

論文

国語読解テストにおける設問設定が能力評価に及ぼす影響

——特色ある読解教育を実践している小中一貫教育校の事例

安永 和央^{*,**}, 石井 秀宗^{*}

^{*}名古屋大学大学院教育発達科学研究科

^{**}日本学術振興会特別研究員

The effects of item presentation style on a Japanese language reading test: The case of an elementary-middle school with particular reading instruction

Kazuhiro Yasunaga ^{*,**}, Hidetoki Ishii ^{*}

^{*} Graduate school of Education and Human Development, Nagoya University

^{**} Research Fellow of the Japan Society for the Promotion of Science

The purpose of this study was to examine if the manner of presentation of questions on a Japanese language reading test affects item difficulty and item discrimination for students attending schools combined through both elementary and middle school levels. A total of 169 students were administered the Gunma Prefecture Achievement Test. Students were given 45 minutes to do the test, and they were graded based on an answer rubric. Four variations of item presentation were used: 1) positioning of blank spaces (left or right) and hint (beside or underneath), 2) item format of choosing a sentence (multiple choice or free-response), 3) number of blanks and figure presentation style (with or without an equal sign between blank spaces), 4) inclusion/exclusion of descriptive phrase before a blank. Item difficulty and item discrimination of items were analyzed. For 1), items composed of two blank spaces presented in the same order as in the original text with hints underneath them were higher in discrimination. Furthermore, when placing the hint beside the blank space, discrimination was seen to be higher when hints were placed in the opposite order from what had been presented in the text. It was concluded that item presentation methods do affect their discriminative ability, and should be examined more closely.

Keywords : Japanese Language Reading Test, Item Analysis, Item Difficulty, Item Discrimination, Item Presentation Style

キーワード : 国語読解テスト, 項目分析, 得点率, 識別力, 設問設定

1. はじめに

日常生活において, 最初はある質問に答えられなかったが, 問われ方が変わることで答えられるようになった, という経験はないだろうか。また逆に, 尋ねられる方法が変わったことで, かえってわかりにくくなったという経験もあ

*〒464-8601 名古屋市千種区不老町 名古屋大学大学院教育発達科学研究科

Correspondence concerning this article should be sent to: Kazuhiro Yasunaga, Graduate school of Education and Human Development, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601 Japan.

E-mail: yasunaga.kazuhiro@d.nagoya-u.jp

るだろう。このように、同じ質問内容でも、その問い方によって答えやすさが変わるといことがしばしばある。これは、会話に限ったことではなく、テストのような回答を求められるものにも当てはまる。

これまで、テストにおける設問の問い方が、受検者の回答に及ぼす影響を検討した研究が行われてきた (Ebel & Frisbie, 1991 ; 石井, 2009 ; 安永・石井, 2010)。Ebel & Frisbie (1991) は、多枝選択式問題において、設問の問い方及び選択枝の差し替えにより、得点率や識別力が変化することを示した。また、石井 (2009) は、記述式問題を教科横断的に検討し、要求される回答の質や量の違いによる設問設定の差異が、無回答率に影響することを明らかにした。そして、安永・石井 (2010) は、空所に相対する 2 つの単語を記述する設問において、本文に出てくる単語の順序通りに空所を配置することが、得点率及び識別力を高くすることを明らかにした。その際、ヒントは空所の横に配置することが性差を生じさせないことも示唆された。これらは、設問のわずかな設定の違いが受検者の回答傾向に影響を与えることを示唆している。

このように、テストにおいても設問の問い方が受検者の回答に影響することが明らかにされている。しかしながら、わが国で用いられているテストは、設問設定についてそれほど慎重に検討されていない。その理由として、わが国には、テスト研究が非常に少なく、テスト作成に関する知見が蓄積されていないことが挙げられる。そのため、設問を実証的に検討する必要性が周知されておらず、わが国のテストの多くは、教科の専門家や現場で働く教師の経験のみに頼って作成されている。受検者の能力が、経験的に作成されたテストで適切に測定されるのであれば問題はない。しかしながら、教師や教科の専門家は、テストの項目得点率をある程度予測できるが、その個人差は大きく、実測値ほど正確ではないことが示されている (Hambleton & Jirka, 2006 韓・林訳 2008; Impara & Plake, 1998; Lorge & Kruglov, 1953; Quereshi & Fidher, 1977)。これを踏まえると、受検者の能力を適切に測定するテストを作成するためには、教師や教科の専門家の経験に加え、実際にテストを実施し、設問そのものを検討する必要性も考えられる。また、先に述べたように、わが国にはテスト作成についての知見が少ないことから、様々な条件のもとで設問に関する実証的な検討を行い、その成果をテスト作成に資する知見として蓄積していくことが求められる。

したがって、本研究では、国語読解テストの複数の設問を操作し、それらが受検者の回答に及ぼす影響を検討する。本研究では、小中一貫教育校の生徒を対象とする。小中一貫教育校の中には、国語教育 (読解教育) に力を入れてい

るところがあり、その影響を検討する。

これらを検証するためには、テストが実施される時点で、ある程度適格なものであるという保証が必要である。そこで、本研究では、群馬県教育委員会により作成された群馬県児童生徒学力診断テストの国語読解テストを用いて検討を行う。また、一部の設問においては、安永・石井（2010）と同一の設問操作を行い、異なる受検者に対する設問設定の影響を検討し、テスト作成に資する知見を得ることを目的とする。

2. 方法

(1) 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。問題本文の要旨を表 1 に示す。

表 1 国語読解テスト問題本文の要旨

-
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」の欲望に突き動かされて、利便をもたらすさまざまな新製品がテクノロジーによって開発された。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たな可能性を獲得した。「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。
 2. 現代においては、感性と知的能力の順序が逆転し、テクノロジーという知的能力が感性を支配し始めているのだ。いっそう徹底すれば、人間が自然から切り離され、テクノロジーの中でしか生きている実感を持たなくなってしまうだろう。
 3. テクノロジーにより人間の外的な身体能力を拡張したが、人間がもつ内的な身体能力は衰えた。
 4. 人間は、自然界に適応しながら生き残ってきた「ひと」の側面と、テクノロジーを始めとする文化の創始者としての「ヒト」の側面を持っている。今「ヒト」の側面が突出し過ぎているが、動物としての人間が持つ自然への適応性は欠かすことができない。「ヒト」と「ひと」をいかに調和させるかが、21 世紀の大きな課題である。
 5. そのためのひとつのヒントは、新しいテクノロジーとつきあうとき、身体能力の何が失われるかを考える癖を持つことだ。便利になるということは、「ひと」としての能力の喪失につながるからだ。
 6. 新たな可能性が拓かれるなら良いのではないかという考え方もあるが、やはり、失われるかもしれない能力と、新たに獲得できるかもしれない可能性を秤にかけ、長い時間スケールをかけて得失を判断しなければならない。失われた「ひと」の能力は、「ヒト」が作り出したテクノロジーだけでは完全に代替できないのだ。

池内 了 「科学は今どうなっているの？」 より (1,691 字)

表 2 各設問の原本

設問	原本
10, 11	図1の空所a, bに当てはまる言葉を書きなさい。(2字)
12	図1と図3の空所cに当てはまる文を, 5段落(表3)の中の(ア)～(カ)から見付け, 記号を書きなさい。
13	図1と図3の空所dに当てはまる内容を, 6段落の語句を使って書きなさい。(30字程度)
16	図4の(㉔)の中に当てはまる内容を書きなさい。(30字程度)

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。本テストは, 問題文を読んで各設問に答えるという日本の中学校で一般的に実施されている国語テストの構成となっていた。

本研究では, 設問 10, 11, 12, 13, 16 の検討を行った。各設問の原本を表 2 に示す。設問 10, 11 は, 2 つの空所に相対する「ヒト」と「ひと」という 2 つの単語を補充する求答形式問題である。設問 12 は, 問題本文について話し合ったことをまとめた会話文の空所に入る一文を, 第 5 段落にある 6 つの文の中から選択する多枝選択式問題である。設問 13 は, 第 6 段落において, 問題本文の筆者が断言していることを, 図及び会話文の空所に第 6 段落の語句を使って記述する自由記述式問題である。設問 16 は, 『ひと』としての側面と『ヒト』としての側面を調和させていくためには, (㉔) が大切である」という文の (㉔) の部分を記述する自由記述式問題である。

なお, 本研究では, 多枝選択式問題を「正答が 1 つに定まっており, 多数の選択枝の中から 1 つを選択して答える問題」, 求答式問題を「正答が 1 つに定まっており, 短い語句で答える問題」, 自由記述式問題を「正答が 1 つに定まっておらず, 本文やテキスト, 図について, 様々な知識や思考力などを組み合わせる自由記述する問題」とする。

(2) 設問設定

具体的な設問設定は, 以下に示す通りである。O は原本の設問, N は新たに設定した設問を示す。設問 10, 11 では, 安永・石井(2010)と同様の操作を行った。すなわち, 相対する単語が入る空所の位置(本文と同じ順序・本文と反対)とヒントの位置(空所の横・空所の下)を操作した。具体的には, O :

人類に関する空所を本文と反対にした図，N1：空所を本文と反対にして，「ひと」のヒントを b の下に配置した図，N2：人類に関する空所を本文通りに提示した図，N3：空所を本文通りに提示し，かつ「ひと」のヒントを a の下に付けた図，の4つを設定した。これにより，空所の位置及びヒントの位置が受検者の回答傾向に及ぼす影響を検討する（図1）。

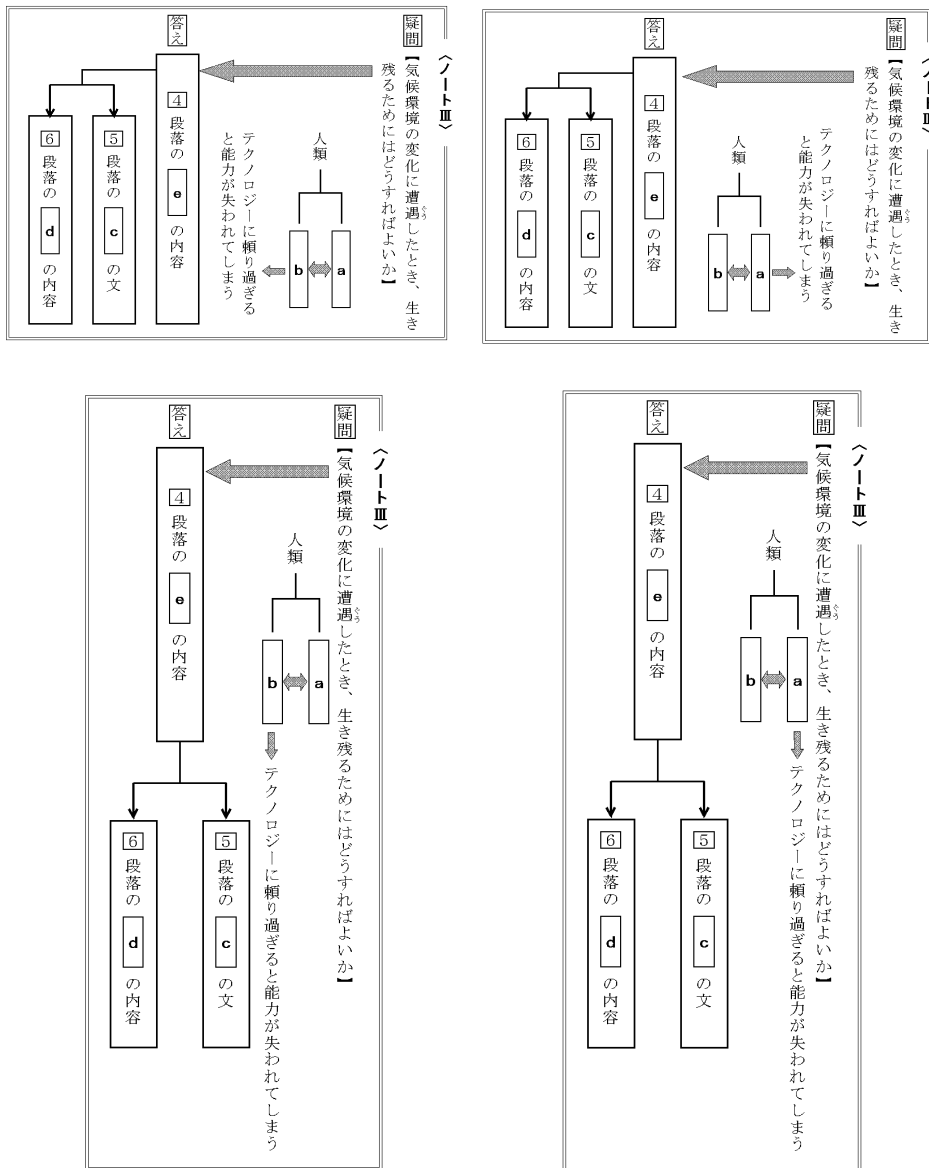


図1 設問 10, 11 における空所の位置及びヒントの位置を操作した図
(左上：0，左下：N1，右上：N2，右下：N3)

表 3 設問 12 において操作した第 5 段落の構造 ^{a)}

<p>5. (ア) そのための一つのヒントは、新しいテクノロジーとつきあうとき、これを使えば自分の持つ身体能力の何が失われていくかを考える癖を持つことではないだろうか。(イ) 便利になるということは、体のどこかを動かさなくなるからだ、必ず「ひと」としての能力の喪失につながるからだ。(ウ) ワープロばかり使っていると、漢字を覚える能力を失っていく。(エ) GPS (全地球測位システム) を使ったカーナビ頼りになると全体的な方向感覚や土地勘が失われていこう。(オ) ロボットに家事をやらせるようになると、包丁の使い方を忘れ、舌は微妙な味を区別できなくなる。(カ) 五万年のホモ・サピエンスの歴史で獲得してきた「ひと」としての能力が衰えていくのだ。</p>

^{a)} Nでは、文頭の記号のないものを提示した

設問 12 では、設問形式を操作した。具体的には、O：第 5 段落の 6 文の文頭に (ア) ~ (カ) の記号を振り当て記号を選択する選択式問題、N：実際に第 5 段落から一文抜き出す記述式問題、の 2 つの設問形式を設定した (表 3)。これにより、設問形式の違いが受検者の回答傾向に及ぼす影響を検討する。

設問 13 では、空所の数 (多い・少ない) と図の提示 (「=」あり・「=」なし) を操作した。具体的には、O：会話文中の空所の数が多く、図に「=」のないテキスト、N1：空所の数が多く、図に「=」のあるテキスト、N2：空所の数が少なく、図に「=」のないテキスト、N3：空所の数が少なく、図に「=」のあるテキスト、の 4 つを設定した (図 2, 3)。これにより、空所の数及び図の提示法の違いが受検者の回答傾向に及ぼす影響を検討する。

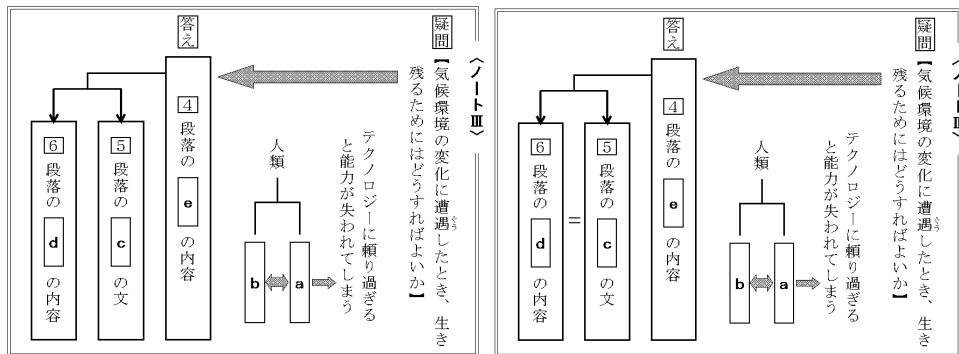


図 2 設問 13 における図の提示を操作したテキスト (左：O、右：N)

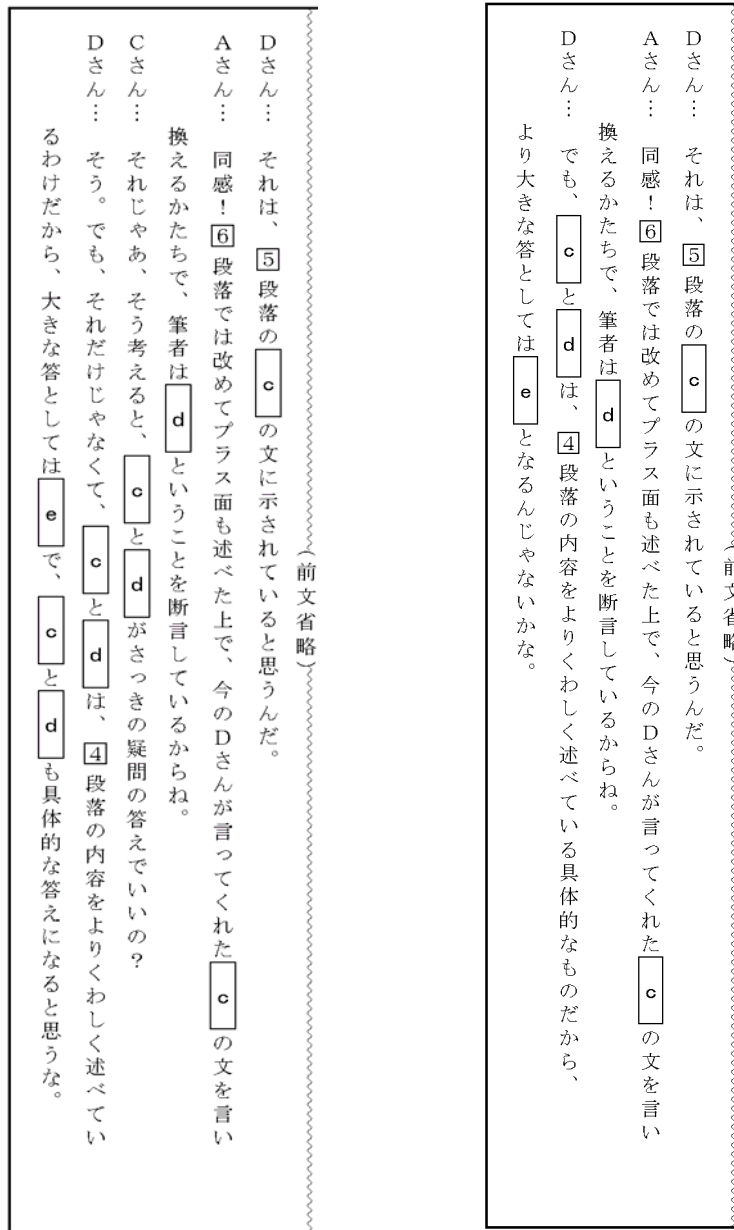


図3 設問13における会話文の空所数を操作したテキスト(左:O,右:N)

設問16では、空所前の単語における説明の有無を操作した。具体的には、O:「ひと」と「ヒト」の説明が空所前がないテキスト、N:「ひと」と「ヒト」の説明が空所前にあるテキスト、の2つを設定した(図4)。これにより、単

<p style="text-align: center;">(ノートⅣ)</p> <p>人類は、自然界に適応しながら生きてきた「ひと」の側面と、知的能力によりテクノロジーを生み出した「ヒト」の側面とを持っている。現代の人間は、その「ヒト」が生み出すテクノロジーに支配されるようになってきた。しかし、人間は、テクノロジーとだけつきあっているだけでは生きていけない。なぜなら、テクノロジーにばかり頼っていると、() ①) からである。</p> <p>そこで、これからの人類が「ひと」としての側面と「ヒト」としての側面とを調和させていくためには、() ②) が大切である。</p>	<p style="text-align: center;">(ノートⅣ)</p> <p>人類は、自然界に適応しながら生きてきた「ひと」の側面と、知的能力によりテクノロジーを生み出した「ヒト」の側面とを持っている。現代の人間は、その「ヒト」が生み出すテクノロジーに支配されるようになってきた。しかし、人間は、テクノロジーとだけつきあっているだけでは生きていけない。なぜなら、テクノロジーにばかり頼っていると、() ①) からである。</p> <p>そこで、これからの人類が自然界に適応しながら生き残ってきた「ひと」としての側面とテクノロジーを生み出してきた「ヒト」としての側面とを調和させていくためには、() ②) が大切である。</p>
--	---

図4 設問16における空所(②)前の単語の説明を操作したテキスト
(左:O, 右:N)

語に関する説明の繰り返し、受検者の回答傾向に及ぼす影響を検討する。

直交表に基づき、設問10, 11, 12, 13, 16を要因が交絡しないように組み合わせる8種類の問題冊子を作成した。直交表とは、要因の数が増えるとその水準の組み合わせが多くなるが、その際、最小の組み合わせで各要因が交絡しないように実験計画を組む方法である(鷺尾, 1974)。直交表に基づいた8種類の問題冊子A, B, C, D, E, F, G, Oの構成を表4に示す。

表 4 直交表に基づいた問題構成

	問題冊子							
	A	B	C	D	E	F	G	O (原本)
設問10,11	左下	右横	右下	左横	左下	右横	右下	左横
設問12	N	O	N	O	N	O	N	O
設問13-1	O	N	O	N	N	O	N	O
設問13-2	N	N	O	O	N	N	O	O
設問16	N	O	O	N	O	N	N	O

Oは原本を表す Nはそれぞれの改訂した条件を表す

設問10,11 左横：左の横にヒントを配置，左下：左の下にヒントを配置，

右横：右の横にヒントを配置，右下：右の下にヒントを配置（図1）

設問12 N：問題本文から一文抜き出す問題（表3）

設問13-1 N：空所dとcの間に=を入れたテキスト（図2）

設問13-2 N：空所の数を減らしたテキスト（図3）

設問16 N：空所の前に単語の説明を加えたテキスト（図4）

（3）受検者及び実施方法

2009年6月に小中一貫教育校に通う中学3年生169名が受検した。Aは22名、Bは21名、Cは19名、Dは21名、Eは21名、Fは21名、Gは21名、Oは23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45分であった。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。評定については、本稿の筆者が解答類型を用いて採点した。また、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と合議のうえ評定を行った。

（4）調査校について

調査校は、2つの公立小学校の児童が進学してくる小中一貫教育校である。同校では、義務教育9年間を一貫して捉えており、義務教育の区切りを小学校6年間、中学校3年間の6・3制から5・4制の区切りに見直す新たな取り組みを導入している。同校では、読解力を各教科、領域に必要なものと捉え、「読解科」という新教科を立ち上げ、読解科を全ての学年で週1時間、年間35時間設定している。読解科では、情報を正しく「読み取り」、それを自ら十分に考えた上で「意見を持ち」、その自分の意見を相手に「伝える」ことを念頭に指導が行われている。教材資料は、文章に限らず、グラフ、音声、映像など、広く日常的なものが用いられており、映像資料を使うときには、映像を単に見るだけにとどめず、文章と比較しながら授業の展開が行われている。

また、国語科では、学習課題を明確にして、情報に含まれる事実を正確に見極めた上で論理的に思考し、表現する力、相手や目的に応じて言葉で伝え合う力の育成を重視している。文章を「読む」ことに加え、報告や討論といった「話す・聞く」場面や、立場を決めて文章を「書く」場面を設けた指導も行われている。

これらの授業は、板書中心の一斉授業だけではなく、複数の場面でグループ学習が取り入れられている。グループ学習は4人を基本としながら、話し合う内容によっては、2人、6人、12人と人数を調整し、討議の司会や報告を全生徒が担当することが可能である。本研究で対象とした小中一貫教育校では、以上のような特色ある読解教育の実践を通して、論理的に考える力、表現する力、わかりやすい言葉で相手に伝える力の育成を目指している。

(5) 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 (Hogan, 2007; 石井, 2007; Musial, Nieminen, Thomas, & Burke, 2009; 野口, 1985)。具体的には、得点率 (難易度)、識別力 (弁別力) を算出した。得点率及び識別力における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%同時信頼区間 (95%CI) を算出した。

(6) 各指標の説明

本研究の難易度は、得点率である。各設問の解答類型に従い、回答が正答であれば1点、準正答であれば回答内容に基づき0.25点または0.5点、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、全受検者の平均得点率を算出した。この指標は、値が1に近いほど易しく、0に近いほど難しいことを示す。本研究の識別力は、I-T相関係数とD指標を用いた。I-T相関は、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である。本研究では、複数の設問を操作しているため、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。D指標は、低群、中群、高群と3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。この識別力は、-1~1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 (石井, 2007)。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; 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。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・安永・金・千・朴・李, 2010)。なお、以下では、解釈のしやすさを考慮し、設問ごとに検討を行う。

3. 結果と考察

(1) 問題冊子及び性別の検討

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。表 5 に、共通項目合計得点における各問題冊子及び性別の平均値 (標準偏差, 以下 *SD*) を示す。共通項目合計得点の平均は、男性が 6.98 (*SD* = 1.39), 女性が 6.91 (*SD* = 1.71) 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 ($F(1, 153) = 0.19, n.s.$), 問題冊子の主効果 ($F(7, 153) = 1.72, n.s.$) は、ともに有意ではなかった。これにより、性別における平均値に関して、男女で差があるとは言えないと判断された。OECD が実施した PISA 調査において、読解力は 2003 年調査のリヒテンシュタイン公国を除く全ての参加国で、男性より女性の方が高いことが示されている (国立教育政策研究所 (編), 2002, 2004, 2007, 2010)。また、わが国の大規模調査においても、中学生の国語力は男性よりも女性の方が高いことが示されている (国立教育政策研究所教育課程研究センター, 2003)。しかしながら、本研究で対象とした小中一貫教育校では、国語力に性差がないことが明らかとなった。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。

表 5 共通項目合計得点における各問題冊子及び性別の平均値 (*SD*)

	問題冊子								合計
	A	B	C	D	E	F	G	O	
男性平均	7.19	6.29	7.63	6.41	6.97	6.19	7.30	7.66	6.98
(<i>SD</i>)	(1.47)	(1.63)	(.94)	(1.39)	(1.25)	(1.58)	(1.28)	(.79)	(1.39)
<i>N</i>	12	12	12	8	9	13	11	14	91
女性平均	6.78	7.64	6.57	7.02	7.40	5.59	7.23	6.58	6.91
(<i>SD</i>)	(1.85)	(1.07)	(2.26)	(1.51)	(1.45)	(2.47)	(1.50)	(1.45)	(1.71)
<i>N</i>	10	9	7	13	12	8	10	9	78
合計	7.00	6.87	7.24	6.79	7.21	5.96	7.26	7.24	6.95
(<i>SD</i>)	(1.63)	(1.54)	(1.59)	(1.46)	(1.35)	(1.93)	(1.36)	(1.20)	(1.54)
<i>N</i>	22	21	19	21	21	21	21	23	169

表 6 共通項目合計得点における群及び性別の平均値 (*SD*)

	低群	中群	高群	合計
男性平均(<i>SD</i>)	5.13 (.87)	7.38 (.55)	8.7 (.30)	6.98 (1.39)
<i>N</i> (%)	25 (27.5)	51 (56.0)	15 (16.5)	91 (100)
女性平均(<i>SD</i>)	4.70 (1.34)	7.26 (.59)	8.69 (.30)	6.90 (1.71)
<i>N</i> (%)	22 (28.2)	36 (46.2)	20 (25.6)	78 (100)
合計(<i>SD</i>)	4.93 (1.12)	7.33 (.57)	8.69 (.30)	6.95 (1.54)
<i>N</i> (%)	47 (27.8)	87 (51.5)	35 (20.7)	169 (100)

次に、群分けの詳細を確認する。共通項目合計得点に基づき、受検者を低群 (6.25 未満)、中群 (6.25 以上 8.25 未満)、高群 (8.25 以上) の 3 群に分割した。各群の受検者人数比は、低群 27.8%、中群 51.5%、高群 20.7%であった。表 6 は、共通項目合計得点における群及び性別の平均値 (*SD*) を表す。低群の平均は 4.93 (*SD*=1.12)、中群の平均は 7.33 (*SD*=0.57)、高群の平均は 8.69 (*SD*=0.30) であった。

(2) 設問 10, 11

O は 21 名、N1 は 43 名、N2 は 42 名、N3 は 40 名が受検した。表 7, 8 に、設問 10, 11 における得点率及び識別力を示す。なお、条件間で人数に違いがあるのは、安永・石井 (2010) の結果と比較するため、一部分析に含めない冊子があったことによる。まず、各条件の得点率と識別力について検討する。設問 10 における得点率の差は、N1-O が 0.157 ポイント (95%CI : -0.141~0.455)、N3-N1 が 0.180 ポイント (95%CI : -0.090~0.450)、N3-N2 が 0.146 ポイント (95%CI : -0.127~0.419)、N2-O が 0.191 ポイント (95%CI : -0.109~0.491) であった。設問 10 における I-T 相関係数の差は、N1-O が 0.12 (95%CI : -0.505~0.677)、N3-N1 が 0.026 (95%CI : -0.481~0.527)、N3-N2 が 0.44 (95%CI : -0.093~0.771)、N2-O が -0.294 (95%CI : -0.759~0.374) であった。設問 11 における I-T 相関係数の差は、N3-N2 が 0.256 (95%CI : -0.289~0.675)、N3-N1 が -0.133 (95%CI : -0.606~0.386) と異なるのみで、他の得点率の差及び I-T 相関係数の差は、ほぼ同じ値であった。これらの結果から、N2 と N3、N2 と O において、識別力に差が認められた。

これにより、相対する 2 つの単語の空所を本文に出てくる順序通りに提示した時、ヒントを空所の横に置くよりも下に置く方が識別力の値が高くなることが示唆された。また、ヒントを空所の横に置く場合は、空所を本文に記載されている順序通りに並べるよりも、本文の順序とは反対に並べる方が識別力の値

を高くする可能性が示された。次に、各群の得点率について検討する。設問10のN2において、低群の得点率は0.5（設問11は0.444）、中群の得点率は0.294となり、得点率の値が通常期待される順序と逆転した結果となった。2つの単語の空所を本文に出てくる順序通りに並べ、かつ、ヒントを空所の横に置いた場合、低群の受検者の得点率が中群の受検者よりも高くなり、テストの識別性としては望ましくない結果になりうる可能性が示された。

安永・石井（2010）は、一般中学校を対象に同様の検討を行い、本文に出てくる単語の順序通りに空所を配置した場合、ヒントの位置にかかわらず、得点率及び識別力の値が高くなることを報告した。本研究では、本文に書かれている単語の順序通りに空所を配置し、空所の下にヒントを置いた場合は、先行研究と同様の結果が得られたものの、ヒントを空所の横に配置した場合には識別力の値が低くなるという異なった結果が得られた。安永・石井（2010）で得られた知見と合わせて考えると、空所は本文に単語が出てくる順序通りに配置し、ヒントを空所の下に置くことにより、得点率及び識別力を高くすることがわかった。

表7 設問10における得点率及び識別力

設問10	各群の平均得点率			得点率 《難易度》	《D指標》	識別力 《I-T相関係数》
	低群	中群	高群			
O (N=21)	.000(.000) N=7	.333(.500) N=9	.400(.548) N=5	.238 (.436)	.400 [-.029, .829]	.260 [-.193, .622]
N1 (N=43)	.200(.422) N=10	.348(.487) N=23	.700(.483) N=10	.395 (.495)	.500 [.123, .877]	.380 [.090, .611]
N2 (N=42)	.500(.515) N=18	.294(.470) N=17	.571(.535) N=7	.429 (.501)	.071 [-.362, .504]	-.034 [-.335, .273]
N3 (N=40)	.250(.462) N=8	.591(.503) N=22	.800(.422) N=10	.575 (.501)	.550 [.161, .939]	.406 [.108, .637]

()内の数字はSD []内の数字は95%信頼区間

表8 設問11における得点率及び識別力

設問11	各群の平均得点率			得点率 《難易度》	《D指標》	識別力 《I-T相関係数》
	低群	中群	高群			
O (N=21)	.000(.000) N=7	.333(.500) N=9	.400(.548) N=5	.238 (.436)	.400 [-.029, .829]	.260 [-.193, .622]
N1 (N=43)	.200(.422) N=10	.348(.487) N=23	.700(.483) N=10	.395 (.495)	.500 [.123, .877]	.380 [.090, .611]
N2 (N=42)	.444(.511) N=18	.294(.470) N=17	.571(.535) N=7	.405 (.497)	.127 [-.306, .560]	-.009 [-.312, .296]
N3 (N=40)	.375(.518) N=8	.546(.510) N=22	.800(.422) N=10	.575 (.501)	.425 [.008, .842]	.247 [-.070, .519]

()内の数字はSD []内の数字は95%信頼区間

表 9 設問 12 における得点率及び識別力

設問12	各群の得点率			得点率 《難易度》	《D指標》	識別力 《I-T相関係数》
	低群	中群	高群			
O (N=86)	.483(.509) N=29	.667(.477) N=42	.800(.414) N=15	.628 (.486)	.317 [.045, .589]	.308 [.103, .488]
N (N=83)	.500(.514) N=18	.733(.447) N=45	.900(.308) N=20	.723 (.450)	.400 [.134, .666]	.287 [.076, .473]

()内の数字はSD []内の数字は95%信頼区間

(3) 設問 12

O は 86 名, N は 83 名が受検した。表 9 に, 設問 12 における得点率及び識別力を示す。得点率の差は, 0.095 ポイント (95%CI : -0.045~0.235), I-T 相関係数の差は, -0.021 (95%CI : -0.319~0.277) であった。これらの結果から, O と N の得点率及び識別力には差がないことが示された。これにより, 設問形式の違いは受検者の回答傾向に影響を及ぼさないことがわかった。

(4) 設問 13

O は 42 名, N1 は 42 名, N2 は 43 名, N3 は 42 名が受検した。表 10 に, 設問 13 における得点率及び識別力を示す。まず, 各条件の得点率と識別力について検討する。得点率の差は, N1-O が -0.114 ポイント (95%CI : -0.385~0.157), N3-N1 が 0.026 ポイント (95%CI : -0.245~0.297), N3-N2 が 0.080 ポイント (95%CI : -0.245~0.297), N2-O が -0.168 ポイント (95%CI : -0.435~0.099) であった。I-T 相関係数の差は, N1-O が 0.063 (95%CI : -0.441~0.557), N3-N1 が 0.050 (95%CI : -0.450~0.549), N3-N2 が 0.180

表 10 設問 13 における得点率及び識別力

設問13	各群の平均得点率			得点率 《難易度》	《D指標》	識別力 《I-T相関係数》
	低群	中群	高群			
O (N=42)	.250(.418) N=6	.603(.409) N=29	.750(.433) N=7	.551 (.432)	.500 [.028, .972]	.404 [.114, .631]
N1 (N=42)	.308(.370) N=13	.319(.418) N=18	.864(.323) N=11	.437 (.568)	.556 [.233, .879]	.467 [.190, .675]
N2 (N=43)	.309(.429) N=17	.441(.455) N=17	.611(.417) N=9	.383 (.649)	.302 [-.085, .689]	.337 [.041, .579]
N3 (N=42)	.136(.303) N=11	.630(.398) N=23	.750(.378) N=8	.463 (.434)	.614 [.252, .976]	.517 [.253, .710]

()内の数字はSD []内の数字は95%信頼区間

表 11 設問 16 における得点率及び識別力

設問16	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=84)	.353(.460)	.606(.478)	.867(.297)	.601	.514	.394
	N=17	N=52	N=15	(.471)	[.229, .779]	[.196, .561]
N (N=85)	.300(.447)	.643(.447)	.825(.373)	.565	.525	.468
	N=30	N=35	N=20	(.474)	[.291, .759]	[.283, .619]

()内の数字はSD []内の数字は95%信頼区間

(95%CI : -0.315~0.646), N2-O が-0.067 (95%CI : -0.555~0.438) であった。これらの結果から, O, N1, N2, N3 の得点率及び識別力には差がないことが示された。次に, 各群の得点率について検討する。O における低群は 0.25, 中群は 0.603, N3 における低群は 0.136, 中群は 0.63 であり, 他の条件に比べて低群と中群の間に高い識別性が認められた。また, N1 における中群の得点率は 0.319, 高群は 0.864 であり, 他の条件に比べて中群と高群の間に高い識別性が観測された。

以上より, 条件ごとの得点率と識別力を検討した場合, それらの間に差は見られなかった。しかしながら, 各条件の群ごとに検討を行うと, 識別性に違いがみられた。

(5) 設問 16

O は 84 名, N は 85 名が受検した。表 11 に, 設問 16 における得点率及び識別力を示す。得点率の差は, -0.036 ポイント (95%CI : -0.185~0.113), I-T 相関係数の差は, 0.074 (95%CI : -0.213~0.378) であった。これらの結果から, O と N の得点率及び識別力には差がないことが示された。これにより, 単語に関する説明の繰り返しは, 受検者の回答傾向に影響を及ぼさないことがわかった。

4. 総合考察

本研究では, 特色ある読解教育を実践している小中一貫教育校を対象に, 設問の問い方が受検者の回答傾向にどのような影響を及ぼすかを検討した。本研究で得られた知見を以下に述べる。

まず, 合計得点について検討する。これまで, 多くの大規模テストにおいて, 国語力(読解力)は男性より女性の方が高いことが示されてきた(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)。また, 安永・石井(2010)は, 一般中学校を対象に本研究

と同一の国語読解テストを用いた検討を行い、男性より女性の合計得点が高いことを示した。しかしながら、本研究では、性差が認められず、従来と異なる結果となった。

次に各設問について検討する。設問 10, 11 では、空所の位置及びヒントの位置が受検者の回答傾向に及ぼす影響を検討した。その結果、2つの知見が得られた。第 1 に、相対する 2つの単語の空所を本文に出てくる順序通りに提示した場合、ヒントは空所の横に置くよりも下に置くほうが、識別力の値を高くすることが示唆された。第 2 に、ヒントを空所の横に置いた場合、2つの単語の空所を本文に出てくる順序通りに提示するより、本文にでてくる順序とは逆に提示する方が識別力の値が高くなる可能性が示唆された。第 1 の結果に対し、安永・石井（2010）では、ヒントの位置にかかわらず、空所を本文に単語が出てくる順序通りに配置した条件において、高い得点率と識別力が得られた。本研究では、ヒントを空所の下に配置した場合は、同様の結果が得られたものの、ヒントを空所の横に配置した場合には識別力の値が低くなる可能性が示唆された。これらの結果を踏まえると、2つの空所に相対する単語を補充する求答形式問題では、空所は本文に単語が出てくる順序通りに配置し、ヒントを空所の下に置くことが得点率及び識別力に最も影響を与えることが示唆された。

設問 12 では、一文抜き出し問題における設問形式の違いが受検者の回答傾向に及ぼす影響を検討した。その結果、一文抜き出し問題の場合、設問形式は得点率及び識別力に影響を及ぼさないことが示された。記述式問題を選択式問題に代用できることは、評定者の採点における負担の軽減につながる知見であると考えられる。

設問 13 では、空所の数及び図の提示の違いが受検者の回答傾向に及ぼす影響を検討した。その結果、空所の数や図における「＝」の有無は、得点率及び識別力に影響を及ぼさないことが示された。しかしながら、群ごとに検討を行った結果、会話文中の空所の数が多く、図に「＝」のない条件のテキストと、空所の数が少なく、図に「＝」のある条件のテキストにおいて、低群と中群の間に高い識別性が認められた。また、空所の数が多く、図に「＝」のある条件のテキストにおいて、中群と高群の間に高い識別性が観測された。以上より、全体として検討した場合、条件間で得点率及び識別力に差は認められなかったが、群ごとに検討を行うと、条件によっては特定の群間において得点率に大きな差が見られ、高い識別性が観測された。

設問 16 では、空所前の単語を説明する語句の有無が受検者の回答傾向に及

ぼす影響を検討した。その結果、空所前の単語の有無は、得点率及び識別力に影響を及ぼさないことが示された。本研究では、テキストにおける、空所前の「ひと」と「ヒト」の説明を操作したが、これらの説明はすでにテキストの前半部に出てきており、一度出ている説明を再び繰り返すことは、受検者の回答に影響を及ぼさないことがわかった。したがって、設問作成の際、空所前の単語の説明が同一テキスト内で既に行われている場合は、その単語の説明に関してそれほど慎重に検討する必要はないことが示唆された。

本研究では、一部の設問において、わずかな設定の違いが受検者の回答傾向に影響を与えることが示された。冒頭で述べた通り、わが国にはテスト作成に資する知見が少ないため、本研究の結果はテスト作成の際、作成者が参考とする一つの指標となりえるであろう。ただし、本研究は、学校独自の特別なカリキュラムを実践している小中一貫教育校を対象にしており、安永・石井(2010)で検討された一般的な中学校よりも国語読解能力が相対的に高かった。そのため、設問10、11においては、安永・石井(2010)と若干異なった結果になったと推察される。これを踏まえると、他の設問に関しても、読解能力の高い受検者を対象としたことにより、設問操作の影響を検出できなかった可能性も考えられる。したがって、今後は本研究で得られた知見をより一般化するために、複数の中学校を対象に検討を進める予定である。

5. まとめ

わが国では、多くの場面でテストが用いられているにもかかわらず、テスト作成に関する研究蓄積が非常に少ない。そのため、テストの多くは、教科の専門家や現場で働く教師の経験のみに頼って作成されている。しかしながら、受検者の能力を適切に測定するためには、教科の専門家や現場で働く教師の経験のみでは十分ではなく、それらに加え、複数の設問について実証的に検討する必要性も考えられる。

したがって、本研究では、小中一貫教育校に通う中学3年生を対象に、群馬県児童生徒学力診断テストの国語読解テストを用いて、設問設定が受検者の回答に及ぼす影響を検討した。具体的には、1) 空所の位置(左・右)及びヒントの位置(横・下)、2) 一文抜き出し問題における設問形式(多枝選択形式・自由記述形式)、3) 空所の数(多・少)及び図の提示(「=」あり・「=」なし)、4) 空所前の単語を説明する語句の有無、を操作した。これらの設問を小中一貫教育校の中学3年生169名に実施した。項目分析の結果、1)において、相対する2つの単語の空所を本文に出てくる順序通りに提示した場合、ヒントは

空所の横に置くよりも下に置くほうが、識別力の値が高くなることが示唆された。また、ヒントを空所の横に置く場合は、空所を本文に記載されている順序通りに並べるよりも、本文の順序とは反対に並べる方が識別力の値を高くする可能性が示された。

謝辞

本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と齋藤 信さん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。そして、本研究にご協力いただきました中学校の前校長先生、教頭先生及び生徒のみなさまに心よりお礼申し上げます。最後に、貴重なコメントをいただきました査読者の先生にお礼申し上げます。

参考文献

- Brown, J.D., *Testing in language programs*, 1996, Upper Saddle River, N.J.: Prentice-Hall, Inc.
- (和田 稔 (訳) 『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. 2003)
- Crocker, L., & Algina, J., *Introduction to classical and modern test theory*, 1986, New York: Holt, Rinehart, and Winston.
- Ebel, R. L. & Frisbie, D.A., *Essentials of educational measurement*, 5th ed, 1991, Englewood Cliff, N.J.: Prentice Hall.
- 群馬県総合教育センター。「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. 2007.
- Hambleton, R. K., & Jirka, S. J., “Anchor-Based Methods for Judgmentally Estimating Item Statistics”, In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development*, 2006, Mahwah, N.J.: Lawrence Erlbaum.
- (韓 太哲・林 規生 (訳). 「アンカー基準法を基にした項目統計量の判断による推定」, 池田 央 (監訳). 『テスト作成ハンドブック－発達した最新技術と考え方による公平妥当なテスト作成・実施・利用のすべて－』. 教育測定研究所. 2008, pp.424-445.)
- Hogan, T. P., *Educational Assessment—A Practical Introduction—*, 2007, Hoboken, N.J.:

John Wiley & Sons.

Impara, J. C., & Plake, B. S., "Teachers' ability to estimate item difficulty: A test of the assumptions in the Angoff standard setting method", *Journal of Educational Measurement*, 35, 1998, 69-81.

石井秀宗. 「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, 2007, 60-70.

(Ishii, H., "What Leads to Nonresponses for Open-Ended Questions? A Data Analysis of the Japanese Test for 6th Grade Students in the Gunma Prefecture Achievement Test for Middle School Students", *Japanese Journal for Research on Testing*, 3, 2007, 59-70.)

石井秀宗. 「無回答および得点に基づいた記述式問題の項目分析—群馬県児童生徒学力診断テストにおける教科横断的検討—」, 『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』, 9, 2009, 1-13.

(Ishii, H., "An Item Analysis of Open-Ended Questions Based on Nonresponses and Scores: A Cross-Subjects Study on the Achievement Test for Middle School Students in Gunma Prefecture", *Bulletin of Center for Secondary Education Studies, Graduate School of Education and Human Development*, 9, 1-13.)

石井秀宗・安永和央・金 由那・千 鎬誠・朴 賢晶・李 正連. 「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, 2010, 29-43.

(Ishii, H., Yasunaga, K., Kim, Y-N., Cheon, H-S., Park, H-J., Lee, J., "Japan - South Korea Comparison Study on the responses for a Reading Literacy Test: An Item Analytic Approach using the Achievement Test for Middle School Students in Gunma Prefecture", *Bulletin of Center for Secondary Education Studies, Graduate School of Education and Human Development*, 10, 2010, 29-43.)

国立教育政策研究所(編).『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—』. 2002, ぎょうせい.

国立教育政策研究所(編).『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—』. 2004, ぎょうせい.

国立教育政策研究所(編).『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—』. 2007, ぎょうせい.

国立教育政策研究所(編).『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—』. 2010, 明石書店.

- 国立教育政策研究所教育課程研究センター. 『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—』. 2003, ぎょうせい.
- Lorge, L., & Kruglov, L., “The improvement of estimates of test difficulty”, *Educational and Psychological Measurement*, 13, 1953, 34-46.
- Musial, D., Nieminen, G., Thomas, J., & Burke, K., *Foundations of Meaningful Educational Assessment*, 2009, New York: McGraw-Hill.
- 野口裕之. 「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. 1985, 67-83.
- Quereshi, M. Y., & Fisher, T. L., “Logical versus empirical estimates of item difficulty”, *Educational and Psychological Measurement*, 37, 1977, 91-100.
- 鷺尾泰俊. 『実験計画法入門』. 日本規格協会. 1974.
- 安永和央・石井秀宗. 「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, 2010, 87-101.
- (Yasunaga, K., & Ishii, H., “The effects of variations in placing diagrams within items of a Japanese language reading test”, *Japanese Journal for Research on Testing*, 6, 2010, 87-101.)