

Some Problems in Absolute Evaluation

金沢大学大学院
教育学研究科 英語教育専攻
水上晃一

はじめに

2002年度に(絶対評価を加味した)相対評価が絶対評価に改訂されてから4年が過ぎようとしている。その間、各学校や各教師においては、絶対評価に基づく評価方法の質を高めるための様々な工夫・改善がなされてきたが、多くの教師が難しさを感じたり、学校で混乱が起こったりした。要因としては、

- ①筆記テストでは測定しにくい「関心・意欲・態度」を評価しなければならないため、主観が入りやすく、評価の客観性・信頼性を保つのが難しいこと。
- ②評定や観点別評価についての表現が「十分満足」や「おおむね満足」などのように極めてあいまいなため、その表現に対する理解の程度が教師間で異なること。
- ③筆記テストを各問いごとに見た場合、「観点」と「内容のまとめり」のどこに位置づけられるかの判断は教師によって異なること。つまり、同じテスト結果にもかかわらず評価結果が異なることが起こりうること。
- ④「評価規準」・「評価基準」の作成と3段階の観点別評価を5段階の評定に変える方法が各学校の裁量となっているので、学校間で差異が現れる可能性が高いこと。

などが挙げられる。

妥当性・信頼性のある評価方法の確立が緊急の課題である。

1章:修士論文の目的

本論文の目的は、評価やテストについて【定義・特徴(長所・短所)など】詳細に研究した後、絶対評価における問題点、特に観点や内容のまとめりに関する教師の判断の妥当性、英作文の採点の信頼性、学校間の評定の信頼性について検証することである。

2章:目標基準準拠テスト(CRT)と集団基準準拠テスト(NRT)

日本の教育界では CRT を絶対評価、NRT を相対評価と呼ぶ場合が多いが、その理由は、J. D. Brown が示すように“absolute”という用語が CRT の特徴を、“relative”という用語が NRT の特徴をうまく説明しているからである。CRT の目的は各学習者がどの程度学習内容を習得したかを知ることであり、NRT の目的はある学習者の能力を他の学習者の能力と比較することによってその差異を知ることである。すなわち、CRT と NRT の違いを理解する鍵は“percentage”と“percentile”という用語にある。CRT は受験者が習得した学習事項の量を“percentage”の形で捉え、NRT は特定の

受験者の成績を他の受験者の成績との関係で“percentile”という形で捉えるのである。両者の詳細な比較^{*1}により、絶対評価においては CRTの方が NRTより適しているということが判明した。

3章:絶対評価におけるテスト

3.1. 妥当性と信頼性

妥当性とはテストが測定しようとしている能力を実際に測定しているかどうかを示すテストの特性である。次のような場合がある。リスニングテストの答えでスペリングミスを見つけると、教師はそれに対して減点しがちである。しかしこの場合英文を聞いて内容を理解する能力だけではなくスペリングの能力もチェックすることになる。1度に2つ以上の能力を測定するのは妥当ではない。

信頼性とは測定結果の安定性である。つまり、同じ条件でテストが実施された場合、常に同じ結果を出さなければならないのである。そのために、教師は、テストのアイテム数を増やしたり、客観的に採点可能な問題の作成や明確な採点規準(基準)の設定など、可能な限りテストの信頼性を高める工夫・改善を行わなければならない。

3.2. 筆記テスト

評価方法の1つが筆記テストである。一般に、その特徴を踏まえて適切に作成・実施すれば、信頼性の高い測定が可能になるといわれている。絶対評価において筆記テストを使用する際には次のような点に留意しなければならない。

- ①4観点の何を評価するのかを決め、観点に合った適切な問題を作成しなければならないこと。
- ②合計点だけでなく、各問いごとに得点を記録しなければならないこと。
- ③どのようにカッティングポイントを設定するか。
- ④筆記テストが唯一絶対の方法ではないことを十分に理解し、いろいろな方法との組み合わせを工夫しなければならないこと。

筆記テストには様々な種類があるので、それぞれの長所・短所を十分に理解した上で、教師は目的に合った最も適切なテストを作成し、実施しなければならない。

4章:仮説と研究

絶対評価における問題点に関して、次の3つの仮説を立て、順に検証した。

*1 : J.D.Brown (1996) Testing in Language Programs

仮説1: 筆記テストの各問いが、どの観点とどの内容のまとまりを測定しているかに関する判断は、教師間で大きく異なるだろう。

◇方法: 「アンケート用紙(回答用紙)」, 「4観点と内容のまとまりごとのマトリックス」(**Appendix 1**), 「テスト問題の見本^{*2}」(**Appendix 2**)を配付し、テスト問題の各問いが「観点」と「内容のまとまり」のマトリックスの①～⑬のどこに位置づけられるかを判断してもらい、教師の判断の妥当性を調査する。

◇対象者: 奥能登教育事務所管内の20名の英語教師(初任者・講師は除く)

仮説2: 英作文の採点結果は、教師間で大きく異なるだろう。

◇方法: 同じ英作文に対して2回の採点を実施し、採点者内信頼性と採点者間信頼性を調査する。

①まず所属する中学校の3年生25名に「私の趣味」という題で5文以上の英作文を書いてもらい、その中から3名の英作文を選ぶ。

②各自の採点基準で、3人の英作文それぞれ10点満点で採点してもらう。

③次に、同じ英作文に対して、与えられた採点基準(文法点5点、内容と文の数で5点の合計10点満点)で採点してもらう。

◇対象者: 奥能登教育事務所管内の10名の英語教師(初任者・講師は除く)

仮説3: 観点別評価結果が同じにもかかわらず、学校間で評価結果に差異が現れるだろう。

◇方法: 各学校の評価の仕方について調査し、学校間の評価の信頼性を調査する。

◇対象校: 奥能登教育事務所管内の4中学校

*2: 根岸雅史(監修) 単元別プリント観点別評価 (正進社)

5章:結果と分析・・・【仮説2・仮説3は省略】

アンケート結果(回答者は全部で20人;空欄は0人を表す)

	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	⑪	⑫	⑬
Q 1											3	2	15
Q 2											5		15
Q 3					7		6				4		3
Q 4							2				6		12
Q 5					4		5				3		8
Q 6							13						7
Q 7									20				
Q 8								19		1			

英語の筆記テストでよく使われる問題には、語彙問題・語形変化問題・会話表現問題・空所補充文法問題・並べ換え問題・条件作文問題・文章読解問題・聴解問題など(**Appendix 2**参照)がある。英語科教師20人のアンケート調査の結果、会話表現問題・並べ換え問題・条件作文問題の3つで教師の判断が大きく異なり、「表現の能力」と「言語や文化についての知識・理解」の2つに判断が分かれていた。

結論

3つの仮説に基づく調査結果から次の3つのことが検証された。

第1に、絶対評価において筆記テストが用いられるとき、各問いがどの観点とどの内容のまとまりを測定しているかに関する教師の判断は、問いによって大きく異なった。教師によって判断が異なった問いに関しては、同じテスト結果にもかかわらず異なる観点の評価がなされることになり、教師の判断の妥当性に関して問題があることが判明した。

第2に、英作文の採点も教師間で異なることがわかった。特に条件を満たさない場合(この場合は5文以上という条件)、採点者間・採点者内共に採点結果に大きな開きが見られ、採点の信頼性に関して問題があることが判明した。

第3に、観点別評価結果が同じであっても学校によって評価結果が異なっていた。学校間の評価の信頼性に関して問題があることが判明した。

成果と課題

本論文の成果としては、絶対評価における「評価の妥当性・信頼性」についての問題点がデータに基づいて具体的に検証できたことである。

課題としては、「より妥当性・信頼性のある評価を実施するための適切な評価方法の確立」、「評価方法として筆記テストを用いる場合、評価しようとする観点に合った適切なテスト問題の作成」が挙げられる。

自分自身の評価能力の向上を目指して、今後も研鑽に励みたいと思う。