

日本人の読み書き能力1948年調査の非識字者率に対する新解釈

横山詔一， 前田忠彦（統計数理研究所）， 野山 広， 福永由佳， 高田智和

2020年10月24日（土） 16時15分～16時55分
日本語学会2020年度秋季大会



国立国語研究所(NINJAL)



統計数理研究所(ISM)

きょうの流れ

1. はじめに【予稿集137頁】
2. 1948年調査の方法【137～139】
 - (1) 調査の体制・組織, (2) 調査時期・地点, (3) 事前のサンプリングと達成率, (4) 問題用紙, (5) 教示
3. チャンスレベルを考慮した非識字者率の推定法【139～141】
 - 付録：チャンスレベルでの正答数分布のための確率分布【143～144】
4. 考察【141～143】

注

1. 本研究では『日本人の読み書き能力調査』（1951, 読み書き能力調査委員会, 東京大学出版部）における「文盲」を「非識字者」, 「文盲率」を「非識字者率」と言い換える
2. 『日本人の読み書き能力調査』（1951）を「報告書（1951）」と略称することがある

きょうの流れ

1. はじめに【予稿集137頁】
2. 1948年調査の方法【137～139】
 - (1) 調査の体制・組織, (2) 調査時期・地点, (3) 事前のサンプリングと達成率, (4) 問題用紙, (5) 教示
3. チャンスレベルを考慮した非識字者率の推定法【139～141】
 - 付録：チャンスレベルでの正答数分布のための確率分布【143～144】
4. 考察【141～143】

1. 『日本人の読み書き能力』（読み書き能力調査委員会, 1951, 東京大学出版部）の内容の一部を検討する
2. 連合国最高司令官総司令部（General Headquarters, Supreme Commander for the Allied Powers, 略称 **GHQ/SCAP**）の民間情報教育局（**CIE**）の指示により**1948年に実施された調査の報告書**
3. この調査は『アメリカ教育使節団報告書』（マックアーサー司令部公表, 1946）に端を発する
4. 言語政策に関する第1級の史料

きょうの流れ

1. はじめに【予稿集137頁】
2. 1948年調査の方法【137～139】
 - (1) 調査の体制・組織, (2) 調査時期・地点, (3) 事前のサンプリングと達成率, (4) 問題用紙, (5) 教示
3. チャンスレベルを考慮した非識字者率の推定法【139～141】
 - 付録：チャンスレベルでの正答数分布のための確率分布【143～144】
4. 考察【141～143】

2. 1948年調査の方法：（1）調査の体制・組織

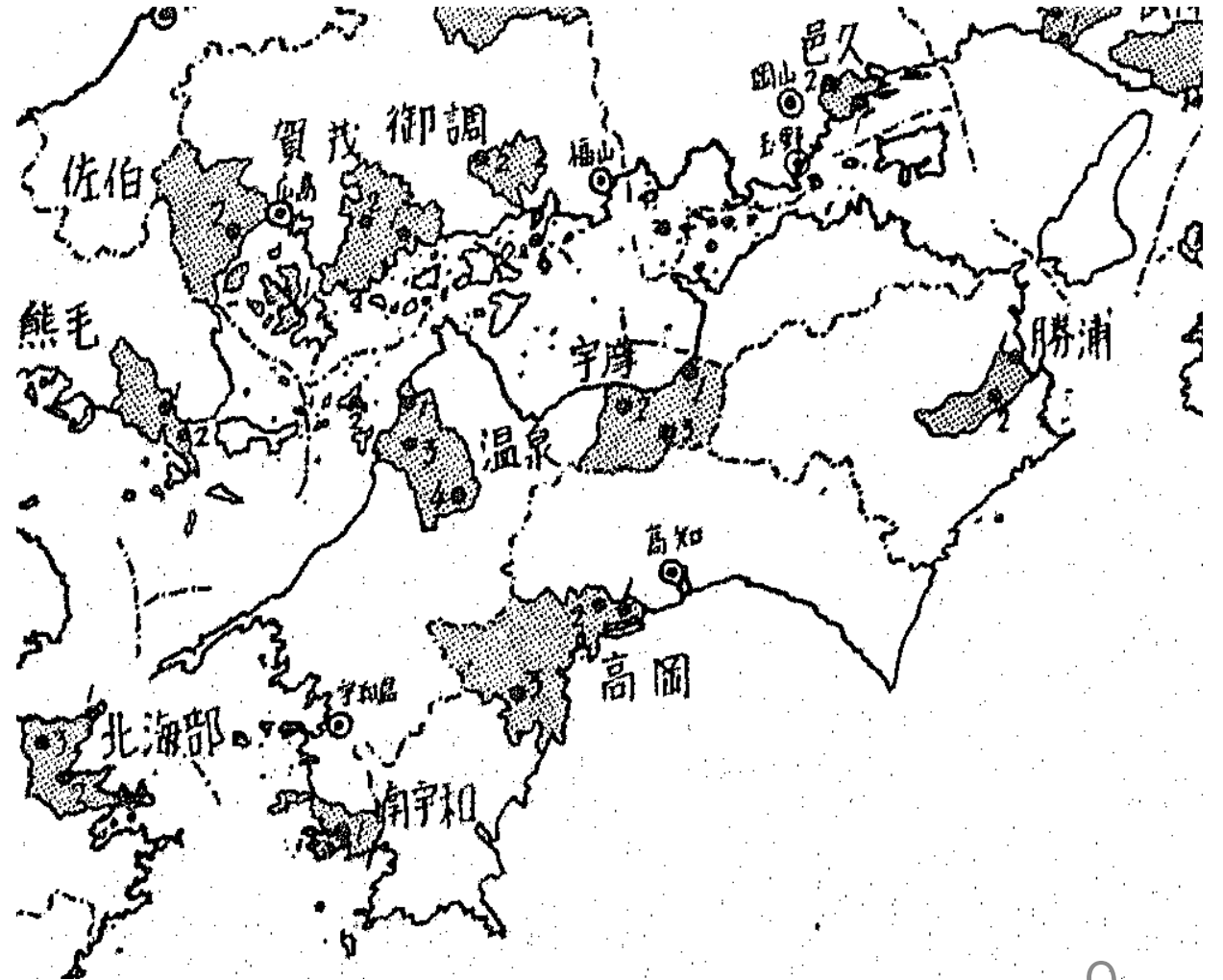
→ 時間の関係で割愛します，予稿集をご覧ください

2. 1948年調査の方法：（2）調査時期・地点

・いつ、どこで調査したのか？

1. 1948年8月と9月，全国270地点で
2. ただし沖縄県や香川県などでは実施していない（先行研究に指摘なし）

図第30 調査市郡，地点地図



2. 1948年調査の方法：（3）事前のサンプリングと達成率

だれに対して調査を実施したのか？

配給台帳等に基づくランダムサンプリングによって抽出された「sampleさん」に対して実施

- sampleさんは何人だったのか？

21,008人, 21,000人, 17,100人, 16,820人, 16,814人と諸説あるが, 実際は以下の通り

1. 事前に21,008人を無作為抽出
2. 本調査に参加したsampleさんは 16,814人
3. 横須賀市でサンプリングをやり直してデータを入れ替えた
4. その結果, 本調査より6名増えて16,820人に (達成率80.0%)

- sampleさんの年齢層や性別などは？

15歳から64歳までの男女, 全国規模で女性のデータを初めて収集 (先行研究に指摘なし)

2. 1948年調査の方法：（4）問題用紙，（5）教示



ネ	ネ	ネ	ネ	ネ	△
ヨ	コ	ビ	ロ	ユ	・

ガ	カ	ア	カ	ザ	△
ラ	ラ	ラ	ラ	ラ	・
ス	ス	ス	ス	ス	・

ミ	ミ	ミ	ミ	ミ	△
シ	シ	シ	シ	シ	・
ツ	レ	い	シ	ン	・

オ	オ	オ	オ	オ	△
モ	モ	モ	モ	モ	・
ツ	香	香	香	夕	・

マ	マ	マ	マ	マ	△
ア	チ	チ	ツ	ン	・
チ	チ	チ	チ	チ	・

九	入	八	七	四	×
円	円	円	円	円	・

8キ	3キ	4キ	5キ	1キ	×
口	口	口	口	口	・

あ	あ	あ	あ	あ	○
ひ	ら	ゆ	た	な	・
ま	ま	ま	ま	ま	・

さ	も	ち	な	き	□
る	る	る	る	る	・

た	た	た	た	た	□
は	ば	ば	が	ゾ	・
こ	こ	こ	こ	こ	・

み	み	み	み	み	□
か	か	か	か	か	・
む	し	ソ	ん	あ	・

こ	こ	こ	こ	こ	□
ん	ん	ん	ん	ん	・
に	に	に	に	に	・
よ	く	く	く	く	・

あ	あ	あ	あ	あ	□
さ	さ	さ	さ	さ	・
て	ん	て	て	て	・
て	て	て	て	て	・

(三)

たいしょう に ねん はちがつ にじゅうはち にち
 ○大正2年8月 日

めいじ にじゅうはちねん くがつ じゅうろくにち
 明治28年9月 日

(二)

●●●●	●●●	●●	●	○	(一)

●●●●	●●●	●●	●	○	(一)

○ 三丁目 さんちやうめ 番地 ばんち

○ 五丁目 ごちやうめ 番地 ばんち

26.2cm

36.9cm

非識字者でも選択肢の配置は理解できるようになっていた

- 選択式問題に回答する際、インフォーマントは選択肢の一つに○を付けるよう試験官から口頭と黒板への板書で説明を受けた

のうちの、どれにあたるかを答えてもらいます。たとえば、わたくしが「あたま」(調査者ジシンノ頭ヲ示ス)といいましたら、この五つのうち(黒板デ示シナガラ)から、「あたま」という字を見つけて○^{マル}でかこむのです(黒板ニ示ス)。それでは、これはためしですから、みなさまも鉛筆を持って「あたま」という字を○でかこんでみてください。いいですか。一度つけた○を直したいときには、それを×^{バツ}じるしで消し、正しいと思うほうを○でかこんでください。

村民運動会

6月5日
10時-16時

青草村小学校校庭
雨天順延
青草村青年会

大阪では朝鮮からの引揚者中村三郎さんに百万円当った。娘さんから貰った小遣いで銀行から買った二枚の宝くじの中一枚が当たったもので家の者にも知らせずしまっておいたが、出してみたら当たっていたというので大サワギ奥さんと娘さんとむすさんの四人ぐらし。

十八日午後十時ごろ品川区大崎一ノ三八四山田栄吉方で一むね十一坪を全焼、電熱器の不始末らしい。

東京都内十七の職業安定所に押しかけた求職者は四月中に三万余で前月より一割の増加だが求人逆減る一方で男女を通じてまともなのはわずか二〇%。五月は一そう深刻で、三日の某職業安定所の窓口には赤ちゃんを背にした婦人などまじえて約二千名の失業者群が早朝から長い列をつくらせていた。

(例) このどうは何のしらせてすか。

(答) 村会
運動会
おまつり
卒業式
草かり

(問) この運動会は五日に雨が降たら、どうなりますか。

(答) 今年はやらない雨が降ってもやるようになるかわからない。天気の良い日にはすひと月のはず

(問) 百万円当った中村さんはぜんぶで宝くじを何枚買いましたか。

(答) 一枚
二枚
三枚
四枚
五枚

(問) その宝くじはどこから買いましたか。

(答) タバコ屋
大阪
娘さん
銀行
中村さん

(問) 上の文章は何のことをいっていますか。

(答) 火事
土地
料理
たきび
停電

(問) 上の文章は何のことをいっていますか。

(答) 結婚がまじまる
配給の行列
子供の育て方
賃金をあげる要求
仕事がないからつらい

(問) 三日とは何月の三日のことですか。

(答) 前月
四月
五月
来月
某月

(問) 約二千名はどのような人たちのならんでいる列ですか。

(答) 東京都民
職業安定所の役人
男女
仕事をみつけた人たち
赤ちゃんをおぶった人

問題は90問，うち**選択式は65問（72%）** 4肢択一が19問， 5肢択一が46問

- 問題1 試験官が発音した語をひらがな，カタカナで書く（8問）
- 問題2 試験官が発音したアラビア数字，漢数字を書く（2問）
- 問題3 試験官が発音した語を選択肢から選ぶ：表記はひらがな，カタカナ，アラビア数字，漢数字，**5択（12問）**
- 問題4 試験官が発音した語を選択肢から選ぶ：表記は漢字，**5択（10問）**
- 問題5 漢字の書き取り（15問）
- 問題6 意味が通じる語を選択肢から選ぶ：表記は漢字，**4択（15問）**
- 問題7 語の意味を選択肢から選ぶ：問題語の表記は漢字，選択肢はひらがな，カタカナ，ルビ付き漢字，**5択（15問）**
- 問題8 読解問題で正答を選択肢から選ぶ：問題文は漢字仮名交じり，選択肢は**5択（9問）**，**4択（4問）**

問題の実像から分かること

1. 問題3の**アラビア数字は簡単そう**に見える
2. 横山が2020年2月に問題に取り組んだところ89点であった（**満点ではない**）

きょうの流れ

1. はじめに【予稿集137頁】
2. 1948年調査の方法【137～139】
 - (1) 調査の体制・組織, (2) 調査時期・地点, (3) 事前のサンプリングと達成率, (4) 問題用紙, (5) 教示
3. **チャンスレベルを考慮した非識字者率の推定法【139～141】**
 - **付録：チャンスレベルでの正答数分布のための確率分布【143～144】**
4. 考察【141～143】

どのような結果だったのか？

配点は90問すべて1問1点， すなわち正答数が得点

1. 得点ゼロだった人は1.7%， 90点満点だった人は4.4%（下の表）
2. 全体（N=16,820）の得点分布はJ字型

code		19	18	17	16	15	14	13	12	11	10	09	08	07	06	05	04	03	02	01	00	
点数		90	89 85	84 80	79 75	74 70	69 65	64 60	59 55	54 50	49 45	44 40	39 35	34 30	29 25	24 20	19 15	14 10	09 05	04 01	00	計
全国	市部	372 6.9	1921 35.6	1042 19.3	594 10.9	355 6.6	247 4.6	143 2.7	117 2.2	81 1.5	62 1.2	63 1.2	57 1.1	62 1.2	53 1.0	41 0.8	36 0.7	40 0.7	31 0.6	15 0.3	50 0.9	5382 100.0
	郡部	360 3.1	2559 22.4	1981 17.3	1381 12.1	1004 8.8	717 6.3	547 4.8	443 3.9	357 3.1	271 2.4	251 2.2	220 1.9	226 2.0	199 1.7	175 1.5	177 1.5	134 1.2	92 0.8	101 0.9	243 2.1	11438 100.0
	地域	732 4.4	4480 26.7	3023 18.0	1975 11.7	1359 8.1	964 5.7	690 4.1	560 3.3	438 2.6	333 2.0	514 1.9	277 1.6	288 1.7	252 1.5	216 1.3	213 1.3	174 1.0	123 0.7	116 0.7	293 1.7	16820 100.0

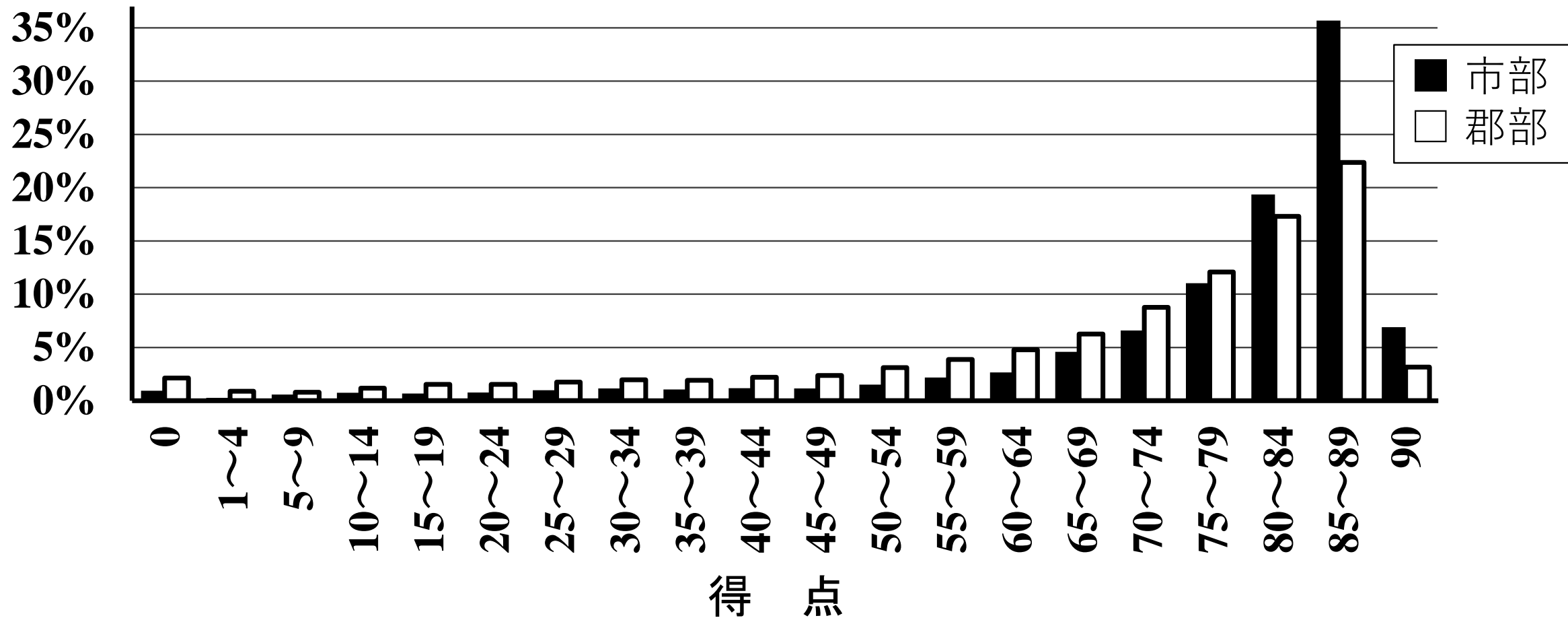
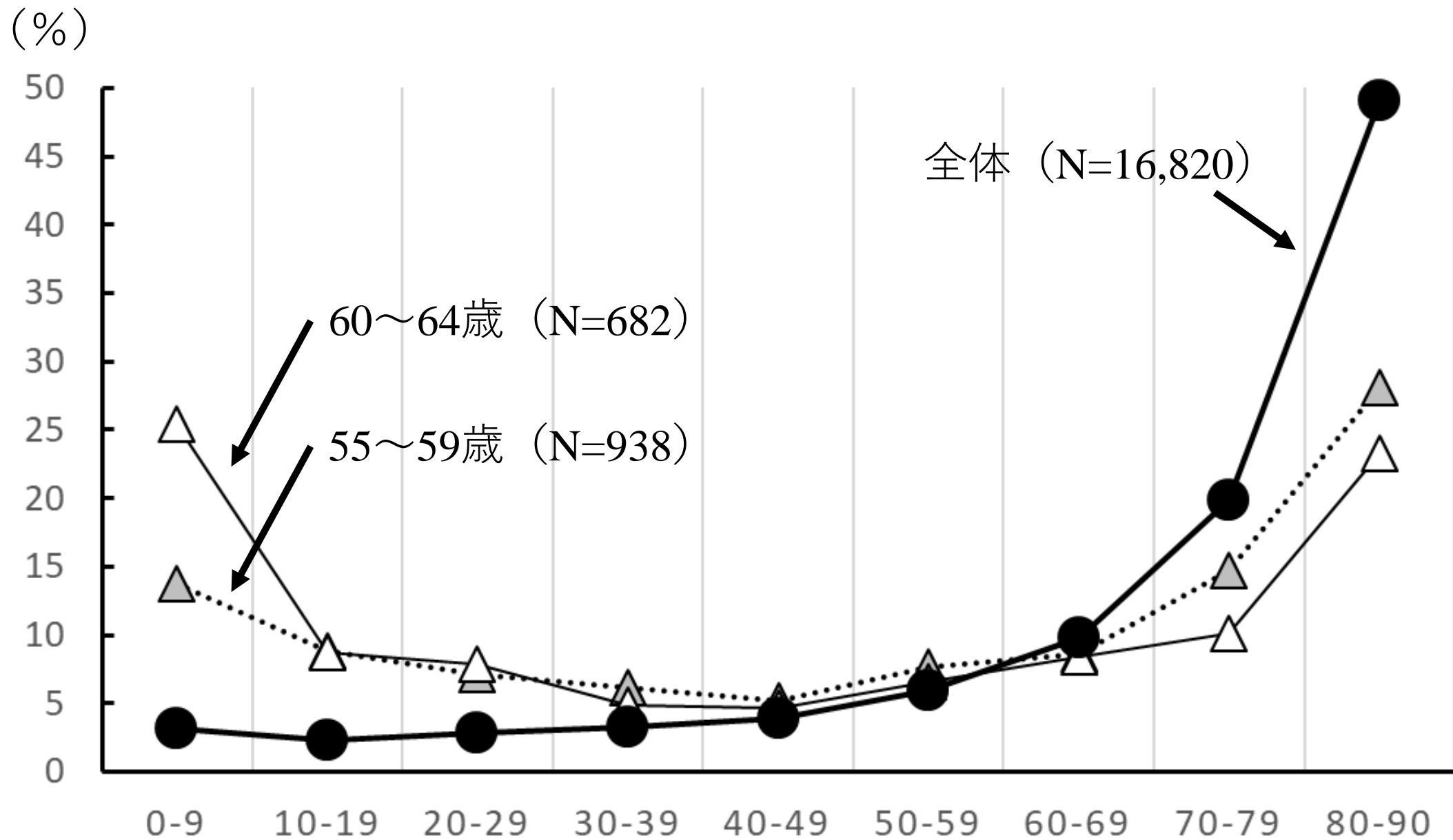


図3 市郡別の得点分布



予稿集にない図 得点分布 (全体 / 55歳から59歳 / 60歳から64歳)

・ 報告書に示された結果は？

要点は以下の2つ

1. 日本でliteracyを持つと見なせる識字者は「90点満点」の4.4%であるが、不注意などによる失点を考慮して割合を補正した結果、6.2%となった
2. 非識字者（完全文盲）は「ゼロ点の人」で、その割合は1.7%であった

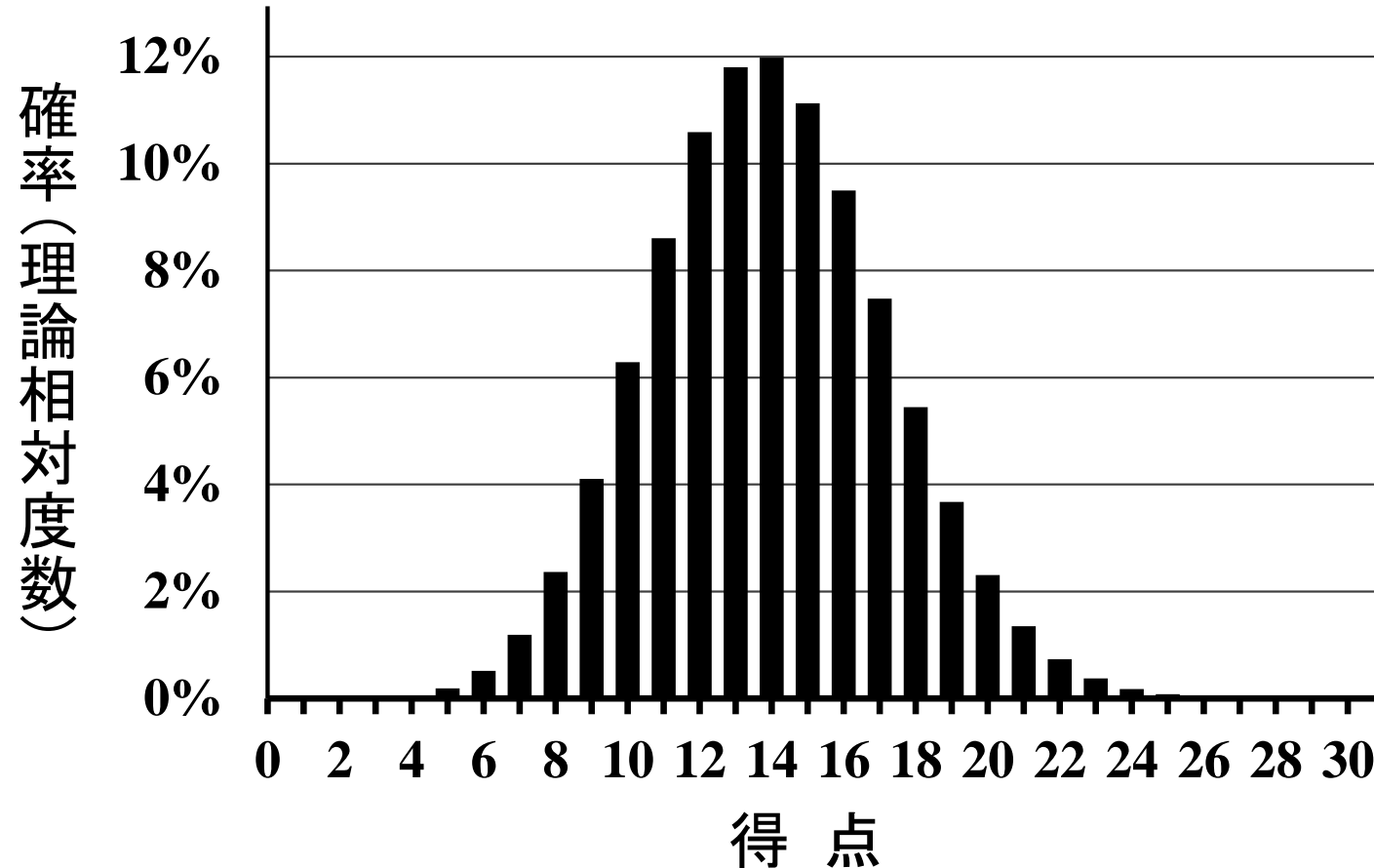
きょうは次の点を検討する

ゼロ点の人を非識字者と定義した結果、非識字者率は1.7%であったとされているが、その判定基準は妥当性を欠くのではないか？

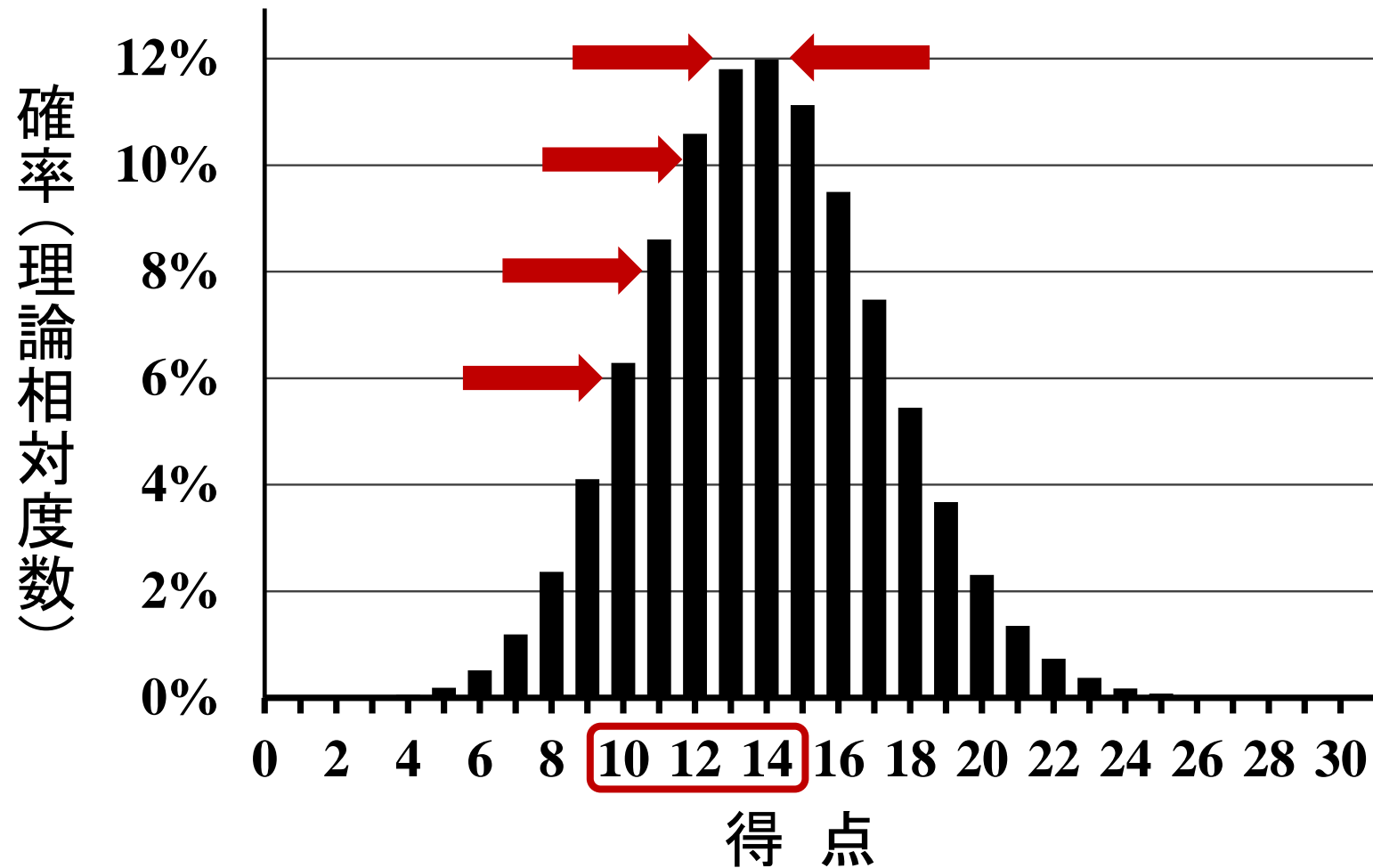
- 「非識字者」と判定すべき人を見逃しているのでは？

- 選択式問題は当て推量（guessing）や勘で選択肢を選んだ場合でも偶然に正答することがある
 - 「難問の選択肢は適当に選ぶべし」という方略は小学生でも知っている
 - 偶然に正答する確率の程度を「チャンスレベル」という

1. 選択式問題65問のすべてにおいて当て推量で選択肢を選んだ場合，得点は何点ぐらいになるかを複合2項分布で計算 → 予稿集143～144の付録を参照
2. 分布は正規分布（ベルカーブ）に近い形 → 予稿集の図2



- 例題として，10～14点になる確率をおおまかに求めてみよう
- 10点6%，11点8%，12点10%，13点12%，14点12%として合計すると48% → 正確な確率は49.256220%：次の頁の数表を参照



- **累積確率**は完全な当て推量による回答によって当該得点以下の得点を得る確率
 1. たとえば, 14点までの累積確率は57.698640%
 2. 15点以上を取れる確率は42.301360% (100から57.698640を減じる)
 3. これを**上側確率**という

表1 完全な当て推量の場合の理論分布と得点の度数分布表

得点	完全な当て推量の分布			全国			
	確率	累積確率	上側確率	度数	相対度数	累積度数	累積相対度数
0	0.000015%	0.000015%	100%	293	1.7%	293	1.7%
1~4	0.072266%	0.072280%	99.999985%	116	0.7%	409	2.4%
5~9	8.370139%	8.442420%	99.927720%	123	0.7%	532	3.2%
10~14	49.256220%	57.698640%	91.557580%	174	1.0%	706	4.2%
15~19	37.224105%	94.922745%	42.301360%	213	1.3%	919	5.5%
20~24	4.943395%	99.866140%	5.077255%	216	1.3%	1135	6.7%
25~29	0.133072%	99.999212%	0.133860%	252	1.5%	1387	8.2%

非識字者の判定基準

1. 複合2項分布を用いた確率計算により、4割以上の人が偶然に15点以上を取れることがわかった
2. 統計的検定論に当てはめて解釈すると、15点の人は**非識字者のチャンスレベルにある＝非識字者のカテゴリーに含めるのが妥当**

非識字者の判定基準

3. この方法によると、25点以上を偶然に取れる確率は0.133860%
4. 偶然に25点以上を取れるのは800人に1人ぐらいなので珍しいことが起きたと解釈、つまり有意
5. このような場合、統計的検定論では24点以下を非識字者と判定する基準を立てる

非識字者率の推定

1. 報告書に示されているデータから該当する人の割合を求めると**6.7%**に
2. これが当時の**非識字者率の推定値の上限**を与える

表1 完全な当て推量の場合の理論分布と得点の度数分布表

得点	完全な当て推量の分布			全国			
	確率	累積確率	上側確率	度数	相対度数	累積度数	累積相対度数
0	0.000015%	0.000015%	100%	293	1.7%	293	1.7%
1~4	0.072266%	0.072280%	99.999985%	116	0.7%	409	2.4%
5~9	8.370139%	8.442420%	99.927720%	123	0.7%	532	3.2%
10~14	49.256220%	57.698640%	91.557580%	174	1.0%	706	4.2%
15~19	37.224105%	94.922745%	42.301360%	213	1.3%	919	5.5%
20~24	4.943395%	99.866140%	5.077255%	216	1.3%	1135	6.7%
25~29	0.133072%	99.999212%	0.133860%	252	1.5%	1387	8.2%

非識字者率の推定

3. もちろん、非識字者全員が選択式問題のすべてに回答したとは考えられない
 4. しかし「非識字者のチャンスレベルを統計的に有意に上回っている人だけを非識字者ではないと解釈する立場」を排除することは合理性を欠く
 5. 非識字者率は定説の「1.7%」ではなく「6.7%」と考えるべきでは
- 以上のことを「チャンスレベル問題」と名付ける

きょうの流れ

1. はじめに【予稿集137頁】
2. 1948年調査の方法【137～139】
 - (1) 調査の体制・組織, (2) 調査時期・地点, (3) 事前のサンプリングと達成率, (4) 問題用紙, (5) 教示
3. チャンスレベルを考慮した非識字者率の推定法【139～141】
 - 付録：チャンスレベルでの正答数分布のための確率分布【143～144】
4. 考察【141～143】

今回の研究のまとめ

1. ゼロ点の人を非識字者と定義することの問題点を「チャンスレベル問題」と命名
2. 当時の非識字者率は上限で6.7%程度と決して低くなかった可能性があるという見方も成立することを明らかにした
3. 報告書が採用した判定基準は、非識字者の割合を過小評価する方向に設定されていたと結論づけて大過ない

今後の課題

1. チャンスレベルを考慮する必要があることはよく知られている
 - この問題を林知己夫や肥田野直は熟知していたはず
 2. しかし、報告書公刊の1951年から2020年現在に至るまで、その点を指摘した研究は皆無
- ✓ なぜこの問題が69年間も放置されてきたのか？

今後の課題

- 報告書（1951）に接する機会がない，読もうとしても難解な統計学用語・表現が頻出するので諦めてしまう
1. 国語研の蔵書をデジタル化して公開してはどうか（2021年4月に出版から70年をむかえる）
 2. 難解な統計学用語・表現の言い換えや説明が必要（注釈など）
 3. 歴史学者や政治学者の協力を得ながらGHQ内部文書（英文）を分析して事実 にせまる必要がある

さらに歴史学や政治学と連携して人文学の総合的研究を展開すべきか？

1. 勝岡寛治（1986）は，この調査に関する GHQ/SCAP/CIE の動向を時系列で明らかにしている
2. しかし，GHQ上層部，とりわけ**マッカーサー元帥**がどのような判断をしたのかなどについての言及はない



1945年8月30日14時05分，厚木海軍飛行場に到着
これから横浜市のホテルニューグランドに向かう



1954年ニューヨーク市にて
吉田茂と腕を組んでいる

さらに歴史学や政治学と連携して人文学の総合的研究を展開すべきか？

3. この調査の発端は、**アメリカ教育使節団**が1946年3月31日にマッカーサー元帥に提出した**報告書の第2章「国語の改革」**にある
4. 上記報告書が公表された1946年4月7日の段階で、マッカーサー元帥は以下の声明を付している → **国語改革に積極的ではない**ように見える
 - 「国語の改革に関する勧告のなかには、あまりにも遠大であるため、長期間の研究と今後の計画に資するに過ぎないものもある」（英文原文は以下で閲覧可）

<https://babel.hathitrust.org/cgi/pt?id=pur1.32754081234191&view=1up&seq=7>

実像をつかむには日本側の報告書だけではまったく不十分

- 国立国会図書館憲政資料室などが所蔵している文書（マイクロフィッシュ）の調査が必要，以下は文書情報の例
 1. Literacy Research: **Final Test Questions** (文書名:GHQ/SCAP Records, Civil Information and Education Section = 連合国最高司令官総司令部民間情報教育局文書) (課係名等:Public Opinion & Sociological Research Division) (シリーズ名:General Subject File, 1946-51) (ボックス番号:5913 ; フォルダ番号:35)
 2. Literacy Research: **Instructor's Manual** (文書名:GHQ/SCAP Records, Civil Information and Education Section = 連合国最高司令官総司令部民間情報教育局文書) (課係名等:Public Opinion & Sociological Research Division) (シリーズ名:General Subject File, 1946-51) (ボックス番号:5913 ; フォルダ番号:33)

実像をつかむには日本側の報告書だけではまったく不十分

- CIE以外の部局の動向も調査する必要がある

1. 普通選挙における女性参政権との関係など（GHQ/SCAP文書の課係名に Public Opinionという文言あり）
2. 米国戦略爆撃調査団文書などにも手がかりがあるかもしれない

1948年読み書き能力調査の遺物は国語研資料庫にも1箱眠っている

<https://rnr.ninjal.ac.jp/materials/fo0161/>

<p>資料作成組織の履歴 Administrative / Bibliographical history</p>	<p>専門員長: 教育研修所研修員 石黒修 専門員: 柴田武 金田一春彦 城戸幡太郎 梅津八三 白石一誠 林知己夫 影山三郎 白石大二 助手: 北村甫 都竹通年雄 満田新一郎 松樹美代治 島津一夫 肥田野直 村瀬隆二 馬場四郎 岩井弘融 堤光臣 藤沢大仁 針ヶ谷正男 高山二郎</p> <p>1948年にGHQの民間情報教育局(CIE)と教育研修所が中心になり読み書き能力調査委員会をつくり、日本人の読み書き能力調査を行ない、15歳-64歳までの男女16,820人を調査した。 1951年に『日本人の読み書き能力』を出版。編集出版委員は、石黒修 柴田武 島津一夫 野元菊雄 林知己夫。</p>
<p>資料作成年月日 Dates of accumulation of the material in the unit of description</p>	<p>1948 - 1949</p>
<p>管理歴 Custodial history</p>	<p>西が丘庁舎 第1資料庫から中央資料庫へ</p>
<p>入手情報 Immediate source of acquisition</p>	
<p>資料内容 Scope and content / Abstract</p>	<p>発端 今までの調査及びふりかえてみた経過等</p> <ul style="list-style-type: none"> ・読み書き能力委員会(編成) ・リテラシー企画 ・統計日誌 ・読み書き能力の報告の構成 1949.7.8 ・行政区画変動による素図索引訂正表 ・昭和14年度壮丁教育調査概況 文部省社会教育局 ・ギリシャにおける調査, 専門員業務経過, アルバイト使用状況(原稿) ・LITERACY RESEARCH PROGRAM ・プリント(SCHEDULE, Report Construction of Reading and wrighting Ability Test), (SAMPLING OF RESPONSE ANALYSIS) <p>Design(リテラシー調査に於けるサンプリングの計画 等)</p> <ul style="list-style-type: none"> ・概要 リテラシー調査に於けるSamplingの計画 ・読み書き調査抽出都市町村名 ・[年齢層別 地方別等割当枚数] ・Report関係(一般全般, 小田原 等) <p>Hityousasha no Kosei to Sono Heikinten(TUZUKI)'49n7gt (第29-54表) 被調査者の構成とその平均点</p>

附記)

1. 原稿作成において相澤正夫氏（国立国語研究所名誉教授）から多くのアドバイスをいただいた。記して感謝の意を表する
2. 本研究は科学研究費補助金19H00627基盤研究（A）「基礎教育を保障する社会の基盤となる日本語リテラシー調査の開発に向けた学際的研究」（研究代表者：野山広）の成果である

ご参加くださり，ありがとうございました
これからも，どうぞよろしく願いいたします