

算数学力評価ルーブリックの妥当性検討の試み —面積学習領域を例として—

○ 佐藤誠子 (石巻専修大学)

柴山 直 (東北大学大学院)

問題と目的

近年, 心理計量・教育測定分野において, IRT モデルを用いて「学力」を単一の尺度上で表現する試みがなされてきている。しかし, 大規模テストを前提とするこの手法から得られた尺度値 (θ) そのものだけでは教育現場における実際の指導に活用しにくいという問題点が指摘できる。これにこたえるため, 佐藤・柴山 (2013) では, 算数面積学習領域を例に, IRT モデルにより算出された尺度値のレベル毎のパフォーマンス評価ルーブリックの作成を試みた。本報告ではそれに引きつづき, ルーブリックの内容と実際の学習指導との整合性をさらに向上させ, 調査対象人数を増やすとともに, 教員へのアンケート実施を行うことにより, ルーブリックの妥当性を検討することを目的とする。

方 法

調査対象者: 山形県内公立 A, B, C 小学校の第 6 学年児童 121 名。**手続**: 算数学力テスト, 面積テストの問題冊子への解答を求めた。また B, C 小学校の担任 3 名を対象に, ルーブリックの妥当性に関する教員アンケートを実施した。**課題構成**: 算数学力テスト平成 18 年度の新潟県学力テストを使用し, 面積に関する問題のうち 2 問を削除した (全 23 問)。**面積テスト**等積変形課題 (対提示, 連続性), 等周長変形課題 (対提示, 連続性) よりなる (工藤・白井, 1991 など参照)。いずれも底辺, 高さの相対的大小を読み取り, そこから面積の大小を導出する必要がある問題である。**教員アンケート**: 佐藤・柴山 (2013) でみられた面積課題の判断理由 (a 高さへの着目, b 公式算出, c 切り貼り変形, d 周長への着目, e 動作への着目) に対して, それぞれ面積の理解度を 5 段階で評定を求め, また a~e について理解が深い順に並べてもらった。**各パラメタの推定**: 項目パラメタ (識別力, 困難度) の推定には, 平成 18 年度新潟県学力テストのデータ (受検者: 小学 5 年生 21,735 名, 2006 年 1 月実施) を用いた (推定方法は周辺最尤法, 事前分布は標準正規分布)。受検者パラメタ (尺度値 θ) の推定は最尤法による。算数テスト全問正答者は分析から除外した。これらのパラメタ推定には EasyEstimation

(熊谷, 2009) を使用した。

結果と考察

算数テスト全問正答者 7 名, テスト欠席者 2 名, 回答不備 1 名を除外した 111 名が分析対象者となった。IRT により算出された尺度値 (θ) の平均は 0.04 (SD1.17, Min.-2.57, Max.2.45) であった。これについてレベル毎 L (-2.57~0.59), M (0.59~0.43), H (0.47~2.45) の 3 群 (N=37) に分類し各面積課題の正答率を算出した (Figure1)。学力レベルが上がるにつれて正答率も高くなる傾向にあること, 特に M 群で課題間の正答率の開きが大きいことがうかがえた。そこで各面積判断の理由について a 高さへの着目, b 公式算出, c 切り貼り変形, d 周長への着目, e 動作への着目に分類し, レベルごとに出現頻度を算出した。その結果, 適切属性 (a,b) への言及は $L < M < H$ であった一方, 不適切属性 (d,e) は $H < M = L$ であった。学力レベルが高いほど, 不適切属性への着目を回避し高さや公式を利用した面積比較が可能になるといえる。以上より, L では不適切属性に着目しがちであり面積の大小の導出が困難であること, M では図形の底辺・高さを読み取り公式代入による面積比較が可能であるが, 課題によっては不適切属性が関与すること, H では具体的数値が示されていないととも, それらの相対的な大小関係に着目して面積比較が可能であることが示唆される。また, 教員アンケートの結果, 3 名中 2 名が判断理由について理解度の高い順に a, b, c, d, e と示していたことから, 作成したルーブリックについてある程度の妥当性が得られたといえる。

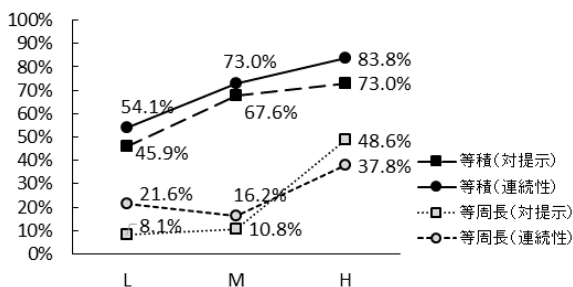


Figure1 θ の各レベルにおける面積課題の正答率

*謝辞: 本研究は JSPS 科研費 25380867 の助成を受けた。