スキルテストへの項目応答理論の適用 (2) — Rasch 系モデルによる幼児の運動技能の推定—

青柳 領

Osamu Aoyagi: Application of Item Response Theory to Motor Skill Tests: Estimation of Motor Skills of Young Children Using Selected Rasch Models. *Japan. J. Phys. Educ. Hlth. Sport Sci.* 53: 409-421, November, 2008.

Abstract : This study investigated the applicability of item response theory to motor skill tests. In particular, we focused on the evaluation of children's motor skills using three Rasch models. Motor skill tests included the number of successes in a given trial, the number of successes in a given time, and a subjective rating regarding the form of the motor skill. We used Andrich's Binomial Trials Model, Rasch's Poisson Counts Model, and Masters' Partial Credit Model, respectively, to derive a synthesized motor skill ability using the item response theory from two or more motor skill test scores. A synthesized motor skill ability was computed by combining these three models and applying them to children's motor skill test items. The sample consisted of 205 kindergarten-aged children, and 8 motor skill items were tested, including (1) side jump, (2) jumping over a 30-cm high rope with hands on a board, (3) kicking and aiming a ball, (4) catching a ball, (5) rolling and aiming a ball, (6) forward roll, (7) side roll, and (8) jumping and turning 90 degrees until a complete 360-degree turn is completed. The following results were obtained:

1) The correlational structure among motor skill test items for this sample had high homogeneity and was one-dimensional. High correlations were found between two sets of Difficulty Parameters, which were computed from two randomly divided subgroups. Obtained Difficulty Parameters were not affected by the sample used.

2) Significant correlations were found between synthesized ability parameters and age, standing height, and body weight, indicating that the motor skill ability parameters measured a latent motor skill, which developed with advancing age.

3) A high correlation was found between synthesized motor skill ability and the Principal Component Scores computed by Principal Component Analysis. However, ability parameters showed lower correlations between motor skill test items and age and physique than Principal Component Scores. In contrast, high correlations between motor skill ability parameters and motor skill test items were found.

Finally, the likelihoods for obtaining a synthesized motor skill parameter and an expected score for each test item were obtained, and an explanation for using these two values was provided.

Key words : child, motor skill, item response theory, Rasch model

キーワード : 幼児, 運動技能, 項目応答理論, Rasch モデル

Ⅰ 緒 言

スキルテストには、ボールの正確投げのように目標からどれだけはずれているかを cm (長さ) で記録したり、バスケットボールのドリブルテストのように一定の動作を完了するまでの秒 (時間) で測定したりするものがある。これらの測定値は

間隔尺度として統計処理することが可能である。しかし、スキルテストの中には、a) バスケットボールのシュートテストのように一定試行回数内での成功回数を測定したり、b) バスケットボールやバレーボールの壁パステストのように一定時間内に成就した回数を記録したり、c) 新体操競技やフィギュアスケートの採点のように一定の課題動作フォームを主観的評価 (subjective rating) によ

り点数化する方法もある (Baumgartner & Jackson, 1982; Clark, 1967; Collins & Hodges, 1978; Larson & Yocom, 1951; Morrow et al., 2000). a) の場合, 信頼性をあげるために数回くり返されるが, 被検者が飽きや疲労のためにパフォーマンスが落ちたり, 逆に学習・練習効果により徐々にパフォーマンスが上がるのが考えられるので, その回数はせいぜい数回の試行が多い. 試行回数が少ない場合は, 一定回数の中での成功回数は二項分布すると考えられ (肥田野ほか, 1961, pp. 68-72), そして, 一定時間内の成就回数はポアソン分布する (肥田野ほか, 1961, p. 37) と考えられる. また, c) の場合も, 人間が主観的に再現性のある判断をするためには細分化した点数を判断することも難しいと思われる. したがって, これらの測定値は完全な間隔尺度というよりは順序尺度の性質も併せ持つものとみなすこともでき (池田, 1976)^{注1)}, 体力測定や実験室で得られたデータとは異なり, スキルテストの中には正規分布および間隔尺度を仮定した統計処理を施すにはかならずしも適切でない場合も少なくない.

従来, 上記のような測定値も間隔尺度とみなし, 因子分析や主成分分析のような多変量解析を用いて尺度化が行われてきた (松浦, 1969). これらに共通しているのは各項目に一定の重みを掛け, 加算していく線型モデルという点である. しかし, この場合, 推定された能力得点^{注2)}と能力得点の推定に用いられた項目間に対応はなく, 例えば「この程度の能力を持った者は, この項目ではこの程度の測定値を得るはずである」といった項目と能力得点とを対応させた予測はできない. この点, 項目応答理論は項目の難易度^{注2)}と能力得点 (項目応答理論では能力特性値^{注2)}と呼ぶことにする) が同一の尺度上に定義することが可能である. このことから項目応答理論は response-centered approach (Zhu and Cole, 1996) と呼ばれている. このような観点から青柳 (2007) は, 様々なシュート方法やシュート角度からなるバスケットボールシュートのスキルテストに項目応答理論の二項分布モデルを適用して, 能力特性値を求める簡易診断シートを作成している. しかし, ここでのモデ

ルは二項分布モデルに限定されている.

よって, 本研究では能力得点と項目難易度を同一尺度上に定義できる項目応答理論を用いて, 能力の推定を行う. ただし, 先に述べたようにスキルテストでは多様な分布・尺度水準を扱わなければならないという実情がある. そこで, 二項分布モデル以外にも近年さまざまな分布や尺度水準を対象にモデルが構築されている (池田, 1994; 芝, 1991; 豊田, 2005) 項目応答理論の中で, その特性を反映したモデルを組み合わせて用い, その適用可能性を検討する. そして, その一事例として, 幼児の運動技能の測定 (スキルテスト) を取り上げ, 能力特性値を求め, その妥当性を検討する. さらに古典的テスト理論に基づき線型モデルで求められた能力得点との比較・検討を行う.

特に, 本研究で採用される Rasch 系モデルは, 能力特性値 θ を項目困難度パラメタから分離して推定できる (ラッシュ, 1985, pp. 198-203) ので, 古典的テスト理論のみならず項目応答理論のその他のモデルよりも, 測定項目の難易度に依存しない長所を持っている.

II 研究方法

(1) 項目応答理論

Andrich (1978) は, 一定数くり返し試行された場合の成功回数が二項分布 (肥田野ほか, 1961, pp. 68-72) することを利用して, 式 1 に示す項目特性関数をもつ項目応答理論のモデルを提案している.

$$P(x, \theta, n, b) = {}_n C_x \frac{\exp\{x(\theta - b)\}}{\{1 + \exp(\theta - b)\}^n} \dots\dots\dots (1)$$

(ただし, n : 試行回数, x : 成功回数, b : 困難度^{注2)}, θ : 能力特性値)

これは, 1パラメタのラッシュモデル (式 2) を二項分布 (式 3) に代入して求められたもので, 二項分布モデル (Binomial Trials Model) と呼ばれる.

$$P(\theta) = \frac{\exp(\theta - b)}{1 + \exp(\theta - b)} \dots\dots\dots (2)$$

$$P(x, n) = {}_n C_x P(\theta)^x Q(\theta)^{n-x} \dots\dots\dots (3)$$

{ただし, $Q(\theta) = 1 - P(\theta)$ }

図1は困難度 b が -1.0 の場合の二項分布モデルのカテゴリ特性曲線である。

同様に, ラッシュ (1985, pp. 147-190) は, 一定時間内に成就した回数に関してポアソン分布 (肥田野ほか, 1961, p. 37) を仮定し, 式4に示す項目特性関数を提案している。

$$P(x, \theta, b) = \frac{\exp[x(\theta - b)]}{x! \exp[\exp(\theta - b)]} \dots\dots\dots (4)$$

(ただし, x : 一定時間内の成就回数, b : 困難度, θ : 能力特性値)

これは, 式5に示したポアソン分布の中のパラメタ λ を $\exp(\theta - b)$ に置き換えたもので, ポアソン分布モデル (Poisson Counts Model) と呼ばれる。

$$P(x, \lambda) = \frac{\lambda^x}{x! \exp(\lambda)} \dots\dots\dots (5)$$

図2は困難度 b が -1.0 の場合のポアソン分布モデルのカテゴリ特性曲線である。

これとは別に Masters (1982), Masters & Wright (1997) は, 順序尺度のデータに関して, 各項目が最大カテゴリ数 m の中で, θ がカテゴリ x をとる確率 P_x を式6のように定義した部分採点モデル (Partial Credit Model) を提案している。

$$P_x = \frac{\exp\left\{\sum_{i=1}^x (\theta - \delta_i)\right\}}{\sum_{m=0}^m \exp\left\{\sum_{k=1}^m (\theta - \delta_k)\right\}} \quad (\text{ただし, } P_0 = 0) \dots\dots\dots (6)$$

これは, お互いに隣接するカテゴリ x と $(x+1)$ の確率の比を式2によって表現し, 全体について加算することによって求められている。特に, δ は隣接するカテゴリ特性曲線が交差する点を示し, ステップ困難度と呼ばれる。図3はステップ困難度が各々 $-2.0, 0.0, 1.8$ のカテゴリ特性曲線を示している。

図1 二項分布モデル・カテゴリ特性関数

図2 ポアソン分布モデル・カテゴリ特性関数

図3 部分採点モデルのカテゴリ特性関数

困難度パラメタはいずれのモデルも最尤法 (Lord, 1980, pp. 184-225) により求めた。

上記の項目応答理論のモデルはいずれもラッシュ (Rasch) モデル (Hambleton et al., 1991) から導出されたもので, 項目パラメタが困難度パラメタのみからなり, 比較的少ない標本数でも安定したパラメタが推定できるという長所がある (Embretson & Reise, 2000)。

(2) 統計分析

項目応答理論の困難度パラメタの推定に先立ち、分析対象となった項目群が1つの運動技能を測定し、1次元尺度を想定することが妥当であるかという点を検討するために、主成分分析を行い、その第1主成分の固有値を第2主成分以降の固有値と比較を行う。そして、1次元尺度の成立を確認後、能力特性値 θ と困難度パラメタの推定を行う。ただし、同時に項目応答理論を用いて推定された能力特性値 θ が、標本に依存せず、不変性を持つものであるかという点を検討するために、標本を無作為に2分割し、各々別に困難度パラメタを求め、その一致の度合いを相関係数により評価する。また、推定された能力特性値 θ が、加齢や成熟とともに向上していくであろう幼児の運動技能を正しく反映しているかという問題を検討するために、暦年齢や形態測定値との関連を検討する。

さらに、古典的テスト理論から導かれた能力得点と本研究で得られた能力特性値 θ との比較検討を行う。古典的テスト理論では、能力得点に正答数を用いる場合があるが、本研究のように正誤(1-0データ)でなく、多様な値を取る場合は標準化された得点の単純和を用いることもある。しかし、この線形結合による単純和は「用いられた項目との相関が最大になる」「得点の分散が最大になる」といった統計学上好ましい特質は持っていない。そこで、本研究では、古典的テスト理論からの能力得点を得るために主成分分析を用いることにする。つまり、測定値が間隔尺度であることを前提に、最も一般的な一次元尺度の合成変量である主成分得点を求め、項目応答理論で得られた能力特性値 θ と主成分得点との比較・検討を行う。その際、主成分分析は「項目間との相関が最大にする」「主成分得点の分散を最大にする」という2つの観点から、各々相関行列と分散共分散行列の2つのデータに対して行う。

(3) 対象および測定項目

対象はF市内のS幼稚園とS保育園の園児205名(男児105名、女児98名)である。年齢別・性別の標本数の内訳は表1に示した。

表1 男女・年齢別標本数

	男児	女児	小計
4歳児	4	7	11
5歳児	48	37	85
6歳児	53	54	107
小計	105	98	203

項目は①両足とび(勝部, 1971; 体育科学センタ調整力専門委員会, 1976, 1981) ②手つきゴムとび(青柳・松浦, 1982) ③キック的あて(青柳・松浦, 1982) ④フライボールキャッチ(青柳・松浦, 1982) ⑤ころがしのあて(青柳・松浦, 1982) ⑥前転(青柳ほか, 1980) ⑦横転(青柳ほか, 1980) ⑧一回旋とび(青柳ほか, 1980)の8項目である。①②は一定時間内で成就できた回数、③④⑤は一定試行回数内での成功回数、⑥⑦⑧は成就の状態を評定し、点数で記録したものであり、各々ポアソン分布モデル、二項分布モデル、部分採点モデルを適用した。測定方法の簡潔な説明は表2に示した。

III 結 果

(1) 1次元尺度の適用可能性

8項目から1次元の尺度を作成することに意味があるかどうかを検討するために、8項目間の相関行列の固有値を求めた。図4は求められた固有値を示したものである。第1主成分の固有値2.291は第2主成分の1.097の固有値と比較して著しく大であった。

(2) 困難度パラメタ

同時最尤法により求めた困難度パラメタを表3に示した。

また、得られた困難度パラメタが標本に依存しない、不変性を持つ値であるかを検討するために、標本をランダムに2分割して、困難度パラメタを別々に求めた。そして、それらの相関係数を求めると0.854であった。

表 2 測定方法

項目名	測定方法
両足とび	35 cm の間隔で平行に貼られた 2 本のビニールテープの右側のラインの上に右足をのせて立ち、「始め」の合図で左側のラインまで両足跳びで跳び、ラインを右足で踏むか、踏み越すようにする。次に、同様右に跳ぶ。これを左右繰り返す。10 秒間に反対側のテープの上、またはそれを越した回数を数える。
手つきゴムとび	大型積み木の上に手をつけて立ち、高さ 30 cm に貼られたゴムひもに触れないよう左右に跳躍する。15 秒間の回数を数える。
キック的あて	5 m 前に、60 cm 間隔で置かれた大型積み木の間を静止した中型ボールを蹴って通過させる。5 回実施して成功した回数を数える。
フライボールキャッチ	高さ 2 m 50 cm の高さから落とした中型ボール落とさず捕球する。5 回実施して成功した回数を数える。
ころがしのあて	5 m 前に、60 cm 間隔で置かれた大型積み木の間をテニスボールをころがして通過させる。5 回実施して成功した回数を数える。
前転	体操マットの上に 25 cm 間隔で 3 本のビニールテープを貼り、その中央のラインに沿って前転させる。「前転自体できない・やらない」は 1 点、「まっすぐ前に回転できない、当初からの不完全な前転で基準線からはみ出た」は 2 点、「基準線内で前転そのものはできたが、その後の動作が不完全で最後寝たままの状態であった」は 3 点、完全なフォームでできた場合は 4 点を与える。
横転	体操マットの上に 25 cm 間隔で 3 本のビニールテープを貼り、頭の上で腕を伸ばし手を握った状態から、中央のラインに沿って左右に 3 回横転させる。「基準線をはずれず横転した場合」「横転中手を離さなかった場合」に各 1 点を加算し、左右 2 回ずつの合計点を記録する。
一回旋とび	床の上に十字にビニールテープを貼り、その上を跳躍して、90°, 180°, 270°, 360° に向きを変える。成功した場合はそれぞれ、1 点、2 点、3 点、4 点とする。左右 2 回ずつ行い、合計点を記録する。

図 4 主成分による固有値の変化

(3) 能力特性値

表 4 は、「得られた能力特性値 θ がそれを推定する際に用いられた各項目の特性をどの程度反映しているか」という問題を検討するために求められた「能力特性値 θ と各項目との相関係数」である。また、「加齢や成熟とともに発達していくであろう幼児期の運動技能を本研究で求められた能力特性値 θ がいかに反映しているか」という点を検討するために求められた「能力特性値 θ と年齢、身長、体重との相関係数」である。

結果、能力特性値 θ は、年齢とは 0.660、身長

表3 困難度パラメタ

no.	項目名	モデル	困難度	
1	両足とび	ポアソン分布	-1.936	
2	手つきゴムとび		-2.059	
3	キック的あて	二項分布	0.968	
4	フライボールキャッチ		-0.955	
5	ころがしのあて		0.799	
			ステップ困難度	
6	前転	部分採点	-2.774	「1点」→「2点」 ^{注)}
			-2.058	「2点」→「3点」
			-1.317	「3点」→「4点」
			-3.050	「3点」→「4点」
7	横転		-2.189	「4点」→「5点」
			-0.836	「5点」→「6点」
			-0.334	「6点」→「7点」
			-2.841	「1点」→「2点」
8	一回旋とび		-2.222	「2点」→「3点」
			-1.090	「3点」→「4点」
			-1.588	「4点」→「5点」
			0.548	「5点」→「6点」
			0.363	「6点」→「7点」
			0.851	「7点」→「8点」
			0.216	「8点」→「9点」
			2.519	「9点」→「10点」
	0.923		「10点」→「11点」	

注) 1点と2点の境界でのステップ困難度

や体重とも 0.469, 0.269 と有意な相関を示した。また、各測定項目のいずれとも有意な関連を示した。しかし、性別との点双列相関係数は 0.030 とほとんど関連を示さなかった。

(4) 主成分得点との比較

従来、いくつかの項目の結果を総合して1つの尺度を作成する場合、最も頻繁に用いられるのは主成分分析 (Hotelling, 1933a, 1933b) であろう。松浦・中村 (1977) や中村・松浦 (1979) は、スピアマンの二因子モデルを根拠に幼児の基礎運動技能を操作的に定義するために、測定項目群の主成分得点を用い、それらの性差や発達傾向を明らかにしている。これは対象になった項目から得られた得点の分散が最大になるように求めたり、すべ

ての項目と最も相関が高くなるように合成得点を求めることによって、最も項目の特性を反映できる (芝, 1975) という利点のためである。前者は分散共分散行列を対象にした主成分分析により、そして後者は相関行列を対象にした主成分分析により行うことができる。本研究でも、測定値を間隔尺度であるという前提のもとに主成分分析を行ってみる。図5は分散共分散行列から求められた主成分得点、そして図6は相関行列から求められた主成分得点と能力特性値 θ の散布図である。前者の相関係数は 0.931、後者は 0.966 であった。

表5は分散共分散行列および相関行列に対して求められた主成分負荷量 (主成分得点と項目との相関係数) と、主成分得点と体格・年齢・性などとの相関係数を示したものである。前者は能力特

定値 θ と比較すると、「両足とび」「手つきゴムとび」「体重」を除くすべての項目で相関が低く、等質性(芝, 1972)・基準関連妥当性ともに能力特定値 θ の方がより項目の変動をより反映し、成熟の指標とみなすことができる年齢や体格とも相関が高かった。

表 4 能力特性値と測定項目・諸要因との相関係数

項目	能力特性値 θ
両足とび	0.711**
手つきゴムとび	0.837**
キック的あて	0.206**
フライボールキャッチ	0.465**
ころがしのあて	0.263**
前転	0.417**
横転	0.269**
一回旋とび	0.619**
年齢	0.660**
性別	0.030
身長	0.469**
体重	0.269**

注) **1% 水準で有意

図 5 分散共分散行列からの主成分得点と能力特性値 θ

図 6 相関行列からの主成分得点と能力特性値 θ

表 5 主成分得点と諸要因との相関係数

項目	主成分得点 (相関行列)	主成分得点 (分散共分散行列)
両足とび	0.651**	0.767**
手つきゴムとび	0.809**	0.934**
キック的あて	0.265**	0.120
フライボールキャッチ	0.532**	0.351**
ころがしのあて	0.308**	0.156*
前転	0.539**	0.317**
横転	0.278**	0.152*
一回旋とび	0.630**	0.492**
年齢	0.662**	0.614**
性別	0.006	0.048
身長	0.499**	0.244**
体重	0.290**	0.439**

注) **1% 水準で有意 *5% 水準で有意

(5) 期待値

表6は能力特性値 θ と対応させた各項目の期待値を示している。期待値はカテゴリー特性曲線から各カテゴリーの出現確率を求め、その出現確率が最大となるカテゴリーを選んでいく。したがって、同じ期待値間であっても、出現確率が大である方が、対応する能力特性値 θ においては「尤もらしい(likely)」予測値であることを意味している。この表を用いて、推定された能力特性値 θ に見合った各項目の期待値を求める場合は、表の最左列の θ の値を基準に、右側へ平行線を引き、その線に対応する値が求めるべき期待値とその出現確率になる。ただし、個々の項目の期待値から最左列の能力特性値 θ を推定することを意図したものではないので、右側の項目期待値から左側へ平行線を引き、能力特性値 θ を求めることはできない。

表7は、 $\theta = -1.0$ と能力を推定された者の各項目の期待値を推定する事例を示している。この表から、この者が両足とびをくり返し測定した場合、6回が最も多く測定され、その出現確率は25.4%であることがわかる。以下同様に、手つきゴムとびは10回(出現確率=23.3%)、キック的あては0回(52.0%)、フライボールキャッチは2回(31.9%)、ころがし的あては0回(46.5%)、前転は4点(49.4%)、横転は6点(38.0%)、一回旋とびは6点(39.8%)という測定値が、この者の典型的な測定結果として予測することができる。

IV 考 察

(1) 1次元尺度の適用可能性

図4に示したように、第1主成分の固有値は第2主成分以下と比較しても著しく大であり、かなりの分散を第1主成分で説明することが可能であり、構造的に1次元(等質性)を仮定することが妥当であることがわかる(Hambleton & Swaminathan, 1985; 芝, 1972)。

(2) 困難度パラメタ

ポアソン分布モデルと二項分布モデルは1つの

困難度パラメタが求められ、前者の困難度は0回と1回のカテゴリー曲線の交点に相当する横軸上の値であり、後者は平均的カテゴリーの確率が最大となる横軸上の位置を意味している(豊田, 2005, pp. 185-192)。また、部分採点モデルは2つの隣接するカテゴリー間ごとにステップ困難度が求められる。このように、異なるモデル間では困難度の意味が異なるので比較は不可能で、かつ同一モデル内でもとりうるデータの範囲(カテゴリー数)が異なるので直接比較検討はできない。しかしながら、同一モデル内では、手つきゴムとびの方が両足とびよりも成就回数が多く、容易な項目であり、二項分布モデル内では、捕球の方が、「ころがす」や「蹴る」よりも容易で、的あてでは蹴るよりも手で転がす方が容易であることがわかる。

また、部分採点モデルでは、Samejima (1969, 1997)の段階反応モデルのようにかならずしも単調に困難度は増加しない。それは、該当するカテゴリーへの反応数が少なければ、そのカテゴリーに反応する能力を持つ者は同時に次のカテゴリーに反応してしまうからである。本研究の場合でも、一回旋とびは、全体では困難度は増加傾向にあるが、(-1.090と-1.588)(0.548と0.363)(0.851と0.216)(2.519と0.923)のように隣接するステップ困難度の大小が逆転する場合が交互にみられた。これは測定値が左右方向の合計点で、右方向で可能な角度であれば、反対方向の左でも可能であることが原因と考えられる。

次に、標本をランダムに2分割して求められた困難度パラメタ間の相関係数には高い相関(0.854)が認められた。このことから、どの標本を用いても同様のパラメタを推定することができるので、標本間でも「不変性」(invariance; Lord, 1980, pp. 34-36)があると考えられる。

(3) 能力特性値

求められた能力特性値 θ は年齢と有意な相関を示し、加齢とともに発達していくであろう幼児の運動技能を反映していると考えられる。加えて、しばしば成熟の指標として用いられる身長や体重などの体格(高石ほか, 1981)とも、

表 6 θ から推定される測定値の期待値と出現確率

θ	両足とび		手つきゴムとび		キック的あて		フライボールキャッチ		ころがしのあて		前転		横転		一回旋とび	
	確率	期待値	確率	期待値	確率	期待値	確率	期待値	確率	期待値	確率	期待値	確率	期待値	確率	期待値
-3.0	0.708	4 回	0.677	8 回	0.911	0 回	0.544	0 回	0.895	0 回	0.462	1 点	0.408	5 点	0.432	2 点
-2.8	0.656		0.621		0.892		0.480		0.874		0.393		0.416		0.377	
-2.6	0.598		0.559		0.870		0.414		0.848		0.387		0.412		0.373	
-2.4	0.533		0.491		0.844		0.409		0.819		0.379		0.396		0.356	
-2.2	0.464	5	0.420	9	0.814	0 回	0.406	1	0.784	0 回	0.358	2	0.369	6	0.332	4
-2.0	0.391		0.367		0.778		0.390		0.744		0.347		0.402		0.352	
-1.8	0.364		0.355		0.738		0.360		0.699		0.373		0.427		0.355	
-1.6	0.345		0.325		0.691		0.334		0.648		0.387		0.438		0.337	
-1.4	0.309	6	0.280	10	0.639	1	0.345	2	0.591	1	0.389	3	0.434	7	0.302	6
-1.2	0.270		0.263		0.582		0.340		0.530		0.426		0.414		0.336	
-1.0	0.254		0.233		0.520		0.319		0.465		0.494		0.380		0.398	
-0.8	0.224		0.215		0.455		0.333		0.403		0.559		0.350		0.444	
-0.6	0.204	7	0.193	11	0.404	2	0.345	3	0.410	1	0.619	4	0.366	8	0.469	10
-0.4	0.186		0.174		0.410		0.341		0.404		0.674		0.371		0.469	
-0.2	0.168		0.158		0.402		0.342		0.384		0.723		0.416		0.440	
0.0	0.150		0.142		0.380		0.378		0.351		0.767		0.485		0.387	
0.2	0.138	8	0.129	12	0.345	2	0.401	4	0.338	2	0.804	5	0.551	9	0.314	12
0.4	0.124		0.117		0.340		0.409		0.346		0.836		0.613		0.232	
0.6	0.113		0.106		0.345		0.405		0.337		0.864		0.670		0.236	
0.8	0.102		0.096		0.334		0.450		0.313		0.887		0.721		0.314	
1.0	0.092	9	0.087	13	0.317	3	0.516	5	0.337	3	0.906	6	0.765	10	0.376	15
1.2	0.083		0.078		0.339		0.578		0.346		0.923		0.803		0.411	
1.4	0.075		0.071		0.345		0.636		0.338		0.936		0.836		0.412	
1.6	0.068		0.064		0.335		0.688		0.352		0.947		0.864		0.385	
1.8	0.062	10	0.058	14	0.357	4	0.735	6	0.384	4	0.957	7	0.887	11	0.395	18
2.0	0.056		0.052		0.388		0.776		0.404		0.964		0.907		0.489	
2.2	0.050		0.047		0.406		0.812		0.410		0.971		0.923		0.578	
2.4	0.046		0.043		0.409		0.842		0.402		0.976		0.937		0.657	
2.6	0.041	11	-	15	0.409	5	0.869	7	0.466	5	0.980	8	0.948	12	0.726	20
2.8	-		-		0.476		0.891		0.530		0.984		0.957		0.782	
3.0	-		-		0.540		0.909		0.592		0.987		0.965		0.828	

注) 最左列の θ から表の右側へ期待値を探し、単一の期待値から θ を求めることはできない。

表7 $\theta = -1.0$ から推定される測定値の期待値と出現確率を求めた例

θ	両足とび		手つきゴムとび		キックのあて		フラインボールキャッチ		ころがしのあて		前転		横転		一回旋とび	
	確率	期待値	確率	期待値	確率	期待値	確率	期待値	確率	期待値	確率	期待値	確率	期待値	確率	期待値
-3.0	0.708		0.677		0.911		0.544		0.895		0.462	1点	0.408		0.432	2点
-2.8	0.656		0.621		0.892		0.480	0回	0.874		0.393		0.416		0.377	
-2.6	0.598	4回	0.559	8回	0.870		0.414		0.848		0.387		0.412	5点	0.373	3
-2.4	0.533		0.491		0.844		0.409		0.819		0.379	2	0.396		0.356	
-2.2	0.464		0.420		0.814		0.406		0.784		0.358		0.369		0.332	
-2.0	0.391		0.367		0.778	0回	0.390	1	0.744	0回	0.347		0.402		0.352	4
-1.8	0.364		0.355		0.738		0.360		0.699		0.373	3	0.427		0.355	
-1.6	0.345	5	0.325	9	0.691		0.334		0.648		0.387		0.438	6	0.337	
-1.4	0.309		0.280		0.639		0.345	2	0.591		0.389		0.434		0.302	
-1.2	0.270	6	0.263	10	0.582		0.340		0.530		0.426		0.414		0.336	
-1.0	0.254		0.223		0.520		0.319		0.465		0.494		0.380		0.393	
-0.8	0.224	7	0.215	11	0.455		0.333	3	0.403		0.559		0.350	7	0.444	6
-0.6	0.204		0.193	12	0.404		0.345		0.410		0.619		0.366		0.469	
-0.4	0.186	8	0.174	13	0.410		0.341		0.404	1	0.674		0.371		0.469	
-0.2	0.168	9	0.158	14	0.402	1	0.342		0.384		0.723		0.416		0.440	
0.0	0.150	10	0.142	15	0.380		0.378	4	0.351		0.767		0.485		0.387	
0.2	0.138	12	0.129	17	0.345		0.401		0.338		0.804		0.551		0.314	
0.4	0.124	14	0.117	19	0.340		0.409		0.346	2	0.836		0.613		0.232	
0.6	0.113	16	0.106	22	0.345	2	0.405		0.337		0.864		0.670		0.236	
0.8	0.102	19	0.096	25	0.334		0.450		0.313		0.887	4	0.721		0.314	10
1.0	0.092	22	0.087	29	0.317		0.516		0.337	3	0.906		0.765		0.376	
1.2	0.083	27	0.078	34	0.339	3	0.578		0.346		0.923		0.803	8	0.411	
1.4	0.075	32	0.071	39	0.345		0.636		0.338		0.936		0.836		0.412	
1.6	0.068	38	0.064	46	0.335		0.688		0.352		0.947		0.864		0.385	
1.8	0.062	45	0.058	-	0.357		0.735	5	0.384		0.957		0.887		0.395	
2.0	0.056	-	0.052	-	0.388	4	0.776		0.404	4	0.964		0.907		0.489	12
2.2	0.050	-	0.047	-	0.406		0.812		0.410		0.971		0.923		0.578	
2.4	0.046	-	0.043	-	0.409		0.842		0.402		0.976		0.937		0.657	
2.6	0.041	-	-	-	0.409		0.869		0.466		0.980		0.948		0.726	
2.8	-	-	-	-	0.476	5	0.891		0.530	5	0.984		0.957		0.782	
3.0	-	-	-	-	0.540		0.909		0.592		0.987		0.965		0.828	

注) 最左列の θ から表の右側へ期待値を探索。単一の期待値から θ を求めることはできない。

各々有意な相関を示し、成熟とともに発達する運動技能を表現していると思われる。しかし、性別とはほとんど関連を示さず、性差はみられなかった。これらは「幼児の前半の運動技能 (motor performance) の経年的発達は著しいが、性差は小さい」という Malina & Bouchard (1991) の見解と一致する。

また、各測定項目との相関係数もいずれとも有意な関連を示し、推定に用いられたすべての項目の特性を反映した推定値であると考えられる。

(4) 主成分得点との比較

相関行列と分散共分散行列からの両主成分得点ともに能力特性値 θ との相関は非常に高かったが、幾分、相関行列を対象に求められた方が高かった。分布については分散共分散行列から求められた主成分得点の方が曲線状で、能力特性値の高い、あるいは低い水準の者に関しては、主成分得点は高めに推定し、中間的な者の能力に関しては比較的lowめに推定していることになる。ただし、それらのバラツキは比較的一様であった。

それに対して、相関行列から求められた主成分得点はやや直線的で、能力水準の高い部分ではバラツキは少ない (関連は高い) が、能力水準の低い部分では比較的バラツキがみられた (関連は低かった)。これは能力水準の低い者は得る測定値の値が低く (例えば、0 点、1 点、2 点など)、連続量および正規分布を前提とした主成分分析の結果との乖離がみられ、逆に能力水準の高い者は測定値のとるべき値が大 (例えば、10 点、11 点、12 点など) で、比較的連続量に近似した値となっていることが原因と考えられる。これは、主成分得点は、例えば、「0 点と 2 点の差である 2 点」と「10 点と 12 点の差の 2 点」と同じ能力差に勘案して求められているのに対して、今回、項目応答理論で求められた能力測定値 θ は必ずしもそのような制約がないことの違いであろう。

主成分負荷量からみると、分散共分散行列から求められた主成分得点は能力特定値 θ と比較すると、「両足とび」「手つきゴムとび」「体重」を除くすべての項目で相関が低く、等質性 (芝、

1972)・基準関連妥当性とともに能力特定値 θ の方がより項目の変動をより反映し、成熟の指標とみなすことができる年齢や体格とも相関が高く、それらを反映した能力得点と考えることができる。しかし、標準化した相関行列を対象に求めた主成分得点に関しては逆の結果であり、「両足とび」「手つきゴムとび」を除くすべての項目で関連は大であった。当然のことながら、最もすべての項目との相関が高くなるように主成分得点は求められているので、主成分負荷量は高くなるであろうが、実際にも能力特性値 θ よりも全体的に幾分相関は高かった。また、年齢・性や体格などの相関もほぼ同様の値であるが、主成分得点の方が能力特性値 θ よりも多少高い値であった。しかし、その中でも両足とびと手つきゴムとびは能力特性値 θ の方が相関は高く、この項目に関しては能力特性値 θ の方がその特性をより反映していると考えられる。

(5) 期待値

能力特性値 θ が求められれば、それに応じた期待値がわかるので、実際の個人の測定値がその期待値以上であれば予想以上にその項目が得意であることがわかり、逆に、その期待値以下であれば、何らかの原因で期待値まで達しなかったことになる。特に、その出現確率が大きい場合にもかかわらずそのような事態が起こった場合は、幼児に関しては運動経験不足や適切な運動課題が提供されていないなどの問題が考えられ、幼稚園での指導カリキュラムで個人的にその点を補う必要があるであろう。さらに、特定の運動技能や身体部位の深刻な発達遅滞を発見する発達スクリーニングテスト (Frankenburg & Dodds, 1967; Frankenburg et al., 1971) として用いることも可能であろう。

V ま と め

スキルテストには、①一定試行回数内での成功回数、②一定時間内に成就した回数、③課題動作フォームの主観的評価による方法などがある。これらのスキルテストに項目応答理論を適用しよう

とすると、①については Andrich の二項分布モデル、②はラッシュのポアソン分布モデル、そして③については Masters の部分採点モデルを用いることができる。そこで、本研究では、これらの3モデルを組み合わせ、幼児のスキルテストに適用し、運動技能の推定を行った。対象は幼稚園児205名、測定項目は①両足とび②手つきゴムとび③キックのあて④フライボールキャッチ⑤ころがしのあて⑥前転⑦横転⑧一回旋とびの8項目である。そして、求められた能力特性値を検討した結果、次のような知見を得た。

1) 対象となった標本は等質性が高く、1次元尺度を作成することが可能であり、求められた困難度パラメタは、無作為に2分割された標本から別々に求められた困難度パラメタの相関は高く、標本に依存しない項目パラメタであると考えられた。

2) 求められた θ は、年齢、身長、体重と有意な相関を示し、加齢や成熟とともに向上していると考えられる幼児期の運動技能を測定していると考えられた。

3) 主成分分析によって得られた主成分得点と比較すると、お互いに高い相関を示したが、項目間および年齢や体格との相関はいずれも能力特性値 θ の方がやや低かった。しかし、高い項目も2項目認められた。

最後に、能力特性値 θ に対応する各項目の出現確率と期待値を求め、その活用例を示した。

注

注1) 人間が実際に100点満点で評定を行ったとしても、100点と90点の差が、60点と50点の差と同じ能力の差を反映しているという明確な保証を見出すのは困難である。

注2) 分析方法やモデルにかかわらず一般的な個人の能力推定値は「能力得点」、項目応答理論上での能力の推定値は「能力特性値」と区別して用いている。同様に、一般性のある概念として「難易度」、項目応答理論のパラメタ名としては「困難度」を用いることにする。

文 献

- Andrich, D. (1978) Scaling Attitude Items Constructed and Scored in the Likert Tradition. *Educational and Psychological Measurement* 38: 665-680.
- 青柳 領・松浦義行 (1982) 幼児の運動能力構造について. *体育学研究* 26(4): 291-303.
- 青柳 領・松浦義行・出村慎一・M. アンワール・パサウ・服部隆・田中喜代次 (1980) 幼児の平衡運動に關与する調整力の因子分析的研究—妥当な項目の選択について—. *体育学研究* 25(3): 197-206.
- 青柳 領 (2007) 成功回数によるスキルテストへの項目応答理論の適用: バスケットボールシュートの事例. *体育学研究* 52: 259-271.
- Baumgartner, T.A. and Jackson, A.S. (1982) *Measurement for Evaluation in Physical Education*. Wm. C. Company Publishers: Dubuque, pp. 311-365.
- Clarke, H.H. (1967) *Application of Measurement to Health and Physical Education*, Fourth Edition. Prentice-Hall, Inc.: Englewood, pp. 300-337.
- Collins, D. R. and Hodges, P. B. (1978) *A Comprehensive Guide to Sports Skills Tests and Measurement*. Charles C. Thomas Publishers: Springfield.
- Embretson, S. E. and Reise, S.P. (2000) *Item Response Theory for Psychologists*. Lawrence Erlbaum Associates, Publishers: New Jersey. P. 124.
- Frankenburg, W. K. and Dodds, J. B. (1967) The Denver Developmental Screening Test. *The Journal of Pediatrics* 71(2): 181-191.
- Frankenburg, W. K., Camp, B. W., van Natta, P. A., Demersseman, J. A. and Voorhees, S. F. (1971) Reliability and Stability of The Denver Developmental Screening Test. *Child Development* 42(5): 1315-1325.
- Hambleton, R. K. and Swaminathan, H. (1985) *Item Response Theory: Principles and Applications*. Kluwer Nijhoff Publishing: Boston, pp. 306-307.
- Hambleton, R. K., Swaminathan, H. and Rogers, H. J. (1991) *Fundamentals of Item Response Theory*. Sage Publications: London, pp. 12-14.
- 肥田野直・瀬谷正敏・大川信明 (1961) *心理教育統計学*. 培風館: 東京.
- Hottelling, H. (1933a) Analysis of A Complex of Statistical Variables into Principle Component (Section one to seven).

- Journal of Educational Psychology, 24-6: 417-441.
- Hottelling, H. (1933b) Analysis of A Complex of Statistical Variables into Principle Component (Section eight to twelve). Journal of Educational Psychology, 24-7: 498-520.
- 池田 央 (1976) 社会科学・行動科学のための数学入門 2 統計的方法 I 基礎. 新曜社: 東京. pp. 9-16.
- 池田 央 (1994) 現代テスト理論. 朝倉書店: 東京. pp. 166-170.
- 勝部篤美 (1971) 幼児体育の理論と実際. 杏林書院: 東京, pp. 40-47.
- Larson, L. A. and Yocom, R. D. (1951) Measurement and Evaluation in Physical, Health, and Recreation Education. The C. V. Mosby Company: St. Louis.
- Lord, F. M. (1980) Applications of Item Response Theory to Practical Testing Problems. Lawrence Erlbaum Associated, Publishers: New Jersey. pp. 179-192.
- Malina, R. M. and Bouchard, C. (1991) Growth, Maturation, and Physical Activity. Human Kinetics: Champaign, pp. 187-204.
- Masters, G. N. (1982) A Rasch Model for Partial Credit Scoring. Psychometrika 47: 149-174.
- Masters, G. N. and Wright, B. D. (1997) The Partial Credit Model. In: van der Linden, W. J. and Hambleton, R. K. (Eds.) Handbook of Modern Item Response Theory. Springer: New York, pp. 101-121.
- 松浦義行 (1969) 運動能力の因子構造. 不昧堂書店: 東京. pp. 24-105.
- 松浦義行・中村栄太郎 (1977) 基礎運動能力の発達に関する研究—4~8歳の男児について—. 体育学研究 21(5): 293-303.
- Morrow, J. R., Jackson, A. W., Disch, J. G. and Mood, D. P. (2000) Measurement and Evaluation in Human Performance, 2nd Edition. Human Kinetics: Champaign. pp. 307-319.
- 中村栄太郎・松浦義行 (1979) 4~8歳の幼児・児童の基礎運動能力の発達に関する研究. 体育学研究 24(2): 127-135.
- ラッシュ: 内田良男ほか訳 (1985) 心理テストの確率モデル. 名古屋大学出版会: 名古屋. <Rasch, G. (1980) Probabilistic Models for Some Intelligence and Attainment Tests. The University of Chicago Press: Chicago.>
- Samejima, F. (1969) Estimation of Latent Ability using a Response Pattern of Graded Scores. Psychometrika Monograph 17: 1-100.
- Samejima, F. (1997) Graded Response Theory. In: van der Linden, W. J. and Hambleton, R. K. (Eds.) Handbook of Modern Item Response Theory. Springer: New York, pp. 85-100.
- 芝 祐順 (1972) 項目分析. 肥田野直編 心理学研究法 第7巻テスト I. 東京大学出版会: 東京, pp. 53-91.
- 芝 祐順 (1975) 行動科学における相関分析法 (第2版). 東京大学出版会: 東京, pp. 48-63.
- 芝 祐順 (1991) 項目反応理論 基礎と応用. 東京大学出版会: 東京 p. 33-50.
- 体育科学センター調整力専門委員会 (1976) 調整力テスト実施要領およびその基準値. 体育科学, 4: 207-217.
- 体育科学センター調整力専門委員会 (1981) 体育科学センター調整力フィールドテストの最終形式—調整力テスト検討委員会報告—. 体育科学, 9: 207-212.
- 高石昌弘・樋口 満・小島武次 (1981) からだの発達. 大修館書店: 東京, pp. 35-36.
- 豊田秀樹 (2005) 項目反応理論 理論編—テストの数理—. 朝倉書店: 東京.
- Zhu, W. and Cole, E. (1996) Many-Faceted Rasch Calibration of A Gross Motor Instrument. Research Quarterly for Exercise and Sport, 67(1): 24-34.

(平成 20 年 4 月 7 日受付)
(平成 20 年 9 月 17 日受理)