

期待効用仮説による学習者の自己評価計量技法(4)

- 確信度と正答率から出題の妥当性を量る -

Quantitation of Learner's Self-Evaluation by Expected Utility Hypothesis (4)

- Validity of Questions by Conviction and Correctness of Students -

田中 規久雄^{*1}, 養老 真一^{*1}, 西本 実苗^{*1}, 下倉 雅行^{*2}
 Kikuo TANAKA^{*1}, Shin-ichi YORO^{*1}, Minae NISHIMOTO^{*1}, Masayuki SHIMOKURA^{*2}

*1 大阪大学

*2 大阪電気通信大学

*1 Osaka University

*2 Osaka Electro-Communication University

あらまし：期待効用仮説に基づき，学習者の自己評価を理解度として計量化する実験を行った．本報告では，学生の回答の確信度とその正答率による設問の妥当性について検討する．

キーワード：自己評価，教育評価，期待効用仮説，累積プロスペクト理論，出題の妥当性

1. はじめに

我々は，学習者の理解度を自己評価させる際に，主観をできるだけ正確に表示させる評価システムについて考案し⁽¹⁾，さらに主観確率 q と記入確信度 p の関係について，累積プロスペクト理論を用いて考察した⁽²⁾．

我々の実験は，○×の2択問題において，回答に対する確信度 p を表示させることによって得点を算出し，回答者の内心の主観的確信度 q を記入確信度 p として記載するときに回答者の得点期待値が最大となるように， $T(p,q) = -(p-q)^2 + (q^2 - q + 1)$ となるような得点関数を与えたものである．(表1. ただし期待値は学生には示していない．)

表1 得点関数値

記入確信度 p	正解の場合の得点	不正解の場合の得点	期待値
100%	100点	0点	100点
90%	99点	19点	91点
80%	96点	36点	84点
70%	91点	51点	79点
60%	84点	64点	76点
50%	75点	75点	75点

実験は2011年7月下旬に，2大学の大学生91名を対象に，○×問題全20問にそれぞれについて，○×の解答，および確信度(5%刻みの50%~100%)を記入してもらった形式で実施された．(ただし，記入確信度50%の場合は原則として解答は△とした．)

なお，以下「記入確信度」を「確信度」とする．

2. 確信度と正答率

この実験では20問の○×問題を提示したが，それぞれの設問について，学生が記入した確信度 p とその正答率でクラスター分析したところ，4グループに分類できた．(図1. 変数は正答率と確信度，距離はユークリッド距離，手法はWard法．)

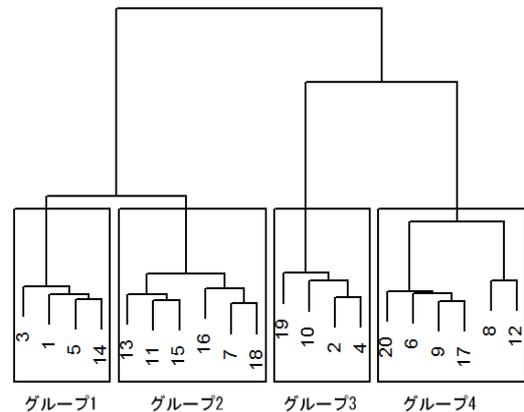


図1 設問のデンドログラム

正答率と確信度の散布図上でこのグループ分けを示すと，図2の様になる．(直線はグループ1，4とグループ2，3それぞれの回帰直線，●はそれぞれのグループの重心位置．)

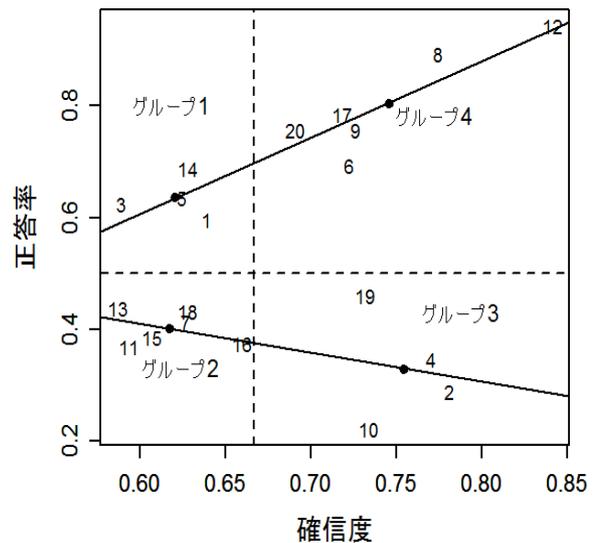


図2 設問毎の確信度と正答率の散布図

概ね、正答率が1/2を越えるかどうかで、良問 - 悪問が、確信度が2/3を越えるかどうかで、易問 - 難問を分けた。

良問とは、講義に対する学生(全体、以下同じ)の「正しい理解(以下理解とする)」が「誤った理解(以下、誤解とする)」を上回っている問題群であり、それゆえ確信度が上がれば正答率も上がる問題群であり、悪問とは学生の理解が誤解を下回る問題群であり、それゆえ確信度が上がるほど正答率が下がる問題群ということができる。なお○×問題の場合、良問悪問の線引きが正答率1/2となるのは素直であると考えるが、多数の問題の正答率が1/2の上下どちらかに偏る場合等については検証の必要がある。また、良問群ではほぼ「確信度平均=正答率」であるが、これが法則性を持つのかも今後検証したい。

次に難問とは、学生が主観として「難しい」と思った問題(主観的難問)である。概ね確信度が2/3を越える問題は学生に易問と判断されたものと思われる。

3. 典型例の経験的評価

結論的にいえば、良問 - 悪問は「直感的判断の理解度」、難問 - 易問は「専門用語の理解度」のトレードオフで4群に分かれているといえる(表2)。

表2 設問分類

	直感的判断を理解	直感的判断を誤解
専門用語を理解	良易問	悪易問
専門用語を誤解	良難問	悪難問

以下、上記の各グループの重心に近い問題をそれぞれのグループの典型問題として評価する。

(1)良易問(グループ4): 典型は問17「違法コピーソフトを販売しても違法ではない場合がある。」

違法コピーソフトを販売することは悪であるという直感的判断が容易で正しくでき、「違法」という専門用語の判断も容易で正しくできる傾向にある。

(2)良難問(グループ1): 典型は問5「名誉毀損の免責要件である真実誤信の相当理由は一般人には軽減される。」

一般人であっても名誉毀損は悪であるという直感的判断は容易で正しくできるが、「真実誤信の相当理由」という専門用語の理解が難しいが一応は正しくできる傾向にある。

(3)悪易問(グループ3): 典型は問4「誰かを特定して「バカ」とネットの掲示板に書き込むことは、名誉毀損罪となる。」

直感的判断が容易な割に誤りがちであり、刑法上の「犯罪」となるかの専門用語を安易に正しいとする傾向にある。

(4)悪難問(グループ2): 典型は問7「個人情報保護法は、個人の自己情報コントロール権を明確に規定している。」

直感的に判断が難しい上に誤りがちの上、「自己情

報コントロール権」という専門用語の誤解が大きい傾向にある。

4. 考察

上記の様に、本研究においては、良問とは講義に照らして学生が正しく直感的判断ができるものであって、試験問題としては学生の主観的難易感に関わらず、良問が望ましいことが示唆される。

直感において誤解を与える傾向にある問題は悪問といえようが、競争試験等で差をつけるためにあえてその様な所謂「引っかけ問題」を出題することを否定する訳ではない。しかし通常授業での教育評価でその様な問題を出す必要はないであろう。

とはいえ、試験問題として良問だけを作成するのは至難の業である。本実験においても、良問と評価されるのが10問、悪問と評価されるのが10問と半分したので、出題教員も自らの教育効果に対しては結局「わかっていなかった」ともいえる。

幸いなことに、本実験における良問、悪問、全問のそれぞれの期待値得点による学生の分布を見る限りにおいては(図3)、悪問に対する正答率が低いため、悪問の影響は比較的少なくなっており、良問評価は全問評価に10点程所謂「下駄」を履かせた格好になっている。しかし、悪問の量と正答率の高さがどの程度までならそれ程悪い教育評価ではないのか等、検証すべきことは多い。

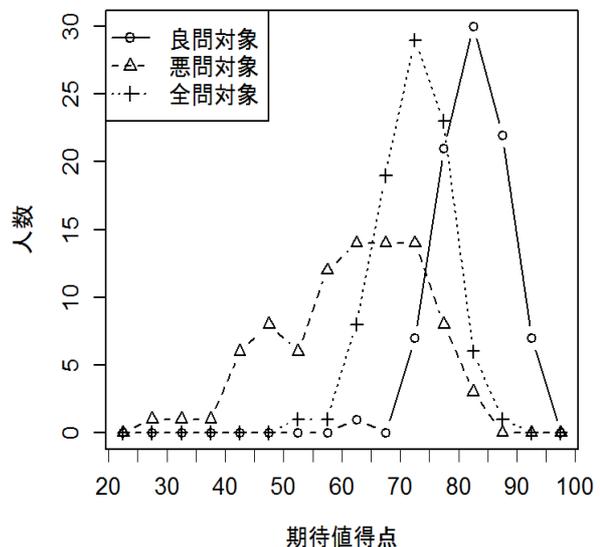


図3 良問 - 悪問 - 全体の評価分布

参考文献

- (1) 田中規久雄, 養老真一, 下倉雅行, 西本実苗: 期待効用仮説による学習者の自己評価計量技法(1) - 2択問題を例として -, 教育システム情報学会第36回全国大会講演論文集, pp.388-389 (2011).
- (2) 養老真一, 田中規久雄, 下倉雅行, 西本実苗: 期待効用仮説による学習者の自己評価計量技法(2) - 累積プロスペクト理論による一分析 -, 情報処理学会第74回全国大会講演論文集, vol.4, pp.421-422 (2012).