

高校入試に出題される数式問題のデータベース化と分析

坂谷内 勝

国立教育研究所
saka@nier.go.jp

〔要約〕国立教育研究所では、平成3年度以降の公立高校の入試問題を収集し、問題に関する基本的情報をデータベース化し、教育関係者の利用に供している。本稿では、数学科の高校入試問題の中で特に数式問題に着目し、数式問題の特徴を把握するために、様々な分析を試みたのでその結果について報告する。数式問題の全体的な特徴は、正答率が高い、問題の長さが長いほど正答率が低くなることが明らかになった。また、問題中の数字や演算子等の頻度や関係について調べた結果、正答率と負の相関がある数字や演算子等があること、数式の長さとの相関がある数字や演算子等があること、対でよく出る数字や演算子等があることなど、興味深い結果を得た。

An analysis of the database of numerical formula questions for high-school entrance examinations

Masaru SAKAYAUCHI

National Institute for Educational Research
Zip Code 153-8681, Japan
saka@nier.go.jp

Our institute have been developing the database of questions for high-school entrance examinations since 1991. This report will analysis of numerical formula questions for high-school entrance examinations using this database. It was clear that the numerical formula questions have some characteristics: (1)the ratio of correct answer is high, (2)the longer length of question is, the lower the ratio of correct answer becomes, (3)there are some correlations between characters, and so forth.

1. はじめに

国立教育研究所では、平成3年度より全国都道府県教育委員会の協力を得て、わが国の公立の高等学校入学者選抜学力検査問題（以下、高校入試問題と呼ぶ。）を収集し、問題に関する基本的情報を当研究所教育情報データベース⁽¹⁾に収録している（表1参照）。

文字テキストからなる「高校入試問題情報データベース」は、約600の教育関係機関に対し、インターネットまたはパソコン通信を介してオンライン検索を可能にしている。問題を画像情報として収録した「高校入試問題画像データベース」と、聞き取り（ヒアリング）問題を音声情報として収録した「高校入試問題音声データベース」は、当研究所内で公開している。

また、平成5年度より高校入試問題の質的向上及び改善に資するために、高校入試問題の分析・評価システム⁽²⁻⁴⁾を開発し、高校入試問題データベースを利用して、入試問題の多角的分析・評価を試みている。

教育関係者にとって、テスト問題及び関連情報の蓄積と統計処理等の科学的分析は、①テスト問題の向上や改善、②児童・生徒の学力の実態把握、③教育到達度の国際比較等に有効であり、重要な教育的課題といえる。

表1 高校入試問題データベース

データベース	収録項目	データ	コンピュータ
問題情報データベース	ID: 問題ID	数値	汎用計算機
	YR: 年度	数値	(インターネットまたはパソコン通信を介してオンライン検索が可能)
	KN: 都道府県名	文字	
	SN: 教科名	文字	
	QB: 大問番号	数値	
	QS: 小問番号	数値	
	PC: 正答率	数値	
	WT: 配点割合	文字	
	LV: 難易度	数値	
	QN: 問題番号	文字	
	CA: 正答例	文字	
	NT: 備考	文字	
	KW: キーワード	文字	
問題テキストデータベース	問題テキスト全文	文字	パーソナルコンピュータ
問題画像データベース	問題・正答画像	画像	ワークステーション(電子ファイリングシステム)
問題音声データベース	聞き取り問題音声	音声	ワークステーション

本稿では、数学科の高校入試問題データベースの中から特に数式問題に着目し、数式問題の特徴を明らかにするために、様々な分析を試みたのでその結果について報告する。

2. 高校入試問題に関する先行研究

全国的な規模による高校入試問題の分析結果は、昭和55年度に全国教育研究所連盟⁽⁵⁾から報告されている。分析対象は、全国の公立高等学校入学者選抜で課した試験問題である。数学科の場合、小問単位で合計985題の問題を分析し、教育的ねらい（領域別）の割合、出題率、出題傾向を報告している。また、中学校における学習評価に関する研究⁽⁶⁾では、IEA（国際教育到達度評価学会）国際数学教育調査と高校入試問題を比較している。国・私立高等学校の入試問題の分析⁽⁷⁾では、望ましい入試問題の在り方について難問と良問の観点で分析を試みている。

これまでの高校入試問題の分析報告を概観すると、問題の正答率等の分析については、コンピュータを利用し、統計的手法を用いてきめ細かな分析がなされている。

しかし、問題自体のデータは電子化されていないため、問題を分析する際や調査結果を考察する際に、問題自体の記述・表記についての科学的な分析が困難な状況である。

問題文の記述・表記等は問題分析する際に重要な要素であり、従来の入試問題や調査結果の報告書のように、冊子等の紙面の上でしか確認できないのでは、問題自体の分析研究に限界が生じる。

3. 数学科における高校入試問題の特徴

他の教科と比較して、数学科の高校入試問題の全体的特徴⁽⁸⁻⁹⁾は、次のとおりである。

表2 データ収録件数（小問単位の問題数）

年度	1991	1992	1993	1994	1995	1996	合計
国語	1371	1359	1356	1379	1355	1329	8149
社会	1800	1822	1818	1848	1877	1881	11046
数学	1141	1158	1174	1199	1198	1206	7076
理科	1605	1580	1636	1649	1682	1660	9812
英語	1546	1572	1545	1573	1611	1590	9437
合計	7463	7491	7529	7648	7723	7666	45520

(1) 問題数が少ない(表2参照)。

1991年度から1996年度の6年分の平均問題数を求めると、数学科の平均問題数は、25.1問で、100点満点のテストに換算すると1問当たりの配点は4点になる。一方、全教科の平均問題数は、32.3問で、100点満点のテストに換算すると1問当たりの配点は3点になる。この結果より、数学科の問題数は少ないという特徴を持つ。

(2) 問題数が年々増加する傾向にある。

1991年度から1996年度の6年分の問題数の増加傾向を見ると、数学科の問題数増加率は5.7% $[(1206-1141)/1141]$ である。一方、全教科の問題数増加率は2.7% $[(7666-7463)/7463]$ である。この結果より、数学科の問題数は年々増加する傾向にあるという特徴を持つ。

(3) 問題難易度に大きな偏りが無い。

1991年度から1995年度の問題の正答率を基に、問題難易度を次のように定義する。

- ① 正答率が 80%~100%は、問題難易度「1」
- ② 正答率が 60%~ 80%は、問題難易度「2」
- ③ 正答率が 40%~ 60%は、問題難易度「3」
- ④ 正答率が 20%~ 40%は、問題難易度「4」
- ⑤ 正答率が 0%~ 20%は、問題難易度「5」

公開されている正答率は全教科で37,854問中の16,625問(43.9%)で、その内、数学科は5,870問中の2,614問(44.5%)である。正答率の算出方法は、都道府県によって様々であるが、多くの都道府県では、全ての受験者数の5%~10%を抽出し、算出している。

全教科と数学科の問題難易度の分布を図1に示す。全教科の問題難易度の分布では、正規分布に近い分布を示している。しかし、数学科の問題難易度の分布は、難易度「1」と「2」にほぼ同数の割合(約23%)で多く集ま

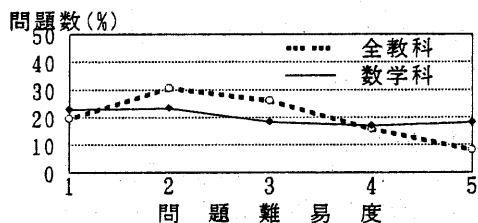


図1 正答率による問題難易度の分布

正答率 (%)

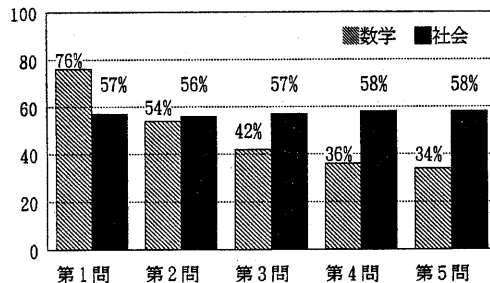


図2 数学科と社会科の大問題別正答率

正答率 (%)

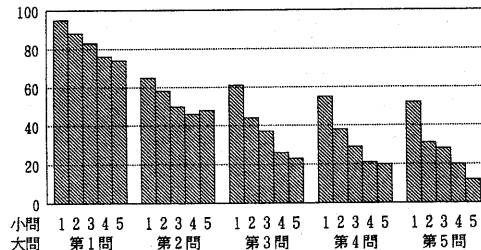


図3 数学科の大問題・小問題別正答率

っており、難易度「3」「4」「5」は、同じ割合(約18%)の分布を示すという、全教科とは異なる分布を示している。これは、数学科の問題難易度の分布は、全体的に大きな偏りが無いという特徴を持つ。

(4) 先頭の問題ほど正答率が高い。

数学科と社会科を比較して、大問題別の平均正答率(1991年度から1996年度まで)を求め、グラフ化したものが図2である。また、数学科のみの大問題と小問題を、各々5番目までを抽出し、平均正答率を求めたグラフが図3である。

これらのグラフから、数学科は先頭の問題ほど(大問題番号、小問題番号が小さいものほど)正答率が高いという特徴を持つ。

特に、数学科の第1問目の平均正答率は95%(最大100%、最小81%)であることから、この問題に関しては、年度と地域に関係なく、ほぼ全員に近い受験者が解答し、正解を導いていることがわかる。正答率が100%近くになることが想定できるにもかかわらず、このような問題が出題されている理由は、合否判別のための問題というより、学力到達度を図るための問題であると考えられる。

4. 数式問題のデータベース化

1991年度から1996年度の高校入試問題テキストデータベースから、数学科の数式に関する問題を抽出し、数式問題データベースを作成した。このデータベースには、テキスト化した数式問題の他に、年度、都道府県名、問題番号、正答率、問題の長さに関する各情報が収録されている。

問題件数（レコード数）は1,596件で、内、正答率の公開不可や正答率を求めていない等の理由で、正答率がないものは878件である。

以下に、数式問題をテキスト化するために用いたルールを記述する。

- (1) 明らかに単純な「数式の問題」であるもののみを対象とし、応用問題と思われるものは抽出しない。

[単純な数式問題の例]

- ・～を計算しなさい。（計算せよ。）
- ・～を解きなさい。（解け。）
- ・～を因数分解しなさい。（因数分解せよ。）
- ・～を展開しなさい。（展開せよ。）

- (2) 数式（数字、演算子、変数等）を抽出し、その他の情報は簡素化する。

[簡素化の例]

- ・計算問題、方程式、不等式等は数式の部分を抽出し、“を計算しなさい。”等の指示文は省略する。
- ・連立方程式は、数式と数式を“&”で連結して表記する。
- ・因数分解の問題には、数式の末尾に“#イ”を付加して表記する。
- ・展開の問題には、数式の末尾に“#テ”を付加して表記する。

- (3) 分数表記は、分子と分母の間に“/”を入れ、かつ分数全体を“[”と“]”で括る。

[分数の例] $\frac{2}{3} \Rightarrow [2/3]$

- (4) ルート表記は、“#ル”の後に、“[”と“]”でルート内の数式を括る。

[ルートの例] $\sqrt{2} \Rightarrow \#ル[2]$

- (5) 指数表記は、“[”を用いる。

[指数の例] $2^3 \Rightarrow 2[3]$

上記のルールにより、テキスト化された数

表3 数式問題の構成要素と総数

番号	分類名	要素	総数	分類計
①	数 <small>(小点を点む)</small> 字数含	1	855	6524
		2	1942	
		3	1177	
		4	585	
		5	604	
		6	351	
		7	275	
		8	280	
		9	164	
		0	242	
		49		
②	算術演算子	+	1064	5363
		-	2070	
		x	278	
		/	227	
		√	561	
		∠	575	
		ル	588	
③	変数	a	462	2662
		b	222	
		c	1	
		m	10	
		n	7	
		x	1511	
		y	449	
④	関係演算子	=	331	501
		>	81	
		<	59	
		≠	9	
		≡	21	
		≡	21	
⑤	括弧	(1136	4582
)	1136	
		{	6	
		}	6	
		[*1]	1149	
]]	1149	
⑥	その他	&	80	932
		イ	111	
		テ	21	
		#*2	720	
合 計			20564	20564

注*1：“[”または“]”は、“/”と“ル”の合計になる。
注*2：“#”は、“イ”と“テ”の合計になる。

式問題の構成要素と総数を表3に示す。

5. 数式問題の分析

以下に、数式問題の特徴を把握するために、様々な分析を試みたので、その結果と考察について述べる。

5.1 同一問題

同一問題で最も多く出る問題は、次の2つの数式問題で、共に過去4回出題されている。極めて易しい数式問題であるが、同じ年度に

2都道府県以上で出題されていないことや、同じ都道府県で2回以上出題されていないことより、同一問題に関しては十分配慮されていると考えられる。

① 3-8

- 平成4年度・長野県 ・正答率 99.1%
- 平成5年度・高知県 ・正答率 98.1%
- 平成6年度・和歌山県・正答率非公開
- 平成7年度・栃木県 ・正答率非公開

② 5+(-8)

- 平成4年度・三重県・正答率非公開
- 平成5年度・愛媛県・正答率 99.0%
- 平成7年度・高知県・正答率 97.6%
- 平成8年度・宮崎県・正答率非公開

5.2 数式問題の正答率

数式問題の平均正答率は83%(718件の平均)で、数学全体の平均正答率である52%(3,168件の平均)と比較して極めて高い値である。この結果より、数式問題が数学科の中で易しい問題であることがわかる。

また、数学科の平均問題数は25.1問であるのに対し、数式問題の平均問題番号は4.0問であることから、数学科の問題の始めに出ていることがわかる。ちなみに、1997年度の数学科の問題で、第1問目が数式問題である都道府県数は45であった。

5.3 数式問題の長さとの正答率

数式問題の長さとの正答率の関係を表4に示す。一般に、数式問題の長さが長いほど、数式が複雑になり、正答率が低くなる傾向が見られる。

なお、数式問題の長さは、データベースを作成する際に用いた「数式問題をテキスト化

表4 数式問題の長さとの平均正答率

長さ (文字数)	平均 正答率	問題数
L < 5	96.59	77
5 ≤ L < 10	85.74	467
10 ≤ L < 15	83.30	629
15 ≤ L < 20	77.69	297
20 ≤ L	74.64	126
全 体	83.23	1596

するためのルール」を適用した後の、テキスト文字列の長さである。

5.4 数式問題中の数字との正答率

数式問題中に1桁の数字(0~9)を1つでも含む数式をグループ化し、そのグループの平均正答率を求めた結果を表5に示す。

“0”と“1”と“2”のグループの平均正答率が他の数字のグループと比較して低いことがわかる。“1”と“2”のグループに関しては、平均文字数が多いことに関係していると考えられるが、“0”のグループに関しては、そのようなことがいえない。“0”を含む数式問題には、他の数字と異なる特徴があると考えられる。

5.5 数式問題中の算術演算子との正答率

数式問題中に算術演算子(加減乗除、指数、分数、ルート)を1つでも含む数式をグループ化し、そのグループの平均正答率を求めた結果を表6に示す。

表5 数式問題中の数字との平均正答率

数字	平均 正答率	問題数	平均 文字数
1	81.51	206	13.98
2	81.70	692	13.67
3	82.93	1279	13.89
4	83.53	976	13.49
5	84.24	546	13.00
6	82.70	540	13.99
7	85.20	336	12.40
8	85.40	264	12.70
9	85.36	263	11.93
0	77.77	155	13.19

表6 数式問題中の算術演算子との平均正答率

算術演算子	平均 正答率	問題数
+	81.78	885
-	83.69	1403
×	84.83	222
÷	86.09	270
/	80.56	343
√	77.07	462
ル	81.48	103

“×”と“÷”のグループの平均正答率が
高く，“Γ”（指数）のグループの平均正答
率が低いことがわかる。これは、複雑な数式
に“×”と“÷”は多用しないためと、指数
が入っている問題（2次方程式等）が一般に
難しいためと考えられる。

5.6 特に正答率が低い数式問題

正答率が低い問題を表7に示す。この表よ
り，“Γ”（指数）及び2次式を扱う問題の
正答率が低いことがわかる。また，“#イ”
（因数分解）の問題の正答率が低いことも読
みとれる。

表7 正答率が低い問題

数式問題のデータ	年度	正答率
$(a-b)^2 - 2a + 2b - 3 \#イ$	1993	16.7
$4x^2 - 15 = 0$	1995	25.1
$mx + 1 - x - m \#イ$	1996	27.1
$x^2 - 2x - 1 = 0$	1992	31.6
$a(x+2y) + bx + 2by \#イ$	1992	32.8
$a(3b+1) + 6b + 2 \#イ$	1996	32.8
$a^2 - 4a + 4 - b^2 \#イ$	1994	32.9
$9x^2 - 8 = 0$	1991	34.6

5.7 数字等の頻度と正答率との相関関係

数式問題中の数字（0～9と小数点）、算
術演算子（加減乗除、指数、分数、ルート）、
関係演算子（等号、不等号）、変数、括弧等
の出現頻度を求め、その数式問題の正答率と
の相関係数を求める。

なお、この分析において、表3に示す要素
の内、いくつかの要素を以下の理由と方法で
削除する。

- ① “c”, “m”, “n”の各変数は総数が
少ないので，“b”の総数にこれらの変数の
総数を加算し、3つの変数を削除する。
- ② “≥”, “≤”の関係演算子も総数が少
ないので，“≥”は“>”, “≤”は“<”と
みなし、これらの総数を各々加算し、2つ
の関係演算子を削除する。
- ③ “{”, “}”の括弧も総数が少ないので、
“{”は“(", “}”は“)”とみなし、
これらの総数を各々加算し、2つの括弧を削
除する。

- ④ “(”と“)”は必ず対で出現するので、
“(”を活かし、“)”を削除する。
- ⑤ “[”, “]”, “#”は、“/”, “ル”
等の要素と必ず対で出現するので、“[”,
“]”, “#”を削除する。
- ⑥ “テ”は総数が少なく、他の要素にも含め
難いので、単純に“テ”を削除する。

表8は、相関係数の有意性検定の結果、危
険率0.1%で有意な要素を、相関係数の高い順
に並べたものである。

この表より、変数“x”, 数字“2”, 算
術演算子“Γ”, そして“イ”（因数分解）
をより多く含む数式問題は、正答率がより低
くなるといえる。これは、先の5.6での考察
を統計的に裏付けるものである。

5.8 数字等の頻度と数式問題の長さとの相関関係

5.7での分析手法と同様にして、数字等の
頻度と数式問題の長さとの相関係数を求める。
表9は、相関係数の有意性検定の結果、危険
率0.1%で有意な要素を、相関係数の高い順に
並べたものである。この表より、数式問題が
より長いものは、算術演算子“ル”と“/”,
数字“2”と“3”, そして“(”をより多

表8 正答率との 表9 長さとの
相関係数 相関係数

要素	相関係数
x	-0.34**
2	-0.31**
Γ	-0.30**
イ	-0.26**
0	-0.21**
1	-0.18**
+	-0.18**
.	-0.17**
=	-0.17**
b	-0.13**

(注: **は、危険率0.1%
で有意である)

要素	相関係数
ル	0.54**
2	0.49**
3	0.35**
(0.34**
/	0.33**
1	0.24**
+	0.23**
y	0.19**
-	0.16**
b	0.15**
Γ	0.15**
6	0.13**
4	0.11**
÷	0.11**
&	0.10**
x	0.10**
<	-0.10**
>	-0.10**

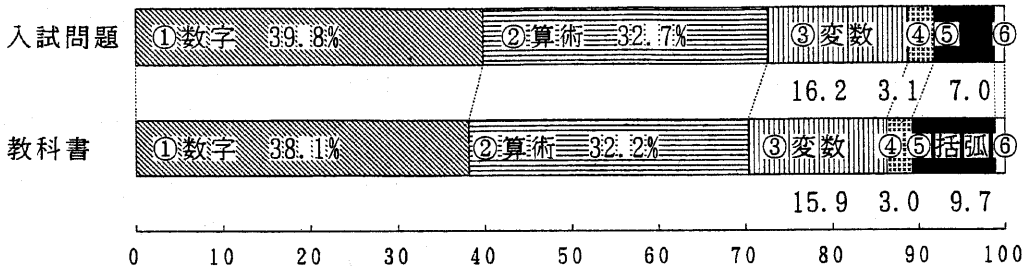


図4 高校入試問題と教科書との比較 (要素分類名総数の構成比)

く含むといえる。

また、関係演算子“<”と“>”を含む数式問題は、長くない(短くなる)ということもいえる。

5.9 教科書の数式問題との比較

高校入試の数式問題と中学校教科書⁽¹⁰⁻¹²⁾の数式問題(教科書に載っている練習問題等の数式問題)と比較してみる。

まず、表3の分類名の総数の構成比を高校入試と教科書で比較すると、図4に示すとおり、大きな差異は見られない。

次に、5.7での分析手法と同様に、数字等の頻度を基に各要素間の相関係数を求めて比較する。表10は、高校入試または教科書のどちらかで、数字等の要素間の相関係数の絶対値が0.3を越えているものを示している。これらの相関係数(絶対値が0.3を越えているもの)は、すべて危険率0.1%で有意である。

“0”と“.”(小数点)が強い相関関係があること、“&”(連立方程式)と“=”, “y”が強い相関関係があること、“2”と“r”, “a”と“b”が強い相関関係があることがわかる。これらのどの対も一緒に出てくる(または、一緒に出ない)という数式問題の特徴がある。

最後に、この相関係数の絶対値を親近性行列とし、要素間の分類(クラスタ分析)を試みる。クラスタ間の距離の計算は最近隣法を用いる。

図5は、クラスタ分析を基に、変数間の距離が近い15結合の様子を、図示したものである。全体的に高校入試と教科書のクラスタに大きな違いは見られない。部分的には、教科

表10 要素間の相関係数

要素対	(1) 入試問題	(2) 教科書
& =	0.79**	0.69**
a b	0.61**	0.57**
& y	0.55**	0.47**
2 r	0.48**	0.47**
x y	0.43**	0.33**
x =	0.40**	0.43**
y =	0.40**	0.31**
0 .	0.37**	0.58**
x +	0.34**	0.14**
b r	0.34**	0.03
a r	0.33**	0.07
イ r	0.31**	0.32**
l /	0.24**	0.34**
a x	-0.29**	-0.30**
x ×	-0.31**	-0.30**
x ル	-0.33**	-0.25**

(注: **は、危険率0.1%で有意である。)

書では、数式問題の長さや分数に強い結びつきがあるのに対して、高校入試では、数式問題の長さやルートに強い結びつきがある点が異なっている。

6. 今後の課題

本データベース開発に関する今後の課題を以下に述べる。

(1) 大量問題の教育的観点による分析

教育的ねらい、内容領域、行動的期待等、

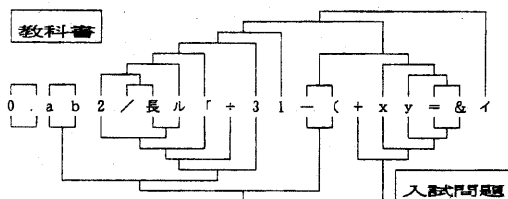


図5 クラスタ分析の結果

教育的観点で問題分析を試みるためには、各問題1問ずつに、これらの教育的属性を付加しなければならない。この人的労力を軽減させるための研究成果は少なく、数学教育の専門家の支援なしでは解決できない課題であると考えられる。

(2) 個人データの分析

問題項目関連度分析、個々の受験者特性による解答分析等を試みるためには、問題に解答した個人データが必要である。しかし、入試問題という制度的な面を考慮すると、全国的な個人データの入手及び分析は困難であると考えられる。

(3) 電子化媒体 (CD-ROM) やインターネット利用

昨今のマルチメディアやネットワーク等の教育界への普及を考慮し、高校入試問題データベースのCD-ROM化とインターネット利用を実現することが課題である。

現在、国立教育研究所では、今後のインターネット普及を考慮し、WWW (World Wide Web) サーバ上に高校入試問題データベースを構築し、一般利用に供することを考えている。しかし、インターネット等による公開可否に関する許諾状況は、すべての問題等の情報が一般に公開できないのが現状である。

今後、高校入試問題に関する情報の公開について、各都道府県教育委員会の積極的な協力支援を得られる体制を整えることが必要であると考えられる。

【参考文献】

(1) 坂谷内勝：「教育情報・資料センター③～高校入試問題データベースの開発～」，

『教育と情報』，文部省大臣官房調査統計企画課編，第一法規出版，No. 432，PP. 36-39，1994年。

(2) 坂谷内勝，浅木森利昭：「高校入試問題の分析・評価システムの開発(1)～高校入試問題データベースの構築～」，日本科学教育学会，年会論文集，Vol. 17，PP. 219-220，1993年。

(3) 坂谷内勝，吉岡亮衛，小松幸廣，浅木森利昭：「高校入試問題の分析・評価システムの開発(2)～問題画像データベースについて～」，電子情報通信学会，技術研究報告，ET93-99，Vol. 93，No. 405，PP. 67-72，1993年。

(4) 坂谷内勝：「高校入試問題の分析・評価システムの開発研究」，文部省科学研究費補助金試験研究(B)，No. 05558022，PP. 1-123，1995年。

(5) 全国教育研究所連盟編：「高校入試問題の分析 [数学] - その出題傾向に関する基本データ」，第一法規出版，PP. 1-267，1981年。

(6) 木田宏：「学習到達度に関する総合的研究」，第2編，中学校における学習評価に関する研究，文部省科学研究費補助金一般研究(A)，No. 56410002，PP. 1-83，1984年。

(7) 三宅征夫：「国・私立中学校・高等学校の入試問題の分析研究」，文部省科学研究費補助金総合研究(A)，No. 01102031，PP. 1-171，1992年。

(8) 坂谷内勝：「高校入試問題データベースの応用(1)～数学入試問題の分析～」，日本科学教育学会，年会論文集，Vol. 20，PP. 323-324，1996年。

(9) 坂谷内勝：「数学科における高校入試問題データベースの開発と課題」，日本数学教育学会，第30回数学教育論文発表会論文集，PP. 109-114，1997年。

(10) 藤田宏，前田昭二他：「新しい数学1」，東京書籍，1993年。

(11) 藤田宏，前田昭二他：「新しい数学2」，東京書籍，1993年。

(12) 藤田宏，前田昭二他：「新しい数学3」，東京書籍，1993年。