

多枝選択式問題作成 ガイドラインの実証的検討

坪田彩乃 石井秀宗

名古屋大学大学院教育発達科学研究科

今日の概要

- ▶ 多枝選択式テストとガイドライン
 - ▶ 本研究の目的
- ▶ テストの作成・実施・結果
- ▶ 個別ガイドラインについて
 - ▶ 作題時に注意したいガイドライン8つ
- ▶ まとめと本研究の限界

多枝選択式テストとは

- ▶ 多くのテスト場面で用いられており、あらかじめ複数用意された選択枝から適切なものを選ぶという形式 (日本テスト学会, 2007)
 - ▶ 大規模試験, 個別学力試験等
 - ▶ 国立大学の一般入学個別学力試験総項目数の12.5%は客観的項目のうち75%は多枝選択式形式 (宮本・倉元, 2017)
- ▶ 一般的に用いられる機会の多い形式

テスト項目にはFlaw(欠点)がつきもの

- ▶ 多枝選択式のテスト項目にFlawがあることでテストが測定道具としての機能を十分に果たさないという指摘
- ▶ 項目のFlawはテストの規模によらず存在する
 - ▶ 学内で用いられた試験問題1,925項目のうち37.3%の項目で1つ以上のFlaw(Rush, Rankin, & White, 2016)
 - ▶ 看護師試験で用いられる多枝選択式項目の46.2%の項目で1つ以上のFlaw(Trrant, Knierim, Hayes, & Ware, 2006)
- ▶ こうしたFlawの影響を抑えるために、項目作成ガイドラインがある。

項目作成ガイドラインとFlaw

- ▶ 項目作成ガイドライン
 - ▶ e.g.)Haladyna, & Rodriguez (2013)
 - ▶ 多枝選択式項目で検討すべき測定内容・配置・記述・設問部分・選択枝に関する27のガイドライン
 - ▶ ガイドラインを守れない項目形式も場合によっては妥当であるという作題者(荒井,2015)
- ▶ 受検者は悪問に気づくのか
 - ▶ 項目のFlawはテストワイズネスにより正答の手掛かりになりうる(Chittooran, & Miles, 2001)
 - ▶ 気づかれやすいFlawであれば対処できる可能性
 - ▶ 気づかれにくいFlawでありながら、正答率が下がるなどの項目は問題

本研究が目 指すところ

ガイドラインの影響を網羅的に検討

- 同一の能力を測定する項目について，項目作成ガイドラインに準拠する作問をした項目と準拠しない作問をした項目を網羅的に比較検討
- 相対的に重視すべきガイドラインを明らかにする。

ガイドラインに受検者がどの程度気が付くのかを検討

- 受検者がガイドラインにどの程度気が付くのかを検討する
- 項目へ影響を与えながらも，受検者に気付かれない(意識することが難しい)ガイドラインを明らかにする。

テストの作成・実施

受検者

愛知県内の国私立大学生477名(有効回答者数453名)

テスト

実験用冊子2冊, アンカーテスト1冊

全て多枝選択式問題

26項目(国語8項目, 数学6項目, 英語12項目)

項目作成ガイドライン(Haladyna & Rodriguez, 2013)への準拠項目/非準拠項目を実験用冊子では13項目ずつランダムに, アンカーテストでは全て準拠項目を割り振った。

手続き

3種類のテスト冊子のうち1種類をランダムに配布・解答を求めた

テスト全体や項目等について, 「気が付いた点」について自由記述でに回答を求めた

テスト実施時間は30分

無回答の

数学のみ全項目に無回答: 無回答として扱った

扱い

その他の無回答: 誤答として扱った

作成したテストについて

指標

受検者の正答率

コメント率

各項目の正誤と受検者の能力推定値の相関：(Item-Theta相関)
1PLMで推定

構成するテスト

準拠項目： $\alpha = .79$

冊子1： $\alpha = .70$

冊子2： $\alpha = .74$

全体的な結果

コメント率

全てのコメントを抽出

受検者の中で1つ以上のコメントをした者は242名で全体の53.4%

能力値とコメント数の間に関連は見られなかった($r=.27$)

正答率

準拠項目全体の正答率の平均は0.61(SD=0.21)

非準拠項目全体の正答率は0.56(SD=0.24)

識別力

準拠項目全体のI-T相関の平均は0.36(SD=0.12)

非準拠項目全体のI-T相関は0.34(SD=0.16)

検討除外

ガイドライン20b:ガイドライン非準拠項目に強意語として不十分な選択枝が存在した

ガイドライン

ガイドライン22:ガイドライン準拠項目に正答選択枝と見なせる項目が複数存在した

特徴的な ガイドライン

- ▶ 作題時に特に注意したい
ガイドライン
 - ▶ ガイドライン3
 - ▶ ガイドライン4
 - ▶ ガイドライン5
 - ▶ ガイドライン9
 - ▶ ガイドライン17
 - ▶ ガイドライン18
 - ▶ ガイドライン19
 - ▶ ガイドライン20e

- ▶ ガイドラインはこちら↓
 - ▶ <http://www.educa.nagoya-u.ac.jp/~ishii-h/materials/guidelines.pdf>

ガイドライン3： 各設問の内容は互いに独立であること

項目の構成

- 数学，二次関数，小問2題
- 準拠項目：独立な項目
- 非準拠項目：(1)の正答を用いて(2)を導出

正答率

- 準拠項目：(1)0.66 (2)0.48
- 非準拠項目：(1)0.73 (2)0.39

I-T相関

- 準拠項目：(1)0.46 (2)0.49
- 非準拠項目：(1)0.41 (2)0.59
- 項目間に依存性が生じているとき，識別力が不当に高まる

コメントについて

- 準拠／非準拠共に低い
- 「よくわからなかった」
- 問題形式に対してのコメントは無

項目の形式が正答率へ影響

- 着目すべきは(2)
- 項目の形式により，識別力，正答率へ影響
- 受検者は項目形式が正答率を低下させる要因だと気づいていない。

ガイドライン4： 重要な事項を問うこと。極端に細かかったり、 逆に一般的すぎる内容にならないこと

項目の構成

- 英語，英単語の意味
- 準拠項目：大学受験レベル
- 非準拠項目：専門性が高く難易度の高い英単語

正答率

- 準拠項目：0.21
- 非準拠項目：0.22

I-T関連

- 準拠項目：0.14
- 非準拠項目：0.28
- 非準拠項目では単語の中に正答を導くヒントを見つけられた可能性

コメントについて

- 非準拠で26.1%と高い
- 一部の受検者「英単語の語源を含め検討すると正答選択枝を絞ることができた」

本項目の目的：「英単語の意味を知っていること」

- 非準拠項目で正答するためには「単語の中にヒントを見つける力」が必要
- 測定したい能力とは別次元の能力を測定する可能性もある

ガイドライン5： 解答が個人の意見に影響されないこと

項目の構成

- 英語、英文を読み文章が示す適切な人物を解答する
- 準拠項目：客観的な正答
- 非準拠項目：主観的な正答

正答率

- 準拠項目：0.86
- 非準拠項目：0.10

I-T相関

- 準拠項目：0.09
- 非準拠項目：-0.28

コメントについて

- 非準拠で35.4%と高い(準拠でも14.2%)
- 多くの受検者「人によって正答が異なる」

正答選択枝は客観的に決められるべき

- 一部の受検者「英文で書かれているテーマについての知識が不足しているため正答選択枝を選ぶことができない」
 - テストに間違いはないという信念が背景にあるか
- 非準拠項目で最も選択率の高いものを正答としてもI-T相関は0.12と低い
 - 正答選択枝は事前に定められるべき
- このガイドラインが守られない場合、受験者の能力を反映しない項目となる

ガイドライン9： 言語レベルを受検者集団に合わせること

項目の構成

- 数学，中学生相当の難易度
- 準拠項目：言語レベルは受検者に合わせられている
- 非準拠項目：全てひらがなによる記述

正答率

- 準拠項目：0.54
- 非準拠項目：0.52

I-T相関

- 準拠項目：0.45
- 非準拠項目：0.42

コメントについて

- 非準拠で62.2%と非常に高い(準拠では0%)
- 受検者の多くは全て平仮名の記述に対し疑問
- 「全て平仮名を用いることで識字障害等への配慮が見られた」等の肯定的な意見を記述した受検者も

ガイドライン9： 言語レベルを受検者集団に合わせること

次の項目へ影響

- 非準拠項目に解答した者のみ、次の設問の正答率が大幅に低くなった
- 漢字仮名交じりの記述と全て平仮名での記述では、読みやすさに及ぼす文脈的・意味的要因が異なる (北尾, 1960)
- 平仮名では漢字よりも深い処理が必要になるとされている (篠塚・窪田, 2012)
- 他の設問に比べて時間ならびに認知資源を多く費やし、次の設問に十分に解答できなかった可能性

言語レベルを下げてでも影響する

- 非準拠項目において受検者に対して大幅に言語レベルを下げた。
- 多くの受検者は設問を理解することは可能
- 正答率への影響が小さかった (準拠0.54, 非準拠0.52) 一因か
- 一方、受検者に対して言語レベルを上げたとき、問題文や選択枝の理解が不十分となる
- 正答率や識別力に影響が出る可能性がある。

ガイドライン17： 各選択枝は互いに独立であること。 内容に重なりがないこと

項目の構成

- 英語。英文が説明する単語を選択。
- 準拠項目：異なる4つの選択枝
- 非準拠項目：正答選択枝と包含関係になるもうひとつの選択枝

正答率

- 準拠項目：0.97
- 非準拠項目：0.90

I-T相関

- 準拠項目：0.36
- 非準拠項目：0.23

コメントについて

- 非準拠項目で17.6%と高い
- 「選択枝2と4はどちらも正しい」

正解になりそうなものを選択する

- 非準拠項目では2枝から絞ることができず、どちらも解答欄に書いた受検者の存在
- 採点基準等を想像して丸をもらえそうな選択枝を探すという出題者の意図を推測するストラテジー(池上,2015)

ガイドライン14「正答枝が唯一であること」との違い

- もうひとつの正答選択枝への選択率は低く、コメント率も低い
- 正答とされる選択枝の「わかりやすさ」が影響か

ガイドライン18： 『上記のいずれでもない』『上記すべてあてはまる』『わからない』などの選択枝を用いないこと

項目の構成

- 英語，多義語の意味
- 準拠項目：提示した語に含まれない意味を選択(選択枝1~3は非準拠項目と同じ)
- 非準拠項目：当てはまる意味を答えさせる項目。選択枝4「上記すべてあてはまる」

正答率

- 準拠項目：0.34
- 非準拠項目：0.39

I-T相関

- 準拠項目：0.35
- 非準拠項目：0.27

コメントについて

- 準拠項目で1.5%，非準拠項目で3.3%

受検者の知識体系は同じ

- 準拠項目：選択枝2が当該英単語のもつ意味だと知っている受検者は多く，選択枝2の選択率は0.10と低い。
- 選択枝1か3のどちらかについてのみ正しいとわかった受検者はそれ以外の選択枝と選択枝4のどちらが正答選択枝かわからなかった
- 非準拠項目：選択枝1か3が当該英単語のもつ意味だと知っている受検者は，選択枝4を選択(選択枝1，3の選択率は低い)
- どちらの項目でも，選択枝2は提示された単語のもつ意味であると理解している受検者が多く，選択枝1，3の意味まで知っている受検者はそれほど多くなかった
- 受検者の持っている知識体系はほぼ同じであるにも関わらず，非準拠項目では正答率・I-T相関が変化した。

ガイドライン18：

『上記のいずれでもない』『上記すべてあてはまる』『わからない』などの選択枝を用いないこと

いたずらに識別力を上げただけ

- 4枝選択式において、1選択枝を「上記すべてあてはまる」とする
 - 残りの3枝のうち2枝が当てはまるとわかった時点で、正答を絞ることが可能
- 全ての選択枝を検討させるために複雑な組み合わせ形式などを使うと…
 - 正答選択枝を全て選ばせたとき、学生の多くは確実に正答と理解する1選択枝を選ぶに留まる傾向がある（遠山・中村, 2013）
 - テストワイズネスを助長させるだけという批判（斎藤・有田, 1981）
 - 同一内容を問う複数のテストにおいて、問題形式によって正答に必要な受検者の知識量レベルはほぼ同じ分布をするのに対し、複数選択式では識別力が高くなる項目が多い（木村・福島・栗原・黒澤, 2000）
- 解答するために必要な知識量にはあまり差はなく、いたずらに識別力を上げるに留まる
 - 高すぎる識別力は、一部の上位層のみを選抜することになる
- 多枝選択式を用いるにあたっては、唯一である正答選択枝を選択させることが項目として最大限に機能する

ガイドライン19： 『でない』『～以外』などの否定表現 を用いないこと

項目の構成

- 国語，和歌の修辞法
- 準拠項目：選択枝に否定表現なし。間違っている選択枝を選ぶ
- 非準拠項目：選択枝の中に多くの否定表現。正しい選択枝を選ぶ

正答率

- 準拠項目：0.60
- 非準拠項目：0.35

I-T相関

- 準拠項目：0.50
- 非準拠項目：0.48

コメントについて

- 非準拠項目で2.2%(準拠項目では0.7%)
- 多くの受検者にとって違和感のない項目

非準拠項目では測定対象を十分に測定できていない

- 同じ内容の項目，正答率は大きな差
- 否定語を含む項目は正確に読み取ることが困難
- 非準拠項目でのコメント「こうした項目の解き方を習った」という主旨のものも有
 - ストラテジーとして解答方略を知っているかどうかで正誤を分けた可能性
 - 測定している能力以外で対処されたか

ガイドライン20e： 明らかに不要・不自然な選択枝は入れ ないこと

項目の構成

- 国語、古典文学作品の選択
- 準拠項目：古典文学作品の選択枝
- 非準拠項目：選択枝4を現代漫画作品のタイトル

正答率

- 準拠項目：0.52
- 非準拠項目：0.58

I-T関連

- 準拠項目：0.48
- 非準拠項目：0.45

コメントについて

- 非準拠項目で3.9%(準拠項目では0.0%)

4枝選択にするか3枝選択にするか

- 適切な誤答枝を用意できない場合、3枝選択にするかを考える
- ガイドライン13「もっともらしく、識別力の高い選択枝のみにすること。多くの場合、3枝で事足りる」
- 選択枝数を5→3, 4→3としても難易度や識別力への影響は少(Rodriguez, 2005)
- 一方、ガイドライン13では準拠項目にのみコメント
 - なぜこの項目だけ3枝選択なのか
 - ガイドライン13を準拠した方が違和感を抱く受検者
- 総合的に考える必要がある

総合的に 見て

ガイドライン間の独立性と依存性

- ガイドライン単体で検討すればよいもの
- 別のガイドラインと同時に検討する必要性のあるもの
- 複数のガイドラインを参照する場合には、同時に準拠すればいいということではない
- 項目の内容と受検者に与える印象を含めて検討する必要性

受検者はテストをあまり疑わない

- 項目に違和感を覚えながらも、敢えてそのように出題された理由を探す受検者もいる
 - ガイドライン5：知識が不足していて正答がわからない
 - ガイドライン9：すべて平仮名で書くことは障害への配慮と認識
- テスト自体に間違いはないという考えが背後にある

本研究の 限界

解答のストラテジーを知っているかどうか
が違和感を抱く理由なのか

- ストラテジーを知っているから違和感を抱く可能性もある
- ストラテジーやガイドラインを事前に周知すれば、受検者は気を付けて解答できるか？

全ての教科について検討したものではない

- 教科特有の問題によって、同じガイドラインでも影響力が違う可能性はある
- 各教科の特徴を考慮した上でガイドラインを検討する必要がある