

原 著

英語歯科用語に関するテスト項目の研究 (2)

—オプションの数と種類の違い—

廣 瀬 浩 二

明倫短期大学 歯科衛生士学科

A Study of Test Items on English Terminologies for Dentistry (2):
Focused on the Differences in the Number and the Kind of Options

Koji Hirose

Department of Dental Hygiene and Welfare, Meirin College

本研究の目的は、平成9年度に作成した英語歯科用語のテスト項目を改善することである。被験者は、歯科衛生士学科に所属する短大1年生78名である。平成9年度と同じ多肢選択式テストではあるが、18のテスト項目(1)―(18)を作成し直した。各項目のstemは英語から日本語に、選択肢は日本語から英語に、また、その数は3から4に増やした。18項目について古典的テスト理論による分析と項目応答理論による分析の両方を行った。各項目を項目困難度・項目弁別力・モデルとの適合度の観点から分析した結果、平易すぎる項目について再検討を加えることとなった。テスト項目を作成し直したことによって、容易な項目数が減り、極端な得点分布の偏りが改善された。

キーワード：英語歯科用語、多肢選択式テスト、項目困難度、項目弁別力指数、モデルとデータの適合度

The purpose of this research is to improve the test problems of the English for Dentistry examination which was made one of the formative tests in 1998. Subjects were 78 college students who belong to the Department of Dental Hygiene. Eighteen multiple choice test items related to English for Dentistry were recreated and reexamined on the grounds that the test problems proved to be too easy in the previous research conducted in 1998. For amelioration, the test problems needed to increase slightly in level of difficulty. From the viewpoint of this difficulty, stem and distractors of each item were taken into consideration. The test type was the same multiple choice format as the test problems in 1998. The stem of each test item, however, substituted the Japanese equivalent for English terminology. One correct English terminology and three other distractors were arranged as options, that is, options were changed to English and the number was expanded from three to four. Each test item was analyzed by both classical test theory and item response theory. Consequently, items which should be reconsidered were confined to three problems, nos. 5, 10, and 12. The revision of the test problems resulted in a decrease of overly simple questions, and an improvement in the radical distribution of scores.

Key words: English for dentistry, Multiple choice test, Item difficulty, Item discrimination power index, Model-data fit

緒 言

明倫短期大学カリキュラム「英語Ⅱ」では、歯科衛生士学科1年生に対して英語歯科用語を教えている。

その教授内容は、主として単語の発音・日本語訳・用語の意味である。また、課程の評価方法として平成9年度では、1つの英語歯科用語に対し選択肢(options)として3つの日本語訳と当て推量を排除す

るための錯乱肢 (distractor) 1つを配したタイプのテスト項目を作成し学習者に課した。

このテストタイプで、錯乱肢 (“?” で表示した) を選択する学習者は実際には、ほとんど見られなかった。実施した 18 のテスト項目すべてについて古典的テスト理論 (Classical Test Theory: 以下 CTT と省略) による分析と項目応答理論 (Item Response Theory: 以下 IRT と省略) による分析を行った結果、適切と考えられる項目は 1 項目、易しいと考えられる項目は 14 項目、不良項目 3 項目となった (廣瀬¹⁾)。このように、平成 9 年度に実施したテストタイプでは易しいと考えられる項目数が多い結果となった。

そこで本研究では、適切な項目数を増やすために平成 9 年度とは異なったタイプのテスト項目を新たに作成実施し、その結果を再び CTT と IRT によって分析した。

言語テスト技術

言語テスト技術は Hughes²⁾ によれば、受験者の能力を引き出す手段であり、次の事項が含まなければならない。

1. テスト実施者が評価しようとする能力に関して、信頼性・妥当性のある指標となるような受験者の反応を引き出すこと。
2. 信頼できる得点を引き出すこと。
3. できる限り時間と労力を節約すること。
4. 有益な波及効果 (backwash effect) があること。

これに関連して本研究では、(1)授業で指導した英語歯科用語の習得の度を評価できる。(2)平成 9 年度実施のテストでは、信頼性係数が 0.825 と高い数値を得ている。(3)採点が比較的容易である。(4)授業においても日本語訳より英語の用語に注目するようになるであろう、との判断から多肢選択式テスト (Multiple choice) の形式に則って問題を作成した。

多肢選択式テストには、一つの「正答」と複数の錯乱肢 (distractors) が必要である。だが、これまで実際の項目の中には正答が複数ある場合、正答がない場合、効果的でない錯乱肢が配列してある場合などが見受けられた。良いテスト項目を作成することは困難なことである。テスト項目を作成する際に最も重要なことは、「正答」はまぎれもない正答でなければならない (Alderson et al.³⁾) ことである。この指摘は、このタイプのテストにこれまで果たして正答かどうか判断できない項目が見受けられたことを示している。この点に関して本研究では、作成した各テスト項目には

「正答」が一つしか存在しないことを確認した上でテストを実施した。

多肢選択式テストには、認識 (recognition) の知識を評価できるが、産出 (production) の知識までは評価できない、といった特徴がある。これは、構成概念妥当性 (construct validity) と深く関わる問題である。特に、応用言語学の分野には直接評価する基準のない構成概念が多く存在する。従って、理論に基づき予測検証しながら有効と考えられる仮説のみが妥当な構成概念となる (Hatch & Lazaraton⁴⁾)。この点に関連して、ESP (English for Specific Purposes) の分野における英語の役割 (Richards et al.⁵⁾)、特に歯科衛生士のためのコースにおける役割を考えた場合、用語指導と診療所における英会話指導が主要な指導事項となる。本研究では、用語指導に焦点化した。尚、指導する用語は正確に綴りを産出できる段階までは要求せず、認識の段階に留めておいて良いと判断した。

対象および方法

対象

本実験の対象は、本短大歯科衛生士学科 1 年生 78 名である。テストは平成 10 年 12 月に通常の授業時間内に実施した。

方法

テスト項目の作成

平成 10 年 11 月までに指導した 165 の英語歯科用語の中から、18 のテスト項目を平成 9 年度の経験に基づき新たに作成した。本研究では、「適切な項目」数を増やすためにテスト形式を変更した。前年度のテスト項目からの変更点は次の通りである。①多肢選択式のテストタイプは踏襲し、各テスト項目とも stem は英語の用語から日本語の歯科用語に変更した。②選択肢は日本語から英語に変えた。③平成 9 年度には “?” で表示していた錯乱肢を除去し、選択肢の数を 3 から 4 に増した。

各項目は以下に示すように、最初に日本語の用語を挙げ、次に日本語に対応する英語による選択肢 4 つ (正しい選択肢 1, 錯乱肢 3) を配列した。受験者には、各項目において日本語に対応する英語の用語を A ~ Eの中から選び、解答欄に記入させた。

- (1)彫刻刀 () ア, carvone knife イ, carvacrol knife ウ, carven knife エ, carving knife
- (2)槌 () ア, mallet イ, mallee ウ, malleolus エ, malleus

- (3)中切歯 () ア, centric incisor イ, centroclinal incisor ウ, centripetal incisor エ, central incisor
 (4)側切歯 () ア, latent incisor イ, lateral incisor ウ, later incisor エ, latter incisor
 (5)定期検査 () ア, recap イ, recall ウ, recast エ, recce
 (6)伝達麻酔 () ア, block anaesthesiology イ, block anaesthetic ウ, block anaesthesia エ, block anesthetics
 (7)他覚症状 () ア, objective symptom イ, objective symptomatology ウ, objective symptomology エ, objective symphysis
 (8)倒立円錐バー () ア, inverted comma bur イ, inverted con bur ウ, inverted cone bar エ, inverted cone bur
 (9)鍍銀法 () ア, silver impregnant method イ, silver imprescriptibility method ウ, silver impregnation method エ, silver imprecation method
 (10)ドライソケット () ア, dry sock イ, dory socket ウ, dry socket エ, dry sock
 (11)塗蠟絹糸 () ア, floss silk イ, flossy silk ウ, florin silk エ, florist silk
 (12)乳歯 () ア, mill teeth イ, milk teeth ウ, milky teeth エ, millcake teeth
 (13)嚢胞 () ア, cycle イ, cyst ウ, cyclist エ, cystine
 (14)膿瘍切開 () ア, incision of abscess イ, incisor of abscess ウ, incisor of absence エ, incision of absence
 (15)排唾器 () ア, salivary ejector イ, salivary ejection ウ, saliva ejection エ, saliva ejector
 (16)麦粒鉗子 () ア, dressed forceps イ, dressing forceps ウ, dressing forcite エ, dressed forcite
 (17)抜歯 () ア, extraction of the tooth イ, extractor of the tooth ウ, extradition of the tooth エ, extraneousness of the tooth
 (18)抜歯鉗子 () ア, dental extortionary forceps イ, dental extortionary forcite ウ, dental extracting forceps エ, dental extracting forcite

分析方法

テスト結果のデータ分析に際し、大友賢二と中村洋一作成の Test Data Analysis Program : TDAP Ver. 1. 0 (Copyright : Kenji Ohtomo & Youichi

Nakamura, 1996) を使用した。また、基礎統計量の算出には、FUNDAMENTAL STATISTICS のプログラムを、CTT による分析には ITEM ANALYSIS のプログラムを、IRT による分析には ITEM RESPONSE THEORY のプログラムをそれぞれ使用した。

結 果

基礎統計量

(1)得点の総計 (Sum of the raw scores) : 1219 (2)最小得点 (Minimum score) : 8 (3)最大得点 (Maximum score) : 18 (4)Median : 16 (5)Range : 10 (6)平均 (Mean) : 15.628 (7)分散 (Variance) : 5.618 (8)標準偏差 (Standard deviation) : 2.370 (9)Skewness : -1.229 (10)Kurtosis : 1.147

前年度の結果と比較すると、得点の総計は減少した (1235 → 1219)。最低得点は高くなった (3 → 8) が、最高得点は変わらない (18)。平均が低くなった (16.467 → 15.628) ため、Median も低くなった (17 → 16)。しかし、平均値が依然として高い位置にあるため、Skewness は負の数値を示している (-2.934 → -1.229)。最高得点と最低得点の差 (Range) は狭まった (15 → 10) ため、分散 (5.876 → 5.618) や標準偏差 (2.424 → 2.370) の数値も小さくなった。Kurtosis の数値がかなり小さくなった (11.438 → 1.147) ことから、得点の分布曲線は平坦なカーブを描き、得点分布の集中が回避される結果となった。

CTT による項目分析結果

表 1 は、CTT により、困難度・点双列相関係数による項目弁別力・実質選択肢数等の数値を求め、標準適切度合計に従って配列したものである。

表 1. CTT による項目分析結果: 標準適切度合計順位

RANK	NO.	DIFF	DISC	AENO	ADIF	ADIS	AAEN	SATOT
1	9	0.692	0.561	2.568	0.865	0.458	0.981	1.880
2	13	0.846	0.737	1.769	0.558	1.000	0.939	1.864
3	14	0.718	0.499	2.375	0.814	0.331	0.934	1.758
4	18	0.885	0.628	1.598	0.481	0.653	0.960	1.708
5	7	0.885	0.560	1.604	0.481	0.457	0.969	1.643
6	6	0.808	0.466	1.925	0.635	0.277	0.911	1.610
7	16	0.897	0.542	1.527	0.455	0.417	0.944	1.589
8	11	0.910	0.468	1.474	0.429	0.280	0.961	1.539

9	4	0.897	0.398	1.527	0.455	0.189	0.944	1.503
10	15	0.821	0.284	1.801	0.609	0.088	0.843	1.459
11	17	0.833	0.606	1.569	0.583	0.579	0.644	1.439
12	8	0.769	0.214	1.904	0.712	0.048	0.746	1.412
13	3	0.962	0.367	1.206	0.327	0.156	0.905	1.379
14	1	0.962	0.083	1.206	0.327	0.007	0.905	1.323
15	2	0.974	0.286	1.127	0.301	0.089	0.797	1.236
16	10	0.821	0.071	1.601	0.609	0.005	0.633	1.228
17	12	0.974	0.182	1.127	0.301	0.034	0.797	1.216
18	5	0.987	0.031	1.071	0.276	0.001	0.824	1.213

(NOTES)

RANK = 標準適切度合計による順位 (Rank in SATOT order)
 DIFF = 項目困難度指数 (Item difficulty index)
 DISC = 項目弁別力指数 (Discrimination power index)
 AENO = 実質選択肢数 (Actual equivalent number of options)
 ADIF = 項目困難度適切度 (Appropriateness of difficulty)
 ADIS = 項目弁別力適切度 (Appropriateness of discrimination power index)
 AAEN = 実質選択肢数適切度 (Appropriateness of actual equivalent number of options)
 SATOT = 標準適切度合計 (Standard appropriateness total)

信頼性

(1)アルファ係数 (Coefficient alpha) : 0.695 (2)標準偏差 (Standard deviation of the test) : 2.370 (3)測定の標準誤差 (Standard error of measurement) : 1.310 (4)得点の総和 (Sum of the raw scores) : 1219 (5)平均値 (Mean of the raw scores) : 15.628 (6)分散 (Variance of the test) : 5.618 (7)項目分散の和 (Sum of the item variance) : 1.933 (8)平均通過率 (Average proportion of passing) : 0.868 (9)平均不通過率 (Average proportion of failing) : 0.132

信頼性係数は予想に反し、平成9年度より低かった (0.825 → 0.695)。これはテスト項目数のためであろう。データ上、もし信頼性係数を上げようとするれば、表2のように新しいテスト項目を増やす必要がある。

表2. アルファ係数とテスト項目数の関係

If you want the following coefficient alpha:	Number of your new test items should be:
0.800	32
0.820	36
0.840	42
0.860	49
0.880	58
0.900	71
0.920	91
0.940	124

0.960	190
0.980	388
0.695	⇐ 18
0.927	⇐ 100

また、測定の標準誤差の数値は大きくなった (1.015 → 1.310)。これは、15点を取った受験者の中で、68%の人の真の得点は13.69から16.31に入り、95%の人の真の得点が12.38から17.62に入ることを示し、平成9年度に比較して誤差が大きくなっている。

IRTによる項目分析結果

表3は、IRTの1パラメーター・ロジスティック・モデルにより、項目困難度パラメーターとモデルとの適合度を検定した結果である。

表3. IRTによる項目分析結果

Item NO.	Final Calib.	Standard Error(d)	Fit(t)
1	-0.919	0.558	0.434
2	-1.765	0.775	-2.468
3	-0.919	0.558	-2.929
4	-0.025	0.410	-0.423
5	-2.589	1.086	3.290
6	0.882	0.324	-0.117
7	0.135	0.391	-1.911
8	1.179	0.307	2.337
9	1.706	0.289	0.161
10	0.775	0.331	4.900
11	-0.203	0.434	-1.458
12	-1.765	0.775	9.722
13	0.543	0.350	-3.285
14	1.537	0.293	0.434
15	0.775	0.331	0.905
16	-0.025	0.410	-1.941
17	0.543	0.350	-1.141
18	0.135	0.391	-2.237

考 察

本研究では、平成9年度第1学年在学生のテスト結果との比較を考慮に入れ、同一の用語に関するテスト項目を新たに作成した。テスト項目の内容を大幅に変

更したため、いくつかの相違が明らかになった。以下で、統計分析結果を踏まえながら検討する。

基礎統計量によると、Skewness は負の数値を得た (negative skew)。即ち、比較的高い得点の位置に受験者が多く集中したことを示している。ただ、その程度は平成 9 年度より緩やかであった ($-2.934 \rightarrow -1.229$)。どちらかといえば問題が比較的容易であったとの解釈もできるが、一方、よく学習して目標の達成に近づいた者が多かった、とも解釈できる。そもそも、本研究で実施したテストは形成的評価 (formative evaluation) の一つとして位置付けられ、目標基準準拠テスト (criterion-referenced test) であったことから判断するとこの結果はむしろ望ましいものと考えられる。

CTT による項目分析では、特に、項目困難度と項目弁別力指数に注目した。項目困難度は 0.500 が最も適切である、といわれる。本研究で作成した 18 項目の中では、項目番号(9)が最も適切なテスト項目であると考えられる。(1)(2)(3)(5)(11)(12)はかなり易しい項目であった、と判断され、項目(4)(6)(7)(10)(13)(15)(16)(17)(18)も易しい部類に属すると考えられる。しかしながら、これら比較的易しい項目も配列の仕方によっては有効に活用できる。また、項目弁別力指数 (discrimination index) に関しては、一説には +0.4 若くはそれ以上が満足する数値といわれるが、特に決まった数値はない。ただ、負の数値を示した場合には、明らかに不良項目で除外若くは改訂の必要がある (Alderson et al.³⁾) が、該当する項目はなかった。

IRT による項目分析では、特に、モデルとの適合度に注目した。モデルとの適合度は、2.00 が危険域である。2.00 あるいはそれ以上の場合、応答妥当性を欠きモデルとは適合しないと考えられる (Henning,⁶⁾)。モデルとの適合度では、項目番号(5)(10)(12)の 3 項目がモデルと適合せず、テスト項目から除外したほうが良いと考えられる。一方、負の適合度を持つ項目が 10 項目あった。 -2.00 あるいはそれ以下の場合、モデルと過剰適合していると考えられるが、テスト項目から除外する理由にはならない。項目(5)(10)(12)は、CTT における項目困難度分析では、どれも易しい項目に属していた。項目弁別力指数では不良項目は見当たらなかったが、少なくともこの 3 項目については作成し直した方が良いと思われる。

テスト形式は CTT や IRT による分析を意図したため、多肢選択形式のテストタイプをとった。平成 9 年度の研究結果からは、若干難易度を増す必要性が生じ

ていた。そこで、各項目とも stem を英語から日本語に、選択肢を日本語から英語に、選択肢の数を増し問題を作成し直した。その結果、既に述べたように平均値は依然として高い位置にあるが、総得点が減少し、得点の集中的な分布が減少した。改善すべき項目数は同数となったが、最終項目困難度 (final calibration) は、Wright and Stone⁷⁾ の例示によると、項目(5)はかなり易しい、項目(1)(2)(3)(4)(7)(11)(12)(16)(18)はやや難しい、項目(6)(8)(9)(10)(13)(14)(15)(17)はより難しい、と考えられる。平成 9 年度と比較して、かなり易しい項目が減少し ($2 \rightarrow 1$)、やや難しい項目が増加した ($4 \rightarrow 9$)。

テスト結果の分析に使用した One-Parameter (Rasch) Model (1 PLM) で必要な受験者は 100 名 (Alderson et al.⁵⁾) と言われる。例えば、ミネソタ大学で実施された入学時と卒業時の学生の言語能力 (フランス語・ドイツ語・スペイン語) の測定の際には、十分な学生数が確保され Two-Parameter Model (2 PLM) の使用が可能であった (Lozier and Chalhoub-Deville⁸⁾)。本研究においては、受験者数はやや不足していたが、測定の精度を高めるため 1 PLM を使用した。実際の教育現場に適用する場合には、必ずその要件を満たせるとは限らない。

結 語

本研究では、英語歯科用語の習得状況を評価するため、新たにテスト項目を作成し学生に課した。テスト結果を CTT と IRT によって分析した。分析では、項目困難度・項目弁別力・モデルとの適合度に注目した。その結果、18 項目のうち項目(5)(10)(12)の 3 項目について検討を加えることにした。テスト項目を作成し直したことによって、容易な項目数が減り、極端な得点分布の偏りが改善された。

今後の課題は、英語歯科用語に関するテスト項目をさらに増やすこと、可能な限り適切な受験者数を確保することである。項目銀行 (item bank) を発展させ活用することは、多大な努力を払うに充分値するものであるが、膨大な時間と専門的知識が必要である (Hughes²⁾)。より適切でより多くのテスト項目に関するデータを蓄積することが、受験者の変化に影響を受けない不変性を備えた項目銀行 (item bank) の形成につながる。

文 献

- 1) 廣瀬浩二：英語歯科用語に関するテスト項目の研究

- (1) — 項目困難度, 項目弁別力指数, モデルとデータの適合度の検討 —. 明倫齒誌, 1: 39-43, 1998
- 2) Hughes A : Testing for Language Teachers. Cambridge University Press, Cambridge, 1989
- 3) Alderson J C, Clapham C and Wall A D : Language Test Construction and Evaluation. Cambridge University Press, Cambridge, 1995
- 4) Hatch E and Lazaraton A : The Research Manual, Design and Statistics for Applied Linguistics. Heinle & Heinle Publishers, Boston, 1991
- 5) Richards J C, Platt J and Platt H : Longman Dictionary of Language Teaching and Applied Linguistics. Longman, Harlow, 1992
- 6) Henning G : A Guide to Language Testing. Heinle & Heinle Publishers, Boston, MA, 1987
- 7) Wright B D and Stone M H : Best Test Design. MESA Press, Chicago, 1979
- 8) Lozier V M and Chalhoub-Deville M : Preliminary Item Response Theory Analysis of the University of Minnesota CLA Language Proficiency Tests in French, German, and Spanish. Center for Advanced Research on Language Acquisition, University of Minnesota, Minneapolis, 1997