

IRTって何？

公益社団法人医療系大学間共用試験実施評価機構：CATO

CATO 先生



共用試験実施評価機構に
協力する〇×大学の先生

共子ちゃん



〇×大学医学部の3年生
来年受験する共用試験 CBT
がちょっぴり不安



CATO 先生、共用試験 CBT に使われている「項目反応理論（IRT）」って何ですか？

項目反応理論（IRT）は、テストの評価のための理論だよ。英語の名称は Item Response Theory というよ。項目（Item）はテストの一つ一つの問題を表し、反応（Response）は回答状況を表しているよ。



うーん・・・なんだか難しそうですね。

まあ、あわてずに順を追ってゆっくりみていこう。

ところで、共子ちゃん、どうして共用試験ではテストの問題が1人1人バラバラかは知っているかな？

臨床実習の時期が大学によって違うから、医師国家試験のように全員が同じ時期に試験を受けることができないからじゃないですか？

その通りだよ。実施時期が違うのにみんな同じテスト問題では、あとから受けた方が有利になってしまうよね。

でも、1人1人のテストが違っていたら不公平じゃないですか？

項目反応理論（IRT）を使えば、1人1人のテストが違っていても公平に評価できるんだよ。

では、共子ちゃん、簡単なテストで100点とったA君と、難しいテストで50点しかとれなかったB君、実際に勉強ができるのはどちらかわかるかい？

テストの難しさが違うのだから、単純に比較はできないような気がします。

これまでの一般的なテストでは、問題の難しさにかかわらず、A君の方が高い評価になってしまっていたよね。テストの点数には、受験者の能力とテストの難しさの両方が影響していて、2つを切り離すことはできなかったんだ。

でも、項目反応理論では、テストの難しさと受験者の能力を別々に把握することができるんだよ。

共子ちゃんの視力はどれくらいなのか？

先生、急にどうしたんですか！？ 1.0ですけど・・・

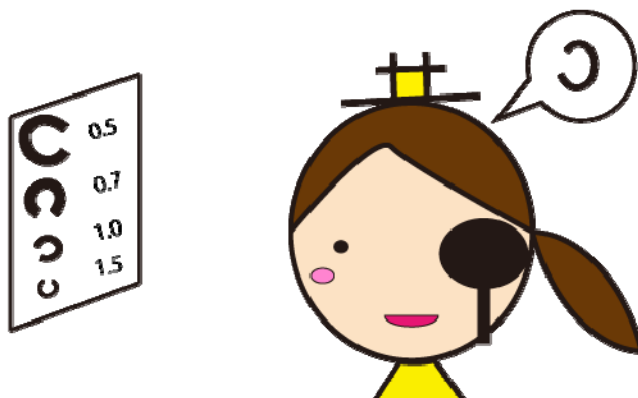
IRTはよく視力検査に例えられるんだよ。視力検査では、Cのマークがいくつ見えたかじゃなくて、どの大きさのCが安定的に見えたかで視力を測っているよね。

共子ちゃんは、視力1.0なら、きっと、視力0.5の人よりは、小さいCがわかるし、視力1.5の人ほどは、小さいCはわからないよね。

今までの素点によるテストだと、Cの大きさにかかわらず、Cがいくつ見えたかで能力を測っていたようなものなんだ。でも、IRTを使えば、どの大きさのCが見えるかで能力を測定することができるんだよ。

ある難易度の問題に対し、ある能力をもった受験生が正答する確率はX%であるというように受験者の能力を求めることができるよ。

注:文部科学省中央教育審議会第15回高大接続特別部会資料内容を引用しています。



試験終了後に問題が公開されないのはどうしてですか？

問題を公開すると、その問題は、その後テストに出題できなくなってしまうよね。問題を再利用することで、事前に問題を評価することができるから、安定した試験が行えるんだ。

そういえば、最近受けた英語試験の TOEIC BRIDGE®でも問題は公開されていませんでした。

こういった形式の試験は増えてきているよね。

他にどんな試験で項目反応理論が使われていますか？

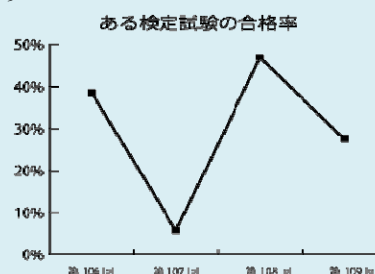
英語圏留学に使用される試験の「TOEFL iBT®」や、一般企業で広く採用試験として利用されている「SPI」、国家試験である情報処理技術者試験「IT パスポート試験」でも使われているよ。

年によって問題の難しさは同じなんですか？ 去年、サークルの先輩が受けたテストのほうが簡単だったら、先輩がずるい気がするなあ。

項目反応理論では、事前に問題の評価を行って、良質と評価された問題のみをテストに使っているよ。事前に評価された問題を使っているから、テストを実施する前にある程度テストの得点を予測することができるし、問題の難易度を合わせることもできるよ。

資格試験なんかでも、毎年、試験の難易度が変わって、年によって合格率が大きく違うことがあるよね。

共子ちゃん、このグラフは、実際に行われたある検定試験の合格率の推移だよ。



試験回によってずいぶん合格率が違いますね。これでは、合格者のレベルが本当に同じかわかりませんね……。マラソンレースで毎回距離が変わっていたら大変なことですけど、試験も同じですよ。

従来の試験方法では、実際に試験を行って見ないと結果を予測することは難しかったからね。

項目反応理論を使えば、こうした問題も解決できるんですね！

X年のテスト



Y年のテスト



年によって距離が違ふレースは不公平



具体的には、どのように項目反応理論を CBT に使っているんですか？

項目反応理論は、CBT の問題の評価と得点の算出の両方に使われているんだよ。

では、CBT の出題方法の流れに沿って説明していくよ。

まず、大規模な集団に問題を出題して、1 問 1 問の問題の性質を検証するんだ。

「問題の性質」ってどのようなことを指しているのですか？

主に、問題の識別力と問題の難しさのことだよ。

「識別力」って何ですか？

識別力は、テスト全体の成績と一つひとつの問題の成績の関連が強いかどうかをみる指標だよ。

テストの成績の良い人が正答し、悪い人が誤答するような問題は、テストの成績とリンクすることになるから、識別力が良い問題として評価されるよ。逆に、テスト全体の成績がよい人ほど、正答率が低くなる問題は、識別力が悪い問題とされるよ。識別力の悪い問題は、問題の表現があいまいだったりして受験者が惑わされる傾向がある問題のことが多いよ。

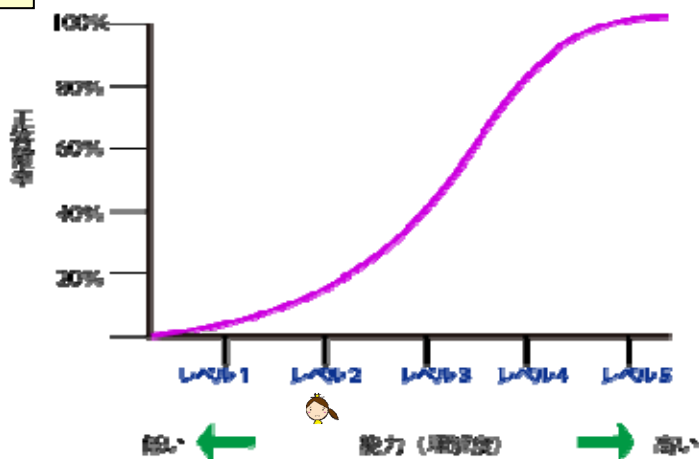
問題の難しさは、正答率で調べているんですか？

正答率では、その問題に答えた受験生のレベルに正答率が影響されてしまうよね。項目反応理論では、項目特性曲線と呼ばれるものを使うんだよ。

項目特性曲線・・・また難しい言葉ができましたね。

項目特性曲線は、横軸に受験者の能力、縦軸に問題の正答率をとったグラフで表すよ。グラフをみると、あるレベルの人がこの問題に正答する確率が何%かがわかるよ。

項目特性曲線のイメージ



これでは、私は、この問題に正解する確率は低そうですね。

問題ごとに、どのようなレベルの人が、どれくらい正解する難しさの問題かが、わかるんですね。

問題の性質がわかったら、質の良い問題だけをたくさん集めた貯蔵庫のようなものを作るんだ。これを「項目バンク」と呼んでいるよ。

CBT では、この項目バンクから問題が出題されているんですね！

では、項目反応理論を使って、どのように受験者のスコアをだしているんですか？

項目特性曲線をみると、1 問ごとに、これくらいの理解力がある人が、この問題に正解する確率は何%かということがわかるよね。この 1 問 1 問の項目特性曲線を各受験者に出題されたテスト別に組み合わせて、その受験者の解答パターン（正答・誤答）に従って計算すると、この解答パターンを示す人の最も可能性の高いスコアを出すことができるんだよ。

スコアは他の年に受験した学生のスコアとも比べることができますか？

共子ちゃんのスコアを、ある基準の年に受験した集団と比較して、自分がどの位置にいるかで比べることができるよ。この基準とした集団を「基準集団」と呼んでいるよ。

基準集団の平均点を 50、標準偏差（平均点からどれくらいの幅にばらついているかを示した数値）を 10 と設定して、受験生の成績を算出したのが IRT 標準スコアだよ。

注：2016 年度実施試験からは、平均点を 500、標準偏差を 100 に改訂します。

IRT 標準スコアは、大学入試の時によく耳にした「偏差値」と似ていますね？

計算方法は偏差値と似ているね。でも、偏差値は、そのテストの平均点を基準としているんだが、年により受験生が違うから年度間の得点の比較が難しいんだ。IRT 標準スコアは基準集団にあわせて評価を行っているので比較ができるんだよ。

基準は安定していないといけませんよね。

毎年、ものさしの目盛が変わってしまったら測定することはできないからね。

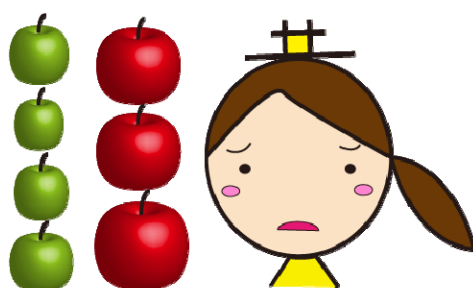
「基準集団」というものさしを使うことで、いつでも同じように公平に成績を測ることができるんですね。

どの年の受験生を基準集団にしているのですか？

これまでは、CBTの正式実施の最初の年の2006年の集団を基準集団としていたよ。ただ、2006年は最初の年ということもあって、あまり成績がよくなかったんだ。2006年以降は成績は安定してきたんだけど、そうすると、2006年の集団は基準としての役目を果たさなくなるよね。それで、2016年度実施試験から基準集団が2012年から2014年までの3年間の受験者集団に変更されたよ。

(注：歯学系の基準集団は2013年受験集団である)

昨年受験した先輩達や、後から受験する後輩達とも成績を比較することができますね。



ものさしの単位がばらばらだと測定はできない



テスト問題は過去に集めた問題を繰り返し使っているんですよね。問題が漏れたら、影響してしまいますよね？

項目バンクには2万問を超える項目が集められていて、問題は、この項目バンクからランダムに出題されているよ。受験生には、どの問題が出題されるかわからないから、もし、共子ちゃんが項目バンクを知ったとしても、2万項目全てを暗記するのは不可能だよ。

2万問！すごい問題数ですね！

もちろん、性質が変化した問題は、項目バンクから取り除いているし、毎年、新たな問題を追加して、項目バンクは常に新陳代謝を図っているよ。問題の漏洩を防止するため、機構では厳しく問題の管理を行っているし、情報を漏洩した場合は、厳しい罰則も設けられているから、問題を漏洩させてはいけないよ。

ざっと説明してきたけど、項目反応理論について少しはわかってきたかな？

項目反応理論は、勉強した成果を正しく、公平に測るために使われているんですね！

私、項目反応理論についてもっと詳しく知りたくなってきました！

医療系大学間共用試験実施評価機構から「項目反応理論についての説明書」が各大学に送られてるから、参考にしてみるといいよ。

なんだか、CBTを受験するのが少し楽しみになってきました。

CATO: 公益社団法人医療系大学間共用試験実施評価機構の略称です

Ver1.1 改訂履歴

1. 7/8 頁 （注：歯学系の基準集団は 2013 年受験集団である）を追加