

論文

D指標を用いた項目分析のためのテスト条件の検討

——より真値に近い状態を目指して

坪田 彩乃*, 石井 秀宗*

* 名古屋大学大学院教育発達科学研究科

Examination of testing conditions to obtain good D-index estimate

Ayano Tsubota*, Hidetoki Ishii*

* Graduate School of Education and Human Development School of Education, Nagoya University

It is necessary to examine quality of tests even when non-test specialists administrate them. A simple measure of item discrimination is the D-index. To calculate D-value, examinees are divided into three groups, and it is obtained by subtracting the rate of correct answer of lower group from that of upper group. It is easier for non-test specialists to use D-index in their education fields. Then, this study investigated adequate ordering methods of test takers and suitable testing conditions to obtain more precise D estimate and more accurate order of examinees. First, several simulation studies were conducted. As the results, it was found that ordering based on item discriminations is better than ordering based on item difficulties. Test takers who take the same total score should be ordered by one or more items which have higher discrimination powers. The desirable test condition is the following; more than 30 items, high item discriminations and moderate item difficulties. Numerical example was also presented and the adequacy of these results were confirmed.

Keywords : D-index, item analysis, classroom test, item discrimination

キーワード : D指標, 項目分析, クラスルームテスト, 項目識別力

* 〒464-8601 名古屋市千種区不老町 名古屋大学教育発達科学研究科

Correspondence concerning this article should be sent to: Ayano Tsubota, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, Japan 464-8601.

E-mail: ayano.melon@gmail.com

1. はじめに

テストとは、教育成果の実態について目に見える形での情報を与えてくれる唯一の道具である(池田,2006)。この道具が適切でない場合、様々な問題が生じる。たとえば、受検者の能力を引き出すのに十分ではない試験問題であったとき、受験者の能力の実際が反映されない評価をつける可能性がある(池田, 1992)。こうした事態を避けるためにも、そのテストが適切なものであったかという評価を行う必要がある。つまり、テストを評価するということは、そのテストがテストの目的に則った測定道具であったかを確認する作業ともいえる。これには、一貫した測定が可能かという信頼性や、狙った能力を測定しているかという妥当性に留まらず、そのテストが測定したい学力について、その学力の多寡により正誤が反映されているかどうかという点も含まれる。

個人を対象として行うテストには、全国的に一斉に行うようなもの(e.g.大学入学共通テスト、全国一斉学力調査)や、各教員が個別に行うクラスルームテストがある(日本テスト学会, 2007)。これらは、受検者集団や作題者など、様々な点で異なる。全国的に行われるテストの多くは、テストの専門家により評価が行われている。一方で、クラスルームで行われるようなテストでは、授業実施者がテストの作成・実施をし、受検者にフィードバックを与えるまでの一連の作業を行う。しかし、こうしたテストの評価はほとんどなされていない。

テストを評価するための特性のひとつに、識別力がある。識別力とは、その項目がどの程度受検者の能力に基づいて正誤を判別できるかどうかというものである。そして、識別力指標とは、能力の低い受検者について正答率が低く、能力が高くなるほど正答率が上がる傾向を示す指標である(e.g.石井, 2007)。識別力は-1 から+1 の数値をとる。+1 のとき、識別力が最大であり、能力の高い受検者は正答するが、能力の低い受検者は誤答することを表す。0 は識別力がなしとして、受検者の能力によらず正答率が同じことを意味する。-1 では負の識別、すなわち、能力低群は正答するが能力高群では誤答することを意味する。

Diederich(1973)は、クラスルームテストでは非専門家がテストの評価を行うことを踏まえ、簡便なアプローチが望ましいとし、項目の識別力を評価する方法を提案している。これは、受検者を上位群と下位群の2群に分け、上位群の正答者数と下位群の正答者数の差を取るものである。また、Brown(1996)は言語テストにおける項目の識別力として、上位群と下位群の正答率の差を用いることを提案している。実際に、Wiyasa, Laksana, & Indrawati. (2019) や Laliamsyah & Apriyanti (2020)などで、テストの項目分析としてこの方法に則り、項目識別力の算出が行われている。

こうした項目の識別力の指標の一つに、*D* 指標(*D-index*)がある。これは、テストの受検者について能力値を上位群・中位群・下位群と分けた際の、上位群と下位群の正答率についての差をとったものである。上位群と下位群の差をとるという考え方は、Johnson(1951) で紹介されており、Ebel & Frisbie(1991) はこれが初出であるという。*D* 指標は簡便な識別力指標のひとつであるが、これは項目の正答率の影響を強く受ける(石井, 2007)。正答率が高い項目や低い項目では、十分に機能をしないという問題点がある。そのため、ある項目における各受検者の正誤情報と、各受検者の合計得点の相関を取ることで識別力の指標とする *I-T* 相関係数が考えられている。*I-T* 相関の算出については、石井(2020)などのソフトウェアを用いることにより、その値を得ることが可能である。しかし、*I-T* 相関を理解するためには点双列相関係数の理解が必要となり、かつ、相関関係を識別力として捉え直す必要がある。このように、*I-T* 相関そのものの概念的な理解が難しいことから、テストの非専門家が実施するテストにおいて扱うことは難しいとされる(Brown, 1996)。一方 *D* 指標は、上位群と下位群の正答率の差を取るのみで算出が可能であり、その数値も理解しやすい。このことから、特にテストの非専門家が行うテストにおいては、*D* 指標は有用な指標となり得ると考えられる。また、受験者能力群を横軸に、それぞれの群における正答率を縦軸にとり図示化する方法として、トレースラインがある。*D* 指標は、このトレースラインの上位群と下位群の高低差を数値化するものであり、視覚的情報との対応も容易である。

さて、*D* 指標を用いるためには、受検者を 3 群に分ける必要があるが、それぞれの群の割合は定められていない。たとえば、赤根・伊藤・林・椎名・大澤・柳井・田栗(2006)では、上位群/下位群を 25%ずつとした群わけを行っており、医学系試験の項目分析では多く用いられる割合だという。また、野口・大隅(2014)では、上位群・中位群・下位群の人数を揃えた例を紹介している。Kelly(1939) は、この割合について、上位群を 27%、下位群を 27%と提案している。これは、変数 X が正規分布するときの標本平均の差の臨界比を最大とする値に由来している。

しかしながら、*D* 指標を用いる際に行う群わけの方法については、十分な検討がなされていない。テストで同点だった受検者について、同じ群とするか異なる群とするかによっても *D* 指標は異なる値となる。また、同点だった受検者について、Kelly(1939)のような割合になるように異なる群に割り振るのであれば、受検者を分ける方法は多様に考えられる。このとき、テスト全体の *D* 指標は群わけの方法によらず同じ値を示すが、テストに含まれる項目の *D* 指

D 指標を用いた項目分析のためのテスト条件の検討

標は群わけの影響を受けることになる。適切な群わけが行われていなければ、そこで表される D 指標は群の分け方という誤差を含んだものとなる。同点の受検者を減らすために、テストの項目数を増やすことも考えられるが、学校現場などで実際に用いられるテストではテスト時間に制限があったり、作問する教員への負担になったりするなど限界がある。そのため、識別力の高い項目のみで構成されたテストであっても、同点の受検者をゼロにすることは容易ではない。

加えて学校現場においては、クラスルームテストに基づいて児童・生徒の評価が行われる。すなわち、不適切な群わけにより行われた評価は、その評価そのものに意味がなくなるばかりか、児童・生徒自身の処遇への影響を及ぼすという観点から望ましくない。梶谷・小林・鈴木・中田・盛本(2013)は、学生に成績順位を通知した際に、成績上位者と成績下位者では次の試験に対する効果が異なっていたという。特に、上位の成績と通知された学生は、次回の試験に対して油断するという油断効果が見られた。つまり、試験の成績そのものよりも、自分が相対的にどの群に入るかどうかによって、次の試験への動機づけが異なるという。また、ピグマリオン効果に代表されるように、教師自身の期待が生徒の学力へ影響を及ぼす要因になりうる(東, 2001)。教育現場で行うテストでは、受検者個人がどの群に入ったかという参照が可能である。そのため、不適切な群わけにより、より下位の群に割り振られた場合、それによって不利益をもたらす可能性も否めない。このように、受検者を群わけすることの効果は、クラスルームテストにおいて教育指導場面にまで及ぶと考えられる。

これらのことを踏まえ本研究では、より真値を反映した生徒の群わけの方法と望ましいテストの条件について、 D 指標という側面から検討を行う。

具体的には、群わけの方法のひとつとして考えられる、同点を同じ群にした際のテストの信頼性ならびに生徒の順位づけについて評価した後、同点を異なる群にするための並び替えの方法について望ましいテスト条件と併せて検討する。その上で、実データを用いたシミュレーションにより、並び替え方法の有用性について検討する。

2. 研究 1: 同じ得点の受検者を同じ群とした条件についての検討

2. 1. 目的

群を分ける方法として、同点の受検者を同じ群とすることが考えられる(e.g. 安永・斎藤・石井, 2010; 安永・石井, 2012)。同点を同じ群とすることで、上位群・中位群・下位群の比率は結果的に Kelly(1939)と異なるが、得点と同じ

者が同じ群であることは直感的にわかりやすいと考えられる。ここでは、同点の受検者を同じ群とした際の D 指標への影響を、受検者数と項目数の観点から検討する。

2. 2. 方法

(1) 群わけの方法

同点を同じ群として3群とするために考えられる以下4パターンで群わけを行った。

パターン1. 上位群と中位群ならびに中位群と下位群のカットオフ値の受験者を全て中位群とする方法

パターン2. 上位群と中位群の間のカットオフ値の受験者は中位群とし、中位群と下位群の間のカットオフ値の受験者は下位群とする方法

パターン3. 上位群と中位群のカットオフ値の受験者は上位群とし、中位群と下位群の間のカットオフ値の受験者は中位群とする方法

パターン4. 上位群と中位群の間のカットオフ値の受験者は上位群とし、中位群と下位群の間のカットオフ値の受験者は下位群とする方法

(2) I-T 相関ならびに正答率の真値

実際に調査等で行われたテストの I-T 相関係数ならびに正答率が公表されている Azeem(2012), Yu-mien(2010), 寺尾・安永・石井・野口(2015)を参考に、正答率ならびに I-T 相関の真値を、ベータ分布を用いて発生させた。

ベータ分布は 2 つの自由度の組み合わせにより分布が変化する確率分布である。 $\beta(\alpha, \beta)$ のとき、

$$E(X) = \frac{\alpha}{(\alpha + \beta)}$$

$$V(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

となることが知られている。

正答率ならびに I-T 相関係数について、具体的にはそれぞれ、以下のベータ分布を用いた。

正答率は、一様分布となる $\beta(1,1)$ ($M=0.5, SD=0.289$)、低い正答率で固まる場合 $\beta(2,6)$ ($M=0.25, SD=0.144$)、中程度の正答率で固まる場合 $\beta(6,6)$ ($M=0.5, SD=0.139$)、高い正答率で固まる場合 $\beta(6,2)$ ($M=0.25, SD=0.144$) の4パターンについて検討した。

D 指標を用いた項目分析のためのテスト条件の検討

I-T 相関は、低い識別力の場合 $\beta(1,9)$ ($M=0.1, SD=0.090$), 中程度の識別力の場合 $\beta(4,12)$ ($M=0.25, SD=0.105$), 高い識別力の場合 $\beta(10,15)$ ($M=0.4, SD=0.096$) の3パターンについて検討した。

これらの組み合わせ計 12 パタンについて検討を行った。

(4) テストの条件

テストの条件として、受験者数と項目数を設定した。受検者数は $N=10, 40, 70, 100, 200$ の5パターン、項目数は $m=10, 20, 30, 40, 50$ の5パターンとした。これらの組み合わせ 25 パタンについて検討を行った。

(5) シミュレーションの手続き

以下の手続きによりシミュレーションを行った。

1. 項目数の正答率ならびに識別力の真値を発生させ、これらを 2 Parameter Logistic Model の識別力と困難度に対応するように変換を行った。
2. 受験者数の潜在特性真値を $N(0,1)$ から発生させ、1 で求めた項目特性値を用い、各項目の正答確率を求めた。
3. 受検者の一様乱数を発生させ、2 で求めた正答確率を用いて、正誤パターンを作成した。
4. 合計得点順に並び替え、パターン 1~4 の方法で群わけを行った。
5. 群わけに基づき、真値との比較として、重み付けカッパ係数ならびに ΔD を算出した。

以上の手続きを全ての受験者数・項目数・正答率・I-T 相関の組み合わせについて 100 回ずつ行った。

(6) 変換の方法

Lord & Novick(1968), 松井(1991) を参考に、項目 j のときの I-T 相関係数 ρ_j と、標準正規分布の上側確率が正答率となる z 値 γ_j を用いて、以下のように変換を行った。

$$a_j = \frac{\rho_j}{\sqrt{1 - \rho_j}}$$

$$b_j = \frac{\gamma_j}{\rho_j}$$

(7) 用いる指標

真値に基づいた群わけと、提案手法による群わけの比較をするために、重み付けカッパ係数、 ΔD という2つの指標を用いた。

重み付けカッパ係数は、真値の群わけと提案手法の群わけがどの程度一致して

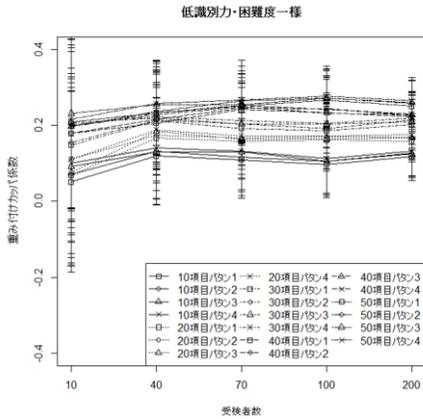


Figure1. 低識別力・困難度一様の重みづけカッパ係数

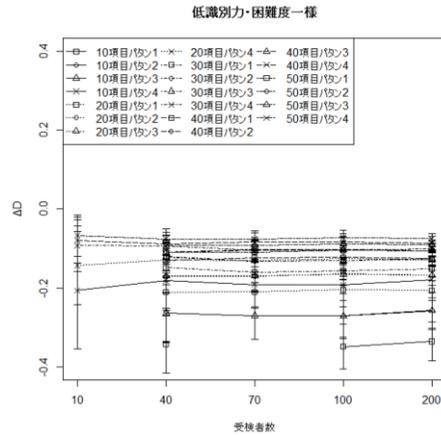


Figure2. 低識別力・困難度一様の ΔD

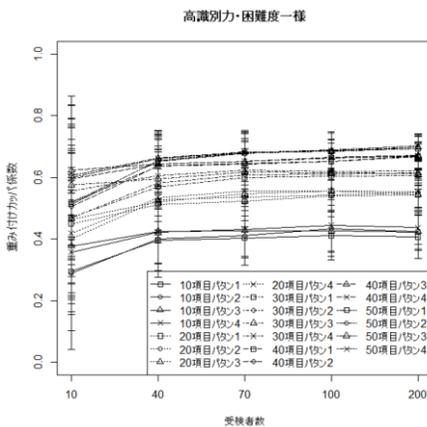


Figure3. 高識別力・困難度一様の重みづけカッパ係数

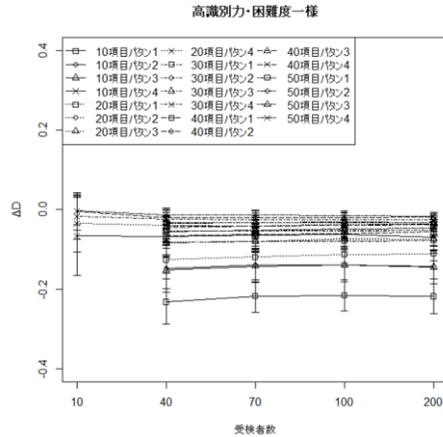


Figure4. 高識別力・困難度一様の ΔD

いるかどうかを表す指標である。

ΔD は、真値の群わけによる D 指標と提案手法の群わけによる D 指標の差であり、両者がどの程度一致しているかの指標である。群わけの方法によって、テストの識別力の差を表すものである。

2. 3. 結果と考察

重みづけカッパ係数ならびに ΔD の結果の一部を Figure1~4 に示した。Figure1,2 は識別力が低く困難度が一樣である条件，Figure3,4 は識別力が高く困難度が一樣である条件である。

D 指標を用いた項目分析のためのテスト条件の検討

群わけを行ったパタンのうち、パタン 4 以外の方法では上位群もしくは下位群の該当者がおらず、そもそも D 指標の算出が不可能であるケースが見られた。パタン 1~3 ではカットオフ値上の受検者を中位群とするため、カットオフ値以上/以下の受検者が居ないとき、上位群/下位群に入る受検者が存在しないということになる。

また、重み付けカッパ係数の値を上げ、 ΔD を小さくする要因として、人数はほとんど影響しないと考えられる。図を確認すると、受験者数が 40 名以上ではグラフの傾きがほぼ横ばいとなった。すなわち、1 クラスの生徒数だと考えられる 40 名程度の受験者数がテストを評価するための最低単位だと考えられる。

一方、真値による群わけに近づけるような、重み付けカッパ係数の値が高くなり、 ΔD の値が小さくなる要因として、テストの識別力条件と項目数が大きく影響すると考えられる。テスト項目の識別力が高く、かつ項目数が多ければ多いほど、重み付けカッパ係数の値は大きくなり、 ΔD の値は小さくなった。しかしながら、クラスルームテストでの項目数を考えたとき、50 項目を超えるテストは少ない。更に、本シミュレーションの識別力条件は、実際のテストで算出される識別力よりも概して大きいと考えられる。

ΔD は、識別力の条件や項目数による影響が小さい。重み付けカッパ係数と合わせて考えると、識別力が低いテストでは、重み付けカッパ係数が低いことから、受検者の群わけがランダムに近いものの、そもそも真値による D 指標自体が小さい値であることで真値との差が小さくなったと考えられる。対して、識別力が高いテストでは、識別力が低いテストよりも重み付けカッパ係数が高いことから、より真値に近い群わけになり、真値との差が小さくなったと考えられる。

このように、 ΔD を 0 に近づけること以上に、重み付けカッパ係数の値を上げるための条件を検討する必要がある。合計得点と同じ受検者を異なる群に分けることで、重み付けカッパ係数を上げることが可能であるかを検討する必要があるだろう。

3. 研究 2: 同じ得点の受検者を異なる群とする並び替え方法についての検討

3. 1. 目的

研究 2 では、上位群と中位群、中位群と下位群のカットオフ値上にいる同じ得点の受検者について、Kelly(1939)の用いた割合になるよう、異なる群に割り

当てる方法について検討する。その上で、より真値に近いテストとなる条件についても検討する。

3. 2. 方法

(1) 並び変える方法

まず、合計得点を降順に並び替える。合計得点が同点だった場合の受検者を並び替える方法として、ID(受検者番号)、I-T 相関、正答率を用いた以下の 9 パタンについて検討した。

Type1. ID の大小について、ID 番号順に並び替える。このとき、ID はランダムに振り分けられている番号である。

Type2.1. 項目内で最も I-T 相関が高かった項目への正誤について降順に並び替える

Type2.2. 項目内で最も I-T 相関が高かった項目への正誤、次に I-T 相関が高かった項目への正誤について降順に並び替える

Type2.3. 項目内で最も I-T 相関が高かった項目への正誤、次に I-T 相関が高かった項目への正誤、さらにその次に I-T 相関が高かった項目への正誤について降順に並び替える

Type2.4. 受検者が正答した項目のうち、I-T 相関がより高い項目に正答していた受検者を上位として降順に並び替える

Type3.1. 正答率が 0.5 に最も近い項目への正誤について降順に並び替える

Type3.2. 正答率が上位群と中位群のカットオフ値に最も近い項目への正誤、中位群と下位群のカットオフ値に最も近い項目への正誤について降順に並び替える

Type3.3. 正答率が 0.5 に最も近い項目への正誤、上位群と中位群のカットオフ値に最も近い項目への正誤、中位群と下位群のカットオフ値に最も近い項目への正誤について降順に並び替える

Type3.4. 受検者が正答した項目のうち、正答率がより高い項目に正答していた受検者を上位として降順に並び替える小見出し

(2) I-T 相関ならびに正答率の真値

設定は研究 1 で用いたものと同じである。

D 指標を用いた項目分析のためのテスト条件の検討

Table1 識別力が低く困難度が一樣のときの各指標(N=40)

項目数	ΔD		type1	type2.1	type2.2	type2.3	type2.4	type3.1	type3.2	type3.3	type3.4
10	-0.254 (0.067)	r	0.186 (0.170)	0.185 (0.175)	0.185 (0.177)	0.185 (0.174)	0.189 (0.173)	0.189 (0.176)	0.189 (0.174)	0.189 (0.174)	0.187 (0.167)
		k	0.107 (0.133)	0.106 (0.139)	0.112 (0.139)	0.110 (0.141)	0.112 (0.134)	0.107 (0.136)	0.105 (0.132)	0.107 (0.130)	0.107 (0.133)
30	-0.124 (0.030)	r	0.338 (0.139)	0.339 (0.143)	0.339 (0.143)	0.341 (0.143)	0.347 (0.141)	0.347 (0.144)	0.336 (0.144)	0.336 (0.143)	0.334 (0.139)
		k	0.216 (0.119)	0.219 (0.122)	0.222 (0.119)	0.223 (0.119)	0.221 (0.113)	0.216 (0.119)	0.215 (0.123)	0.215 (0.121)	0.220 (0.113)
50	-0.085 (0.022)	r	0.409 (0.138)	0.410 (0.137)	0.412 (0.137)	0.412 (0.138)	0.417 (0.135)	0.415 (0.139)	0.415 (0.137)	0.415 (0.137)	0.418 (0.133)
		k	0.255 (0.136)	0.257 (0.135)	0.263 (0.133)	0.265 (0.133)	0.265 (0.128)	0.255 (0.136)	0.253 (0.134)	0.258 (0.132)	0.257 (0.140)

()内は標準偏差を表す

Table2 識別力が低く困難度も低いときの各指標(N=40)

項目数	ΔD		type1	type2.1	type2.2	type2.3	type2.4	type3.1	type3.2	type3.3	type3.4
10	-0.264 (0.062)	r	0.172 (0.190)	0.173 (0.186)	0.169 (0.186)	0.170 (0.186)	0.172 (0.187)	0.175 (0.186)	0.170 (0.185)	0.172 (0.185)	0.170 (0.183)
		k	0.101 (0.130)	0.105 (0.132)	0.095 (0.130)	0.099 (0.135)	0.099 (0.136)	0.109 (0.134)	0.094 (0.132)	0.100 (0.138)	0.103 (0.139)
30	-0.123 (0.038)	r	0.351 (0.157)	0.351 (0.160)	0.352 (0.159)	0.351 (0.160)	0.349 (0.166)	0.348 (0.158)	0.348 (0.164)	0.348 (0.161)	0.345 (0.165)
		k	0.221 (0.131)	0.221 (0.132)	0.223 (0.135)	0.218 (0.132)	0.221 (0.133)	0.219 (0.140)	0.219 (0.134)	0.220 (0.139)	0.215 (0.142)
50	-0.086 (0.022)	r	0.437 (0.135)	0.437 (0.136)	0.438 (0.137)	0.438 (0.137)	0.441 (0.133)	0.438 (0.135)	0.437 (0.134)	0.439 (0.134)	0.437 (0.132)
		k	0.257 (0.117)	0.255 (0.118)	0.261 (0.121)	0.262 (0.120)	0.263 (0.125)	0.262 (0.126)	0.258 (0.124)	0.264 (0.125)	0.264 (0.123)

()内は標準偏差を表す

Table3 識別力が低く困難度が中程度のときの各指標(N=40)

項目数	ΔD		type1	type2.1	type2.2	type2.3	type2.4	type3.1	type3.2	type3.3	type3.4
10	-0.288 (0.071)	r	0.238 (0.148)	0.237 (0.146)	0.238 (0.148)	0.236 (0.148)	0.233 (0.147)	0.238 (0.146)	0.237 (0.142)	0.238 (0.142)	0.238 (0.139)
		k	0.139 (0.129)	0.131 (0.129)	0.133 (0.127)	0.134 (0.125)	0.143 (0.126)	0.139 (0.135)	0.141 (0.127)	0.141 (0.129)	0.142 (0.131)
30	-0.130 (0.042)	r	0.413 (0.158)	0.415 (0.158)	0.417 (0.159)	0.419 (0.158)	0.418 (0.157)	0.414 (0.155)	0.415 (0.157)	0.416 (0.156)	0.409 (0.155)
		k	0.263 (0.141)	0.260 (0.143)	0.262 (0.146)	0.263 (0.148)	0.259 (0.142)	0.260 (0.139)	0.256 (0.146)	0.257 (0.142)	0.255 (0.150)
50	-0.095 (0.024)	r	0.474 (0.126)	0.476 (0.126)	0.477 (0.125)	0.477 (0.124)	0.478 (0.124)	0.473 (0.126)	0.472 (0.123)	0.472 (0.125)	0.472 (0.126)
		k	0.293 (0.123)	0.296 (0.121)	0.295 (0.119)	0.297 (0.120)	0.293 (0.119)	0.289 (0.121)	0.283 (0.120)	0.289 (0.122)	0.287 (0.122)

()内は標準偏差を表す

Table4 識別力が低く困難度が高いときの各指標(N=40)

項目数	ΔD		type1	type2.1	type2.2	type2.3	type2.4	type3.1	type3.2	type3.3	type3.4
10	-0.252 (0.066)	r	0.187 (0.153)	0.191 (0.158)	0.192 (0.159)	0.192 (0.159)	0.191 (0.159)	0.191 (0.158)	0.184 (0.147)	0.187 (0.156)	0.185 (0.152)
		k	0.115 (0.132)	0.112 (0.129)	0.112 (0.129)	0.107 (0.132)	0.104 (0.126)	0.108 (0.133)	0.108 (0.129)	0.100 (0.138)	0.107 (0.128)
30	-0.124 (0.036)	r	0.327 (0.151)	0.328 (0.152)	0.330 (0.154)	0.333 (0.153)	0.336 (0.152)	0.330 (0.152)	0.323 (0.157)	0.327 (0.153)	0.325 (0.154)
		k	0.197 (0.121)	0.201 (0.121)	0.204 (0.122)	0.207 (0.122)	0.205 (0.126)	0.202 (0.122)	0.203 (0.128)	0.203 (0.126)	0.202 (0.118)
50	-0.089 (0.024)	r	0.397 (0.138)	0.398 (0.138)	0.397 (0.137)	0.396 (0.138)	0.399 (0.144)	0.396 (0.142)	0.397 (0.141)	0.398 (0.143)	0.399 (0.139)
		k	0.245 (0.130)	0.247 (0.127)	0.246 (0.124)	0.245 (0.122)	0.245 (0.128)	0.253 (0.127)	0.251 (0.125)	0.253 (0.129)	0.255 (0.116)

()内は標準偏差を表す

Table5 識別力が中程度で困難度が一樣のときの各指標(N=40)

項目数	ΔD		type1	type2.1	type2.2	type2.3	type2.4	type3.1	type3.2	type3.3	type3.4
10	-0.182 (0.053)	r	0.441 (0.151)	0.444 (0.147)	0.446 (0.143)	0.447 (0.145)	0.448 (0.148)	0.439 (0.152)	0.435 (0.148)	0.436 (0.149)	0.437 (0.151)
		k	0.273 (0.145)	0.277 (0.131)	0.282 (0.134)	0.283 (0.129)	0.276 (0.135)	0.271 (0.139)	0.269 (0.131)	0.265 (0.137)	0.276 (0.124)
30	-0.076 (0.023)	r	0.664 (0.091)	0.667 (0.089)	0.670 (0.089)	0.669 (0.090)	0.671 (0.091)	0.662 (0.094)	0.666 (0.094)	0.665 (0.094)	0.664 (0.094)
		k	0.454 (0.105)	0.448 (0.102)	0.456 (0.105)	0.450 (0.116)	0.459 (0.111)	0.449 (0.113)	0.463 (0.117)	0.463 (0.112)	0.454 (0.112)
50	-0.050 (0.017)	r	0.745 (0.089)	0.747 (0.090)	0.748 (0.090)	0.747 (0.089)	0.751 (0.088)	0.747 (0.089)	0.745 (0.088)	0.745 (0.089)	0.747 (0.089)
		k	0.523 (0.110)	0.523 (0.111)	0.522 (0.112)	0.521 (0.111)	0.523 (0.106)	0.523 (0.111)	0.523 (0.112)	0.522 (0.111)	0.526 (0.113)

()内は標準偏差を表す

Table6 識別力が中程度で困難度が低いときの各指標(N=40)

項目数	ΔD		type1	type2.1	type2.2	type2.3	type2.4	type3.1	type3.2	type3.3	type3.4
10	-0.178 (0.051)	r	0.437 (0.131)	0.442 (0.131)	0.444 (0.133)	0.446 (0.133)	0.448 (0.132)	0.435 (0.130)	0.440 (0.131)	0.438 (0.131)	0.441 (0.135)
		k	0.291 (0.117)	0.290 (0.116)	0.291 (0.118)	0.295 (0.119)	0.283 (0.114)	0.279 (0.118)	0.290 (0.121)	0.285 (0.121)	0.286 (0.127)
30	-0.071 (0.025)	r	0.681 (0.106)	0.681 (0.107)	0.682 (0.106)	0.682 (0.106)	0.684 (0.105)	0.682 (0.105)	0.683 (0.109)	0.682 (0.108)	0.679 (0.107)
		k	0.473 (0.130)	0.469 (0.134)	0.471 (0.132)	0.465 (0.138)	0.472 (0.131)	0.474 (0.129)	0.466 (0.132)	0.463 (0.131)	0.459 (0.133)
50	-0.051 (0.018)	r	0.753 (0.090)	0.755 (0.089)	0.755 (0.088)	0.755 (0.090)	0.758 (0.089)	0.753 (0.090)	0.754 (0.090)	0.754 (0.091)	0.753 (0.092)
		k	0.535 (0.121)	0.533 (0.119)	0.538 (0.120)	0.535 (0.121)	0.540 (0.121)	0.529 (0.118)	0.537 (0.122)	0.535 (0.119)	0.533 (0.121)

()内は標準偏差を表す

Table7 識別力が中程度で困難度も中程度のときの各指標(N=40)

項目数	ΔD	type1	type2.1	type2.2	type2.3	type2.4	type3.1	type3.2	type3.3	type3.4	
10	-0.194 (0.063)	r	0.510 (0.153)	0.513 (0.150)	0.515 (0.150)	0.514 (0.153)	0.517 (0.150)	0.510 (0.149)	0.514 (0.148)	0.513 (0.146)	0.511 (0.152)
		k	0.337 (0.135)	0.343 (0.135)	0.339 (0.139)	0.336 (0.144)	0.327 (0.146)	0.336 (0.126)	0.331 (0.136)	0.330 (0.133)	0.337 (0.140)
30	-0.074 (0.024)	r	0.731 (0.081)	0.734 (0.080)	0.736 (0.079)	0.737 (0.079)	0.739 (0.079)	0.730 (0.084)	0.732 (0.083)	0.731 (0.085)	0.734 (0.082)
		k	0.511 (0.121)	0.518 (0.110)	0.519 (0.110)	0.524 (0.111)	0.523 (0.120)	0.508 (0.119)	0.513 (0.116)	0.511 (0.116)	0.515 (0.117)
50	-0.050 (0.019)	r	0.796 (0.062)	0.799 (0.062)	0.797 (0.063)	0.798 (0.063)	0.800 (0.064)	0.796 (0.063)	0.797 (0.063)	0.797 (0.063)	0.796 (0.064)
		k	0.584 (0.098)	0.589 (0.099)	0.589 (0.103)	0.591 (0.104)	0.594 (0.098)	0.586 (0.103)	0.593 (0.097)	0.595 (0.099)	0.594 (0.098)

()内は標準偏差を表す

Table8 識別力が中程度で困難度が高いときの各指標(N=40)

項目数	ΔD	type1	type2.1	type2.2	type2.3	type2.4	type3.1	type3.2	type3.3	type3.4	
10	-0.175 (0.059)	r	0.450 (0.144)	0.459 (0.144)	0.463 (0.142)	0.464 (0.143)	0.461 (0.145)	0.452 (0.142)	0.450 (0.141)	0.450 (0.140)	0.450 (0.135)
		k	0.294 (0.146)	0.304 (0.147)	0.306 (0.144)	0.309 (0.144)	0.301 (0.140)	0.287 (0.144)	0.291 (0.147)	0.284 (0.150)	0.292 (0.134)
30	-0.073 (0.025)	r	0.667 (0.101)	0.669 (0.102)	0.669 (0.102)	0.669 (0.102)	0.674 (0.101)	0.667 (0.100)	0.664 (0.102)	0.666 (0.101)	0.664 (0.100)
		k	0.463 (0.116)	0.460 (0.119)	0.461 (0.116)	0.464 (0.117)	0.467 (0.126)	0.461 (0.121)	0.449 (0.121)	0.457 (0.123)	0.461 (0.125)
50	-0.052 (0.017)	r	0.735 (0.082)	0.735 (0.081)	0.735 (0.081)	0.736 (0.081)	0.739 (0.082)	0.735 (0.082)	0.735 (0.083)	0.735 (0.083)	0.733 (0.083)
		k	0.528 (0.113)	0.527 (0.111)	0.528 (0.113)	0.525 (0.110)	0.527 (0.111)	0.521 (0.115)	0.519 (0.114)	0.516 (0.115)	0.519 (0.114)

()内は標準偏差を表す

Table9 識別力が高く困難度が一樣のときの各指標(N=40)

項目数	ΔD	type1	type2.1	type2.2	type2.3	type2.4	type3.1	type3.2	type3.3	type3.4	
10	-0.129 (0.046)	r	0.645 (0.109)	0.647 (0.110)	0.646 (0.110)	0.649 (0.110)	0.648 (0.106)	0.647 (0.110)	0.648 (0.110)	0.647 (0.110)	0.645 (0.106)
		k	0.437 (0.128)	0.432 (0.126)	0.432 (0.125)	0.435 (0.129)	0.429 (0.134)	0.429 (0.132)	0.433 (0.136)	0.430 (0.134)	0.432 (0.130)
30	-0.047 (0.016)	r	0.826 (0.059)	0.827 (0.059)	0.827 (0.059)	0.828 (0.059)	0.828 (0.059)	0.826 (0.059)	0.825 (0.060)	0.825 (0.059)	0.825 (0.060)
		k	0.632 (0.094)	0.631 (0.090)	0.626 (0.090)	0.625 (0.091)	0.631 (0.095)	0.627 (0.091)	0.625 (0.092)	0.625 (0.094)	0.627 (0.094)
50	-0.031 (0.016)	r	0.880 (0.045)	0.880 (0.045)	0.880 (0.045)	0.880 (0.046)	0.882 (0.045)	0.880 (0.043)	0.881 (0.045)	0.880 (0.044)	0.879 (0.045)
		k	0.683 (0.104)	0.679 (0.106)	0.676 (0.107)	0.677 (0.108)	0.686 (0.104)	0.683 (0.103)	0.681 (0.110)	0.685 (0.107)	0.683 (0.101)

()内は標準偏差を表す

Table10 識別力が高く困難度が低いときの各指標(N=40)

項目数	ΔD	type1	type2.1	type2.2	type2.3	type2.4	type3.1	type3.2	type3.3	type3.4	
10	-0.120 (0.038)	r	0.651 (0.089)	0.653 (0.089)	0.654 (0.091)	0.654 (0.090)	0.656 (0.092)	0.651 (0.098)	0.650 (0.092)	0.648 (0.095)	0.651 (0.092)
		k	0.453 (0.102)	0.454 (0.103)	0.453 (0.102)	0.447 (0.110)	0.455 (0.106)	0.446 (0.107)	0.452 (0.106)	0.447 (0.105)	0.451 (0.109)
30	-0.050 (0.018)	r	0.815 (0.063)	0.815 (0.063)	0.815 (0.063)	0.815 (0.064)	0.817 (0.064)	0.815 (0.063)	0.814 (0.063)	0.815 (0.063)	0.816 (0.062)
		k	0.611 (0.102)	0.611 (0.101)	0.611 (0.102)	0.612 (0.102)	0.611 (0.108)	0.607 (0.104)	0.610 (0.107)	0.612 (0.105)	0.612 (0.106)
50	-0.030 (0.014)	r	0.872 (0.043)	0.873 (0.044)	0.874 (0.044)	0.874 (0.044)	0.875 (0.042)	0.873 (0.043)	0.872 (0.044)	0.873 (0.043)	0.873 (0.043)
		k	0.683 (0.092)	0.683 (0.092)	0.685 (0.092)	0.685 (0.093)	0.686 (0.092)	0.685 (0.089)	0.681 (0.098)	0.684 (0.095)	0.687 (0.091)

()内は標準偏差を表す

Table11 識別力が高く困難度が中程度のときの各指標(N=40)

項目数	ΔD	type1	type2.1	type2.2	type2.3	type2.4	type3.1	type3.2	type3.3	type3.4	
10	-0.131 (0.046)	r	0.711 (0.084)	0.713 (0.085)	0.714 (0.086)	0.717 (0.085)	0.716 (0.084)	0.711 (0.083)	0.711 (0.086)	0.712 (0.084)	0.716 (0.083)
		k	0.501 (0.103)	0.501 (0.108)	0.501 (0.110)	0.500 (0.109)	0.499 (0.106)	0.497 (0.111)	0.499 (0.114)	0.490 (0.111)	0.503 (0.110)
30	-0.054 (0.025)	r	0.866 (0.050)	0.866 (0.051)	0.866 (0.050)	0.867 (0.049)	0.868 (0.048)	0.865 (0.049)	0.866 (0.049)	0.865 (0.049)	0.866 (0.049)
		k	0.662 (0.103)	0.663 (0.108)	0.662 (0.106)	0.659 (0.110)	0.660 (0.106)	0.658 (0.104)	0.662 (0.106)	0.659 (0.103)	0.663 (0.105)
50	-0.031 (0.013)	r	0.908 (0.033)	0.908 (0.033)	0.908 (0.033)	0.908 (0.033)	0.908 (0.034)	0.908 (0.033)	0.908 (0.033)	0.908 (0.034)	0.908 (0.033)
		k	0.739 (0.080)	0.737 (0.080)	0.734 (0.081)	0.731 (0.079)	0.733 (0.082)	0.737 (0.079)	0.735 (0.079)	0.733 (0.080)	0.733 (0.078)

()内は標準偏差を表す

Table12 識別力が高く困難度も高いときの各指標(N=40)

項目数	ΔD	type1	type2.1	type2.2	type2.3	type2.4	type3.1	type3.2	type3.3	type3.4	
10	-0.130 (0.045)	r	0.635 (0.109)	0.637 (0.109)	0.638 (0.109)	0.641 (0.109)	0.644 (0.109)	0.636 (0.112)	0.637 (0.110)	0.638 (0.110)	0.636 (0.110)
		k	0.421 (0.116)	0.426 (0.115)	0.429 (0.109)	0.431 (0.109)	0.435 (0.114)	0.419 (0.123)	0.421 (0.118)	0.421 (0.121)	0.425 (0.116)
30	-0.053 (0.021)	r	0.812 (0.055)	0.811 (0.054)	0.813 (0.054)	0.813 (0.054)	0.817 (0.051)	0.812 (0.054)	0.813 (0.055)	0.813 (0.054)	0.811 (0.057)
		k	0.591 (0.123)	0.589 (0.122)	0.589 (0.120)	0.588 (0.126)	0.594 (0.125)	0.591 (0.123)	0.591 (0.123)	0.591 (0.118)	0.586 (0.117)
50	-0.031 (0.013)	r	0.876 (0.043)	0.877 (0.042)	0.878 (0.042)	0.878 (0.042)	0.878 (0.043)	0.878 (0.042)	0.878 (0.041)	0.875 (0.042)	0.875 (0.044)
		k	0.673 (0.089)	0.671 (0.089)	0.672 (0.100)	0.673 (0.104)	0.674 (0.095)	0.677 (0.095)	0.668 (0.096)	0.671 (0.092)	0.672 (0.097)

()内は標準偏差を表す

D 指標を用いた項目分析のためのテスト条件の検討

Table13 識別力が低く困難度が一樣のときの各指標(N=100)

項目数	ΔD		type1	type2.1	type2.2	type2.3	type2.4	type3.1	type3.2	type3.3	type3.4
10	-0.253 (0.043)	r	0.193 (0.109)	0.193 (0.108)	0.193 (0.109)	0.191 (0.111)	0.190 (0.114)	0.194 (0.113)	0.192 (0.112)	0.194 (0.113)	0.184 (0.111)
		k	0.116 (0.078)	0.114 (0.082)	0.115 (0.079)	0.112 (0.084)	0.113 (0.085)	0.115 (0.084)	0.117 (0.083)	0.119 (0.086)	0.109 (0.080)
30	-0.118 (0.022)	r	0.362 (0.105)	0.366 (0.107)	0.368 (0.107)	0.368 (0.108)	0.369 (0.107)	0.360 (0.106)	0.361 (0.106)	0.361 (0.106)	0.361 (0.106)
		k	0.233 (0.091)	0.236 (0.095)	0.234 (0.092)	0.237 (0.091)	0.239 (0.094)	0.226 (0.092)	0.232 (0.087)	0.228 (0.093)	0.231 (0.092)
50	-0.086 (0.017)	r	0.427 (0.105)	0.430 (0.107)	0.432 (0.106)	0.432 (0.107)	0.435 (0.108)	0.426 (0.106)	0.426 (0.105)	0.426 (0.105)	0.426 (0.104)
		k	0.273 (0.090)	0.277 (0.096)	0.281 (0.093)	0.283 (0.095)	0.285 (0.095)	0.273 (0.094)	0.278 (0.093)	0.277 (0.092)	0.277 (0.090)

()内は標準偏差を表す

Table14 識別力が低く困難度も低いときの各指標(N=100)

項目数	ΔD		type1	type2.1	type2.2	type2.3	type2.4	type3.1	type3.2	type3.3	type3.4
10	-0.240 (0.039)	r	0.188 (0.117)	0.189 (0.120)	0.192 (0.120)	0.192 (0.120)	0.192 (0.121)	0.188 (0.120)	0.191 (0.116)	0.190 (0.120)	0.189 (0.119)
		k	0.112 (0.087)	0.115 (0.089)	0.113 (0.086)	0.114 (0.087)	0.115 (0.086)	0.120 (0.080)	0.112 (0.088)	0.118 (0.085)	0.118 (0.079)
30	-0.123 (0.020)	r	0.329 (0.102)	0.330 (0.102)	0.331 (0.103)	0.332 (0.103)	0.336 (0.103)	0.331 (0.101)	0.330 (0.104)	0.331 (0.102)	0.328 (0.104)
		k	0.208 (0.083)	0.206 (0.084)	0.207 (0.084)	0.207 (0.080)	0.212 (0.083)	0.213 (0.086)	0.208 (0.082)	0.214 (0.085)	0.211 (0.081)
50	-0.085 (0.014)	r	0.424 (0.098)	0.427 (0.100)	0.429 (0.099)	0.431 (0.099)	0.434 (0.098)	0.425 (0.099)	0.425 (0.097)	0.425 (0.098)	0.424 (0.098)
		k	0.274 (0.090)	0.277 (0.088)	0.279 (0.088)	0.280 (0.089)	0.283 (0.087)	0.276 (0.091)	0.278 (0.086)	0.277 (0.087)	0.278 (0.088)

()内は標準偏差を表す

Table15 識別力が低く困難度が中程度のときの各指標(N=100)

項目数	ΔD		type1	type2.1	type2.2	type2.3	type2.4	type3.1	type3.2	type3.3	type3.4
10	-0.282 (0.049)	r	0.226 (0.106)	0.229 (0.109)	0.231 (0.112)	0.231 (0.112)	0.230 (0.111)	0.228 (0.108)	0.233 (0.105)	0.232 (0.107)	0.229 (0.105)
		k	0.135 (0.080)	0.137 (0.085)	0.138 (0.085)	0.136 (0.087)	0.136 (0.087)	0.136 (0.088)	0.131 (0.082)	0.130 (0.084)	0.141 (0.075)
30	-0.141 (0.025)	r	0.380 (0.103)	0.382 (0.104)	0.383 (0.106)	0.384 (0.106)	0.385 (0.107)	0.380 (0.102)	0.378 (0.103)	0.379 (0.102)	0.379 (0.104)
		k	0.236 (0.095)	0.237 (0.094)	0.236 (0.094)	0.235 (0.094)	0.233 (0.095)	0.233 (0.094)	0.235 (0.096)	0.237 (0.094)	0.231 (0.094)
50	-0.094 (0.016)	r	0.481 (0.090)	0.485 (0.090)	0.487 (0.091)	0.488 (0.091)	0.490 (0.092)	0.480 (0.089)	0.481 (0.089)	0.481 (0.089)	0.481 (0.089)
		k	0.304 (0.076)	0.308 (0.079)	0.313 (0.076)	0.313 (0.076)	0.310 (0.077)	0.305 (0.078)	0.307 (0.080)	0.306 (0.081)	0.311 (0.078)

()内は標準偏差を表す

Table16 識別力が低く困難度が高いときの各指標(N=100)

項目数	ΔD		type1	type2.1	type2.2	type2.3	type2.4	type3.1	type3.2	type3.3	type3.4
10	-0.239 (0.043)	r	0.193 (0.111)	0.196 (0.109)	0.195 (0.111)	0.196 (0.112)	0.197 (0.114)	0.190 (0.115)	0.191 (0.110)	0.189 (0.113)	0.193 (0.110)
		k	0.120 (0.083)	0.117 (0.081)	0.119 (0.086)	0.122 (0.088)	0.116 (0.087)	0.115 (0.087)	0.119 (0.082)	0.114 (0.085)	0.120 (0.087)
30	-0.122 (0.021)	r	0.338 (0.109)	0.341 (0.109)	0.342 (0.110)	0.342 (0.109)	0.346 (0.112)	0.335 (0.108)	0.337 (0.109)	0.335 (0.108)	0.340 (0.110)
		k	0.202 (0.091)	0.205 (0.091)	0.208 (0.091)	0.210 (0.095)	0.217 (0.090)	0.209 (0.088)	0.204 (0.095)	0.209 (0.087)	0.208 (0.091)
50	-0.086 (0.015)	r	0.424 (0.101)	0.428 (0.103)	0.429 (0.103)	0.431 (0.103)	0.432 (0.103)	0.425 (0.100)	0.423 (0.100)	0.425 (0.099)	0.425 (0.100)
		k	0.270 (0.086)	0.277 (0.087)	0.277 (0.088)	0.276 (0.086)	0.273 (0.088)	0.270 (0.086)	0.270 (0.083)	0.271 (0.084)	0.274 (0.079)

()内は標準偏差を表す

Table17 識別力が中程度で困難度が一樣のときの各指標(N=100)

項目数	ΔD		type1	type2.1	type2.2	type2.3	type2.4	type3.1	type3.2	type3.3	type3.4
10	-0.178 (0.038)	r	0.444 (0.111)	0.448 (0.110)	0.451 (0.110)	0.452 (0.107)	0.456 (0.108)	0.446 (0.112)	0.446 (0.109)	0.444 (0.111)	0.446 (0.110)
		k	0.282 (0.094)	0.289 (0.097)	0.293 (0.091)	0.294 (0.089)	0.299 (0.092)	0.287 (0.096)	0.288 (0.093)	0.288 (0.099)	0.292 (0.099)
30	-0.072 (0.013)	r	0.678 (0.064)	0.680 (0.064)	0.680 (0.064)	0.681 (0.065)	0.685 (0.064)	0.677 (0.064)	0.679 (0.064)	0.678 (0.065)	0.678 (0.065)
		k	0.470 (0.071)	0.474 (0.067)	0.474 (0.068)	0.476 (0.068)	0.481 (0.072)	0.474 (0.068)	0.475 (0.071)	0.473 (0.070)	0.470 (0.071)
50	-0.047 (0.011)	r	0.767 (0.056)	0.770 (0.056)	0.771 (0.056)	0.772 (0.056)	0.775 (0.055)	0.768 (0.056)	0.768 (0.055)	0.768 (0.056)	0.768 (0.056)
		k	0.552 (0.063)	0.557 (0.060)	0.558 (0.064)	0.559 (0.065)	0.557 (0.065)	0.554 (0.061)	0.554 (0.061)	0.554 (0.060)	0.554 (0.064)

()内は標準偏差を表す

Table18 識別力が中程度で困難度が低いときの各指標(N=100)

項目数	ΔD		type1	type2.1	type2.2	type2.3	type2.4	type3.1	type3.2	type3.3	type3.4
10	-0.172 (0.034)	r	0.453 (0.106)	0.455 (0.108)	0.456 (0.110)	0.458 (0.112)	0.461 (0.114)	0.454 (0.105)	0.453 (0.108)	0.453 (0.107)	0.451 (0.110)
		k	0.287 (0.088)	0.287 (0.090)	0.288 (0.089)	0.291 (0.090)	0.295 (0.084)	0.289 (0.095)	0.287 (0.093)	0.289 (0.095)	0.287 (0.095)
30	-0.073 (0.015)	r	0.670 (0.066)	0.674 (0.065)	0.676 (0.066)	0.677 (0.066)	0.682 (0.068)	0.671 (0.068)	0.669 (0.066)	0.670 (0.067)	0.670 (0.068)
		k	0.462 (0.075)	0.466 (0.077)	0.469 (0.075)	0.471 (0.075)	0.473 (0.080)	0.465 (0.077)	0.461 (0.075)	0.461 (0.078)	0.459 (0.075)
50	-0.048 (0.010)	r	0.752 (0.052)	0.754 (0.051)	0.754 (0.051)	0.754 (0.051)	0.757 (0.051)	0.750 (0.053)	0.751 (0.052)	0.751 (0.053)	0.749 (0.052)
		k	0.548 (0.073)	0.550 (0.071)	0.550 (0.069)	0.549 (0.071)	0.553 (0.069)	0.548 (0.073)	0.549 (0.070)	0.547 (0.073)	0.548 (0.069)

()内は標準偏差を表す

Table19 識別力が中程度で困難度も中程度のときの各指標(N=100)

項目数	ΔD		type1	type2.1	type2.2	type2.3	type2.4	type3.1	type3.2	type3.3	type3.4
10	-0.188 (0.035)	r	0.517 (0.090)	0.522 (0.092)	0.524 (0.095)	0.526 (0.094)	0.530 (0.096)	0.516 (0.093)	0.518 (0.092)	0.517 (0.093)	0.521 (0.088)
		k	0.334 (0.087)	0.337 (0.079)	0.340 (0.078)	0.343 (0.074)	0.343 (0.078)	0.336 (0.082)	0.338 (0.083)	0.334 (0.089)	0.342 (0.076)
30	-0.077 (0.017)	r	0.729 (0.058)	0.732 (0.058)	0.734 (0.057)	0.735 (0.057)	0.737 (0.056)	0.730 (0.056)	0.729 (0.057)	0.730 (0.056)	0.727 (0.059)
		k	0.523 (0.076)	0.523 (0.077)	0.525 (0.077)	0.527 (0.076)	0.532 (0.077)	0.522 (0.078)	0.519 (0.075)	0.519 (0.075)	0.518 (0.075)
50	-0.049 (0.011)	r	0.809 (0.044)	0.811 (0.043)	0.811 (0.043)	0.812 (0.043)	0.815 (0.042)	0.809 (0.043)	0.809 (0.043)	0.809 (0.042)	0.809 (0.042)
		k	0.599 (0.068)	0.599 (0.065)	0.600 (0.063)	0.600 (0.061)	0.607 (0.067)	0.600 (0.063)	0.599 (0.064)	0.599 (0.064)	0.601 (0.064)

()内は標準偏差を表す

Table20 識別力が中程度で困難度が高いときの各指標(N=100)

項目数	ΔD		type1	type2.1	type2.2	type2.3	type2.4	type3.1	type3.2	type3.3	type3.4
10	-0.176 (0.035)	r	0.449 (0.102)	0.451 (0.105)	0.454 (0.107)	0.457 (0.107)	0.461 (0.106)	0.449 (0.102)	0.447 (0.103)	0.447 (0.103)	0.449 (0.100)
		k	0.289 (0.084)	0.292 (0.090)	0.294 (0.086)	0.294 (0.083)	0.298 (0.084)	0.293 (0.086)	0.292 (0.087)	0.292 (0.086)	0.291 (0.088)
30	-0.073 (0.016)	r	0.660 (0.077)	0.662 (0.077)	0.664 (0.077)	0.666 (0.077)	0.669 (0.077)	0.659 (0.077)	0.660 (0.079)	0.660 (0.078)	0.661 (0.078)
		k	0.462 (0.084)	0.458 (0.084)	0.462 (0.082)	0.463 (0.080)	0.467 (0.078)	0.458 (0.081)	0.460 (0.080)	0.457 (0.081)	0.462 (0.078)
50	-0.047 (0.010)	r	0.758 (0.054)	0.761 (0.054)	0.762 (0.053)	0.763 (0.053)	0.765 (0.054)	0.758 (0.054)	0.759 (0.053)	0.759 (0.054)	0.759 (0.054)
		k	0.553 (0.072)	0.551 (0.072)	0.553 (0.069)	0.553 (0.069)	0.558 (0.068)	0.551 (0.071)	0.548 (0.070)	0.550 (0.070)	0.553 (0.068)

()内は標準偏差を表す

Table21 識別力が高く困難度が一樣のときの各指標(N=100)

項目数	ΔD		type1	type2.1	type2.2	type2.3	type2.4	type3.1	type3.2	type3.3	type3.4
10	-0.127 (0.030)	r	0.644 (0.082)	0.646 (0.083)	0.647 (0.083)	0.648 (0.081)	0.651 (0.079)	0.643 (0.079)	0.644 (0.081)	0.642 (0.081)	0.645 (0.079)
		k	0.433 (0.081)	0.439 (0.080)	0.440 (0.082)	0.443 (0.080)	0.452 (0.081)	0.437 (0.080)	0.440 (0.080)	0.438 (0.080)	0.437 (0.084)
30	-0.050 (0.012)	r	0.830 (0.036)	0.831 (0.037)	0.832 (0.037)	0.832 (0.037)	0.834 (0.037)	0.831 (0.037)	0.830 (0.036)	0.830 (0.036)	0.830 (0.036)
		k	0.615 (0.058)	0.619 (0.059)	0.618 (0.059)	0.620 (0.057)	0.622 (0.059)	0.616 (0.060)	0.619 (0.060)	0.616 (0.059)	0.616 (0.058)
50	-0.030 (0.008)	r	0.887 (0.027)	0.888 (0.027)	0.888 (0.027)	0.888 (0.027)	0.890 (0.027)	0.887 (0.027)	0.887 (0.028)	0.887 (0.028)	0.887 (0.027)
		k	0.703 (0.058)	0.706 (0.056)	0.706 (0.055)	0.705 (0.055)	0.702 (0.057)	0.703 (0.057)	0.700 (0.058)	0.701 (0.057)	0.699 (0.057)

()内は標準偏差を表す

Table22 識別力が高く困難度が低いときの各指標(N=100)

項目数	ΔD		type1	type2.1	type2.2	type2.3	type2.4	type3.1	type3.2	type3.3	type3.4
10	-0.125 (0.025)	r	0.644 (0.082)	0.647 (0.084)	0.649 (0.084)	0.650 (0.084)	0.651 (0.085)	0.642 (0.081)	0.644 (0.079)	0.643 (0.080)	0.641 (0.082)
		k	0.440 (0.083)	0.443 (0.085)	0.444 (0.085)	0.450 (0.084)	0.444 (0.084)	0.431 (0.083)	0.437 (0.083)	0.431 (0.084)	0.435 (0.087)
30	-0.049 (0.011)	r	0.820 (0.041)	0.821 (0.040)	0.822 (0.039)	0.822 (0.039)	0.824 (0.039)	0.819 (0.042)	0.820 (0.041)	0.819 (0.042)	0.819 (0.041)
		k	0.613 (0.058)	0.613 (0.060)	0.613 (0.058)	0.615 (0.060)	0.620 (0.060)	0.615 (0.061)	0.613 (0.061)	0.612 (0.060)	0.614 (0.061)
50	-0.031 (0.009)	r	0.884 (0.027)	0.885 (0.028)	0.885 (0.027)	0.885 (0.027)	0.887 (0.027)	0.884 (0.028)	0.884 (0.027)	0.883 (0.027)	0.884 (0.027)
		k	0.681 (0.062)	0.684 (0.059)	0.687 (0.060)	0.687 (0.058)	0.688 (0.058)	0.681 (0.063)	0.683 (0.061)	0.681 (0.060)	0.684 (0.058)

()内は標準偏差を表す

Table23 識別力が高く困難度が中程度のときの各指標(N=100)

項目数	ΔD		type1	type2.1	type2.2	type2.3	type2.4	type3.1	type3.2	type3.3	type3.4
10	-0.135 (0.028)	r	0.705 (0.064)	0.709 (0.064)	0.712 (0.063)	0.712 (0.063)	0.714 (0.062)	0.705 (0.064)	0.704 (0.064)	0.704 (0.063)	0.704 (0.064)
		k	0.496 (0.071)	0.502 (0.066)	0.506 (0.065)	0.503 (0.065)	0.503 (0.067)	0.498 (0.073)	0.501 (0.070)	0.503 (0.070)	0.494 (0.068)
30	-0.051 (0.014)	r	0.867 (0.032)	0.868 (0.032)	0.869 (0.032)	0.869 (0.031)	0.870 (0.030)	0.867 (0.032)	0.867 (0.031)	0.867 (0.031)	0.867 (0.032)
		k	0.674 (0.060)	0.677 (0.058)	0.677 (0.058)	0.678 (0.057)	0.678 (0.058)	0.672 (0.054)	0.675 (0.061)	0.674 (0.057)	0.677 (0.057)
50	-0.031 (0.009)	r	0.914 (0.021)	0.915 (0.020)	0.915 (0.020)	0.915 (0.020)	0.916 (0.020)	0.914 (0.020)	0.914 (0.021)	0.914 (0.020)	0.914 (0.021)
		k	0.739 (0.052)	0.740 (0.053)	0.739 (0.052)	0.738 (0.051)	0.740 (0.050)	0.738 (0.051)	0.741 (0.049)	0.740 (0.050)	0.736 (0.053)

()内は標準偏差を表す

Table24 識別力が高く困難度も高いときの各指標(N=100)

項目数	ΔD		type1	type2.1	type2.2	type2.3	type2.4	type3.1	type3.2	type3.3	type3.4
10	-0.123 (0.030)	r	0.635 (0.085)	0.637 (0.086)	0.639 (0.087)	0.640 (0.088)	0.642 (0.086)	0.633 (0.083)	0.635 (0.088)	0.633 (0.085)	0.636 (0.089)
		k	0.433 (0.086)	0.435 (0.086)	0.434 (0.086)	0.435 (0.090)	0.441 (0.089)	0.429 (0.087)	0.434 (0.087)	0.432 (0.085)	0.434 (0.091)
30	-0.048 (0.012)	r	0.822 (0.035)	0.824 (0.035)	0.825 (0.035)	0.825 (0.035)	0.826 (0.035)	0.823 (0.035)	0.823 (0.034)	0.824 (0.035)	0.823 (0.034)
		k	0.619 (0.063)	0.621 (0.065)	0.624 (0.061)	0.623 (0.061)	0.622 (0.066)	0.618 (0.064)	0.623 (0.059)	0.620 (0.061)	0.621 (0.059)
50	-0.031 (0.007)	r	0.880 (0.024)	0.881 (0.024)	0.881 (0.024)	0.882 (0.024)	0.883 (0.024)	0.880 (0.024)	0.881 (0.024)	0.880 (0.024)	0.880 (0.023)
		k	0.685 (0.055)	0.684 (0.055)	0.681 (0.056)	0.682 (0.054)	0.683 (0.057)	0.682 (0.055)	0.681 (0.054)	0.680 (0.054)	0.681 (0.057)

()内は標準偏差を表す

D 指標を用いた項目分析のためのテスト条件の検討

(3) テストの条件

テストの受験者数は $N=40, 100$ の 2 パタン, テストの項目数は $m=10, 30, 50$ の 3 パタンとした. これらの組み合わせ 6 パタンについて検討を行った.

(4) シミュレーションの手続き

以下の手続きによりシミュレーションを行った.

1. 項目数の正答率ならびに識別力の真値を発生させ, これらを 2 Parameter Logistic Model の識別力と困難度に対応するように変換を行った.
2. 受験者数の潜在特性真値を $N(0,1)$ から発生させ, 1 で求めた項目特性値を用い, 各項目の正答確率を求めた.
3. 受験者の一様乱数を発生させ, 2 で求めた正答確率を用いて, 正誤パターンを作成した.
4. 正誤パターンより, 各項目の正答率ならびに I-T 相関を算出した.
5. Type1~3.4 の方法で並び替え, それに基づき群わけを行った.
6. 群わけに基づき, 真値との比較として, 順位相関係数, 重み付けカップ係数ならびに ΔD を算出した.

以上の手続きを全ての受験者数・項目数・正答率・I-T 相関の組み合わせについて 100 回ずつ行った.

(5) 用いる指標

真値と提案手法による並び順ならびに群わけの比較をするために, スピアマンの順位相関係数, 重み付けカップ係数を用いた. また, 真値による群わけの識別力と, テスト条件による群訳の識別力を比較するために, ΔD を用いた.

スピアマンの順位相関係数は, 真値の並び順と提案手法の並び順がどの程度一致しているかという指標である.

また, 重み付けカップ係数と ΔD は, 研究 1 と同じ指標である.

3. 3. 結果と考察

順位相関係数, 重みづけカップ係数ならびに ΔD が, Table1~24 である. ここで検討している並び替えの方法は, 同点の受験者についての並び替えであり, 上位群, 中位群, 下位群のそれぞれの平均点はどのような並び替えを行っても同じになる. そのため, ΔD は各条件につき一つ算出された.

項目数が増えるほど, 順位相関係数ならびに重み付けカップ係数の値は小さくなった. 識別力条件について検討すると, 識別力が高くなるにつれて順位相関係数, 重み付けカップ係数の値は大きくなる. また, 困難度条件について確認すると, 最も順位相関係数ならびに重み付けカップ係数の値を大きくするの

は、困難度が中程度付近で固まる条件であり、次いで一様分布の条件であった。

Lord(1982) は、平行テスト x, y におけるの信頼性の基準として、 $\rho_{xy} = .90$ としている。真値と並び替え順の順位相関係数を代用すると、その基準を満たす重みづけカッパ係数はおよそ 0.7 を示していることより、信頼できる D 指標を算出する重みづけカッパ係数の条件として、最低 0.7 程度であることが考えられる。これを踏まえると、識別力が高く、50 項目程度の項目数があることが望ましい。しかし、30 項目程度以上でも条件によってはかなり高い r の値を得た。クラスルームテストとして実施するテストでは、妥当な項目数であると考えられる。

また、同点者を並び替える手法としては、多くの条件で識別力を用いることが望ましいことが示唆された。特に、Type2.4 のように、受検者の全ての正答項目の識別力を用いた方法がより良い。一方、実際のテスト現場での実用性を鑑みると、識別力の高い数項目について並び替えを行うだけでも、ランダムである ID 順に並び替えるよりは望ましい結果となることが示された。

しかしながら、識別力が高い条件では、並び替えの影響がほとんど見られず、どのような手法で並び替えても重み付けカッパ係数の値は一貫して大きかった。元々識別力の高い項目で作成されたテストでは、合計得点が算出された時点で十分に受検者の能力の高低を識別できているため、識別力による並び替えの影響は小さいのだと考えられる。

$4D$ を見ると、識別力が大きくなるにつれて、また、項目数が大きくなるにつれて 0 に近い値となった。実際のテスト場面において、識別力の検討はテスト実施後であることを考慮すると、項目分析のためには最低でも 30 項目程度が必要になると考えられる。

以上のことより、項目分析を行うためには、テストの項目数は 30 項目以上であり、識別力が高いテストを作成することが望ましい。また、合計得点が同じ受検者について、識別力の高い項目への正誤を用いて並べ替えを行うことで、より真値に近い順位になることがわかった。

4. 研究 3: 実データを用いたシミュレーションによる検討

4. 1. 目的

研究 3 では、シミュレーションで示された研究 2 の結果について、並び替え方法が実際のテスト場面でも有用であるかどうかの確認をする。具体的には、実際のテストデータについて、潜在特性推定値 θ を真値の代わりに用い、 θ による並び替え(群分け)と、研究 2 で示した素点を用いた並び替え(群分け)の一

D 指標を用いた項目分析のためのテスト条件の検討

致率を検討する。

4. 2. 方法

(1) テストの実施

2019年4月～6月にかけて愛知県内の大学生453名にテストを行った。使用したテストは国語・数学・英語から成る26項目である。

(2) 使用するテスト条件

全受検者を母集団として、10名、40名、100名をサンプリングした。テストは、全項目を用いた26項目と、英語のみを用いた12項目の2パターンについて検討した。

(3) 並び変える方法

研究2と同じ方法を用いた。 **Table25 26項目の項目特性推定値ならびにI-T相関と正答率**

	項目識別力	項目困難度	I-T相関	正答率
項目1	0.372	-1.185	0.141	0.605
項目2	0.258	-1.069	0.117	0.567
項目3	1.039	-1.023	0.342	0.704
項目4	1.019	-1.073	0.338	0.711
項目5	1.235	-0.637	0.383	0.645
項目6	0.896	0.125	0.295	0.475
項目7	1.053	-0.026	0.401	0.503
項目8	1.013	-0.239	0.344	0.547
項目9	0.127	7.687	0.035	0.274
項目10	1.119	-0.805	0.348	0.671
項目11	1.532	0.238	0.443	0.433
項目12	0.795	-0.128	0.297	0.521
項目13	0.955	-0.059	0.321	0.510
項目14	1.036	0.047	0.355	0.488
項目15	0.275	-0.973	0.127	0.565
項目16	0.281	4.701	0.083	0.214
項目17	0.510	1.214	0.191	0.358
項目18	1.190	-2.817	0.241	0.943
項目19	2.062	-2.288	0.259	0.960
項目20	0.903	-0.118	0.326	0.521
項目21	2.400	-1.680	0.359	0.914
項目22	-0.106	2.401	-0.048	0.563
項目23	1.113	-1.618	0.295	0.812
項目24	2.293	-1.691	0.369	0.912
項目25	0.482	-1.056	0.199	0.618
項目26	0.545	2.243	0.168	0.241

(4) シミュレーションの手続き

1. 全受検者、全項目について、2 Parameter Logistic Model を用いて項目特性値ならびに潜在能力特性値を推定した。

2. 全受検者より N=10, 40, 100 についてサンプリングを行い、正答率ならびに I-T 相関を算出した。

3. サンプルを Type1 ～ Type3.4 について順位付けならびに群わけを行い、D 指標を算出した。

4. 潜在特性推定値と比較するため、スピアマンの順位相関係数、重みづけカッパを算出した。

Table26 12項目の項目特性推定値ならびにI-T相関と正答率

	項目識別力	項目困難度	I-T相関	正答率
項目15	0.318	-0.844	0.135	0.565
項目16	0.080	16.259	0.017	0.214
項目17	0.691	0.938	0.184	0.358
項目18	1.705	-2.286	0.306	0.943
項目19	1.978	-2.378	0.253	0.960
項目20	1.069	-0.097	0.247	0.521
項目21	2.052	-1.807	0.287	0.914
項目22	-0.127	1.993	-0.046	0.563
項目23	1.550	-1.325	0.283	0.812
項目24	2.231	-1.731	0.344	0.912
項目25	0.588	-0.884	0.210	0.618
項目26	0.687	1.839	0.151	0.241

5. 2~4 について、全 26 項目ならびに英語のみ 12 項目のそれぞれについて 100 回行った。

(5) 使用する指標

潜在特性推定値と提案手法での順序を比較検討するため、順位相関係数、重みづけカッパ係数について検討した。

4. 3. 結果

まず、使用したテストについて次元性を確認するため、固有値の λ_1/λ_2 を確認した。26 項目のテストにおいて 2.359, 12 項目のテストにおいて 1.731 と、どちらのテストも第一固有値が第二固有値に比べて大きく、次元性が確認された。

それぞれのテストについて 2PLM を用いて推定したテスト項目の識別力、困難度の推定値ならびに I-T 相関と正答率は Table25,26 であった。

26 項目のテストについて、I-T 相関は $M=0.259$, $SD=0.122$, 正答率は $M=0.587$, $SD=0.201$ であり、12 項目のテストについて、I-T 相関は $M=0.198$, $SD=0.113$, 正答率は $M=0.635$, $SD=0.261$ であった。

これらのテストについて、 $N=10, 40, 100$ をサンプリングして行ったシミュレーションの結果を Table27,28 に示した。項目特性推定値に基づき推定した潜在特性推定値を受験者の真値として代用し、順位相関係数ならびに重みづけカッ

Table27 実データ 26 項目

		Type1	Type2.1	Type2.2	Type2.3	Type2.4	Type3.1	Type3.2	Type3.3	Type3.4
N=10	r	0.925 (0.060)	0.925 (0.061)	0.923 (0.062)	0.925 (0.062)	0.932 (0.062)	0.923 (0.062)	0.924 (0.062)	0.921 (0.064)	0.930 (0.060)
	k	0.806 (0.229)	0.806 (0.229)	0.800 (0.232)	0.809 (0.229)	0.809 (0.238)	0.791 (0.226)	0.800 (0.232)	0.797 (0.227)	0.794 (0.251)
N=40	r	0.938 (0.046)	0.939 (0.047)	0.938 (0.048)	0.940 (0.048)	0.947 (0.048)	0.936 (0.048)	0.937 (0.048)	0.936 (0.049)	0.942 (0.047)
	k	0.804 (0.169)	0.806 (0.170)	0.807 (0.172)	0.812 (0.170)	0.823 (0.175)	0.793 (0.168)	0.805 (0.171)	0.801 (0.168)	0.804 (0.186)
N=100	r	0.945 (0.039)	0.946 (0.040)	0.947 (0.041)	0.948 (0.041)	0.956 (0.041)	0.944 (0.041)	0.945 (0.041)	0.944 (0.042)	0.949 (0.039)
	k	0.808 (0.140)	0.812 (0.141)	0.815 (0.143)	0.820 (0.141)	0.834 (0.145)	0.802 (0.139)	0.812 (0.141)	0.809 (0.139)	0.811 (0.154)

()内は標準偏差を表す

Table28 実データ 12 項目

		Type1	Type2.1	Type2.2	Type2.3	Type2.4	Type3.1	Type3.2	Type3.3	Type3.4
N=10	r	0.786 (0.124)	0.787 (0.127)	0.785 (0.132)	0.788 (0.131)	0.815 (0.129)	0.785 (0.129)	0.808 (0.115)	0.804 (0.113)	0.829 (0.091)
	k	0.581 (0.239)	0.588 (0.242)	0.597 (0.247)	0.600 (0.250)	0.619 (0.252)	0.566 (0.242)	0.600 (0.250)	0.578 (0.231)	0.606 (0.215)
N=40	r	0.805 (0.098)	0.809 (0.101)	0.810 (0.106)	0.815 (0.106)	0.847 (0.106)	0.800 (0.103)	0.819 (0.095)	0.811 (0.094)	0.837 (0.073)
	k	0.592 (0.181)	0.601 (0.182)	0.609 (0.188)	0.617 (0.190)	0.653 (0.196)	0.575 (0.187)	0.598 (0.193)	0.582 (0.182)	0.615 (0.166)
N=100	r	0.816 (0.083)	0.820 (0.086)	0.824 (0.091)	0.830 (0.091)	0.865 (0.092)	0.816 (0.091)	0.827 (0.082)	0.825 (0.083)	0.843 (0.063)
	k	0.596 (0.152)	0.604 (0.153)	0.614 (0.157)	0.622 (0.159)	0.667 (0.166)	0.587 (0.161)	0.598 (0.165)	0.593 (0.158)	0.617 (0.140)

()内は標準偏差を表す

D 指標を用いた項目分析のためのテスト条件の検討

パ係数を算出したものである。

26 項目のテストでは、受験者数によらず $r=.921\sim.956$, $\kappa=.791\sim.834$ と信頼性の基準を満たす値が得られた。並び替えのパタンでは、受検者の正答項目の識別力を用いて並び替える方法において最も高い重み付けカッパ係数の値となった。

12 項目のテストでは、受験者数によらず $r=.785\sim.865$, $\kappa=.578\sim.667$ となった。並び替えのパタンでは、受検者の正答項目の識別力を用いて並び替える方法において最も高い重み付けカッパ係数の値となったが、基準となる 0.7 を超える条件はなかった。

4. 4. 考察

使用した 26 項目ならびに 12 項目のテストは、I-T 関連の平均値がそれぞれ 0.259, 0.198 であることから、研究 2 で設定した識別力条件のうち、平均値が 0.25 となる中程度の条件に近いと考えられる。また、正答率については、26 項目、12 項目のどちらも 0.214~0.960 の範囲であり、平均値ならびに標準偏差が 0.587(0.201), 0.635(0.261)であることから困難度が一樣である条件に近いと考えられる。

識別力が中程度で困難度が一樣であるシミュレーションの結果と比較すると、順位相関係数ならびに重み付けカッパ係数のどちらもが、シミュレーションよりも実データについて高い値を示している。ここでは、真値の代わりに 2PLM で推定した項目識別力ならびに項目困難度を算出した上で、潜在能力推定値を用いている。IRT モデルにおける推定値には誤差が含まれる(Chen & Wang, 2007)。推定は項目への反応パタンに基づいて行われており、よりデータに適合するように過剰推定したものとなっている可能性がある。そのため、真値を基準としたシミュレーションの結果に比べて、推定値を基準とした実データを用いたシミュレーション結果の方が、基準に近い並び順になったと考えられる。

また、項目数に着目すると、26 項目のテストでは信頼性の基準を満たす値であったが、12 項目のテストではその基準は満たされず、信頼性を上げるためには項目数の影響が大きいことが実データにおいても確認された。

合計得点が同じ受検者における並び替えの方法として望ましいのは、項目の識別力を用いる方法(Type2.1~2.4)であった。中でも、受検者が正答した全ての項目の識別力について比較をし、識別力がより高い項目に正答した受検者の順に並び替えることが最も望ましい。しかし、テスト項目のうち、識別力の最

も高い 1 項目についてのみの正誤について並べ替えても、ID 順に並べたときに比べて重み付けカッパ係数は大きくなった。また、困難度を用いた並び替えでは、ID 順による並び替えよりも重み付けカッパ係数が低くなる条件もあった。たとえば、正答率が 0.5 に最も近い項目は 26 項目のテストにおいて項目 13, 12 項目のテストにおいて項目 20 であるが、そこで用いられる項目の I-T 相関係数はそれぞれ 0.321, 0.247 と中程度の識別力である。この 1 項目を用いて並び替えた結果が Type3.1 であるが、全ての条件において ID 順による並び替えよりも順位相関係数ならびに重み付けカッパ係数は低い値となった。つまり、識別力が十分でない項目を用いて並び替えると、いたずらに信頼性を下げることがあるといえる。

5. おわりに

本研究では、テストの評価として D 指標を用いる際に望ましいテストの条件を確認すると共に、受検者の能力を適切に反映した群わけを行うための並び替えの手法について検討を行った。まず、同じ得点の受検者について同じ群とする分け方を検討した上で、同点受検者を異なる群とする方法について検討し、実データを用いた確認を行った。その結果、望ましいとされるテストは、識別力が高く、かつ困難度が中程度に固まる項目数が 30 項目以上のときであることがわかった。また、項目数による影響が大きいことから、項目数が 50 項目以上であれば、更に真値に近い群わけを行うことができることがわかった。同点の受検者を分ける方法としては、受検者が正答した項目のうち、より識別力の高い項目に正答した順に並び替えた上で群わけを行うことが、より真値に近い分け方になることが明らかとなった。また、テスト項目の中で識別力が高かった 1~3 項目の正誤について並び替えるだけでも、ランダムである ID 順よりは真値に近い並び順になることがわかった。

実際のクラスルームテスト場面のテスト時間は、50 分程度と、限られた時間の中で実施される。科目の特性はあるものの、真値に近い結果を得るために、テストの項目数を増やすということはあまり現実的ではない。限られた項目数の中でよりテストの精度を上げることが望ましく、そのためには項目の識別力を上げる必要がある。項目の識別力に影響を及ぼす問題としては、赤根他 (2006) で検討がされているものの、実施したテストを用いてどのような項目で識別力が高くなったかというものである。そのため、考えられる要因を操作する等、より実証的な研究によりどのような要因がどの程度識別力を高めるかを検討していくことが今後の課題と考えられる。

D 指標を用いた項目分析のためのテスト条件の検討

謝辞

本研究は JSPS 科研費 17K01078 の助成を受けたものである。

附記

本研究は名古屋大学教育発達科学研究科研究倫理審査にて、承認を受けたものである。

参考文献

- 赤根敦・伊藤圭・林篤裕・椎名久美子・大澤公一・柳井晴夫・田栗正章 (2006). 識別指数による総合試験問題の項目分析. 大学入試センター研究紀要, 35, 19-47.
- Azeem, M. (2012). Development of math proficiency test using item response theory(IRT). PhD Thesis, University of Education, Lahore.
- 東洋 (2001). 子どもの能力と教育評価(第2版), 東京大学出版会
- Brown, J. D. (1996). Testing in language problems. Prentice-Hall, Inc. [和田稔(訳) (1999). 言語テストの基礎知識—正しい問題作成・評価のために 大修館書店]
- Chen, C. & Wang, W. (2007). Effect of ignoring item interaction on item parameter estimation and detection of interacting items. Applied Psychological Measurement, 31, 388-411.
- Diederich, P. B. (1973). Short-cut statistics for teacher-made tests. Educational Testing Service, Princeton, N.J.
- Ebel, R.L. & Frisbie, D.A. (1991) Essentials of Educational Measurement. 5th Edition, Prentice-Hall, Englewood Cliffs.
- 池田央 (1992). テストの科学-試験にかかわるすべての人に— 日本文化科学社.
- 池田央 (2006). 学力テストの科学. 山森光陽・荘島宏二郎(編著) 学力いま, そしてこれから, ミネルヴァ書房

- 石井秀宗 (2007). 記述式問題における無回答に関連する要因の検討—群馬県児童生徒学力診断テスト小学 6 年生国語テストデータ分析の結果から—日本テスト学会誌, 3, 60-70.
- 石井秀宗 (2020). 項目分析システム. 石井研究室 Retrieved from http://www.educa.nagoya-u.ac.jp/~ishii-h/test_system.html (2021 年 3 月 10 日)
- Johnson, A. P. (1951). Notes on a suggested index of item validity: The U-L Index. *Journal of Educational Psychology*, 42, 499–504.
- 梶谷真也・小林健太郎・鈴木史馬・中田勇人・盛本圭一 (2013). 成績順位の通知と学習意欲. 明星—明星大学明星教育センター研究紀要, 3, 101-110.
- Kelly, T.L. (1939) The selection of upper and lower groups for the validation of test items, *Journal of Educational Psychology*, 30, 17-24.
- Lailamsyah, A. F. & Apriyanti, F. (2020). An item analysis of English summative test for the tenth grade students of sma muhammadiyah 3 Jakarta in the 2013/2014 academic year. *Ed-Humanistics*, 5, 610-615.
- Lord, F. M. (1982). The standard error of equipercentile equating. *Journal of Educational Statistics*, 7(3), 165-174.
- Lord, F. M., & Novick. M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison Wesley.
- 松井仁 (1991). 項目反応理論の応用 II ローレルシャッハテストの運動反応の数量化芝祐順(編) 項目反応理論—基礎と応用—東京大学出版会. pp187-194.
- 日本テスト学会 (2007). テスト・スタンダード 日本のテストの将来に向けて. 金子書房.
- 野口博之・大隅敦子 (2014). *テストの基礎理論*. 研究社.
- 寺尾尚大・安永和央・石井秀宗・野口裕之 (2015). 能力別にみた錯乱枝の効果に関する検討. 日本テスト学会誌, 11, 1-20.
- Wiyasa, P.I., Laksana, I.K.D, & Indrawati, N.L.K.M. (2019), *Evaluating Quality of Teacher-Developed English Test in Vocational High School: Content Validity and Item Analysis*. *Education Quarterly Reviews*, 2, 344-356.
- 安永和央・斎藤信・石井秀宗 (2010). 構造的性質を操作した国語テストにおける回答の検討—中学生を対象にしたテストの実証研究— 日本テスト学会誌, 8, 117-132.
- 安永和央・石井秀宗 (2012). テストにおける設問の問い方が回答傾向に及ぼす影響. *教育心理学研究*, 60, 296-309.

D 指標を用いた項目分析のためのテスト条件の検討

Yu-mien, S. (2010). An item analysis of an English achievement test taken by EFL college students in Taiwan. 明道學術論壇, 6(3), 59-82.