

C-CATのスコア解釈及び他のテストとの相関性について

C-CATプロジェクトチーム 代表 侯 仁鋒

0. 概要

独立行政法人日本学術振興会から、科学研究費補助金 基盤研究（B）（一般）を頂き、下記の研究開発が行われた。

研究課題名 コンピュータ適応型中国語テストの開発と検証
課題番号 15H03225
研究期間 平成27年度～平成29年度
テスト名称 コンピュータ適応型中国語テスト（C-CAT：Chinese Computerized Adaptive Test）
URL www.c-cat.top

プロジェクトチーム*

赤木 彌生（山口大学大学教育機構留学生センター）
浅野 雅樹（慶應義塾大学文学部）
今井 新悟（筑波大学人文社会系）
侯 仁鋒（県立広島大学人間文化学部）
曲 明（室蘭工業大学ひと文化系領域）
齋藤 貴志（麗澤大学外国語学部）
松岡 栄志（東京学芸大学教育学部）
丸山 浩明（県立広島大学人間文化学部）

本研究は、3年間かけて、コンピュータを利用した中国語能力テスト（通称：C-CAT）の開発を目的とする。問題アイテムの開発から始まり、プレテストの実施、解答データによるパラメタの算出、そして問題アイテムバンクの構築を行い、アダプティブ形式のコンピュータ適応型中国語テストを完成させた。現在その運用も本格的にスタートしている。

その中において、「C-CATのスコア解釈及び他のテストとの相関性について」以下に報告する。

1. 最高得点と最低得点の確認

C-CATは、測定分野の設置も、能力値の推定も、得点の算出法も、先行研究のJ-CATのシステムを援用しながら開発したものであり、試運転を始めてから、まず、最高得点と最低得点を確認する必要がある。

1.1 4分野の配点

本テストは、聴解、語彙、文法と読解の4分野で、中国語能力を測定する。各分野の配点は満点100

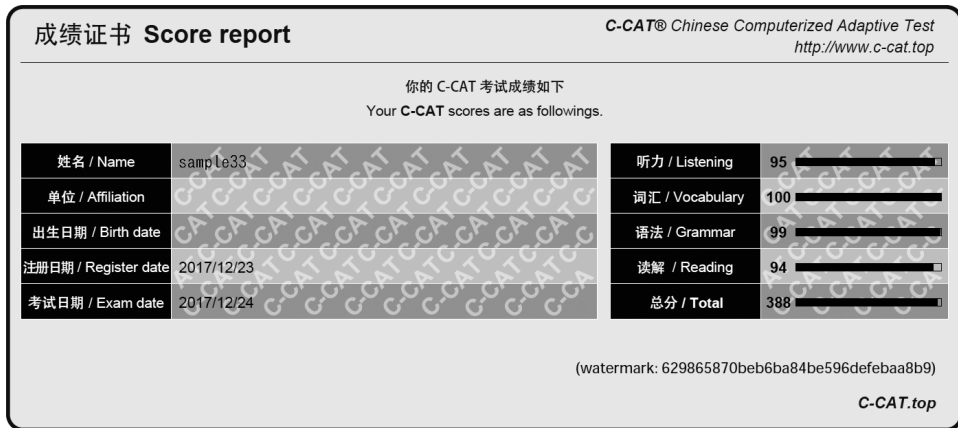
* 諸先生の所属はいずれも研究開発期間中のものである。

点であり、総得点は400点という設定である。

1.2 最高得点

全問正答の場合、聴解は21問、語彙は15問、文法は19問、読解は21問で終了し、その最高得点は図1のようにになっている。

図1 成績書

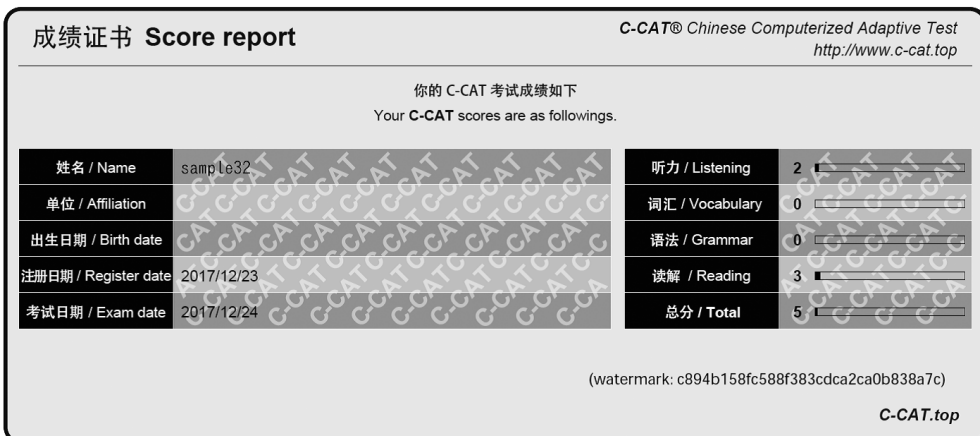


聴解 (21問) 95点
 語彙 (15問) 100点
 文法 (19問) 99点
 読解 (21問) 94点
 最高総得点 388点

1.3 最低得点

全問不正答の場合、聴解は21問、語彙は12問、文法は11問、読解は21問で終了して、その最低得点は以下の通りである。

図2 成績書



聴解 (21問) 2点
 語彙 (12問) 0点
 文法 (11問) 0点
 読解 (21問) 3点
 最低総得点 5点

表1 設定点数と実際点数

分野	設定点数	実際点数
聴解	0-100	2-95
語彙	0-100	0-100
文法	0-100	0-99
読解	0-100	3-94
総得点	0-400	5-388

表1に示したように、C-CATは、デザインした通りに両極端にあるスコアは設定点数とほぼ同じであり、全て測定範囲に見合うので、各レベルの受験者の能力を正確に推定することが可能であると確認できた。

2. シミュレーションによる能力値推定の確認

2.1 2段階式の能力値推定

C-CATは、2段階の形式で受験者の能力値を推定していく。

まず練習問題を1問出し、受験者にやり方を理解してもらう。それから、固定問題アイテムが3問、易しい、中レベル、難しいという順で出題されて、初期のレベル推定を行う。

更にこの第1段階のレベル推定に基づき、第2段階では回答に応じてアダプティブに出題されて本格的に能力レベルを推定する。

イメージとしては以下表2のような構成である。

表2 能力値推定イメージ

初期推定			本格的な推定
易しい	中程度	難しい	アイテムバンク
1	1	1	……

2.2 初期レベル推定の解答パターン

最初の3問は、受験者がすべて解答しなければならない固定問題アイテムである。解答パターンは表3に示した四つが想定される。

表3 初期レベル推定の解答パターン

初期推定			本格的な推定
易しい	中程度	難しい	アイテムバンク
1	1	1	……
×	×	×	3問とも不正答 ……
○	×	×	1問正答 ……
○	○	×	2問正答 ……
○	○	○	3問とも正答 ……

受験者のレベルに基づいて想定すれば、解答はこの四つのパターンになる。もちろん他の解答パターンもあるのは否定できない。例えば、易しい、中程度の問題が不正答したのに、難しい問題が正答したパターンなどである。これはあっても偶然だろうし、まぐれ当たりとしか思えないので、このようなパターンはシミュレーションの対象としない。従って以下の形式でシミュレーションを行った。

2.3 シミュレーションのスコアとレベル判断

ほぼ2ヶ月かけて、以下の解答パターンを繰り返して、シミュレーションを行った。

表4 シミュレーションのスコアとレベル区分

初期推定			本格的な推定		
易しい	中程度	難しい	アイテムバンク		
1	1	1	……	スコア	レベル
×	×	×	○×……正答1(4分野に各1)	30~35	初級
○	×	×	××…… 正答1 ○×…… 正答2 ○○×……正答3	30~35 46~75 99~142	
○	○	×	○×…… 正答3	105~142	初級上
			○○×…… 正答4	149~163	中級
			○○○×…… 正答5	172~186	
			○○○○×…… 正答6	192~199	
			○○○○○×…… 正答7	211~223	
			○○○○○○×……正答8	217~245	中級上
○	○	○	○○○○○×…… 正答8	217~245	中級上
			○○○○○○×…… 正答9	226~249	
			○○○○○○○×…… 正答10	256~270	上級
			○○○○○○○○×…… 正答11	273	
			○○○○○○○○○×…… 正答12	276	
			○○○○○○○○○×…… 正答13	265~279	
			○○○○○○○×…… 正答14	277	上級上
			○○○○○○○×…… 正答15	300	
			○○○○○○○×…… 正答~	~	
			○○○○○○○×…… 正答18	338	母語話者相当
○○○○○○○×…… 正答~	351				
○○○○○○○×…… 正答20	378				

表4では便宜上、○○○○のように表記しているが、アダプティブテストなので実際には、○○×○○××○のような形式はむしろ多い。

シミュレーションの中で、解答した問題アイテムの数が増えるにつれて、問題アイテムの解答を完全にコントロールすることができない(次に出題されるかどうか完全に予測できない)ので、例えば、正答は6の場合、聴解は6、語彙は6、文法は7(或いは5)、読解は6という形で終了される例もある。このように正答9(N)の場合、中に10(8)の分野があるかもしれない。

シミュレーションは、特に正答数8まで繰り返してやってみたので、何れも複数回であった。スコアは概ねその範囲内であった。

また、同じく正答数8問の場合でも、スコアがほとんど多少異なっている。それは、解答した問題アイテムの困難度が違うためである。

各パターンによるシミュレーションでは、色々な受験者のレベルが明確に区別されて測定できたの

で、本テスト（問題アイテム）が良好な困難度と識別度を持っていることが確認できた。同時に、シミュレーションを繰り返した結果としては、250点以上取れるのは、相当難しいということが分かった。

各パターンのスコアは客観的なものであり、レベルの区分は、あくまでも研究者の教育現場の経験に基づきながら、問題アイテムの困難度、問題アイテムの正答数による判断である。相関性について実際の検証をする必要がある。

3. 県立広島大学における受験実施

県立広島大学では、中国語を履修している学生を対象に、10名以上の集団受験2回を含めて、C-CATを受験してもらった。

3.1 実施条件

実施環境と設備

2回とも県立広島大学のマルチメディアラボにおいて、Wi-Fi環境で、Windows 10端末を使って実施した。その他は自宅のインターネットの環境で実施した。

受験者と実施日

表5 受験者と実施日

受験者	人数	実施日
週に2回中国語を履修している1年生 「中国語Ⅲ」	12	平成30年2月8日
週に2回中国語を履修している2年生 「中国語Ⅶ」	21	平成30年2月2日
週に2回中国語を履修している3年生 (中国語母語話者・台湾からの交換留学生) 「中国語文化論演習」	8 3	平成30年2月2日
週に2回中国語を履修している4年生 「ゼミ生」	3	平成30年2月9日

3.2 受験者のスコアと中国語検定

表6 1年生のスコア

ID	スコア					
	聴解	語彙	文法	読解	総得点	中検
sample209	29	32	35	30	126	未受験
sample208	24	30	39	39	132	
sample206	33	37	37	39	146	
sample203	28	40	39	43	150	
sample210	35	42	40	36	153	
sample207	37	41	42	36	156	
sample205	33	48	39	37	157	
sample204	30	50	36	45	161	
sample201	37	53	35	38	163	
sample212	37	50	42	39	168	
sample211	38	49	38	46	171	
sample202	38	42	40	54	174	同上

総得点から見るに、1年間の学習歴を有する受験者であれば、ほとんど「初級上」か「中級」に接近しているレベルにあることが分かる。また、スコアは普段の学習成績とかなり一致している。

表7 2年生のスコアと中国語検定

ID	スコア					
	聴解	語彙	文法	読解	総得点	中検
sample107	26	40	40	36	142	4○ 3×
sample111	35	47	31	36	149	4○ 3×
sample115	28	39	42	41	150	4○ 3×
sample103	21	48	39	45	153	4○ 3×
sample121	31	41	40	45	157	4○ 3○
sample120	33	42	49	41	165	4○ 3×
sample118	31	52	48	35	166	4○ 3○
sample106	38	54	40	36	168	4○ 3×
sample113	35	45	45	45	170	4○ 3○
sample109	35	47	49	40	171	4○ 3○
sample112	40	50	39	45	174	4○ 3○
sample105	38	44	54	41	177	4○ 3×
sample114	49	40	45	45	179	4○ 3×
sample117	38	40	48	54	180	4○ 3×
sample116	39	48	46	49	182	4○ 3×
sample110	44	59	48	32	183	4○ 3○
sample109	53	48	40	63	204	4○ 3○
sample102	37	54	69	45	205	4○ 3○
sample119	57	58	51	44	210	4○ 3○
sample108	57	58	64	58	237	4○ 3○
sample101					休み	
sample104	19	1	40	47	無効	4○ 3×

無効：受験中、ローディングエラーが発生したため、成績は正確ではない可能性あり無効とした。

中検：中国語検定2017年6月4級、11月3級の受験結果。○は合格、×は不合格で、以下は同である。

まとめてみると、

表8 レベル分布

スコア	人数	レベル
142 149	2	初級上
150~183	14	中級
204 205 210 237	4	中級上

総得点から見るに、約2年の学習歴を有する受験者であれば、「中級」か「中級上」にあることが分かる。また、本テストのスコアは、中国語検定4級、3級とは一定の相関性があることが観察できた。

表9 3年生のスコアと中国語検定

ID	スコア					
	聴解	語彙	文法	読解	総得点	中検
sample124	32	43	40	29	144	4○
sample123	37	43	49	53	182	4○
sample133	41	52	48	44	185	3○
sample126	34	51	48	54	187	3○ 2×
sample127	39	60	52	40	191	3○ 2×
sample125	37	58	58	47	200	3○ 2×
sample129	44	65	48	45	202	3○ 2×
sample122	57	58	63	52	230	3○ 2×
sample131	83	77	99	86	345	母語話者
sample132	80	98	89	86	353	母語話者
sample130	82	99	99	87	367	母語話者

1名を除けば、レベル中級上にあることが分かる。また、母語話者の総得点は、IDsample131はピンインが出来ないので、語彙の点数が若干低い、3名ともネイティブスピーカが取れるべき成績である。C-CATは見事にそれを反映しているといってもよからう。また、本テストでは250点を取ることが難しく、中国語検定も2級合格は難しいとされているので、少人数のデータだが、相関性がかなり高いと思われる。

表10 4年生のスコアと中国語検定

ID	スコア					
	聴解	語彙	文法	読解	総得点	中検
sample21	43	70	63	28	204	3○ 2×
sample22	39	50	64	47	200	4○ 3×
sample20	55	63	58	49	225	3○ 2○
sample134	78	99	91	83	351	準1級○

こちらも少人数のデータだが、スコアは見事に学習歴を反映しているし、中国語検定と相関性があると見られる。IDsample20は中国留学半年間の経験があって、リスニングが強いので中国語検定2級に合格したわけである。IDsample134は中国語学習者というより中国語使用者で、中国に長期滞在したこともあり、ネイティブスピーカ相当のレベルを持っていて、取得した中国語検定準1級のレベルにも合っている。

3.3 まとめ

県立広島大学の受験者のスコアは、シミュレーションの結果とほぼ一致していることが分かる。C-CATは、県立広島大学の受験では、4年生、3年生、2年生、1年生と、その成績が歴然と異なっているし、同じクラス（2年生の場合）でも、普段成績がよい学生は得点が高いことから、学習者の実態をよく反映している。

中国語検定との相関については、実施時期にずれがあるものの、C-CATのスコアは、中国語検定4級、3級、特に2級とはある程度相関性があると観察された。これは、あくまでも1大学のデータによる判断で、引き続き広範囲で検証すべきであろう。

4. 得点の解釈（点数とレベルの目安）

受験終了後、画面に結果が示される。その結果は（図1と図2）「成績证书」（Score report）としてPDF形式でダウンロードすることもプリントすることもできる。

この「成績证书」（Score report）の下には、「Interpretation of C-CAT score C-CAT的分数解释与相关参考」（図3）があり、得点とレベルの関係、CEFR、HSK、中国語検定との関連の目安を示したものである。また、利用者の便宜のため、成績書に以下の説明が書いてある。あくまでも目安であり、関連付けを保証するものではない。この点について後述する。

図3 Interpretation of C-CAT score C-CAT的分数解释与相关参考

Interpretation of C-CAT score C-CAT 的分数解释与相关参考					
C-CAT	Proficiency Level		CEFR	HSK	中国語検定
- 100	Basic	初级	A1	Level 1・2	准4級
100 - 150	Pre-Intermediate	初级上	A2	Level 3	4級
150 - 200	Intermediate	中级	B1	Level 4	3級
200 - 250	High-Intermediate	中级上	B2		
250 - 300	Advanced	高级	C1	Level 5	2級
300 - 350	High-Advanced	高级上	C2	Level 6	准1級
350 -	Near Native	接近汉语母语者			1級

・ "Basic" learners can exchange basic ideas.
 ・ 初级, 使用汉语能表达最基本的交际需求。
 ・ "Intermediate" learners can manage daily communication.
 ・ 中级, 使用汉语能进行日常会话和应对大部分交际任务。
 ・ "Advanced" learners can manage academic and professional communication.
 ・ 高级, 使用汉语能就学术、专业进行交流。

C-CAT.top

“Basic” learners can exchange basic ideas.

初级, 使用汉语能表达最基本的交际需求。

“Intermediate” learners can manage daily communication.

中级, 使用汉语能进行日常会话和应对大部分交际任务。

“Advanced” learners can manage academic and professional communication.

高级, 使用汉语能就学术、专业进行交流。

C-CATの結果（得点）は、項目応答理論での能力推定値を元に換算されたものである。プレテストの標本抽出において中国語学習者母集団が反映されるように配慮しているので、得点が各セクションで50点、総合で200点が中国語学習者母集団全体の平均に近く、それは「いわゆる」中級レベルに相当すると考えられる。

C-CATでは、得点を不変的に提供することで、受験者間、受験時期間の比較も行えるようにしている。しかし、それがどのレベルなのかに関しての判断は、利用者にお任せしている。例えば、「中級」

と言ったところで、その意味するレベルは、国・地域によって、また、教育機関によって異なるはずである。

従って、C-CATの得点に基づくレベル分けは利用者が行うべきことで、C-CAT側で決めるわけにはいかない。その意味で、C-CATは検定試験ではない。あくまでも信頼できる得点を提供するに徹している。それでも「目安」として、レベルとCEFR、HSK、中国語検定を示しているのは、個人受験者の場合で得点の解釈が全く出来ないと、困惑するだろうと思われるからである。そのためあくまで、便宜的に例示した「目安」に過ぎない。

なお、特に中国語検定との関係について、相関係数など統計値を示していないのは、中国語検定の難易度が安定せず（「中国語検定試験の信頼性と妥当性に関する研究」による）、テストの度にそれが変化してしまっていたので、相関係数を求めることに意味がないからである。HSKとの相関の考察は、機会を見て調査する予定である。ただし、テストは測定すべき構成概念も内容も形式もそれぞれ異なるので、テスト間の相関の解釈も慎重であるべきであろう。

C-CATを教育機関が利用する場合は、ここに示した「目安」とらわれることなく、C-CATの得点をその機関のレベル分けに合わせて再解釈して使用すべきである。

C-CATが保証しているのは得点の不変性であり、レベル分け自体や解釈ではない。例えば、TOEFLやTOEICの点数をもって、単位認定の条件としたり、入学の要件としたりする大学もあるが、何点以上でその要件を満たすとするかの設定は各教育機関に任せられている。TOEFLやTOEICの実施者であるETS (Educational Testing Service) が提供するのは不変的な指標としての得点だけであり、その解釈と利用方法は各教育機関に委ねられているわけである。C-CATの得点とその解釈は全くこれに当てはまると思われる。

5. 終わりに

本研究の推進実施にあたり、特に問題アイテムの作成・検討、プレテストの実施において、多くの方々と大学から多大なご協力を賜った。ここに改めて厚く御礼を申し上げる。

本テストは運用して以来、世界各地から中国語学習者に利用されており、本テストの利用と改善修正に対する提言があれば、お願いする次第である。

参考文献

今井新悟 編著 赤木彌生・中園博美 著『J-CATオフィシャルガイド コンピュータによる自動採点日本語テスト』ココ出版 2012年3月

一般社団法人 日本中国語検定協会HP 中国語検定試験

<http://www.chuken.gr.jp/tcp/outline.html>

ヨーロッパ言語共通参照枠 (Common European Framework of Reference for Languages : CEFR) フリー百科事典『ウィキペディア (Wikipedia)』

<https://ja.wikipedia.org/wiki>

汉语水平考试 (HANYU SHUIPING KAOSHI : HSK)

<http://www.hskj.jp/>

Abstract

On the correlation between C-CAT score
interpretation and other tests

C-CAT Project Team
Hou Renfeng (Representative)

关于C-CAT的分数解释及与其他考试的相关性研究

摘要

C-CAT的分数（得分），是根据项目反应理论推定出的能力值而换算成的。依据预测标本抽样，首先估算出了汉语学习者群体的平均学习水平。听力、词汇、语法、阅读各部分得分为50分，总分在200分，这个成绩与汉语学习者群体的平均水平接近，所以我们认为大致相当于“中级水平”。

C-CAT提供的是恒定不变的分数，考生之间、考试之间可以做比较。但是，判断分数具体相当于哪个水平，则由利用单位自行决定。因为，就以“中级”而言，也会因国家、地区、或教育机关不同而有异。所以，基于C-CAT的得分划分水平，由利用单位自行判断，C-CAT不提供划分标准，而只提供信度可靠的分数。

尽管如此，为了作为“参照”，C-CAT的分数通过与其他考试平行测试做了相关，这只是便于“参照”理解个人成绩（得分）而为。另一方面，平行测试的结果是，考生的C-CAT得分与其他考试的得分呈较高相关。

C-CAT所保证的是得分的不变性，不提供水平划分和解释。这也正如TOEFL、TOEIC一样，只提供分数，选拔或录取分数线的划定由利用单位自行判断。C-CAT的分数及其解释如同。