

項目分析結果に基づく作問支援システムの開発

林 貴史[†] 高木 正則[†] 山田 敬三[†] 佐々木 淳[†]

岩手県立大学大学院ソフトウェア情報学研究科[†]

1. はじめに

多くの教育現場では、学習者の能力や学習効果を測定するためにテストを実施し、テストの得点から各学習者を評価する。しかし、テストに出題された問題（以下、項目）の良し悪しを評価することは少ない。その要因として、テスト項目の分析に必要なテスト理論や統計などの専門知識が不足していることが考えられる。また、テスト項目の分析結果から、作問時の改善点を導き出すのも難しい。そこで、我々はテスト項目の分析負担の軽減を目的とし、項目分析結果に基づく作問支援システムを提案する。本システムでは、テスト項目の分析結果や次回作問時のアドバイスをシステムから提示することによって、作問経験の少ないテスト作成者でもテスト項目を評価でき、作問スキルの向上が期待できる。

本稿では、岩手県盛岡市で開催されているご当地検定「盛岡もの識り検定⁽¹⁾」（以下、もりけん）を対象とし、過去の被験者データを用いた分析結果の提示方法や、作問アドバイスの生成ルールならびに提示方法について述べる。

2. 関連研究

樋口は⁽²⁾、テスト理論の知見を有さない教授者が容易に使用できるテスト結果分析 Web アプリケーションを開発している。また、熊谷は⁽³⁾、項目反応理論によるテスト分析を行う無料ソフトウェアの Easy Estimation を開発している。これらのシステムでは、テスト結果の分析といった点で本研究と類似するが、次回作問時の作問アドバイスの生成までを対象としていない点で異なる。

3. 研究課題と課題解決へのアプローチ

本研究では、①テスト理論の知識を持たないテスト作成者でも理解できるテスト分析結果の表示方法の解明と、②次回テスト作成時の参考になる作問アドバイスの生成ルールの構築が研究課題となる。課題①については、テスト項目の良し悪しを視覚的に判断できるようにグラフを活用した表示方法を検討する。課題②では古典的テスト理論や項目反応理論によって算出される各項目のパラメータ（項目難易度、項目識別度、S-P 表分析から得られる注意係数

の値など）から次回作問時における改善点（アドバイス）を生成するルールを構築する。

4. 作問支援システム

4.1 システムの概要

提案する作問支援システムの概要を図 1 に示す。システム利用者は被験者の解答データを①本システムに入力し、②項目分析モジュールで古典的テスト理論や項目反応理論などを駆使し、各テスト項目を分析する。分析結果は利用者にフィードバックされるとともに、③作問アドバイス生成モジュール内で表 1 に示すルールに基づいて生成された作問アドバイスとして利用者に提示される。

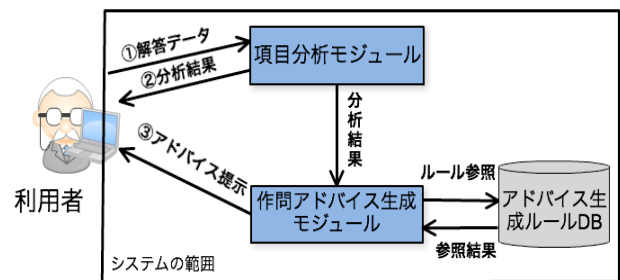


図 1: システム概要図

表 1 テスト項目の評価指標・評価手法・評価基準

指標	手法	基準
信頼性	再テスト法	“信頼性 0.80 以上”であれば、信頼性が高い ⁽⁵⁾
	平行テスト法	
	折半法	“信頼性 0.71 以上”であれば、テスト全体の信頼性が高い ⁽⁶⁾
	内部一貫法(クロンバックの α)	
難易度	古典的テスト理論 項目反応理論	“難易度 0.39 以下”または“難易度 0.81 以上”の項目は難易度が不適切な項目、 “難易度 0.29 以下” または “難易度 0.91 以上”は修正及び検討が必要な項目 ⁽⁴⁾
識別度	古典的テスト理論 項目反応理論	“識別度 0.29 以下”は識別度が低く不十分な項目、“識別度 0.19 以下”は識別度が極端に低く可否判定には直結しない項目 ⁽⁴⁾
注意係数	S-P 表分析	“注意係数 0.5 以上 0.75 未満”は注意すべき項目 “注意係数 0.75 以上”は特に注意が必要な項目 ⁽⁷⁾

Development of a Support System for Creating Test Questions based on Item Analysis Result.

[†]Takafumi HAYASHI [†]Masanori TAKAGI [†]Keizo YAMADA [†]Jun SASAKI

[†]Graduate School of Software and Information Science, Iwate Prefectural University

4.2 項目分析モジュール

一般に、テストは信頼性と妥当性によっては評価されるが、妥当性は定量的評価が難しいことから、本研究ではまず信頼性を評価する機能について検討した。信頼性を評価する方法には内部一貫法、折半法、再テスト法、平行テスト法など数多くの手法がある⁽⁴⁾。今回は、テスト全体の信頼性を評価するために一般的に広く用いられている内部一貫法のクロンバックの α 係数を算出することとした。また、個々のテスト項目を評価するパラメータとして、古典的テスト理論で算出した難易度と識別度、S-P 表分析を元に算出した項目注意係数についても分析して提示する。表 1 に文献(4)～(7)から得られたテスト項目の評価指標・評価手法・評価基準を示す。

4.3 アドバイス生成モジュール

図 2 に作問アドバイス生成の流れを示す。項目番号 1 の難易度 0.26、識別度 0.16、注意係数 0.76 の問題の場合、①表 1 で示した基準値に当てはめ、難易度/識別度/注意係数の特徴の判断を行う。次に、②その問題に関する特徴をまとめた文章を作問アドバイスとして生成する。生成されるアドバイス条件と文例を表 2 に示す。

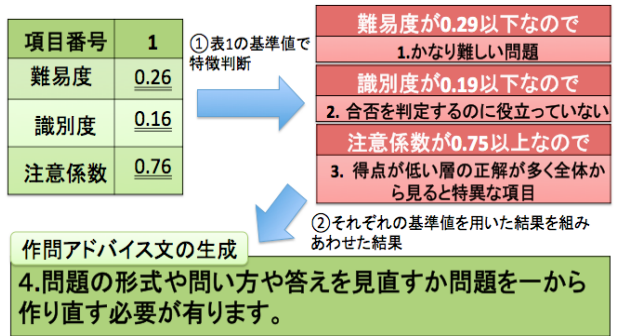


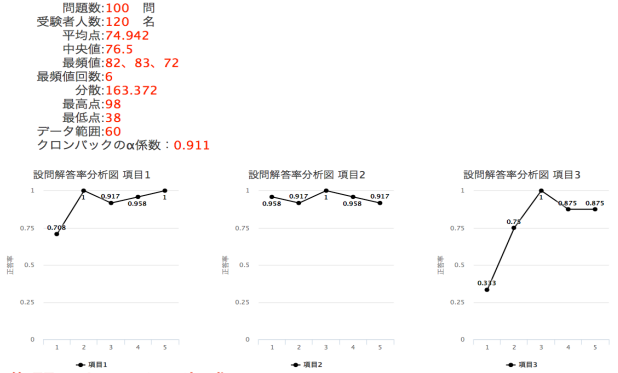
図 2：作問アドバイス生成の流れ

表 2：作問アドバイス生成ルール例

条件	作問アドバイス例
難易度 0.29 以下 識別度 0.19 以下 注意係数 0.75 以上	・ 問題の形式や問い方や答えを見直すか問題を一から作り直す必要が有ります。
難易度 0.29 以下 識別度 0.19 以下 注意係数 0.5 以上 0.75 未満	・ 合否判定には機能していない問題であり、問題の問い方や答えがあっているか注意してみる必要が有ります。
難易度 0.29 以下 識別度 0.39 超過 注意係数 0.5 未満	・ 難しい問題を作る際はこの問題をベースに問題を作るか、修正すると次のテストでも使えるでしょう。
難易度 0.39 超過 及び 0.81 未満 識別度 0.39 超過 注意係数 0.5 未満	・ 最も理想的な問題であり、次回以降のテストにおいても使えるでしょう。この問題をベースにすると理想的な問題ができそうです。

5. システムの開発および動作実験

上記の考え方にに基づき作問支援システムを開発した。開発言語は PHP, JavaScript, HTML, データベースには MySQL を用いた。2012 年のもりけん 3 級（四択問題 100 問）の全被験者 120 名分の解答データを活用し、本システムでテスト項目を分析した。図 3 にシステムから提示された分析結果とアドバイスの例（一部）を示す。これにより、システムが正常に動作することを確認した。



作問アドバイス生成

1項目	作問アドバイス 1. 難易度が極端に高い問題 2. 識別度が極端に低く、合否判定には機能しない問題 3. 注意係数は適切な問題 4. 問題文や選択肢などを難しめな表現や類似した選択肢を入れるなどして問題を作ると良いでしょう。
2項目	作問アドバイス 1. 難易度が極端に高い問題 2. 識別度が極端に低く、合否判定には機能しない問題 3. 注意係数が非常に高く、要注意な問題 4. 問題の形式や問い方や答えを見直すか問題を一から作り直す必要が有ります。
3項目	作問アドバイス 1. 難易度が適切な問題 2. 識別度が高く判別に適している問題 3. 注意係数が適切な問題 4. 理想的な問題であり、次回以降のテストにおいても使えるでしょう。この問題をベースにすると理想的な問題ができそうです。

図 3：分析結果と作問アドバイス生成画面例

8. まとめと今後の課題

本稿では、テスト項目の分析結果とアドバイスが生成される作問支援システムの開発について述べた。また、本システムを用いて実際のテスト項目を分析し、改善点など作問アドバイスが生成されることを確認した。今後は、もりけんの作問委員会に対してシステムの利用実験を行い、システムの改良を図る。

参考文献

(1) 盛岡商工会議所, 盛岡もの識り検定
<http://www.ccimorioka.or.jp/jinzai/moriken.htm>
(2) 樋口三郎: 『テストおよびアイテム分析 Web サービスの開発』, 教育システム情報学会第 39 回全国大会講演論文集, pp.377-378, 2014.
(3) 熊谷龍一: 項目反応理論と EasyEstimation., <http://irtanalysis.main.jp>, 2011.
(4) 大友賢二: 『言語テスト・データの新しい分析法 項目応答理論入門』, 大修館書店, 1996.
(5) 山森光陽, 前田啓朗(編)『英語教師のための教育データ分析入門』, pp.4-12, 東京:大修館, 2004.
(6) Nunnally, JumC. Psychometric Theory 2nd Edition. New York: McGraw-Hill, 1978.
(7) 藤垣雅司, 藤垣康子, 中島光洋: 「注意係数の規格化: S-P 表における反応パターンの指数について」, 日本科学教育学会, Vol.9, pp.260-261(1985)