

論述式テストと客観式テストの関連

熊谷 賢*・阿久津洋巳**

(2005年2月7日受理)

Satoshi KUMAGAI, Hiromi AKUTSU

The Relationship between Essay Tests and Objective Tests

はじめに

筆記試験を大別すると、客観式テストと論述式テストに分けられる。多肢選択式や〇×式のような客観式テストは、学力の客観的評価はできるが、「学力の深い部分」までは把握できない(梶田、1992, pp.165-166)。他方、論述式テストは総合的な学力の評価は可能だが評定者間の一致度が低く信頼性が低い(梶田、1992, pp.170; 池田、1992)という欠陥が指摘されてきた。そこで、客観式テストと論述式テストの欠点を補い、両者の利点を利用する方法として、2つのテストを併用して総合的に学力を評価する方法がある。多くの大学で入学試験に小論文を課す理由の一つは、客観式テストでは測定が難しい思考力と表現力を評価する意図があり、2つのテスト形式(客観式テスト:センター入試、論述式テスト:小論文)を併用する一例である。このような場合、論述式テスト採点において評定者間の一致度が低いという欠陥は、複数の評定者が同一の答案を採点し、その得点を合計する(あるいは平均する)ことによって補うことができ、結果はほぼ客観的な評定値である、と仮定される。

客観式テストと論述式テストを併用する背景には、この2つのテストは学力の点で関連はあるがやや異なった側面を測定するという仮定がある。もし客観式テストと論述式テストが学力のまったく異なった側面を測定するならば、両者の得点には関連がない。反対に二つのテストが学力のまったく同じ側面を測定するならば、両者の得点には強い関連がある。しかし、客観式テストと論述式テストが学力のまったく同じ側面を測定するならば、二つのうちどちらか一つを行えば、学力は測定できるから、2つのテストを行う必要はない。

さて、大学の教科における筆記試験はどのようなものかと考えてみると、その評価方法が多様であることに気づく。論述式テストを用いる教員が最も多いかもしれない。レポートを提出させて、その評価に基づいて成績を決める教員も多い。中には客観式テストの方が信頼できるという理由で、客観式テストを採用する教員もいるであろう。宿題の課題レポートと学期末テスト(論述式あるいは客観式)を総合して成績を決める教員もいるであろう。一つの科目において客観式テストと論述式テストを比較するならば、両テストの得点の関連はどのようなものかに筆者たちの興味はある。そこで、本研究は、課題レポートの成績と論述式テストの成績の関連を検討した。

試験 1

方法

岩手大学教育学部の選択科目（1科目）の受講生を対象に調査を行った。論述式の課題は、宿題のレポートであった。レポートの課題は、授業で学習した知識に基づいて、自分の意見をまとめるものであった。筆者たちをふくめ4人が独立に30点満点で全てのレポートを採点した。1人は岩手大学教育学部の教員、他の3人は教育学部研究科の大学院生である。客観式テストの資料は、中間テスト（33問）と期末テスト（54問）の得点を用いた。宿題のレポートを提出し、中間テストと期末テストの両方を受験した学生のデータ（12人分）のみを分析対象とした。

結果

論述式課題の得点は4人分あるので、各採点者の得点を偏差値に変換し（平均50、分散1）、これらの平均を求め最終的な論述式課題の得点とした。客観式テストの得点についても、中間テストと期末テスト各々の得点を偏差値に変換し（平均50、分散1）、その平均を求め最終的な客観式テストの得点とした。中間テストと期末テスト間の得点の相関係数（Pearsonの r ）は0.50であった。論述式課題の得点と客観式テストの得点との間の相関係数 r は -0.317 （ $p>0.3$ ）と小さかった（Fig. 1）。

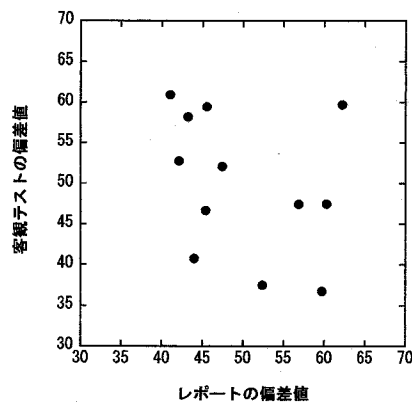


Fig. 1 試験1の結果

レポート評価の平均偏差値と客観式テストの平均偏差値の関連を散布図で示す。関連は小さく、しかも負の相関がうかがえる。

試験 2

レポートの評価1の標本数は12と小さかったため、偶然の要因が強く働いた可能性がある。試験2では、標本数を大きくした。

方法

岩手大学教育学部の選択科目（1科目）の受講生を対象に調査を行った。調査2の科目は調査1の科目と異なり、受講生も異なる。論述式の課題は、宿題のレポートであった。レポートの課題は、授業で学習した知識に基づいて、自分の意見をまとめるものであった。筆者たちをふくめ4人が独立に10点満点で全てのレポートを採点した。4人の評定者は試験1と同じである。客観式テストの資料は、中間テスト（50問）と期末テスト（72問）の得点を用いた。宿題のレポートを提出し、中間テストと

期末テストの両方を受験した学生のデータ（38人分）のみを分析対象とした。

結果

調査1と同じように、論述式課題の得点は各採点者の得点を偏差値に変換し（平均50、分散1）、これらの平均を求め最終的な論述式課題の得点とした。客観式テストの得点についても、中間テストと期末テスト各々の得点を偏差値に変換し（平均50、分散1）、その平均を求め最終的な客観式テストの得点とした。中間テストと期末テスト間の得点の相関係数（Pearsonのr）は0.768であった。論述式課題の得点と客観式テストの得点との間の相関係数rは-0.075（ $p > 0.66$ ）ときわめて小さかった（Fig. 2）。

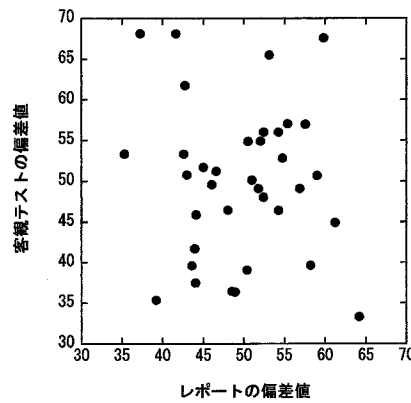


Fig. 2 試験2の結果

レポート評価の平均偏差値と客観式テストの平均偏差値の関連を散布図で示す。レポート評価と客観式テストの得点に関連がないことがわかる。

総合考察

試験1と2ともに、中間テストと期末テストの得点の相関係数rは高く、2つのテストの成績には関連があった。これに対し、客観テストとレポート得点の相関係数は低く関連が低かった。もし客観テストとレポート得点の相関が高ければ、客観テストの高得点者はレポートでも高得点であった、という結果が予想される。散布図（Fig. 1と2）を見ると、実際には（1）客観テストの得点が高くレポートの得点が低い人、（2）客観テストの得点が低くレポートの得点が高い人が何人もいた。

客観テストとレポート得点の相関が低いことから、これらが異なった学力を測定していると推測できる。客観テストは、当該科目の基礎事項を理解し記憶するという意味の学力を測定していると考えてよいであろう。レポートはどのような学力を測定しているのでしょうか。

①課題に関する事実の理解、②自分の解釈や論理的な判断、③レポートの構成、④文章表現、⑤記述の統一とフォーマットの確認、などがレポートの作成には含まれている。このうち①と②は当該科目と直接関連があり、③から⑤までは、一般的なレポート作成能力（もしくは文章能力）とみることができる。

客観テストとレポート得点の関連が低かった理由として、①課題に関する事実の理解と②自分の解釈や論理的な判断がレポートの得点に影響していた、と推測することができるであろうか。レポートの得点に③、④、⑤の要因が影響するであろうから、①と②の影響が直接、散布図と相関係数に現れ

ることではない。しかし、確率的にみれば、もし、①と②がレポートの得点に影響したならば、客観テストの得点とレポートの得点の間に低いけれども正の相関があると予想できる。(①、②の要因と③～⑤の要因間に相関がないと仮定すれば、③～⑤は平均0で分散を固定した正規分布からランダムに選んだ値を加算したとみなせる。そこで、この分散が①と②の要因の分散より数倍大きいのでなければ、①と②の要因の効果は散布図に姿を残すはずである。) そのような傾向を Fig. 1 と 2 にみることはできない。

客観テストとレポート得点の関連が低かった理由として、レポート得点は、一般的なレポート作成能力が強く影響していた、と推測することもできる。例えば、学習したことをよく理解し記憶しているが、文章能力が低いため、うまくレポートに記述できなかった ((1)のケース) かもしれない。また、暗記することは得意ではないが、多くの情報から自分に必要な情報を適切に選択し、自分の考えを適切に文章に表す能力が高い ((2)のケース) のかもしれない。しかし、ここで一般的なレポート作成能力がレポートの得点に影響する主要な要因であると仮定すると、レポートの得点を教科の得点とすることが妥当か否か問題となる。教科の成績はその科目に関して学習した事柄の習得の程度を表すはずであるからである。学習した事柄を文章で表現できることも科目の目標に含めるならば、レポートの得点を教科の得点とすることは妥当であろう。(ただし、その場合は、授業の中に「学習した事柄を文章で表現できる」能力を養成する要素が組み入れられている必要がある。)

客観テストとレポート得点の関連が低い、という本研究の結果は、成績評価にレポート得点を含めることにより総合的評価を行える可能性を示唆するが、大きな制限条件がある。それは、レポート得点に、一般的なレポート作成能力が強く影響していない、という条件である。そのような条件が満たされるか否かを知るためには、レポートの分析的評価を行うようなデータを集める必要がある。

さらに、客観テストとレポート得点の関連が低い、という本研究の結果は、もしどちらか1つの得点に基づいて、成績を決めるとしたら、どうすればよいのかという問題を投げかける。客観テストの得点による順位とレポート得点による順位が異なるからである。今後検討すべき問題の1つである。

学力を評価する際に、総合的な学力と深い学力を評価できることが好ましいといわれる。総合的学力には、基礎知識の習得、適切な問題の設定、その知識に基づいて問題を考察する能力、解答を明瞭に表現する能力、などが含まれる。論述式テストやレポートによって総合的学力を評価するためには、論述式テストの評価要因をまず明らかにする必要がある(例えば、渡部ほか、1988)。

【引用文献】

- 渡部 洋 平由美子 井上俊哉 1988. 小論文評価データの解析 東京大学教育学部紀要 第28巻
池田 央 1992. テストの科学 日本文化科学社
梶田毅一 2002. 教育評価 [第二版補訂版] 有斐閣