

原 著

英語歯科用語に関するテスト項目の研究 (6) - テスト信頼性と妥当性の2年間の検証 -

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 (6) : A Two-Year Investigation into Test Reliability and Validity

Koji Hirose

Department of Dental Hygiene and Welfare, Meirin College

本研究では、平成12年度と13年度に実施した英語歯科用語のテスト項目の信頼性と妥当性を検証することを目的とした。テストは、50の歯科用語に関するものと7つの対話文から構成された。信頼性はクロンバック・アルファとカッパ係数(13年度のみ)によって、妥当性は内部構成概念証明法によって検証した。被験者は、歯科衛生士学科に所属する短大1年生69名(12年度)と64名(13年度)である。クロンバックの α 係数(信頼性)は、0.886(12年度)と0.826(13年度)であった。カッパ係数は $\kappa=0.69$ (13年度のみ)でほぼ満足できる数値であった。この結果から、テストの信頼性は証明された。内部構成概念証明法では、ピアソンの積率相関係数によった。12年度では2項目を除き、13年度では10項目を除いて、歯科用語に関するテスト得点の相関は対話文に関するテスト得点の相関より高かった。内部構成概念証明法では証明されなかった項目数は増えたが、運用上で工夫が必要であろう。また、今後は、内容的妥当性についても検討したい。

キーワード：英語歯科用語、構成概念妥当性、内部構成概念証明法、信頼性

The purpose of this research was to examine the validity and reliability of the English test for the Dentistry examination which was made and administered as one of the formative tests in the each academic years of 2000 and 2001. The test battery comprised of a subtest of 1 of 50 dental terms and a subtest of 2 of 7 dialogues. Reliability was estimated by the Cronbach's coefficient alpha and the kappa coefficient, only for 2001. Validity was estimated by internal construct validation.

Subjects were 69 (2000) and 64 (2001) college students who belonged to the Department of Dental Hygiene and Welfare. The reliability coefficients for the subtest 1 were 0.886 (2000) and 0.826 (2001). The kappa coefficient (2001) was a satisfactory value. These results revealed that the subtest 1 was highly reliable. Internal construct validation was determined to employ among some validation methods in order to investigate the test validity. Both categories of the subtest 1 and subtest 2 were analysed by the Pearson's product moment correlation coefficient. The correlation coefficients between an individual score and total scores on dental terms were consistently higher than those between an individual score and total scores on dialogue, except two (2000) and ten (2001) test items. The number of items that did not prove that they were valid increased. Refinement in arranging test items will be reviewed when creating the test. Another validation, content validity will be examined in future research.

Key words : English for dentistry, Construct validity, Internal construct validation, Reliability

緒 言

本研究において、筆者は本学歯科衛生士学科「英語Ⅱ」における歯科用語に関するテストを作成し、実施した。そのテスト形式は、伝統的な筆記テスト (paper-and-pencil language test) であり、多肢選択式 (multiple choice format) である。また、テストの目的からすると指導内容の評価のためのものであり、到達度テスト (achievement test) であり、目標規準準拠テスト (criterion-referenced test) の領域に属するものであった。平成13年度は、その信頼性と妥当性を検証した¹⁾。信頼性は満足できる数値を得たが、妥当性が若干の問題が残った。そこで、改良を加えたテストを実施し、その結果を分析し、信頼性と妥当性の点で変化がみられたか検証することとした。

テストの信頼性に関しては、目標規準準拠テストで有効な「しきい値損失一致度の推定 (the estimating threshold loss agreement)」を使用することにした。この分析に際しては、分割点 (the cut-point score) とクロンバック・アルファ (Cronbach alpha) が必要になる。

テスト妥当性に関しては、①テスト内容とその効果の研究に関連する内的妥当性 (internal validity) 、②テスト得点をテスト以外の評価結果と比較研究する外的妥当性 (external validity) 、③テストが構成概念を測定しているかを研究する構成概念妥当性 (construct validity) の3つに大きく分類される²⁾が、本研究では目標規準準拠テスト (CRT) に適応可能な構成概念妥当性を上げる。その検証手法は、内部構成概念証明法 (internal construct validation)³⁾ である。

平成12年度と13年度の2年分のテスト問題を信頼性と妥当性の両面から総合的に評価した。

対象および方法

1. 対象

本研究の対象者は、平成12年度本短大歯科衛生士学科1年生69名と平成13年度1年生64名である。テストは平成13年1月と平成14年1月の後期試験期間内に実施した。

2. 方法

1) テストの目的

本テストの目的は、本学歯科衛生士学科「英語Ⅱ」の指導内容に関して、学生の学習状況を客観的に評価

することである

2) テストの構成

本研究で使用したテストは、2種類のテストからなる。50のテスト項目からなる英語歯科用語に関するものと対話文の穴埋め問題である。

英語歯科用語に関するテスト項目の特徴は次の通りである。①テストタイプは多肢選択式で、各テスト項目とも基幹部分 (stem) は日本語の歯科用語名である。②選択肢 (options) は英語の歯科用語名である。③選択肢の数は4である。

対話文の穴埋め問題では、授業中、学習した歯科衛生士と患者の会話の一部を取り上げた。7つの短い対話例を提示し、合計20か所に適する語句を挿入する形式である。挿入すべき語句は選択肢として示してある。また、空欄を含む英文には日本語訳も付与した。

3) 分析方法

本研究で分析しようとするテストは、目標規準準拠テスト (criterion-referenced test) の領域に属するものである。従って、集団基準準拠テスト (norm-referenced test) に有効な手法を利用しても正確さに欠けるため、分析に適用できる手段は制限される。

信頼性の推定に関しては、クロンバックのアルファ係数 (Cronbach's coefficient alpha) とカッパ係数 (13年度) を利用する。アルファ係数は次の計算式で求められる。

$$2 \left(1 - \frac{\text{奇数番号のテスト項目の標準偏差の二乗} + \text{偶数番号のテスト項目の標準偏差の二乗}}{\text{テスト総得点の標準偏差の二乗}} \right)$$

カッパ係数算出のために次の計算式を用いた。

$$\frac{\text{二分割点の素点} - 0.5 - \text{平均点}}{\text{標準偏差}}$$

妥当性の統計的分析では、テストの下位項目間の相関を調べる内部構成概念証明法を採用し、ピアソンの積率相関係数 (Pearson's product moment correlation coefficient) を使用する。

分析に際し、大友賢二と中村洋一作成のTest Data Analysis Program: TDAP version.1.0 (Copyright: Kenji Ohtomo & Youichi Nakamura, 1996) 及びTDAP version.2.0 (2002) を使用した。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。また、ピアソンの積率相関係数の算出には、田中敏作成のSTAR version 2.8.3bを使用した。

結果と考察

1. 基礎統計量

表1は、英語歯科用語のテスト項目に関する基礎統

表1. 基礎統計量

	12年度	13年度
受験者数 (Number of examinees) :	69	64
得点の総計 (Sum of the raw scores) :	2274	2205
最小得点 (Minimum score) :	4	17
最大得点 (Maximum score) :	48	48
中央値 (Median) :	34	36
範囲 (Range) :	44	31
平均 (Mean) :	32.957	34.453
分散 (Variance) :	67.694	42.248
標準偏差 (Standard deviation) :	8.228	6.500
歪み (Skewness) :	-0.823	-0.522
尖度 (Kurtosis) :	1.249	-0.008

計量を示したものである。

平均値は32.957 (12年度), 34.453 (13年度)と比較的高い位置にあるため、Skewnessはいずれの年度も負の数値を示している (-0.823, -0.522)。得点分布は得点の高い方に偏っていたことがわかる。最高得点と最低得点の差 (Range) はかなりの幅があり (44, 31), 分散 (67.694, 42.248) の数値も大きい。13年度では数値が小さくなっている。得点分布集中の傾向がみられるが、Kurtosisの数値がかなり小さい (1.249) か負の数値 (-0.008) を示した。このことから、得点の分布曲線は比較的平坦なカーブを描き、極端な得点分布の集中が回避される結果となっている。ある科目の終了後に実施した良いCRTの得点分布は負の方向に歪むのが理想的であろう。その意味では、本テストでは、Skewnessは負の数値を示し、負の方向 (高得点側) に歪んでいるので良いCRTであるといえる。

2. 信頼性

次に、信頼性 (Reliability) ・構成概念妥当性 (Construct validity) について順を追って検討する。

表2. 信頼性係数

	12年度	13年度
アルファ係数 (Coefficient alpha)	0.886	0.826
標準偏差 (Standard deviation of the test)	8.228	6.500
測定の標準誤差 (Standard error of measurement)	2.781	2.708
得点の総和 (Sum of the raw scores)	2274	2205
平均値 (Mean of the raw scores)	32.957	34.453
分散 (Variance of the test)	67.694	42.248
項目分散の和 (Sum of the item variance)	8.934	8.031
平均通過率 (Average proportion of passing)	0.659	0.689
平均不通過率 (Average proportion of failing)	0.341	0.311

信頼性係数は0.886 (12年度), 0.826 (13年度)であった。観察された得点の分散の約89% (12年度), 約83% (13年度)は真の得点の分散に依存し、説明のつかない誤差分散は11% (12年度), 17% (13年度)に過ぎなかった。このことから、信頼性係数は十分に高い数値であり、信頼性のあるテストであったといえる (表2)。

測定の標準誤差は2.781 (12年度), 2.708 (13年度)であった。もし、40点を取った受験者がいたと仮定すると、その中で、68%の人の真の得点は37.219から42.781 (12年度), 37.292から42.708 (13年度)に入り、95%の人の真の得点が34.438から45.462 (12年度), 34.584から45.416 (13年度)に入ることとなる。許容範囲の誤差であるといえる。

さらに、13年度では、カッパ係数を算出した。分析方法で既に記述した計算式による数値と一致度係数とカッパ係数の近似値一覧表であるSubkoviak⁴⁾のTable 2 (1988:49) およびTable 3 (1988:50)を参照することによって、 $\kappa=0.69$ を算出した。この数値は控えめなものであり、一般に、50項目で信頼性係数0.70といわれることからほぼ満足のいく数値が得られたといえる。

3. 妥当性

ここでは、以下のように内部構成概念証明法 (Internal Construct Validation) によって、構成概念妥当性を検証する。

表3. 用語各テスト項目と下位テスト総得点との相関係数一覧

12年度		13年度	
・V1VT=.095	> ・V1DT=.023	・V1VT=.452**	> ・V1DT=.293*
・V2VT=.228*	> ・V2DT=.177	・V2VT=.304*	> ・V2DT=.059
・V3VT=.200*	> ・V3DT=.000	・V3VT=.331**	> ・V3DT=.130
・V4VT=.276*	> ・V4DT=.174	・V4VT=.152	< ・V4DT=.235*
・V5VT=.245*	> ・V5DT=.223*	・V5VT=.208*	< ・V5DT=.216*
・V6VT=.417**	> ・V6DT=.366**	・V6VT=.235*	> ・V6DT=.227*
・V7VT=.544**	> ・V7DT=.359**	・V7VT=.573**	> ・V7DT=.378**

・V8VT=.171	>	・V8DT=.059	・V8VT=.314*	>	・V8DT=.260*
・V9VT=.232*	>	・V9DT=.196*	・V9VT=.518**	>	・V9DT=.279*
・V10VT=.259*	>	・V10DT=.044	・V10VT=.433**	>	・V10DT=.387**
・V11VT=.348**	>	・V11DT=.215*	・V11VT=.000	=	・V11DT=.000
・V12VT=.566**	>	・V12DT=.276**	・V12VT=.580**	>	・V12DT=.303*
・V13VT=.492**	>	・V13DT=.277**	・V13VT=.350**	>	・V13DT=.144
・V14VT=.338**	>	・V14DT=.234*	・V14VT=.355**	>	・V14DT=.207
・V15VT=.252*	>	・V15DT=.162	・V15VT=.000	=	・V15DT=.000
・V16VT=.627**	>	・V16DT=.555**	・V16VT=.388**	>	・V16DT=.205
・V17VT=.572**	>	・V17DT=.457**	・V17VT=.588**	>	・V17DT=.324**
・V18VT=.526**	>	・V18DT=.273*	・V18VT=.270*	>	・V18DT=.152
・V19VT=.506**	>	・V19DT=.423**	・V19VT=.496**	>	・V19DT=.371**
・V20VT=.325**	<	・V20DT=.369**	・V20VT=.405**	>	・V20DT=.257*
・V21VT=.504**	>	・V21DT=.315**	・V21VT=.369**	>	・V21DT=.102
・V22VT=.378**	>	・V22DT=.309**	・V22VT=.344**	>	・V22DT=.225*
・V23VT=.294**	>	・V23DT=.112	・V23VT=.447**	>	・V23DT=.346**
・V24VT=.545**	>	・V24DT=.381**	・V24VT=.431**	>	・V24DT=.320*
・V25VT=.410**	>	・V25DT=.270*	・V25VT=.483**	>	・V25DT=.311*
・V26VT=.542**	>	・V26DT=.328**	・V26VT=.536**	>	・V26DT=.339**
・V27VT=.490**	>	・V27DT=.298**	・V27VT=.622**	>	・V27DT=.482**
・V28VT=.250*	>	・V28DT=.187*	・V28VT=.369**	>	・V28DT=.232*
・V29VT=.466**	>	・V29DT=.298**	・V29VT=.476**	>	・V29DT=.301*
・V30VT=.225*	>	・V30DT=.168	・V30VT=.000	=	・V30DT=.000
・V31VT=.342**	>	・V31DT=.215*	・V31VT=.054	<	・V31DT=.108
・V32VT=.107	>	・V32DT=.041	・V32VT=.147	<	・V32DT=.168
・V33VT=.309**	>	・V33DT=.232*	・V33VT=-.017	>	・V33DT=-.110
・V34VT=.735**	>	・V34DT=.527**	・V34VT=.364**	<	・V34DT=.394**
・V35VT=.281**	>	・V35DT=.130	・V35VT=.256*	>	・V35DT=.019
・V36VT=.410**	>	・V36DT=.240*	・V36VT=.347**	<	・V36DT=.381**
・V37VT=.322**	>	・V37DT=.179	・V37VT=.066	>	・V37DT=-.009
・V38VT=.339**	>	・V38DT=.247*	・V38VT=.325**	>	・V38DT=.203
・V39VT=.493**	>	・V39DT=.419**	・V39VT=.137	>	・V39DT=.116
・V40VT=.396**	>	・V40DT=.286**	・V40VT=.533**	>	・V40DT=.356**
・V41VT=.536**	>	・V41DT=.311**	・V41VT=.465**	>	・V41DT=.111
・V42VT=.496**	>	・V42DT=.427**	・V42VT=.198	>	・V42DT=.150
・V43VT=.335**	<	・V43DT=.396**	・V43VT=.570**	>	・V43DT=.342*
・V44VT=.342**	>	・V44DT=.214*	・V44VT=.178	>	・V44DT=.070
・V45VT=.652**	>	・V45DT=.348**	・V45VT=.575**	>	・V45DT=.403**
・V46VT=.486**	>	・V46DT=.383**	・V46VT=.487**	>	・V46DT=.228*
・V47VT=.318**	>	・V47DT=.268*	・V47VT=.025	>	・V47DT=-.005
・V48VT=.345**	>	・V48DT=.181*	・V48VT=.228+	>	・V48DT=.188
・V49VT=.303**	>	・V49DT=.189*	・V49VT=.316*	>	・V49DT=.162
・V50VT=.270*	>	・V50DT=.174	・V50VT=-.043	<	・V50DT=-.002

「英語II」の試験における歯科用語に関するテスト部門と同一試験の他のテスト部門の関係を検討した。例えば、歯科用語の項目1の得点をV1、歯科用語の総得点をVTとし、対話文問題の総得点をDTとすると、 $V1VT > V1DT$ の関係が成立すれば、構成概念妥当性が証明できる。

表3に示されたように、12年度では、歯科用語テスト項目のno.20(現病歴)とno.43(小白歯)以外の項目では、 $V1VT > V1DT$ の関係が成立し、構成概念妥

当性が証明された。また、13年度では、no.4(アマルガム計量器)、no.5(アマルガム修復)、no.11(エキスプローラー)、no.15(鎌型スケーラー)、no.30(歯科衛生士)、no.31(歯垢)、no.32(歯根膜)、no.34(歯痛)、no.36(自覚症状)、no.50(問診)の10項目で $V1VT > V1DT$ の関係が成立しなかった。

12年度の項目no.20では誤答が38あり、項目no.43では誤答が19であった。no.20はかなり困難な項目で、no.43は適度な困難さを伴った項目であったと考えられる。

13年度では、no.11, no.15, no.30は全員正解の項目で非常に容易な項目であった。これら3項目は、到達度テストという観点からみるとむしろ望ましいと言えるかもしれない。残る7項目の誤答人数を挙げると、12(no.4), 5(no.5), 2(no.31), 55(no.32), 23(no.34), 32(no.36), 38(no.50)となる。この中でも、CRTであるという本テストの性格上特に問題となるのは、誤答人数の多かったno.32, no.36, no.50であろう。内部構成概念証明法では証明できなかった妥当性も別の観点から検討する必要があるだろう。また、それらの項目も活用次第では十分にテスト項目として機能できるはずである。項目配列上で工夫することとなる。

以上の分析をふまえて、選択肢の再検討が必要となろう。また、本テストは目標規準準拠テスト(criterion-referenced test)であるという性格上、指導の際にもこれらの項目に配慮し正の波及効果(washback)を期待したい。

結 語

本研究では、英語歯科用語に関するテストの有用性を検証するため、本学歯科衛生士学科で実施した「英語II」の試験結果を分析した。信頼性(reliability)・構成概念妥当性(construct validity)について考察を加えた。信頼性は、高い数値のもと証明できた。構成概念妥当性については、用語に関する50のテスト項目の中、2項目(12年度)、10項目(13年度)を除いて証明できた。内部構成概念証明法では証明できない項目が増えたが、活用する際の工夫次第では機能すると思われる。

テストの妥当性を証明することはたいへん困難であるが、妥当性を高めることはさらに困難なことである。実際、妥当性は低くても信頼性は高いというテストは存在する。しかし、逆に、信頼性が低くても妥当性が高いテストは存在しない。今後は、妥当性の中でもCRTに適用できるもう一つの妥当性である内容的妥当性

(content validity) についても検証し、より良いテスト項目の作成に取り組みたい。

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 (5) - テスト妥当性の検証 - . 明倫歯誌, 5(1) : 30-33, 2002.
- 2) Alderson J C, Clapham C and Wall A D : Language Test Construction and Evaluation. pp183-185, Cambridge University Press, Cambridge, 1995.
- 3) Henning G : A Guide to Language Testing. pp97-100, Heinle & Heinle Publishers, Boston, MA, 1987.
- 4) Subkoviak M J : A practitioner's guide to computation and interpretation of reliability indices for mastery tests. J Educational Measurement, 25(1) : 47-55, 1988.