらが間隔尺度に変換したときに等間隔性を持っているかをラッシュ・モデルの Rating Scale Model を使って検証した。

3. 結果

全体として, 認知方略(頻度と適合度の相関 $r=.58$)とメタ認知方略($r=.66$)の両方で適合度の方の得点が高く(t 検定で有意差あり), 検証的因子分析では適合度の方がモデルに当てはまっているという結果となった。その他の詳しい結果は発表時に報告する。