項目反応理論による汎用的能力に関する尺度の垂直尺度化

○並木 雄大¹・須田 望^{2*}・長谷川 康代^{3*}・牧野 直道^{3*}・岸本 きよら^{3*}・川端 一光⁴(明治学院大学大学院心理学研究科¹・株式会社日経リサーチ²・株式会社ベネッセコーポレーション³・明治学院大学心理学部⁴)

キーワード: キー・コンピテンシー, 項目反応理論, 垂直尺度化

Vertical scaling of generic skills by item response theory

Yudai Namiki¹, Nozomi Suda², Yasuyo Hasegawa³, Naomichi Makino³, Kiyora Kishimoto³ and Ikko Kawahashi⁴(Graduate School of Psychology, Meiji Gakuin Univ.¹, Nikkei Research Inc.², Benesse Corporation Inc.³, Faculty of Psychology, Meiji Gakuin Univ.⁴)

Key Words: Key-competency, Item Response Theory, Vertical Scaling

1. 問題と目的

近年, これからの社会で必要な資質・能力の育成に向けて, 教科学力だけではなく汎用的能力が求められている(文部科学省, 2017)。汎用的能力はこれまでも学校活動で育まれてきたが, 教科学力と比べて見えにくい力であった。そのような背景をもとに, ベネッセコーポレーションは汎用的能力としての思考力を, 批判的思考力・協働的思考力・創造的思考力の観点から測定する高校生用の尺度であるGPS-Academicを開発し, 2016年にリリースしている。これまで, 低学齢から高学齢までの思考力を比較可能にする研究は行われていない。そこで, GPS-Academicによって, より低学齢からの思考力の成長を同一尺度上で評価し, 可視化することを目的として, 小学生・中学生を対象としたモニター調査を実施した。

本研究では, 小学生・中学生・高校生の受検データに対して, 項目反応理論による垂直尺度化とそれに向けた分析を行った。

2. 方法

2017年6月に, 3種類のGPS-Academicのモニター試験(各30項目)を実施した。中学生用の問題(以下G2)には2770人, 小学校高学年(5・6年)用の問題(以下G3)には534人, 小学校中学年(3・4年)用の問題(以下G4)には441人が回答した。各項目反応行列に対して, 古典的テスト理論(Classical Test Theory; CTT)に基づく分析を実行し尺度の一次元性を確認したうえで, 項目反応理論(Item Response Theory; IRT)に基づく分析に進んだ。実際の分析では, 各項目反応行列において欠測値を含む受検者を除外し, G2, G3, G4それぞれ2763人, 532人, 440人を分析対象とした。また, 正答率・識別力, GP分析図, 因子負荷量, 信頼性係数などの観点から多面的に分析し, テストの信頼性に影響を与える可能性のある項目を削除して, G2は22項目($\alpha = .82$), G3は20項目($\alpha = .77$), G4は20項目($\alpha = .80$)を分析対象とした。そして, 共通受検者法を用いて, 高校生用の問題(以下G1), G3及びG4の尺度を以下のような手順でG2尺度上に等化した。等化

(equating)とは, 2つの尺度について, どちらか一方の尺度に, 他方の尺度の原点の位置や目盛りの幅を揃えることである(加藤・山田・川端, 2014)。なお, 本研究においては2母数ロジスティックモデルを用いた。また, 本分析では, 統計解析環境Rのirttoolsパッケージ(Partchev, 2014)とフリーソフトICL(IRT Command Language; Hanson, 2002)を用いて, 項目母数及び能力母数の推定を行った。

Step1: 等化前の尺度上で共通項目の項目母数 p 及びそれ以外の項目母数 p^+ , そしてそれらを用いて能力母数 θ を推定する。

Step2: G2尺度上で推定されている共通項目の項目母数 p^* を用いて, 等化前の尺度の能力母数 θ^* を推定する。

Step3: Step1で推定した θ 及びStep2で推定した θ^* の平均

と標準偏差を用いて等化係数 k , l を推定する。

Step4: Step3で推定した k , l を用いて, Step1で推定した p^+ をG2尺度上に等化し p^{**} を得る。

Step5: Step2で用いた p^* とStep4で得た p^{**} を用いて, 能力母数 θ^{**} を推定する。

3. 結果と考察

前述した5つのStepを経て, G1, G3, G4の3つの尺度をG2尺度上に等化した。その結果, G1, G2, G3, G4の順で段階的に困難度母数が小さくなっており易しい項目数が増加していた。一方, 識別力母数は段階的に大きくなっており識別力が高い項目数が増加していた。また, G1, G2, G3, G4の順でテスト情報曲線が左にシフトしていき, 情報量のピークが能力の高い尺度域から能力の低い尺度域へ移っていた。さらに, G1, G3, G4をG2尺度上へ等化することで, θ の値を比較可能にすることができた(図1)。具体的には, G1, G2, G3, G4の順で全体的に θ の分布は左にシフトしていき, 段階的に能力が低く推定されていることが確認できた。以上のことから, GPS-Academicを用いて低学齢から高学齢までの思考力の成長を同一尺度上で評価できることが示唆された。

引用文献

Hanson, B. P. (2002). *IRT Command Language*.<http://sourceforge.net/projects/ssm/>.

加藤健太郎・山田剛史・川端一光(2014). Rによる項目反応理論 オーム社

文部科学省(2017). 新しい学習指導要領の考え方—中央教育審議会における議論から改訂そして実施へ—。
http://www.mext.go.jp/a_menu/shotou/new-cs/_icsFiles/afiedfile/2017/09/28/1396716_1.pdf
 (2019年4月23日取得)。

Partchev, I. (2014). irttools: Simple interface to the estimation and plotting of IRT models. R package version 0.1.7,

<http://CRAN.R-project.org/package=irttools>. 図1. G2尺度上に等化後のG1~G4の θ のヒストグラム