

[参考資料] Web テストにおける測定条件が信頼性に及ぼす影響

この研究では、各項目で測定された得点の信頼性を調べるために I-R 相関の代わりに「ある頁の得点と残りの頁の得点との相関係数の平均値」を用いた。以下は測定条件が信頼性に及ぼす影響について調べたものである。

筆者が公開している中学校数学の Web 教材について 2010 年 2 月 1 日から 2010 年 5 月 30 日までの期間に回答・返信のあった答案のうち、学年が中 1～中 3 および卒業生と答えたものうち同一人物の答案を除く 8,322 人の答案について集計した。(学年欄が空欄の答案は解答欄も空欄であるものが多く、精度が下がるため集計に含めなかった。)

各答案にはその頁(以下項目ともいう)の小項目の採点結果とともにそれ以前に行った頁の得点も自動的に記録されるものとした。実際には同一頁や同一小項目を何度も繰り返して行う者が多いが、同一頁についても同一小項目についても第 1 回目の答案だけが記録されるようにした。

一般に行われている e ラーニングとは異なり、単なる Web ページに採点・返信機能を追加しただけのものであるので、厳密に個人を識別することは難しいが、IP アドレスの上位 16 ビットと回答パターンが一致するものは同一人物とみなした。学校のパソコン教室のように 1 教室分のパソコンが対外的には同一の IP アドレスとなって現れる場合には、回答パターンが同じ答案は集計されない。

2 つの項目の両方に回答している答案が 10 人未満のときはデータなしとした。

また、無相関検定によって有意と認められる値のみを使用した。

1 小問数と信頼性の関係

一般に、項目数が増えれば測定の信頼性が高くなると考えられている。しかし、回答者の成績や資格取得に全く影響しない任意参加の Web テストにおいては回答者の持久力に限界があり、小問数を増やすと後半の小問で無答案が増える傾向がある。

図 1 は多くの頁に見られる正誤情報の例を示したもので、上端の区分が無答案を表す。第 10 問あたりになると無答案が多く、能力の測定がうまく行われていない可能性がある。

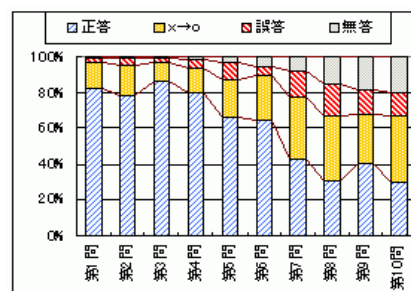


図 1

そこで測定結果の信頼性が最大となる小問数は何問かという問題意識から小問数と信頼性の関係を調べたものである。

図 2 は横軸を各頁の小問数、縦軸を平均相関係数としたときの散布図である。これによれば、10 問程度までなら小問数が増えれば信頼性がわずかによくなるが、その回帰予測の決定係数は 0.01 と小さい。

なお、白紙答案が送信されるのを防ぐため少なくとも 3 問の小問を行わなければ送信できない設定にしており、小問数の最小値は 3 になっている。全問回答しなければ送信できない設定にすることもできるが、そのような設定ではほとんど答案が回収できないと考えられる。また、このグラフには掲載していないが小問数 31 題という場合には、無答案が非常に多くなり、他の頁と負の強い相関が見られる。

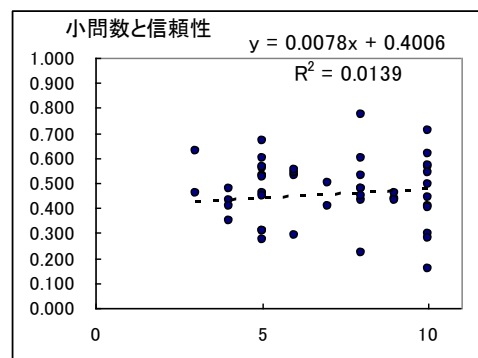


図 2

2 回答者数と信頼性の関係

実際の測定結果においては、まれに負の相関が表れることがあり、そのまま受け止めると「ある項目でよくできる者ほど他の項目ではできない傾向がある」ということになるが、次の図 3 のように偏った標本が切り出されたときには全体の傾向と異なる傾向となることが知られている。

回答者数が多くなると下位群から上位群まで幅広い回答が得られるため測定の信頼性が高くなると考

えられるため、回答者数の多寡と信頼性の関係について調べた。

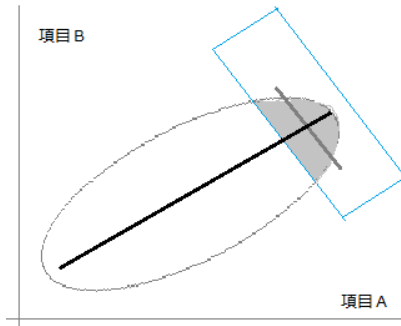


図 3

次の図 4 は横軸を両方の頁に回答した人数、縦軸を平均相関係数としたときの散布図である。これによれば、回答者数の多寡は平均相関係数にほとんど影響していないが、回答者数が増えると平均相関係数が 0.4 付近に収束してくる傾向が見られる。

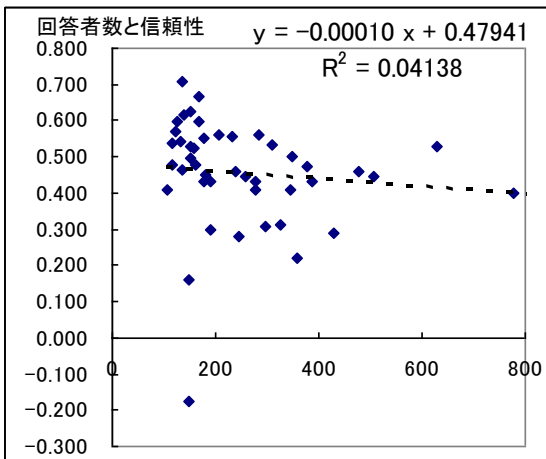


図 4

3 正答率と信頼性の関係

現在、学校では目標準拠測定が行われているが、集団準拠測定のテストにおいては正答率が中程度のときに個人差が大きくなると考えられている。

例えば各項目の正答率が 90%を超えるようなテストにおいては各自の得点が上端に張り付いているため他の項目の得点に連動してその項目の得点に変化することは期待できない。したがって、正答率があまり高いとその項目による測定の信頼性が低くなるというのが理論的な立場だと考えられるが、実データではどうなるのかを調べたものである。

次の図 5 は横軸を各頁の正答率、縦軸を平均相関

係数としたときの散布図で、正答率 50%から 90%の間では正答率と信頼性はほとんど関係がないと考えられる。

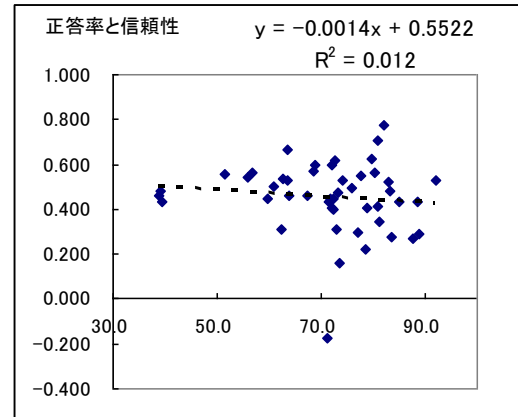


図 5

入学試験、資格試験、学年末試験のように達成度の測定を目的とするテストに比べて解きながら学ぶことが目的である練習問題においては、正答率 50%では難しすぎる。

また、資格や成績と結び付いてはほぼ全員が回答する動機が存在する場合に比べて自由参加の Web テストにおいては脱落・途中放棄が起こりやすい。

以上 2 つの傾向を考慮すると、正答率は高く設定したくなるが、上記の調査から正答率が 80% 台でも測定という側面から見たときの信頼性も失われなると考えられる。

4 入力方式と信頼性の関係

選択問題は偶然でも正解となることがあるため能力の測定には適さないという考え方がある。一方、選択問題は多人数の答案を客観的に採点するのに適しているためマークシート方式などの形で広く利用されている。

選択問題において選択肢の個数が少ないと測定が不正確になると予想されるが、誤答の選択肢を数多く作ることは容易でない。

他方で、空欄書き込み問題にすると教材作者が予想しなかった意外な誤答も回収できるよさがあるが、学習者から見れば入力作業の煩わしさのため思考がしばしば中断され教科の内容に集中できない弱点も考えられる。

以上のような問題意識に立って、入力方式が測定の信頼性に及ぼす影響を調べた。

以下において入力方式の分類に用いた用語の意味

及びその他の参考事項は次の通り。

- ・ 空欄書き込み方式は単純な文字列照合によって採点しているが、別解もしくは許容範囲が学校での教え方によって変わる可能性があり、取り扱いが難しい。またプログラム上も全角文字・半角文字、英字大文字・小文字の区別、余分なスペースの有無の取り扱いなど工夫すべきことが多い。
- ・ 択一とは、小項目ごとに数個の選択肢を並べてそのうちの1つを選ぶものである。ラジオボタン方式がよく用いられるが当調査では文字をクリックする方式を多く用いている。選択問題のうちで選択肢が2個であるものは二択として別に集計した。
- ・ 混合とは、選択と空欄書き込みを併用したものである。
- ・ チェックボックス方式は該当するものを複数個選択するもので、正しいものがすべてチェックされているかつ正しくないものはすべてチェックされていないことが正答の要件となるので、すべての選択肢について正確な判断を要する方式である。
- ・ クリカブルマップはx y平面上で該当する点をポイントするような場合に用い、あらかじめ定めた許容範囲内をクリックされたときに正答とするものである。
- ・ 多対多とは選択問題の弱点を克服するために筆者がよく用いているもので、カード合わせの要領で問題と答えを一致させる形式をいう。選択問題では「誤答を必要数作ることが難しい」「選択肢の相互関係から正答が分かりやすい」「何度も行くと誤答の方を覚えてしまう」といった弱点が考えられるが、他の問題の正答と幾つかのジョーカーを混ぜて選択肢の個数を増やしておく、どの問題に対しても数多くの選択肢があることになり慎重な選択が期待できると考えたものである。ある小項目で正答となった後もその選択肢が消えずに残る形式、異なる問題には異なる選択肢が対応する場合において、正答が選択されれば問題と選択肢が消えるようにした形式、ゲーム的に遊んでしまうことを防ぐために、1つでも間違えば振り出しに戻る形式がある。

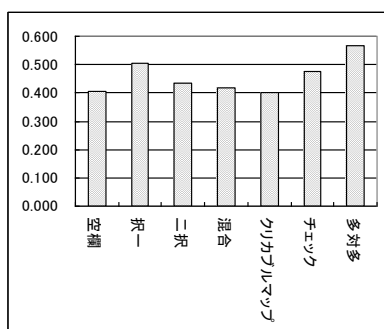


図6

図6は入力方式ごとの平均相関係数のグラフである。図6によれば、空欄書き込みの問題は意外に信頼性が低い。他方で、多対多選択方式は概して信頼性が高い。

5 1題当たり所要時間と信頼性の関係

1題当たり所要時間とはあらかじめ定めたものではなく、多数の回答者の所要時間を平均した結果を表す。

次の図7は横軸を1題当たり所要時間、縦軸を平均相関係数としたときの散布図である。これによれば、1題当たり所要時間が短い(数秒など)ものでは玉石混交となるが、1題当たり所要時間が30秒～50秒程度の問題で安定的に良い結果が得られている。

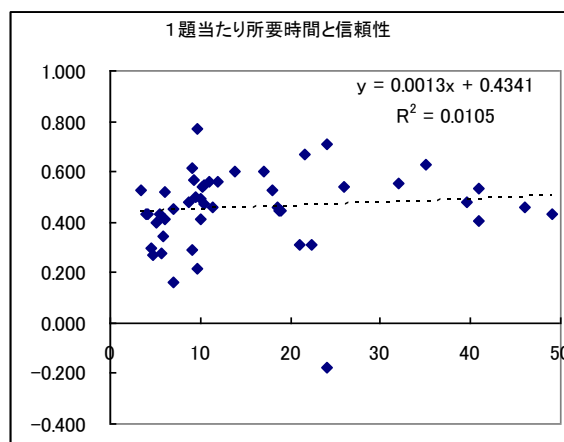


図7

6 一覧表示・順次表示と信頼性の関係

ここでは、問題の一覧表示とはすべての問題が一覧できるようになっている形式を示す。これに対して、順次表示とは、1題答えれば次の問題が出てくる形式になっていて前の問題に回答しないと後の問題が見えない形式のこととする。

次の図8のようになり、順次表示の信頼性が少し高くなる。

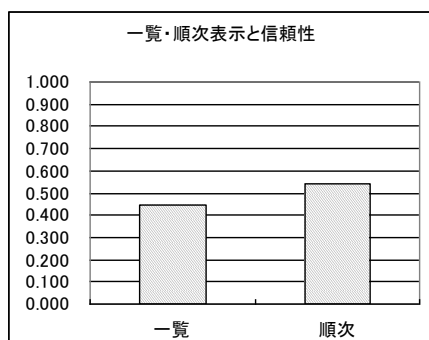


図8

7 解説文の有無と信頼性の関係

各頁に解説文がある場合とない場合の答案の信頼性を比較する。

中学校の数学では教科書レベルの問題でも解説文が役に立たないことがある。たとえば中学校1年生の文字の利用の単元において「掛け算の記号は省略できるが、足し算の記号は省略できない」という解説文があるときでも、成績下位群の生徒においては「一度に一つの原理しか身に付かない」者が多く、「足し算の記号も省略できる」と考える者が多い。このように反復練習で身につけるべき実技のようなものは、まとめて文章として示すだけでは効果が期待できない。このような問題意識から解説文の有無（解説文があるときはその行数）と測定の信頼性について調べた。

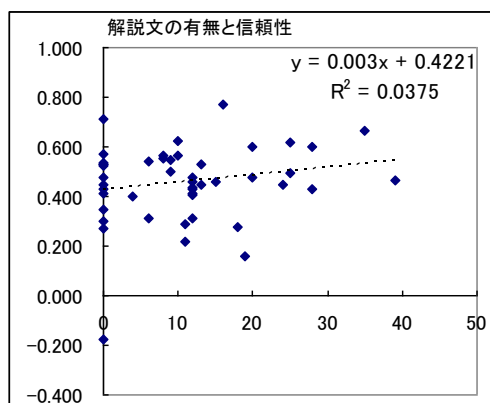


図9

図9は横軸を解説文の行数（0は解説文がないことを示す）、縦軸を平均相関係数としたときの散布図である。これによれば、解説文の有無・その行数と答案の信頼性はあまり関係がないが、解説文が長くなると信頼性が安定してくる傾向が見られる。

8 ヒントの有無と信頼性の関係

ここではヒントという用語によって個々の問題を解く上で参考となる助け舟もしくは解答そのものを表し、一般論を述べた講義に相当する解説と区別した。

回答後にのみヒントを見ることができるようにした問題についてはヒントなしとして集計する。解く前からヒントを見ることができるようになっているものは、それ以前の集計において非常に無答答案が多く、そのままでは無理だと判断した問題に限られる。

次の図10は問題を解く前にヒントを見ることができる問題とそうでない問題との信頼性を比較したものである。これによれば、両者の差異はほとんどなく、ヒントの有無は信頼性に影響していないと考えられる。（子どもなりにプライドがあって解く前からヒントを見る者は少ないのかもしれない。）

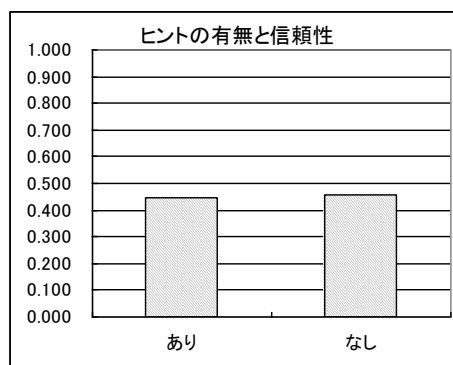


図10

9 各要因の複合的な効果と信頼性の関係

以上の各要因を組み合わせる重回帰分析した。カテゴリーデータについてはダミー変数に変換した。

* $p < .05$ となるものは、一覧表示と順次表示の別のみであった。

(注*) 中学校2年・図形の証明問題で小問数が31問、平均正答率48%というデータはほとんどが無答答案であるため他の項目と強い負の相関があり、この項目は以上のすべての集計に含めなかった。