

国語読解テストにおける設問文中の単語の難しさが 能力評価に及ぼす影響

—具体例を回答させる設問の検討—

安永和央¹⁾ 石井秀宗

問題と目的

読書をしているとき、易しい文章で書かれていれば理解できる内容であったにもかかわらず、ある特定の表現が難しく、文意がうまく読み取れなかった、という経験はないだろうか。また、日常会話においても、初めは難しく理解できなかった話が、わかりやすい言葉に言い直されたことにより理解できた、という経験もあるだろう。このように、ある文章（会話）における単語の難しさや文の構造は、少なからずわれわれの理解に影響を及ぼす。このことと同様のことが、テストにも当てはまる(e.g., Loftus & Suppes, 1972)。

Loftus & Suppes (1972) は、設問文中の単語の数、構文の複雑さ等が問題解決に影響を及ぼすことを示唆している。また、Ebel & Frisbie (1991) は、多枝選択式問題において、設問の問い方をより具体的に、かつ、それに関係する選択枝の変更を行うことで、難易度を保ったまま識別力を高めることが可能であることを示した。このように、欧米においては設問文の難しさが受検者の回答に及ぼす影響についていくつか検討されているものの、わが国においてはこれらの知見は見受けられない。

テストの作成を考える際、その性質は大きく2つに分けられる(安永・石井, 2010)。第1に、内容的性質(内容的妥当性)である。これはテストで「何を問うか」ということに関わってくる。第2に、構造的性質である。これは「どのように問うか」ということに関わってくる。具体的には、先に述べた設問の問い方(受検者に回答を求める際の問いの立て方)や設問形式(記述式問題、選択式問題)などである。前者の内容的性質は、その領域の専門家の知識や経験が重要となってくる。これに対し、

後者の構造的性質は、専門家の知識や経験に加え、実証的に検討された知見が必要となってくる。しかしながら、わが国においては、まだテスト研究が盛んであるとは言いがたく、構造的性質に関する知見が少ない。

したがって、本研究では構造的性質の中でも、設問文中の言葉に着目した検討を行い、テスト作成に資する知見を得ることを目的とする。具体的には、設問文中の言葉の難しさの差異が、受検者の能力評価にどのような影響を及ぼすかを検討する。

研究1では、設問の評価を行う。項目特性値を検討する際、日本のデータのみを検討したのでは、これが日本特有の問題であるのか、それとも普遍的に起こりうる問題であるのかを判断することができない。これらを検討するためには、同じテストを他国で実施し、他国の結果と比較する必要性が考えられる。したがって、研究1では、対照国として韓国のデータを用いて検討する。韓国は日本と同じ東アジア圏に属しており、2000年のPISA読解力では日本と同程度でありながら、2006年にはフィンランドを抜いて1位となっている。これを踏まえると、韓国の読解力は日本の読解力と同程度以上と推測され、比較対照国としては最適であると考えられる。そして、研究2では研究1で改訂された設問が受検者の能力評価にどのような影響を及ぼすかを実証的に比較検討する。

研究1

方法

国語読解テスト 群馬県児童生徒学力診断テストの国語読解テスト(記述式12問、選択式6問、合計18問)を用いた。群馬県児童生徒学力診断テストは、群馬県全県の小学校6年生及び中学校3年生の児童生徒を対象に群馬県教育委員会が2006年4月に実施した大規模テストである(群馬県総合教育センター, 2007)。本テストは、教育経験が豊富な群馬県総合教育センターの指導主事ら複数の教科の専門家によって作成され、開発や実施にあたり教育測定学及び計量心理学の専門知識が活かされて

1) 名古屋大学大学院教育発達科学研究科博士課程(後期課程)(指導教員:石井秀宗准教授), 日本学術振興会特別研究員

いた。また、実施までに予備調査も行われ、設問及び解答類型の修正や検討が行われた。韓国においては、バックトランスレーション法により、日本と韓国の教育に精通した者が韓国語に翻訳したものを用いた。本研究では、設問8と9について検討する。設問8と9は、設問文中の記述について、具体例を挙げて説明する自由記述式問題である。設問8と9に関わる本文の一部と実際の設問及び解答類型をTable 1, 2, 3, 4に示す。

実施 群馬県においては、2006年4月下旬に群馬県下の中学校38校で実施され、3892名が受験した。本研究では、石井・安永・金・千・朴・李（2010）から新たに作成した項目分析の結果に基づいて検討を行う（Table 5）。

韓国においては、2008年4月に近郊部にある公立中学校2校で実施され、232名が受験した。解答時間は、問題冊子の配布と記名に要する時間を含め50分で実施された。回答は、日韓の教育に精通し韓国語を母語とする日本の大学院（博士後期課程）を修了した者が、生徒の記述になるべく忠実に日本語に翻訳した。評定は、本稿

の第一筆者が解答類型に従って採点を行い、類型の判断に迷う箇所については教育測定学及び計量心理学を専門とする大学教員1名と合議のうえ評定を行った。

分析 回答の分析には、古典的テスト理論における項目分析を用いた（Crocker & Algina, 1986；Ebel & Frisbie, 1991；Hogan, 2007；石井, 2007；Kubiszyn & Borich, 2009；Musial, Nieminen, Thomas, & Burke, 2009；野口, 1985）。具体的には、各設問の解答類型分類率及び得点率（難易度）、識別力を算出した。解答類型分類率とは、受検者の回答を評定基準である解答類型に従って振り分けた割合を示す指標である。これにより、各類型の回答を記述した受検者の割合を把握することが可能となる。得点率とは、値が1に近いほど易しく、0に近いほど難しいことを示し、本研究では難易度指標として用いた。算出方法は、各設問の解答類型に従い、回答が正答であれば1点（Table 3, 4における解答類型1）、準正答であれば回答内容に基づき0.5点（解答類型2）、または0.25点（解答類型3）、誤答もしくは無回答であれば0点として各受検者の得点率を求め、全受検者の平

Table 1 国語読解テスト本文の第一段落と第三段落

- ① かつて「必要は発明の母」と言われたが、今や「発明は必要の母」となっている。「必要」とは、より安全で、より便利で、より小型で、より省資源・省エネルギーで、より手に入れ易くて、より能率的で、というような人間が持つ欲望のことである。その欲望に突き動かされてさまざまな新製品がテクノロジーによって開発され、人々の生活に利便をもたらしてきた。その意味では、人間という知的好奇心を持つ動物特有の能力の一つがテクノロジーであることは確かである。その結果、手の延長としての道具、足の延長としての車や飛行機、眼の延長としての望遠鏡や顕微鏡、脳の延長としてのコンピュータなど、人間の身体能力を格段に拡張することができた。これによって、人間は文化という他の動物には見られない新しい可能性を獲得した。「必要」という感性が、「発明」という知的能力を駆動してきたのだ。
- ③ テクノロジーは、確かに人間の外的な身体能力を拡張したが、見方を変えれば、個々の人間がもつ内的な身体能力は衰えてきたことになる。自家用車の使用によって足が衰えただけでなく、糖尿病が増えたというデータがある。着物を縫い直したり、揉み洗いで洗濯したり、小刀で鉛筆を削ったり、柄杓で水を汲んだり、というような手を使った労働をしなくなった結果として、手が持っていた能力も失っているのだ。エアコンで環境温度を一定のまま過ごす生活を続けていけば寒暑に応じて体温を調節する能力も衰えていくかもしれない。そうなれば、気候環境の変化に遭遇したとき、人類は果たして生き残ることができるのだろうか。

池内 了著「科学は今どうなっているの？」より

Table 2 設問8と設問9の原本

設問	文章
8	「人間の外的な身体能力の拡張」とはどういうことですか。具体的な例を一つ挙げて説明しなさい。
9	「人間がもつ内的な身体能力の衰え」とはどういうことですか。具体的な例を一つ挙げて説明しなさい。

Table 3 設問8の解答類型

8	[条件] を基に、下記の各類型に該当する処理番号を記入する。 ※表記や漢字等の誤りは許容する。		
	・下記の条件の①・②の関連性が明確であり、その関連から③の条件を満たした説明の文を書いているもの。	1	◎
	・下記の条件の①・②いずれかを取り上げ、その関連から③の条件を満たした説明の文を書いているもの。	2	
	・下記の条件の①・②の関連性について触れているが、③の条件を満たしていない説明の文を書いているもの。	3	※
	・上記以外の解答	9	
	・無解答	0	
	[条件] ① テクノロジーの具体例（車、飛行機、望遠鏡、顕微鏡、コンピュータ等）のいずれかに着目している。 ② 人間の手、足、目、脳のいずれかに着目している。 ③ 人間の生活の向上に（安全、便利、能率的、利便性等）に言及している。「拡張」をそのまま記述しているものは不可。		

Table 4 設問9の解答類型

9	[条件] を基に、下記の各類型に該当する処理番号を記入する。 ※表記や漢字等の誤りは許容する。		
	・下記の条件の①～③を満たした説明の文を書いているもの。	1	◎
	・下記の条件の①・②いずれかを取り上げ、その関連から③の条件を満たした説明の文を書いているもの。	2	
	・下記の条件の①・②の関連性について触れているが、③の条件を満たしていない説明の文を書いているもの。	3	※
	・上記以外の解答	9	
	・無解答	0	
	[条件] ① 手、足、目、脳等の人間の身体能力のいずれかに着目している。 ② 人間の生活を便利で快適にしている具体的なテクノロジーを取り上げている。 ③ ①と関連した人間の身体的機能（手、足、目等）の低下・衰退に言及している。		

Table 5 設問8と設問9の項目分析結果

設問	国	解答類型分類率					得点率 《難易度》	識別力 《I-T相関係数》
		1 ^{a)}	2	3	9 ^{b)}	0 ^{c)}		
8	日本	.138	.117	.239	.363	.143	.256	.336
	韓国	.103	.164	.147	.530	.056	.222	.393
9	日本	.505	.173	.021	.211	.090	.597	.386
	韓国	.634	.129	.000	.181	.056	.698	.332

石井・安永・金・千・朴・李 (2010) から新たに作成

a) 正答

b) その他の回答

c) 無回答

均得点率を算出した。識別力とは、 $-1 \sim 1$ の値を取り、値が1に近いほど、合計得点が高い受検者はその設問に正答する傾向が強く、低い受検者はその設問に正答する傾向が弱いことを示す指標である(石井, 2007)。研究1では識別力指標として、I-T相関係数を用いた。I-T相関係数とは、その設問の得点率と全設問から当該設問を除いた合計得点との相関係数である。

結果と考察

石井他(2010)から新たに作成した項目分析の結果をTable 5に示す。受検者の回答は、具体例の記述に必要な3つの条件を全て満たした際に正答の解答類型1に分類され、条件を2つ満たした場合に準正答の解答類型2と3に分類される。そして、両設問とも正答以外の回答を記述した場合、その他の解答類型9に分類され、回答を記述していない場合、無回答の解答類型0に分類される(Table 3, 4)。

設問8と9を比較すると、日本と韓国ともに設問8では解答類型1に分類された受検者が少ないが、設問9では多くの受検者が分類された。これに対し、設問9では、解答類型3に分類された受検者はほとんどいないが、設問8では日本で約24%、韓国で約15%の受検者が分類された。これらの分類の得点化に伴い、設問8では日韓ともに低い得点率となっている。Crocker & Algina (1986)の項目選択基準では、項目難易度は中程度(0.4~0.6)が適切とされている。この基準に従うと設問9は難易度の安定した項目であると考えられるが、設問8は得点率の観点から項目改訂の必要性が考えられる。したがって、設問9の内容及び受検者の回答に基づいて、設問8の改訂を試みる。

設問8は「人間の外的な身体能力の拡張」の具体例を記述する問題であり、設問9は「人間がもつ内的な身体能力の衰え」の具体例を記述する問題である。設問9に対して、日本では約50%、韓国では約63%の受検者が①手や足など身体能力に着目し、②生活を便利にする具体的なテクノロジーを取り上げ、③身体能力の衰退に言及できている。しかし、設問8において、日本では約14%、韓国では約10%の受検者が①人間の手や足に着目し、②テクノロジーを取り上げ、③生活の向上に言及をしているものの、日本では約24%、韓国では約15%の受検者が③に言及できている。③は、設問9では「衰え」の言い換えであり、設問8では「拡張」の部分の言い換えることとなる。石井他(2010)は、設問8と9における項目特性値の違いの原因について、1. 設問の細かい文言の違い、2. 解答類型における要件の違い、3. 本文における記述のされ方の違い、の3点を挙げている。これを踏まえると、設問9では③に言及でき、設問

8ではできていない理由として、「衰え」と比べて「拡張」という、1. 設問の細かい文言の違いが項目特性値の違いの原因である可能性が推察される。したがって、「衰え」に合わせて「拡張」を「拡がり」に変換し、「人間の外的な身体能力の拡張」から「人間の外的な身体能力の拡がり」と設問中の文言を変えた設問を作成した。

以上のように、研究1では分析結果及び受検者の回答に基づき設問の改訂を行った。しかしながら、設問中の文言の違いが、実際に受検者の回答に影響を及ぼすかについては検証していない。したがって、研究2では、研究1で設定した設問文中の言葉の難しさの差異が、受検者の能力評価にどのような影響を及ぼすかを実証的に比較検討していく。

研究2

方法

国語読解テスト 研究1と同様に「群馬県児童生徒学力診断テスト」の国語読解テスト(記述式12問、選択式6問、合計18問)を用いた。

設問操作 設問8では、「拡張」を「拡がり」という単語に変えることで、言葉の難しさを操作した。具体的には、O:「人間の外的な身体能力の拡張」の具体例を記述する自由記述式問題、N:「人間の外的な身体能力の拡がり」の具体例を記述する自由記述式問題、の2つを設定した。これらの条件を踏まえ、A, B, C, D, E, F, G, Oと8種類の問題冊子を作成した。問題冊子における条件をTable 6に示す。

実施 テストは2009年5月~6月中旬に、愛知県と京都府の中学3年生703名に実施された。Aは85名、Bは88名、Cは83名、Dは87名、Eは85名、Fは92名、Gは95名、Oは88名が受検した。問題冊子はクラス単位ではなく、ランダムに個々の生徒に割り当てられた。解答時間は、問題冊子の配布と記名に要する時間を含め1単位時間の50分で実施された。テストは、実施中学校の教師により指示書に従って実施された。評定は、本稿の第一筆者が解答類型に従って採点を行い、類型の判断に迷う箇所については教育測定学及び計量心理学を専門とする大学教員1名と合議のうえ評定を行った。

分析 研究1と同様、項目分析を用いた(Crocker & Algina, 1986; Ebel & Frisbie, 1991; Hogan, 2007; 石井, 2007; Kubiszyn & Borich, 2009; Musial et al., 2009; 野口, 1985)。具体的には、得点率及び識別力を算出し、設問ごとにその差の検討を行った。差の検討に関しては、統計的検定ではなく統計的推定を用いた。その際、ボンフェロニ法を用いて95%信頼区間(95%CI)を算出した。本研究では、識別力指標としてD指標とI-T相関係数を用

Table 6 問題冊子ごとの条件

設問8	問題冊子							
	A	B	C	D	E	F	G	O (原本)
	O	O	N	N	N	N	O	O

O：オリジナルの条件 N：改訂した条件

Table 7 共通項目合計得点における問題冊子及び性別の平均値 (SD)

	問題冊子								
	A	B	C	D	E	F	G	O	合計
男性平均	5.75	6.00	5.61	5.58	6.36	5.71	5.87	6.66	5.94
(SD)	(2.15)	(2.11)	(2.31)	(2.18)	(2.16)	(2.23)	(2.05)	(1.48)	(2.11)
N	43	43	39	49	46	52	39	43	354
女性平均	6.67	6.53	6.37	6.69	6.61	6.90	6.31	6.57	6.57
(SD)	(1.80)	(1.74)	(1.89)	(1.70)	(1.62)	(1.92)	(1.74)	(1.37)	(1.72)
N	42	45	44	38	39	40	56	45	349
合計	6.20	6.27	6.01	6.06	6.47	6.23	6.13	6.61	6.25
(SD)	(2.03)	(1.94)	(2.12)	(2.05)	(1.93)	(2.17)	(1.87)	(1.42)	(1.95)
N	85	88	83	87	85	92	95	88	703

いた。D指標とは、共通項目合計得点により3群（低群、中群、高群）に分けられた高群の得点率と低群の得点率の差である。3群の分け方は、古典的テスト理論で用いられる方法に基づき、低群約27%、中群約46%、高群約27%の割合で行った（Brown, 1996 和田訳 2003；Crocker & Algina, 1986；Ebel & Frisbie, 1991；Hogan, 2007）。具体的には、共通項目合計得点を算出し、その分布を低群27%、中群46%、高群27%に分割する境界値を求め、その値を基準に受検者を3群に分けた。また、I-T相関係数を算出する際、研究2で用いられたテストは設問8以外の設問も操作していたため、各問題冊子に共通している9つの設問（1, 3, 4, 5, 6, 7, 9, 14, 15）の得点を共通合計得点として用いた。I-T相関係数は項目正答率の影響を受けにくい性質を持ち、D指標は項目正答率との関連が比較的強いという性質を持つ（石井, 2007；石井他, 2010）。さらに、性別及び群別の得点率をグラフ化した分析も行った。この手法は理論的には古典的であるが、各群における特性を捉えやすく、小規模データにも適用可能という利点を持つ（Council of Europe, 2004；石井他, 2010）。

結果と考察

まず、問題冊子間で受検者能力に差がないことを確認するため、問題冊子及び性別の検討を行う。問題冊子と性別に関する平均値を Table 7 に示す。共通項目合計得

点の平均（標準偏差, 以下SD）は、男性が5.94（SD = 2.11）、女性が6.57（SD = 1.72）であった。性別（男性、女性）×問題冊子（A, B, C, D, E, F, G, O）の2×8の二要因分散分析を行ったところ、性別の主効果が有意であり（ $F(1, 687) = 19.20, p < .01$ ）、問題冊子の主効果は有意でなかった（ $F(7, 687) = 0.99, n.s.$ ）。これにより、性別の平均値において、男性よりも女性の方が高いことが示された。日本においては、一般的に中学生の国語力は男性よりも女性の方が高いことが多くの大規模テストにより示されており（国立教育政策研究所（編）、2002, 2004, 2007, 2010；国立教育政策研究所教育課程研究センター、2003）、本研究においても同様の結果が示された。また、各問題冊子の平均値に関しては、各問題冊子間に差が認められなかったことから、それぞれの問題冊子間における受検者能力に差があるとは言えないと判断された。

次に、設問の検討を行う。Oは改訂前の「人間の外的な身体能力の拡張」、Nは改訂後の「人間の外的な身体能力の拡がり」である。Oは356名、Nは347名が受検した。OとNの得点率及び識別力を Table 8 に示す。得点率に関して、Oは0.185（SD = 0.295）、Nは0.147（SD = 0.249）であり、その差は-0.038ポイント（95%CI:-0.093~0.017）であった。識別力に関して、Oは0.315、Nは0.339であり、その差は0.033ポイント（95%CI:-0.105~0.177）であった。設問8における各群及び性別の得点率を Figure 1 に示す。

Table 8 設問8におけるOとNの項目分析結果

設問8	各群の平均得点率			得点率	識別力	
	低群	中群	高群	《難易度》	《D指標》	《I-T相関係数》
O (N = 356)	.038 (.157) N = 92	.210 (.303) N = 180	.292 (.331) N = 84	.185 (.295)	.254 [.149, .358]	.315 [.218, .406]
N (N = 347)	.041 (.117) N = 103	.154 (.251) N = 149	.250 (.303) N = 95	.147 (.249)	.209 [.114, .304]	.326 [.229, .417]

() 内の数字はSD [] 内の数字は95%信頼区間

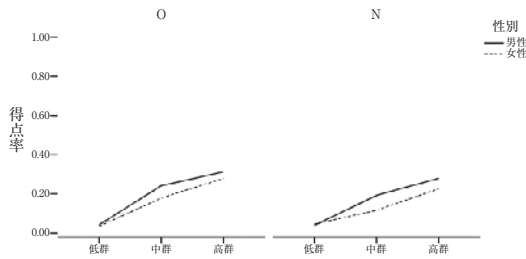


Figure 1 各群及び性別における得点率

OとNにおいて、群別及び性別による大きな差は見られなかった。

設問8は、設問9と同じ問題構造でありながら、設問9よりも得点率が非常に低かった。そのため、設問9の設問文との比較から設問文中の文言を操作し、受検者の回答に影響を及ぼすかを検討した。その結果、具体例を記述する設問において、設問文中の単語の難しさは得点率及び識別力に影響を及ぼさないことが示唆された。

以下では、設問9と得点率の違いを生じさせる要因について考察する。設問8と9に対する回答は、本文の内容から具体例を記述しても、自分で具体例を考えて記述しても、求められている3つの条件を満たしていれば正答となる。その3つの条件とは、設問8は①人間の手や足に着目し、②テクノロジーを取り上げ、③生活の向上に言及することである。一方、設問9は①手や足など身体能力に着目し、②テクノロジーを取り上げ、③身体能力の衰退に言及することであった。設問9の「人間がもつ内的な身体能力の衰え」に回答する際、本文の内容から回答を考えると、第3段落に内的な身体能力の衰えの例が複数記述されている。受検者はこの例を見つけて回答すれば、3つの条件を全て満たすこととなる。これに対し、設問8は、「人間の外的な身体能力の拡張」の例が本文中にいくつかあるものの、例をそのまま抜き出しただけでは条件のうちの2つしか満たさず、3つ満たすためには第1段落の内容を理解しまとめることが要求される。つまり、設問9では、PISAでいうところの「情

報の取り出し」で正答できるが、設問8では「解釈」をしなければ正答とはならないことを表している。「情報の取り出し」とは、本文を正確に理解したうえで、テキストの中の情報を正確に取り出すことであり、「解釈」とは、本文を正確に理解したうえで、本文を根拠にして推論し、本文との関わりを明らかにして自分独自の解釈を述べることである(有元, 2005)。PISAにおいて、日本の生徒は「情報の取り出し」はできるものの、「解釈」を苦手としていることが示されており(国立教育政策研究所(編), 2002, 2004, 2007)、本研究の結果も読解プロセスの観点からはPISAと類似した結果となった。これを踏まえると、設問に回答する際に求められる読解プロセスの差異が受検者の回答に影響を与える可能性が考えられる。

本研究では、設問中の文言の違いが得点率の差異を生じさせていると予測し、実証的に検討を行ったが、設問8と9における得点率の違いは、回答する際に要求される読解プロセスの違いが原因であると推察された。つまり、石井他(2010)が項目特性値の違いの原因として挙げた、1. 設問の細かい文言の違い、2. 解答類型における要件の違い、3. 本文における記述のされ方の違い、における3番目の可能性が考えられた。

しかしながら、これらは推測の段階にとどまっているため、実際に設問が要求する読解プロセスを操作し、検証する必要性が考えられる。今後は、本文中に設問の解答となる具体例を一文入れた本文と、入れない本文を作成し、設問が要求する読解プロセスの違いが受検者の回答傾向に及ぼす影響を検討する予定である。

引用文献

- 有元秀文(2005). PISA 読解力調査でわかったこと. 「中等教育段階における多面的教育測定—PISAを超えて—」. 東京大学大学院教育学研究科教育測定・カリキュラム開発講座, 公開シンポジウム報告書, *Sokutei Report*, 1, 16-29.
- Brown, J.D. (1996). *Testing in language programs*.

- Prentice-Hall, Inc.
(和田 稔(訳) (2003). 言語テストの基礎知識—正しい問題作成・評価のために—. 第2版 大修館書店)
- Council of Europe. (2004). *Reference supplement to the relating language examinations to the common European framework of reference for languages: Learning, teaching, assessment*. Strasbourg: Language Policy Division, Council of Europe.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart, and Winston.
- Ebel, R.L., & Frisbie, D.A. (1991). *Essentials of educational measurement* (5th ed.). Englewood Cliff, N.J.: Prentice Hall.
- 群馬県総合教育センター (2007). 思考のつまずきに視点を当てた授業改善策の提言—『群馬県児童生徒学力診断テスト』結果の分析を通じて—。「児童生徒学力向上調査研究」報告書。
- Hogan, T.P. (2007). *Educational assessment – A practical introduction –*. Hoboken, N.J.: John Wiley & Sons.
- 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・安永和央・金 由那・千 鎬誠・朴 賢晶・李 正連 (2010). 読解テストへの応答傾向に関する日韓比較研究—群馬県児童生徒学力診断テストを用いた項目分析的検討—. 名古屋大学大学院教育発達科学研究科中等教育研究センター紀要, 10, 29-43.
- 国立教育政策研究所 (編) (2002). 生きるための知識と技能—OECD 生徒の学習到達度調査 (PISA) 2000年調査国際結果報告書—. ぎょうせい.
- 国立教育政策研究所 (編) (2004). 生きるための知識と技能—OECD 生徒の学習到達度調査 (PISA) 2003年調査国際結果報告書—. ぎょうせい.
- 国立教育政策研究所 (編) (2007). 生きるための知識と技能—OECD 生徒の学習到達度調査 (PISA) 2006年調査国際結果報告書—. ぎょうせい.
- 国立教育政策研究所 (編) (2010). 生きるための知識と技能—OECD 生徒の学習到達度調査 (PISA) 2009年調査国際結果報告書—. 明石書店.
- 国立教育政策研究所教育課程研究センター (2003). 平成13年度小中学校教育課程実施状況調査報告書—中学校 国語—. ぎょうせい.
- Kubiszyn, T., & Borich, G. (2009). *Educational testing & measurement: Classroom application and practice*. Hoboken, N.J.: John Wiley & Sons.
- Loftus, E.R. & Suppes, P. (1972). Structural variables that determine problem-solving difficulty in computer-assisted instruction. *Journal of Educational Psychology*, 63, 531-542.
- Musial, D., Nieminen, G., Thomas, J., & Burke, K. (2009). *Foundations of meaningful educational assessment*. New York: McGraw-Hill.
- 野口裕之 (1985). テストをテストする. 海保博之 (編). 心理・教育データの解析法10講—基礎編—. 福村出版. pp.67-83.
- 安永和央・石井秀宗 (2010). 図の提示法が能力評価に及ぼす影響に関する一考察—中学国語読解テストを用いて—. 日本テスト学会誌, 6, 87-101.

謝辞

本論文は、第一筆者が名古屋大学大学院教育発達科学研究科に提出した修士論文の一部を再分析し、加筆修正したものです。論文の執筆にあたり、名古屋大学大学院教育発達科学研究科の高井次郎先生には有益なご助言をいただきました。深く感謝いたします。また、群馬県児童生徒学力診断テストに携っておられた、元群馬県総合教育センター所長の飯野眞幸先生、元主任指導主事の岡野健先生に深謝致します。最後に、本研究にご協力いただきました多くの先生方及び生徒のみなさまに、心よりお礼申し上げます。

(2011年9月30日受稿)

ABSTRACT

The effects of difficulty of item wording on a Japanese language reading test:
Open-ended questions requiring writing concrete examples

Kazuhiro YASUNAGA and Hidetoki ISHII

The purpose of this study was to examine if the difficulty of item wording on a Japanese language reading test affects item difficulty and item discrimination. We altered the difficulty of item wording in open-ended questions that required writing concrete examples. Junior high school students in South Korea (N = 232) were administered a translated version of the Gunma Prefecture Achievement Test (GPAT), for the purpose of comparison with the 2006 GPAT data gathered in Japan. Item analysis was used in this study. We examined the specific effects of using the more difficult expression of “the expanse of man’s external physical performance” (condition O) versus the simpler “the expanding of man’s external physical performance” (condition N) had any effects on responses. Students in Japanese junior high school (N = 703) were assigned to one of condition O or N. It became obvious that difficulty and discrimination did not vary across the two conditions. It was suggested that it is not so much the difficulty of item wording, but the existence or non-existence of concrete examples in original text that affects item difficulty.

Key words: Japanese language reading test, item analysis, item difficulty, item discrimination, the difficulty of item wording