

口頭発表A⑨

ルーブリック評価を活用した項目特性分析による 記述式試験問題の評価

三笠洋明¹, 赤池雅史², 西村明儒¹

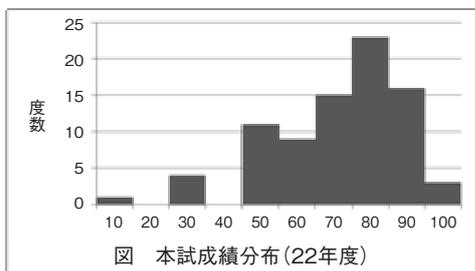
1: 徳島大学医学部教育支援センター, 2: 徳島大学大学院ヘルパバイオサイエンス研究部医療教育学

目的 近年、成績評価における試験にはMCQ(Multiple Choice Qestion)が用いられる事が増えてきたが、想起レベルの知識(タクソミー I)を問う事は出来ても深い知識には向かない。良問の作成には技術と時間を要し、問題の質が結果に影響するという欠点も知られている。一方、記述式問題は、採点の自動化は難しく試験後の問題解析は容易ではないなどの欠点も有しているにもかかわらずタクソミー II (解釈)およびIII (問題解決)レベルの知識を問う事が出来る、また、MCQに見られる偶然により正解する確率が無視できる事などの長所も有している。そこで、記述式試験の採点にルーブリック評価を適用し、採点の半自動化を行い項目特性分析を行い、記述式問題の特性の検討を行った。

対象と方法 2010年度のT大学医学部保健学科一年生(83名)の統計学の最終試験の記述式問題において必要な解答38項目各々の評価基準(ルーブリック)を予め作成し採点を行ったデータを用いた。項目特性分析は試験の総得点順に5群に分け、各問題における各群の正答率を縦軸にとりグラフ化した。また、比較に用いたMCQによる試験は、医学科A分野の最終試験のデータ(89名分)を用いた。統計解析は、SPSS Ver19.0Jを用いた。

表1 各問の正答	
問1	偶然の影響を取り除く
問2	偏りの影響を取り除く
問3	標準偏差の式
問4	偏差平方和を自由度で割ったものの平方根
問5	\bar{X}_i 平均値、 i 等使った記号が説明できている
問6	N の平方根 +1
問7	この式で計算した数を元に区切りの良い階級数を決める。
問8	対応が無い
問9	student's t-test
問10	welch's t-test
...	...

ブリック)を予め作成し採点を行ったデータを用いた。項目特性分析は試験の総得点順に5群に分け、各問題における各群の正答率を縦軸にとりグラフ化した。また、比較に用いたMCQによる試験は、医学科A分野の最終試験のデータ(89名分)を用いた。統計解析は、SPSS Ver19.0Jを用いた。



結果 記述式の試験得点分布を図1に示した。低値に裾野を引く不対称な分布であり、Shapiro-Wilk法により有意確率 0.006 で正規性は否定された。中央値 71、25%値 54、75%値 79、60点未満のものは25名であった。

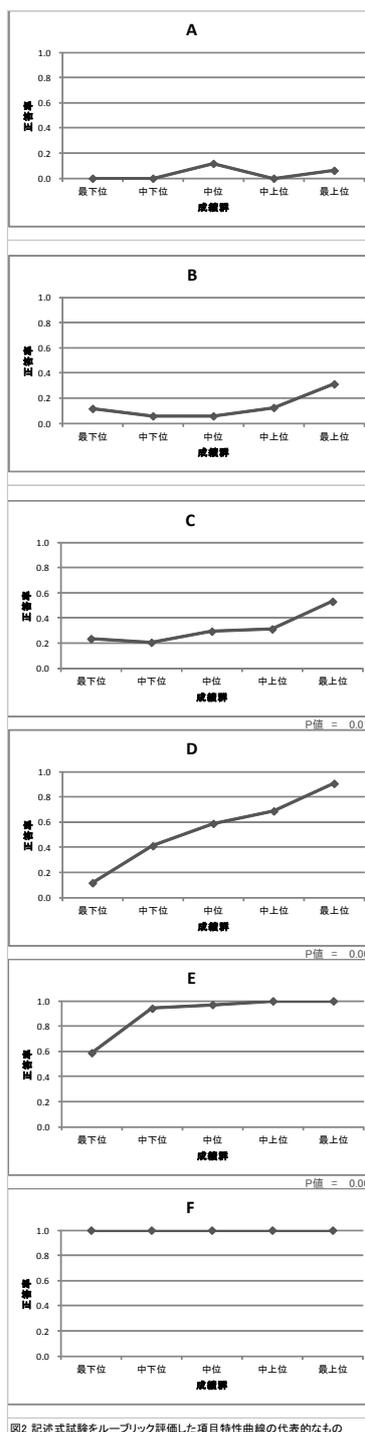
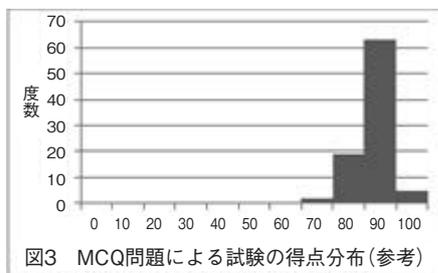


図2 記述式試験をルーブリック評価した項目特性曲線の代表的なもの

中央値 71、25%値 54、75%値 79、60点未満のものは25名であった。左に記述式試験をルーブリック評価した場合の代表的な項目特性曲線を示す。グラフの右下に Mantel-Haenszel 法による線形と線形の連関の P 値を示した(0.05 未満の場合のみ)。この検定で有意であったものは 31 であり全体(38)の 81.6%を占めた。形状としては C,D,E に類似したものが大部分であり、成績全領域に識別力を持つもの D、中上位以上に識別力を持つもの C、中下位以下に識別力を持つもの E などの違いはあれ、成績の良いものと悪いものを識別する能力を有す問題が大部分であった。最下位群の

正答率の最低値は、Aのような全成績群で正答率が零である問題を除けば、0.1と偶然に正答する確率は小さかった。

一方MCQによる試験問題(46問)では80点台が21.3%、90点台が70.8%、100点が5.6%と比較的易しい試験問題であったと考えられた。



項目特性曲線の一部を右に示す。線形と線形の連関で有意であったものは11、全体の24.4%であった。全体的に水平に近いものが多く、最上位群と最下位群の正答率の差が小さいものがほとんどであった。また全群で正答率がゼロに近い、つまり難しい問題は無かった。また最下位群の正答率の最小値は0.45と5択の場合の理論値0.2より大きかった。

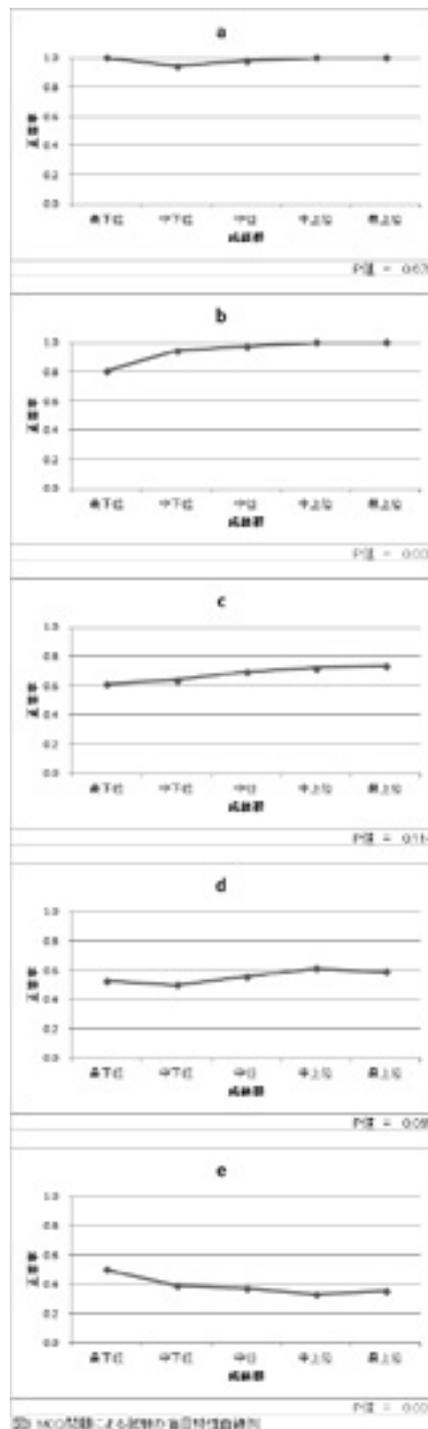
考察 記述式試験問題による試験をルーブリック評価を行い項目特性分析を行いMCQ問題で行われた試験の結果と比較したところ(後者を()で示す)、成績の上位者と下位者を識別する能力を持つ問題が全体の81.6%(24.4%)を占め、また偶然による正答と考えられる確率は0.1(0.45)と小さかった。

試験の対象者も試験内容も異なるので単純な比較は困難であるが、記述式問題は、偶然による正答が殆ど無いために識別力が大きくなる事は十分予想できる。また今回の採点では、2つの問題にあてずっぽうで同じ解答をした場合は一方が正解であってもどちらも零点としたので偶然の正答はかなり小さくなっていると考えられる。

一方で、項目特性曲線を教育の到達度を評価する指標と見ることも可能である。その視点に立てば、医学科のA分野では多くの項目で学生が高い到達度を示しており、また最下位群での正答率が45%と所謂学習困難者も少ないと考えられる。

項目特性分析は、システムティックに行えば、試験問題の難易度と、学生の学習到達度を独立して分析する事が可能であるが、今回の分析では記述

式で行った試験もMCQで行った試験何れも試験問題の難易度を独立して評価する事は出来ていない。しかしながら、この分析により、問題の難易度あるいは対照集団の学生の学習到達度を視覚的に把握する事が可能であり、問題の改善並びに、講義や実習実施へのフィードバックとして有用であると考えられる。



結論 特に記述式の試験問題をルーブリック評価法により半自動的に定量化し、項目特性分析を実施した。その結果大きな識別力を持つ問題が多く(81.6%)学生の学習到達度の評価の改善に役立つと考えられた。