

## 「全民英檢」九十一年高級複試測驗

第一次「全民英語能力分級檢定測驗」（以下簡稱「全民英檢」）高級複試（含寫作與口說測驗）已於九十一年八月二十四及二十五日在台北考區舉辦。199 名初試通過之報名考生中，194 名出席應考，其中 32 名通過寫作與口說測驗，取得高級測驗合格證書。參加本次測驗之 194 名考生的平均年齡為 24.39 歲，其中屬於「全民英檢」高級測驗檢測對象的考生（國內大學英語主修系所或曾赴英語系國家大學或研究所取得學位者）佔所有考生的 45%。

考生的成績表現及作答分析，分寫作測驗及口說測驗詳述如下。

### 一、寫作測驗

高級寫作測驗共分兩大題，測驗時間全長 105 分鐘。第一大題考生須先閱讀兩篇相關主題的文章（verbal input），再依題目指示寫一篇 250 字的作文。作文內容須包含兩篇文章重點摘要及個人看法，作答時間為 60 分鐘。第二大題，考生須先解讀（interpret）兩個相關主題的圖表（non-verbal input），再依題目指示寫一篇 250 字的文章。文章必須描述兩個圖表所傳達的重點訊息或現象、臆測可能原因及提出個人建議，作答時間為 45 分鐘。

評分時採用 1 至 5 級分評分量表，評分人員須根據考生文章的內容適切性（Relevance and Adequacy）、組織（Coherence and Organization）、遣詞用字（Lexical Use）及文法結構（Grammatical Use）的表現評分。考生的兩篇作文必須皆達 3 級分，始能通過寫作測驗（量表請參閱附錄一）。

為求評分標準的一致與穩定，評分人員在正式閱卷前，須先參加評分講習，先熟悉評分程序、範例文章及評分重點，並進行試評。試評結果須達 80% 的一致性，始得參與正式閱卷。每份答卷分別由二位評分者評閱，成績取兩閱平均。如兩閱總分差距大於四或「通過與不通過」結果不一致時，交由第三位評分人員複閱，成績取第三閱結果。

### 1. 各項統計分析

#### 1.1 考生部份

九十一年高級寫作測驗應考人數為 194 名，最高得分為 4 級分，最低得分為 1 級分，平均成績為 2 級分。達通過標準 3 級分以上者（含）共 35 名，通過率為 18%。各級分分布情形請參見表一。

表一 九十一年高級寫作測驗  
成績分布情形

級分	人數	累計人數	累計百分比
5	0	0	0%
4	10	10	5%
3 (Pass)	25	35	18%
2	87	122	63%
1	72	194	100%



分項表現上，考生在內容適切性（Relevancy & Adequacy）與組織（Coherence & Organization）兩部份的平均分別為 2.19、2.27，稍高於遣詞用字（Lexical Use）及文法結構（Grammatical Use）—1.89、1.98（表二）。

表二 九十一年高級寫作測驗  
分項平均成績

	內容	組織	遣詞用字	文法結構
第一大題	2.34	2.36	2.10	1.99
第二大題	2.39	2.41	2.20	2.09
整卷	2.19	2.27	1.89	1.98

本次考生年齡最長 48 歲，為大學畢業在職者；年齡最小 12 歲，為國小學生。全民英檢高級的目標施測對象—「國內大學英語主修系所」或「曾赴英語系國家大學或研究所進修並取得學位者」—年齡 20~29 歲的考生約佔本次測驗人數四成六。成績表現方面，以「25~29 歲」及「35~39 歲」兩年齡層考生表現最好，平均值皆為 2.00，通過情形亦最高，均在三成以上（表三）。

表三 九十一年高級寫作測驗  
年齡層分布情形

年齡層	人數	百分比	平均級分	通過人數	通過率
14 歲以下	10	5%	1.90	1	10%
15~19 歲	50	26%	1.88	8	16%
20~24 歲	56	29%	1.88	7	13%
25~29 歲	33	17%	2.00	11	33%
30~34 歲	22	11%	1.59	3	14%
35~39 歲	10	5%	2.00	3	30%
40~44 歲	7	4%	1.29	1	14%
45 歲以上	6	3%	2.17	1	17%

另「14 歲以下」年齡層考生表現相當接近平均值，但通過率卻最低。經分析發現，該年齡層考生在「遣詞用字」與「文法結構」方面雖表現不錯，皆高於該分項平均值。但在「內容適切性」與「組織」方面卻明顯不如其他較年長考生，皆低於該分項平均值，以致無法通過寫作測驗。

推測「14 歲以下」考生在「內容適切性」與「組織」方面較弱的原因，可能與題目要求及考生心智年齡成熟度有關。高級寫作測驗除要求考生摘要文章及圖表重點外，亦要求考生對題目所提出的社會現象或問題，表達個人意見及深入探討。此對仍處於認知發展階段且社會經歷尚淺的國中、國小學生而言，困難度較高。尤其是在抽象概念的表達或邏輯舉證上的掌握，其精確度確實略遜於其他較年長的組群。此推測亦可從部分評分老師的考後意見中得到佐證。部分評分老師曾表示有些作文語言部分的掌握不錯，但其內容多為童言童語，因而無法在“Relevancy & Adequacy”或“Coherence & Organization”部分給予較高分數。以第二大題為例，其讀者設定為報社編輯，寫作目的為「讀者投書」，此年齡層考生雖在圖表部份



多能以正確的語言描述，但在舉證或提供解決問題時，「太天真」、缺乏邏輯或因果關係倒置，無法有效支持其論點。此類作品，在真實生活中亦無法說服報社編輯採納，刊載於報章中。

194 名考生中，通過率最高者為具國外大學學歷考生，平均級分 2.63、通過率 56%。26 位「國內研究所英語主修系所」考生次之，平均級分 1.93、通過率 36%。36 位「國內大學英語主修系所」考生之平均級分居第三位，1.86、通過率 17%。此三組考生的表現及通過率皆優於「國內大學/研究所非英語主修系所」考生之表現（表四）。

表四 九十一年高級寫作測驗  
學歷背景分布情形

教育背景	人數	平均級分	通過人數	通過比率
國內大學				
- 英語主修系所	36	1.86	6	17%
- 非英語主修系所	28	1.69	3	10%
國內研究所				
- 英語主修系所	14	1.93	5	36%
- 非英語主修系所	12	1.75	1	8%
具外國學歷				
- 大學學歷	16	2.63	9	56%
- 研究所學歷	22	1.56	3	14%

## 1.2 評分人員部份

本次測驗寫作試卷評分第一閱與第二閱評分 90%達一致。

## 1.3 題目部份

194 名考生在第一大題的表現平均值為 1.89，第二大題平均值為 1.98。第一大題的通過比例為 18%，較第二大題 23%稍高，但難度上無顯著差異（ $p > 0.05$ ）。第一大題與第二大題成績相關係數為 0.79（表五）。依據 Many-Facet Rasch Tasks Measurement 分析，兩部份寫作的 Separation 值為 .71，Reliability 為 0.33，Significance 為 0.08，顯示第一大題、第二大題兩個題型所檢測的寫作能力非完全相同的能力，故兩種題型確有同時存在的價值（表六）。

表五 九十一年高級寫作測驗  
第一大題、第二大題成績分布表

級分	第一大題			第二大題		
	人數	累計人數	累計%	人數	累計人數	累計%
5	1	1	1%	2	2	1%
4	10	11	6%	11	13	7%
3	24	35	18%	31	44	23%
2	90	125	64%	87	131	68%
1	69	194	100%	63	194	100%
平均		1.89			1.98	
相關係數		0.79				



表六 九十一年高級寫作測驗  
第一大題、第二大題測驗題型統計分析

Obsvd	Obsvd	Obsvd	Fair-M	Model	Infit	Outfit	
Score	Count	Average	Average	Measure	S.E.	MnSq ZStd	MnSq ZStd   N TASKS
4473	1060	4.2	4.00	.80	.03	1.0 0	0.9 -1   2 2
4404	1065	4.1	3.92	.87	.03	1.0 0	1.0 0   1 1
4438.5	1062.5	4.2	3.96	.83	.03	1.0 -0.2	1.0 -0.8   Mean (Count: 2)
34.5	2.5	0.0	0.04	.04	.00	0.0 0.5	0.0 0.6   S.D.
RMSE (Model) .03 Adj S.D. .02 Separation .71 Reliability .33							
Fixed (all same) chi-square: 3.0 d.f.: 1 significance: .08							

## 2. 建構效度分析

### 2.1 研究目的

依「全民英檢」高級寫作測驗之設計，評分者須就考生作文的內容適切性、組織、遣詞用字及文法結構等四項進行評量，各部分表現均須達量表中 3 級分的敘述，始得通過寫作測驗。其中有關遣詞用字及文法結構兩項係評量英文用語的字彙量（range）、複雜度（complexity）及正確性（accuracy）。以下三個研究題目係針對「遣詞用字」及「文法結構」兩項重點提出：

- 各級分考生在字彙量上是否存在差異？
- 各級分考生在句構複雜度上是否存在差異？
- 各級分考生在文字使用的錯誤量上是否存在差異？

### 2.2 研究方法

#### 2.2.1 抽樣

從各級分中，抽出等量且各部分表現相當的考生作文，進行分析比對。囿於通過寫作測驗人數不多，最後篩選出符合上述條件的考生共八名，其「遣詞用字」及「文法結構」兩項成績分布請參見表七。

表七 九十一年高級寫作測驗  
抽樣考生「遣詞用字」及「文法結構」兩項成績分布

考生代碼	最後成績 (級分)	第一大題成績		第二大題成績	
		遣詞用字	文法結構	遣詞用字	文法結構
A	4	4	4	4	4
B	4	4	4	4	4
C	3	3	3	3	3
D	3	3	3	3	3
E	2	2	2	2	2
F	2	2	2	2	2
G	1	1	1	1	1
H	1	1	1	1	1



### 2.2.2 字彙量分析

字彙量分析部分採用兩種方式分析，一為比較每篇文章的相異詞比例（type-token ratio），以了解各級數考生使用的字彙量是否明顯不同。理論上，程度愈佳者，詞彙應相對較豐富，表達意見時，詞彙變化應較多（variety），而重覆使用相同字詞的比率應較低（Huang, 1999; Richards, Platt & Weber, 1985）。

相異詞比例的定義如下：

$$\text{Type-token ratio} = \frac{\text{number of separate words}}{\text{total number of words in the text}} \times 100$$

以“have”及“has”為例，“have”及“has”在計算文章字數時，視為兩個字（two tokens）。但因“have”及“has”皆為“to have”的衍生字，故在計算相異字數時，視為同一個字（one type）。相異詞比例愈高時，表示文章中字彙重覆使用的情形愈少。

字彙量分析的另一個方式為比對每篇考生作文的字彙（types）與 Collin's COBUILD 字表間的差異。Collin's COBUILD 將常用字彙（frequent words）分成五級（Bands 1~5），此五級字彙涵蓋了所有英語口說及書寫使用字彙之 95%的。

根據 Collin's COBUILD 統計，五級約有 680 字，四級約有 1,040 字。四、五級囊括了 75%的英語使用（all English usage）字彙，屬核心字彙（the core vocabulary of the English language）。三級約有 1,580 字，三級字彙可增廣使用者討論議題的範圍及空間。二級約有 3,200 字，一級約有 8,100 字。一、二級的字彙為較不常用但仍非常重要的字（Collin's COBUILD, 2001）。

依據「全民英檢」高級寫作分項能力說明：通過高級寫作測驗者，應具有足夠的字彙，針對各類主題表達看法及深入探討。因此通過高級寫作測驗的考生，Collin's COBUILD 三級字彙量應較未通過者多。

### 2.2.3 句構複雜度與錯誤量分析

句構部分採用 T-Unit 方式分析句子的複雜度（linguistic complexity of sentences）。本分析所採用的 T-Unit 定義，係依據 J. Richards et al.所編 Dictionary of Applied Linguistics（J. Richards, 1985）：

*"...the shortest unit (the Terminable Unit, Minimal Terminal Unit, or T-Unit) which a sentence can be reduced to, and consisting of one independent clause together with whatever dependent clauses are attached to it."*

根據文獻上記載（S. C. Weigle, 2002; E. Homburg, 1984; E. Barham et al., 1979; F. O'Hare, 1971; J. C. Mellon, 1969; R. O'Donnell et al., 1967; K. Hunt, 1965），T-Unit 的數量、長度及 error-free T-Unit 數量是偵測語言學習者句



構發展成熟度的指標 ( signs of maturity in syntactic development ) 。程度愈佳者，T-Unit 長度愈長，error-free T-Unit 數量亦較高。

本研究錯誤計算採計次方式，亦即計算錯誤發生的次數 ( number of incidences )，而非錯誤的字數 ( number of words )。如例一，有一個介系詞和動詞的使用錯誤，故在計算錯誤量時，數目為 2，而非 4。另獨立子句 ( 例二 ) 或片段不完整的字串 ( fragment )、語意不清的字串 ( garble；例三 )，在本分析亦視為一個錯誤量。

例一：... a survey on people's ... has been revealed that ...

例二：Since I am an office worker who needs to work.

例三：I'd like to suggest conducting a new survey geared the service so much.

## 2.3 結果

### 2.3.1 相異詞比例統計分析結果

八位抽樣考生中，除考生 G 外，其餘考生作文字數皆達「250 字」長度要求，每篇作文平均相異字數 ( types ) 為 153 字，平均長度為 300 字 ( tokens )，平均相異詞比例為 51。惟考生 G 因其作文長度僅 180 字左右，造成計算時分母較小，相異詞比例高出其他考生許多的情形，故其數值並不足以證明該考生字彙量較其他考生廣。如剔除此筆資料，則各級考生間的相異詞比例皆落在 50 上下，並無明顯差異。換言之，各級考生間，在詞彙運用的變化上 ( variety ) 差異不大 ( 表八 )。此結果與原假設略有差距。

表八 九十一年高級寫作測驗  
抽樣考生作文相異詞比例分析

級分	題型	Types 平均值	Tokens 平均值	相異詞 比例	該級平均值
4	第一大題	150	293	51.5	50.5
	第二大題	148	299	49.5	
3	第一大題	157	303	51.9	49.7
	第二大題	176	371	47.6	
2	第一大題	148	288	51.3	49.3
	第二大題	169	349	48.4	
1	第一大題	143	263	58.5	58.1 ( 含 G 考生 )
	第二大題	135	239	57.7	49.4 ( 不含 G 考生 )
總平均		153	300	51.0	

但如以「題型」分析則發現，第一大題的相異詞比例，不論級分，皆高過



第二大題（表八）。顯示考在第一大題作答時，語言使用的變化較第二大題大。此現象與部份評分老師的觀察相符。

### 2.3.2 Collin's COBOULD 字表比對結果

八位抽樣考生作文所使用的字彙，無論級數，約有 98% 落在 Collin's COBOULD 的五級字表。其中約 77% 落在四、五級，11% 在三級，10% 在一、二級（表九）。此與 COBOULD 所記載的分布比例接近：一至五級為 95%；四、五級為 75%。但原先假設：通過高級寫作測驗者（3、4 級分者）在三級部份的比例應較未通過者（1、2 級分者）高的情形並未出現。

表九 九十一年高級寫作測驗  
抽樣考生字彙量與 Collin's COBUILD 字表比對結果

級分	題型	四、五級	平均	三級	平均	一、二級	平均	一至五級	平均
4	第一大題	72%	75%	14%	12%	10%	10%	95%	97%
	第二大題	78%		10%		11%		99%	
3	第一大題	76%	78%	12%	12%	12%	10%	99%	100%
	第二大題	81%		11%		9%		100%	
2	第一大題	77%	78%	13%	11%	7%	9%	97%	98%
	第二大題	79%		9%		11%		99%	
1	第一大題	73%	76%	13%	11%	10%	10%	95%	97%
	第二大題	79%		9%		11%		99%	

但如以「題型」分析則發現，各級數所有抽樣考生在第一大題部份使用的三級字彙比例皆比第二大題高。第一大題比例落在 12%~14% 間，而第二大題比例則落在 9~11% 間（表十）。換言之，第一大題可引出（elicit）較廣的字彙。

表十 九十一年高級寫作測驗  
抽樣考生第一大題、第二大題字彙與 Collin's COBUILD 字表比對結果

題型	四、五級	三級	一、二級	一至五級
第一大題	72~74%	12~14%	7~12%	96%
第二大題	78~81%	9~11%	9~11%	99%



### 2.3.3 T-Unit 統計分析結果

八位抽樣考生作文 T-Unit 平均個數為 16.8，平均長度為 18.6 字。極端能力值考生（4 級分與 1 級分）T-Unit 平均數量較低，分別為 14.8、12.8，但平均字數較多，皆在 20 字以上。中間能力值的考生（3 級分與 2 級分）T-Unit 平均數量較高，分別為 19.0、20.5，但平均字數較少，3 級分為 18.1 字，2 級分為 15.8 字（表十一）。

表十一 九十一年高級寫作測驗  
抽樣考生作文 T-Unit 個數與長度分析統計

級分	題型	T-Unit 平均個數	該級平均	T-Unit 平均長度	該級平均
4	第一大題	15.5	14.8	18.8 字	20.4 字
	第二大題	14.0		22.0 字	
3	第一大題	17.5	19.0	18.2 字	18.1 字
	第二大題	20.5		18.0 字	
2	第一大題	17.5	20.5	16.7 字	15.8 字
	第二大題	23.5		14.9 字	
1	第一大題	13.5	12.8	19.5 字	20.0 字
	第二大題	12.0		20.6 字	
總平均			16.8		18.6 字

以上數據顯示，大致上寫作能力與 T-unit 長度呈正比。程度愈佳者，T-unit 愈長，此與文獻所載結果相似。本分析另一發現為：極端能力值考生（4 級分與 1 級分）皆有企圖以「較長句構」展現其寫作能力的傾向，惟差別在於 4 級分者能成功地達成其目的（successful attempt），1 級分則力不從心（unsuccessful attempt）。

錯誤量分析方面，error-free T-Units 比例以四級分者最高，46%；二、三級分者居中，分別為 22%及 17%；一級分者最低，僅有 6%（表十二）。

表十二 九十一年高級寫作測驗  
抽樣考生 Error-Free T-Unit 平均數量與錯誤量統計分析

級分	題型	Error-free T-Unit 個數	Error-free T-Unit 百分比	該級平均
4	第一大題	7.5	49.5%	46%
	第二大題	6.5	43.0%	
3	第一大題	4.5	20.5%	17%
	第二大題	3.0	14.0%	
2	第一大題	4.5	21.5%	22%
	第二大題	5.5	23.5%	
1	第一大題	1.0	8.5%	6%
	第二大題	0.5	3.0%	



以上數據顯示，大致上寫作能力與 error-free T-unit 數量呈正比。換言之，程度愈佳者，error-free T-unit 佔全文的比例愈高，此與文獻所載結果相似。

從本分析亦可發現：中間能力值考生（2、3 級分者），其 error-free T-unit 比例差異並不明顯。經調閱考生作文，分析錯誤，發現 3 級分的錯誤多屬獨立字、詞（individual words or phrases），如介系詞、冠詞或搭配詞方面的錯誤等，對文意並無重大影響。2 級分的錯誤則多為連續性字串，如句型結構上的錯誤、不完整字串或語意不清字串（fragment or garble）等，造成語意不清。又，因 2 級分的錯誤多呈帶狀，分布集中，3 級分的錯誤則較分散，故而形成兩組 error-free T-unit 比例比差異不大的情形。

### 3. 問卷調查結果

此次高級寫作測驗共收集到有效問卷 191 份，綜合摘要如下。問卷詳細內容與統計數字請參閱附錄二。

題目說明方面，超過九成五以上的考生認為試卷作答的說明清楚或非常清楚。多位考生表示題目「有趣、生活化、易發揮」。

整卷難度方面，七成七以上的考生認為此測驗可以或大致可以測出其寫作能力，較預試時之七成高。題目難度方面，九成以上的考生認為第一大題、第二大題题目的難易度適中、容易或非常容易。其中，第一大題部分更高達九成七，與預試時的結果相同。另，第一大題题目中的兩篇閱讀文章，亦有八成七的考生認為難度相當或非常相當。

作答時間方面，第一大題有近六成的考生認為作答時間適中。第二大題有七成六的考生認為作答時間適中。作答內容部分，第一大題部分有六成二的考生認為容易或非常容易發揮，第二大題則有八成表示容易或非常容易發揮。

其他改進意見方面，部分考生對第一大題閱讀文章的字體及擬稿空間有意見，希望能增加試題冊版面或提供擬稿紙。

### 4. 討論與建議

#### 4.1 考生表現

「全民英檢」高級寫作測驗係為「國內大學英語主修系所」或「曾赴英語系國家大學或研究所進修並取得學位者」所設計的測驗，九十一年首次測驗近七成為 20 歲以上的考生，20 至 30 歲間考生亦有四成六。成績表現以大學畢（25~29 歲組）最好，平均級分為 2.00，通過率三成三。以主修系所分，以國外學歷及國內大學英語主修科系兩群組表現最好，平均 2.03、1.86，通過率為三成二、一成七。顯示高級寫作測驗內容設計與測驗目標相符。

分項成績表現上，考生普遍在內容適切性與組織兩項表現稍優於遣詞用字與文法結構，分別為 2.19、2.27 與 1.89、1.98。此現象與預試時結果相同。惟「14 歲以下」群組呈現不同現象；該群組考生在「遣詞用字」與



「文法結構」方面平均優於其他較年長群組，但在「內容」與「組織」方面不如其他較年長群組，以致形成平均值（1.9）最接近總平均值（2.00），但通過率卻最低（10%）的現象。根據評分老師觀察，此現象可能與考生心智年齡成熟度有關。高級寫作測驗除要求考生摘要文章及圖表重點外，同時要求考生必須對題目所提出的社會現象或問題，進行申述並表達個人看法。此對仍處於認知發展階段、社會經歷尚淺的考生而言，困難度較高。

#### 4.2 字彙量分析

在相異詞比例及 COBOULD 五級字表比對分析中發現，各級分考生在詞彙變化及常用/非常用詞彙的使用數量兩項變數上並無明顯不同；但如以「題型」區分，則有明顯差異。考生在第一大題部份的詞彙變化較多且範圍較廣。

此現象或與高級考生特質及測驗型態有關，參加高級寫作測驗的考生為已通過高級聽力與閱讀測驗，兩項成績皆達 80 分以上者，其能力值已相當集中且位於金字塔頂端，雖寫作成績有高有低，但已在一定水準以上（表十三）。

表十三 九十一年高級寫作測驗  
各級抽樣考生字彙量、句構複雜度及錯誤量比較分析

級分	Type-Token Ratio	COBOULD 四/五級	五級字彙分布 三級	一/二級	T-Unit 平均數目	T-Unit 平均長度	Error-Free T-Units 百分比
4	50.5	75%	12%	10%	14.8	20.4 字	46%
3	49.7	78%	11%	10%	19.0	18.1 字	17%
2	49.3	78%	11%	9%	20.5	15.8 字	22%
1	49.4*	76%	11%	8%	12.8	20.0 字	6%

\*不含 G 考生

#### 4.3 句構複雜度分析

本分析結果大致與文獻所載相似，亦即寫作能力愈佳者，T-Unit 長度愈長，error-free T-Unit 比例愈高。此外，本分析亦發現極端能力值（1、4 級分者）與中間值考生（2、3 級分者）有不同作答傾向。前者考生傾向使用較長的 T-Unit（平均每一 T-Unit 20 字）展現其寫作能力，差別在於四級分者能正確地運用句構或連接詞（linking devices），而 1 級分者則相反，或因句構、連接詞、標點符號的不當使用，造成錯誤的長句（表十三）。

能力值愈接近通過標準的中間考生作答則相對謹慎保守，傾向使用較短的 T-Unit。3 級分 T-Unit 平均長度為 18.1 字，二級分為 15.8 字。錯誤量方面，兩組的 error-free T-Unit 比例相差不大（表十三）。惟 3 級分多屬獨立字、詞錯誤，對文意影響不大；2 級分則多為連續性字串上的錯誤，影響文意。



#### 4.4 問卷調查結果

九成五以上的考生認為試卷作答說明清楚。七成七認為本測驗可有效評量其英語寫作能力。九成以上的考生認為題目的難度適中或容易，並表示題目「有趣、生活化、易發揮」。六成考生認為作答時間適中或稍長。除最後一項指數較預試時低外，其餘皆較預試結果佳，顯示考生對高級寫作測驗的設計大致滿意。

#### 5. 結語

全體考生成績統計分析及問卷調查結果顯示，「全民英檢」高級寫作測驗在題型與難易度設計符合測驗目標——適合「國內大學英語主修系所」或「曾赴英語系國家大學或研究所進修並取得學位者」。建構效度分析結果顯示，高級寫作測驗評分量表中「遣詞用字」及「文法結構」兩分項的能力敘述大致恰當，可鑑別不同能力值考生的寫作表現。惟囿於此次報考人數及符合比對條件的樣本較小，日後仍需持續觀察，以確認所得結果。

### 二、口說測驗

#### 1. 測驗題型

第一部分為考生與主考之間的交談，以問答的方式進行，約五分鐘，主要評量考生自我介紹及回答問題的口語能力。

第二部分包括考生訊息交換、討論及回答主考提問等，約七分鐘，主要評量考生口語互動與討論之能力。

第三部分為考生依據主考所提問題思考兩分鐘後發表，另一考生則針對該生之意見發表作口頭摘要，歷時約十分鐘。此部分主要評量考生針對主題作較深入之表述及口頭摘要的能力。

#### 2. 測驗方式

測驗全長約二十五分鐘，每場有兩至三名考生參加。測驗全程錄音/影，試場內有一位主考（Interlocutor）及一位評分人員（Assessor）。Interlocutor 負責向考生提問並依據整體式評分量表給分，Assessor 則負責依據分項式評量表給分（兩項量表請參閱附錄三、四）。

#### 3. 評分標準

分項式評分與整體式評分的優、缺點於文獻上多有論述（Hughes, 1989; Bachman & Palmer, 1996）。本口說測驗採整體式及分項式評分並行制，兩項計分由低至高分為 1 至 5 五個級分。採並行制可達到兩種評分方式相互驗證的目的（Hughes, 1989, p. 110），同時提高評分信度。分項式評量表評分項目包含發音語調（Pronunciation）、內容適切性（Relevance & Adequacy）、遣詞用字（Lexical Use）、文法結構（Grammatical Use）、流利度（Fluency）與連貫性（Coherence）。

#### 4. 評分方式

為使評分標準一致，Interlocutor 及 Assessor 在正式測驗前均接受數階段的標準化訓練（standardization）及演練，以熟悉評分標準、級分說明及測驗進行方式。原則上，考生必須達到整體式評量表之 3 級分及分項式評分



量表中每項評分標準的 3 級分，才能通過高級口說測驗。當兩項評分有差距時，Interlocutor 與 Assessor 則必須加以討論以達共識，若無法達成共識，則由第三位資深評分人員根據錄影內容作最後的確認。

## 5. 評分者信度

本次測驗 194 位考生中，因兩位評分人員所給級分相差兩級分以上而須再評者，有 31 位考生，重評率（discrepancy rate）為 16%（亦即評分一致性達 84%），顯示評分人員的評分標準穩定。

## 6. 考生成績

### 6.1 考生表現

#### 6.1.1 整體式成績

所有考生的平均成績為 2.54 分，依性別、年齡統計之平均成績請參見表十四。由表十四可知，男、女考生之平均表現相當，分別為 2.53 與 2.54 分（如陰影部分），而各年齡層中，以 15 至 19 歲之平均分數最高（如斜線部分）。

表十四 九十一年高級口說測驗  
依性別、年齡分布之整體式平均成績

類別		人數	人數 %	平均成績
性別	男性	80	41%	2.53
	女性	114	59%	2.54
年齡	14 歲以下	10	5%	2.40
	15-19 歲	50	26%	2.68
	20-24 歲	56	29%	2.57
	25-29 歲	33	17%	2.58
	30-34 歲	22	11%	2.45
	35-39 歲	10	5%	2.40
	40-44 歲	7	4%	2.00
	45 歲以上	6	3%	2.17

各級分考生人數分布請參見表十五。3 級分以上的考生共計 75 人，約佔所有考生人數之四成（如陰影部分）。另外，五個級分中，以 2 級分考生人數最多，共 110 人（如斜線部分）。

表十五 九十一年高級口說測驗  
各級分考生人數分布

級分	人數	累計人數	累計%
5	5	5	3%
4	28	33	17%
3 (通過)	42	75	39%
2	110	185	95%
1	9	194	100%

另外，本測驗所檢測的目標考生（亦即「英語程度相當於國內大學英語主



修系所或曾赴英語系國家大學或研究所進修並取得學位者」)之整體式平均成績及通過率(與寫作測驗成績現象相同)請參見表十六。

由表十六可看出,平均表現最佳者為具有國外大學學歷之考生,平均級分 3.25、通過率 75%,國內大學英語主修考生次之,平均級分 2.58、通過率 39%,其他依序為國外研究所考生、國內研究所英語主修考生、國內大學非英語主修考生、國內研究所非英語主修考生。

表十六 九十一年高級口說測驗  
目標考生之整體式平均成績及通過率

學歷	人數	平均級分	通過率
國內大學			
英語主修	36	2.58	39%
非英語主修	28	2.21	25%
國內研究所			
英語主修	14	2.36	29%
非英語主修	12	2.00	8%
國外大學	16	3.25	75%
國外研究所	22	2.41	32%

#### 6.1.2 分項式成績

考生依年齡層區分之分項式各項平均成績及通過率請參見表十七。

表十七 九十一年高級口說測驗  
依年齡分佈之分項式平均成績及通過率

年齡 \ 評分項目	發音 語調	內容 適切性	遣詞 用字	文法 結構	流利度	連貫性	通過率
14 歲以下	3.50	3.00	2.90	3.10	2.90	2.90	30%
15-19 歲	3.46	3.01	3.10	3.10	3.31	3.09	48%
20-24 歲	3.27	3.02	2.83	2.61	3.16	3.10	41%
25-29 歲	3.32	3.18	2.97	2.71	3.24	3.20	39%
30-34 歲	2.95	3.23	2.86	2.45	2.82	3.00	27%
35-39 歲	3.30	3.10	3.10	2.60	3.20	3.30	40%
40-44 歲	2.57	2.57	3.00	2.29	2.57	2.43	14%
45 歲以上	2.67	2.50	2.67	2.17	2.83	2.33	17%
所有考生	3.25	3.03	2.93	2.73	3.12	3.04	39%

如表十七所示,15-19 歲年齡層之通過率較高(48%)。此外,該年齡層的遣詞用字、文法結構及流利度之平均成績均為所有年齡層中最高者,發音語調則僅次於 14 歲以下考生。

此外,年齡層愈低考生的發音語調平均成績大致愈高(如斜線部分)。而就所有考生之分項成績而言,以發音語調之表現最佳,其他各項依序(由高分至低分)為流利度、連貫性、內容適切性、遣詞用字、文法結構(如菱線部分)。此現象與預試結果一致。



目標考生六項分項式評分項目之平均成績請參見表十八。

表十八 九十一年高級口說測驗  
目標考生之分項式平均成績

學歷	發音 語調	內容 適切性	遣詞 用字	文法 結構	流利度	連貫性
國內大學						
英語主修	3.31	3.11	2.81	2.58	3.17	3.14
非英語主修	3.07	2.71	2.50	2.18	2.68	2.71
國內研究所						
英語主修	2.93	2.93	2.86	2.43	3.00	2.86
非英語主修	2.42	2.67	2.42	2.08	2.58	3.00
國外大學	3.97	3.81	3.81	3.53	4.13	3.59
國外研究所	3.05	3.05	3.05	2.55	2.95	3.09

由表十八可看出，具國外大學學歷考生之各項平均成績均居目標考生之首（如陰影部分），國內研究所非英語主修考生則除連貫性外，其他五項之表現均不及其餘各類組目標考生（如斜線部分）。此現象很可能是由於在各類組目標考生中，國內研究所非英語主修學生接受英語課程與接觸英語環境之機會最少所造成。

## 6.2 各級分之考生表現分析

### 6.2.1 目的

藉由比較各個級分的考生表現，了解各級分考生之表現差異，並同時驗證通過與否分界點（cutting point）之適切性。研究題目如下：

- 各級分考生在字彙的表現上，是否有量的差異
- 各級分考生在字彙的表現上，是否有質的差異
- 各級分考生在文法句構的表現上，是否有所不同

### 6.2.2 對象

為減少影響考生表現的變數，因此選取考同一套試題的考生作為研究對象。由於口說測驗屬於複試階段，因此報名考生皆為通過高級聽力與閱讀測驗之考生，亦即為英語能力已達相當程度的考生，1 級分之考生實屬難尋，且此次所選之對象考生中，並無 1 級分者，因此僅就 2 至 5 級分考生作分析。各級分考生中挑選男、女各一名作分析，共計七名考生：男性三名、女性四名（五級分僅有一位考生）。另外，挑選原則為整體式與分項式評分一致者，不一致而經複閱者，則不列入此次分析。七名考生之性別及級分請參見表十九：

表十九 九十一年高級口說測驗  
抽樣考生基本資料

口說成績（級分）	性別
5	女一名
4	男、女各一名
3	男、女各一名
2	男、女各一名



### 6.2.3 方法

由於考生於本測驗之第三部分，須針對主考官所給的題目準備兩分鐘後，再以兩分鐘的時間表述所準備的內容，因此每位考生於第三部分的表現最為獨立（independent），並無如第一或第二部分作答時之互動影響，而且發表的時間也最長、最完整。故將七名考生於第三部分申述之作答謄錄成文字，以便研究比對。研究分以下三項：

- Collins Cobuild 字典將搜集於 The Bank of English 的字彙，依出現頻率分為五級（Bands 1~5），其中頻率最高的字分別為第五、第四級，約有 2,000 字，多為一般的文法用字及常用的字彙（"common grammar words and very frequent vocabulary items"），而這兩級區（word band）的字，據統計，約佔每位英語使用者所用英語之 75%，第三級約有 1,500 字，屬於可對議題作較深入表達或發揮的字彙，而第二、第一級共有 11,300 字，為使用頻率較低（low-frequency words）的重要字彙。通常語言能力愈佳者，認識的英文字愈多，愈可以使用低頻率字彙表達看法或陳述事件。因此本研究將考生表現與 Collins Cobuild 五級字表作比對，計算每位考生作答時所使用之字彙與所屬字彙級區，以便了解考生字彙使用情形是否因級分而有所不同。
  - 根據 Longman 語料庫（Longman Grammar of Spoken and Written English, 1999, p. 53）之數據分析結果顯示，通常相異詞比例（type-token ratio）的高低代表該語料之重複字的寡多，亦即相異詞比例愈高，重複字愈少，而重複字愈多表示字彙量愈低、變化較少，使用重複字表達意見的趨勢愈明顯。因此利用 Nation 軟體（Nation, 1996）計算考生作答時所使用字彙之相異詞比例，以了解考生之字彙量（lexical range）是否因級分而有所不同。
  - 如 Foster, Tonkyn, and Wigglesworth（2000）所言，"subordination is frequently used as a measure of complexity"（Foster et al., 2000, p. 366），句構的複雜程度與使用者語言能力有直接的關係，通常語言能力愈佳者，使用複雜句表達意見的情形愈普遍，且由於複雜句的結構使然，每句複雜句的字數也會比簡單句多。因此計算複雜句的數量及句子長短，就成為了解考生語言能力的方法之一。由於所分析的是口語語料（spoken text），故採用 Foster et al. (p. 365) 建議的 AS-Unit（the Analysis of Speech Unit: "An AS-unit is a single speaker's utterance consisting of an independent clause, or sub-clausal unit, together with any subordinate clauses(s) associated with either."）計算每位考生的 AS-Unit 數量及平均每 AS-Unit 的字數，以了解考生所用句構之複雜性（complexity of sentence structure）是否因級分而有所不同。
- 在此必須說明的是，由於考量分析所可能耗費的人力與時間，因此在計算 AS-Unit 時，並未將考生發音、文法之錯誤或其所作的自我更正（self-correction）等字詞剔除。

### 6.2.4 結果與分析

- 經比對字表，七名考生依級分分類之第三部分作答所使用字彙數量與級區請參見表二十：



表二十 九十一年高級口說測驗  
各級分考生所用字彙量與級區分佈

口 說 級 分	Collins Cobuild Word List					平均總字數
	Band 5	Band 4	Band 3	Band 2	Band 1	
	較常用字 ←————→ 不常用字					
5	70 ( 77% )	10 ( 11% )	5 ( 5% )	4 ( 4% )	2 ( 2% )	91
	93%			7%		
4	123 ( 76% )	23 ( 14% )	3 ( 1% )	10 ( 6% )	3 ( 1% )	81
	92%			8%		
3	155 ( 73% )	29 ( 14% )	16 ( 8% )	8 ( 4% )	5 ( 2% )	107
	94%			6%		
2	97 ( 75% )	20 ( 15% )	6 ( 5% )	5 ( 4% )	2 ( 2% )	65
	95%			5%		

- 由上表約略可以看出，七名考生之平均字彙量中，3 級分以上考生之字彙使用量較高，約為 80 字以上。比較特別的是，4、5 級分考生之總字彙並無多過 3 級分考生之現象。經檢視 4、5 級分共三位考生及 3 級分兩位考生之分項成績發現，4、5 級分三位考生之各項成績均為 4、5 級分以上，而 3 級分兩位考生之各項成績則為 3 級分或 4 級分。因此除了字彙量的多寡外，是否能正確使用字彙及其他分項成績之好壞，都會影響考生的整體表現，也因此字表比對結果才会有上述的情況發生。而依級分由高至低，考生使用較高級字彙（即較不常用字—low frequency words）之百分比有依級分高低逐漸降低的現象（請見表二十陰影部分）。亦即，級分較高之考生所用較高級字彙之字數及其佔總字數的百分比均較高。
- 經 Nation 軟體計算，考生第三部分作答之相異詞比例請參見表二十一：

表二十一 九十一年高級口說測驗  
抽樣考生字彙相異詞比例計算結果

口說級分	Type/Token	Average Type/Token	Ratio	Average Ratio
5	115/302	115/302	38	38
4	109/250	103.5/232.5	44	45
	98/215		46	
3	120/404	117.5/363	30	33
	115/322		36	
2	67/213	85.5/207.5	31	41
	104/202		51	

由上表可看出 2 級分之一考生的相異詞比例最低，3 級分的兩位考生次之，4 級分及 5 級分考生之相異詞比例則接近或高於 40。亦即 4、5 級分考生之字彙量較豐富、字彙使用較有變化，因此趨向於使用不同的字表達意見與看法，3 級分考生次之，而 2 級分考生之字彙量則較少。



然而較特別的是，5 級分的考生之相異詞比例似乎稍低（如陰影部分），而 2 級分之另一考生則稍高（如斜線部分），似與以上發現不符。然經比較兩位考生的分項成績發現 5 級分考生之分項成績中，除遣詞用字及連貫性表現未達 5 級分外，其餘四項均為 5 級分之滿分，經兩位主考討論達成共識後，給予 5 級分的最後成績。而 2 級分之該位考生之分項成績則是發音語調、流利度與連貫性為 3 級分，但遣詞用字、文法結構及內容適切性均為 2 級分，經兩位主考討論達成共識後，給予 2 級分的最後成績。由此可知，僅以相異詞比例詮釋考生表現，有其不足之處，也因此，本分析進一步計算考生之 AS-Unit 數量，以期得到更完整的分析結果。

- 經統計，七名考生第三部分作答所使用的 AS-Unit 數量及平均字數請參見表二十二：

表二十二 九十一年高級口說測驗  
抽樣考生所用 AS-Unit 數量及平均字數

口說級分	AS-Unit 數量	各級分平均 AS-Unit 數量	平均字數	各級分平均字數
5	24	24	11.8	11.8
4	17	15	13	13.2
	12		13.4	
3	21	21	11.6	11.2
	21		10.8	
2	12	12	10.4	7.8
	13		5.2	

由表二十二可發現，除了 AS-Unit 之數量較多外，通過考生之每 AS-Unit 的平均長度（平均字數）亦較長，3 至 5 級分之平均字數約在 11 個字以上（如陰影部分），顯示通過考生使用之句子較複雜。上表中較特殊的是，2 級分的一名考生所使用的 AS-Unit 平均字數雖不少（10.4 字），但檢視其分項成績發現，該考生除了發音語調達 3 級分外，其他各項成績均僅達 2 級分，顯示該考生的遣詞用字、文法結構正確性待加強。

#### 6.2.5 總結

請參見表二十三匯整之分析數據。

表二十三 九十一年高級口說測驗  
抽樣考生分析數據總表

口說級分	Collins 三至五級 所佔總使用 字彙百分比	Collins 一至二級 所佔總使用 字彙百分比	Average Type/Token	Average Ratio	各級分平均 AS-Unit 數量	各級分平 均字數
5	93%	7%	115/302	38	24	11.8
4	92%	8%	103.5/232.5	45	15	13.2
3	94%	6%	117.5/363	33	21	11.2
2	95%	5%	85.5/207.5	41	12	7.8



## 7. 試題建構效度（construct validity）分析

一套測驗的使用價值，取決於該測驗效度的適當性（郭生玉，1991，第 19 頁）。因此，可以檢測出考生真實語言能力的語言測驗，才是有效的測驗。

### 7.1 目的

藉由比對本測驗之預設語言功能及考生測驗時之表現，檢驗本測驗內容的建構效度，亦即檢驗本測驗是否確實可引導考生使用符合高級測驗口說能力指標的語言功能。依據本級測驗之能力指標，（亦即「對於各類主題皆能流暢地表達看法、參與討論，能在一般會議或專業研討會中報告或發表意見等」），設定本測驗所要求考生達到的語言溝通功能（O'Sullivan, Weir, and Saville, 2002）如下：

- Providing personal information (Present, Past)
- Expressing opinions
- Elaborating
- Justifying opinions
- Staging
- Describing (sequence of events, scene, objects, people)
- Summarizing
- Expressing preference
- Agreeing/Disagreeing
- Reporting what people say
- Reciprocating
- Deciding/Bringing to a close

### 7.2 對象

本測驗之預設目標為檢測考生較高階之語言功能，因此可通過本測驗之考生應在其語言表現中顯現此一現象。也因此自七名分析之對象考生中，選出三名口說測驗通過考生，分別為 3、4、5 級分考生，觀察其表現，藉以了解本測驗能否引導考生使用預設之語言功能。

### 7.3 方法

由三位測驗專業人員（兩位以英語為母語人士及一位本國籍人士），一齊觀看考生錄影帶，並將考生各部分表現之語言功能標示於功能表（採用 O'sullivan et al. 於 2002 年發表 "Using observation checklists to validate speaking-test task" 中所使用者）。觀看時若有異議，則討論、交換意見後達成協議。

### 7.4 結果與分析

三位專業人員所觀察到的語言功能請參見表二十四。由表二十四可看出，除了少數幾項功能外（如陰影所示），考生使用了表中所列大部分的語言功能，與本級測驗所預期達到之各語言功能相符，足見本測驗之建構效度確已達到，而本測驗也確實可引導屬於高級程度之考生使用符合該程度的語言功能。



表二十四 九十一年高級口說測驗  
抽樣考生使用之語言功能

Informational Functions		Part I	Part II	Part III
Providing personal information	Present	✓	✓	
	Past	✓	✓	
	Future			
Expressing opinions		✓	✓	✓
Elaborating		✓	✓	✓
Justifying opinions			✓	✓
Comparing		✓	✓	✓
Speculating		✓	✓	✓
Staging			✓	✓
Describing	Sequence of events	✓		
	Objects	✓	✓	
	People	✓		
Summarizing				✓
Suggesting			✓	✓
Expressing preferences		✓	✓	
Interactional Functions				
Agreeing			✓	
Disagreeing				
Modifying				
Asking for opinions				
Persuading				
Asking for information				
Making and responding to requests			✓	
Reporting what people say			✓	✓
Conversational repair				
Negotiating meaning	Checking understanding		✓	
	Indicate understanding		✓	
	Establish common ground			
	Ask for clarification		✓	
	Correct utterance			
	Respond to request for clarification		✓	
Managing Interaction				
Reciprocating			✓	
Deciding/Bringing to a close			✓	✓

Adapted from O'Sullivan et al., 2002

## 8. 問卷調查

194 名應考考生中，測驗後，共計有 191 名考生填答問卷，佔出席考生之 98%。問卷內容及詳細的統計結果，請參閱附錄五。

### 8.1 結果

以下依問卷內容之類別，摘要問卷結果。

#### 8.1.1 測驗方式

考生普遍表示喜歡面談式的口說測驗及試場的安排，並表示不會受到試場內錄音、錄影設備及同一組考生表現之影響。另外，91%的考生認為主考官的態度友善。



### 8.1.2 測驗各部分的時間

近六成考生認為第二部分（Part II, Information Exchange）的作答時間（兩分鐘）充裕，近八成考生認為第三部分（Part III, Presentation）的作答時間（兩分鐘）充裕。

### 8.1.3 測驗表現

整體而言，考生普遍認為測驗題目容易發揮；就各部分之難易度而言，則第一部分較第二部分容易，第三部分最難。而六成六的考生認為，在測驗時的表現可以反映出其實際英語口說能力。

## 9. 結論與建議

### 9.1 將本分析結果分類摘要如下：

#### 9.1.1 考生表現

- 字彙

語言程度較好的考生，所使用的字彙較多，且所使用低頻率（low-frequency words）的字亦較多。

- 句構

語言程度較好的考生，使用較多的子句，且句子較複雜。

#### 9.1.2 測驗內容

- 建構效度

絕大部分的語言功能與測驗設計相符，詳見二、口說測驗之 7. 建構效度。

- 表面效度

問卷結果顯示，大部分考生對本測驗抱持肯定態度，且能認同施測之方式，足證本測驗具表面效度（face validity）。

雖然考生普遍認為第三部分的題目難度最高，但高級口說測驗的分項能力標準為「對於各類主題皆能流暢地表答看法、參與討論，能在一般會議或專業研討會中報告或發表意見等」，因此第三部分確實有必要存在且具有足以鑑別高級程度考生語言能力之功能。

### 9.2 由本分析發現可看出，各級分考生在字彙之質與量方面的表現，或所使用句構的複雜程度，確有所不同，可說明本測驗所使用的評分量表應屬適切。惟本次並未針對正確性（accuracy）做分析，且所採樣本有限，日後仍有賴更多資料的收集與驗證。



附錄一：寫作測驗整體式評分量表

1	2	3 (Pass)	4	5
<ul style="list-style-type: none"> <li>◆ Parts of the task are not addressed or are unclear. The content is inadequate or lacks relevance.</li> <li>◆ The text lacks logical organization and coherence which may lead to confusion.</li> <li>◆ The range of vocabulary is inadequate to complete the tasks and/or lexical items are often used inappropriately.</li> <li>◆ Structures used are often inaccurate and/or inappropriate.</li> </ul>		<ul style="list-style-type: none"> <li>◆ All parts of the task are clearly addressed. The content is relevant and adequate.</li> <li>◆ The text is coherent and logically organized in general.</li> <li>◆ The range of vocabulary is adequate to complete the tasks and lexical items are used appropriately most of the time.</li> <li>◆ Structures used are accurate and appropriate for the most part.</li> </ul>		<ul style="list-style-type: none"> <li>◆ All parts of the task are effectively addressed. The entire content is relevant and more than adequate.</li> <li>◆ The text is coherent and logically organized throughout.</li> <li>◆ A wide range of vocabulary is used to complete the tasks effectively.</li> <li>◆ A wide range of structures is used to complete the tasks accurately, appropriately and effectively.</li> </ul>

\*plagiarism: more than three consecutive words are copied from the Task 1 input without appropriate quotation

***Non-ratable compositions***

A composition is considered "non-ratable" when

- ◆ the length is shorter than 80 words, or
- ◆ the content is totally irrelevant, or
- ◆ the content has apparently been copied or memorized from other known materials, or
- ◆ the handwriting is completely unintelligible.



## 附錄二：寫作測驗考生意見調查問卷結果

有效問卷 191 份

### 第一大題

- 1、 我認為 Task 1 的題目說明
  - A、 非常清楚 (48%)
  - B、 清楚 (48%)
  - C、 不清楚 ( 4%)
  - D、 非常不清楚 ( 1%)
- 2、 我認為 Task 1 文章的難易度，整體而言
  - A、 非常容易 ( 7%)
  - B、 容易 (32%)
  - C、 適中 (58%)
  - D、 非常難 ( 3%)
- 3、 我認為 Task 1 的兩篇文章，在難易度上
  - A、 非常相當 ( 8%)
  - B、 相當 (79%)
  - C、 不相當 (13%)
  - D、 非常不相當 ( 0%)
- 4、 我認為 Task 1 的寫作題目
  - A、 非常容易發揮 ( 5%)
  - B、 容易發揮 (57%)
  - C、 稍難發揮 (38%)
  - D、 非常難發揮 ( 1%)
- 5、 我認為 Task 1 的作答時間
  - A、 稍長 ( 4%)
  - B、 適中 (54%)
  - C、 稍短 (37%)
  - D、 太短 ( 6%)
- 6、 我認為 Task 1 可否測出我的英語寫作能力？
  - A、 可以 (29%)
  - B、 大致可以 (48%)
  - C、 勉強可以 (20%)
  - D、 無法測出 ( 3%)



## 第二大題

7、 我認為 Task 2 的題目說明

- A、 非常清楚 (47%)
- B、 清楚 (48%)
- C、 不清楚 ( 5%)
- D、 非常不清楚 ( 1%)

8、 我在解讀 Task 2 圖表中的訊息時

- A、 毫無困難 (42%)
- B、 無困難 (49%)
- C、 稍有困難 ( 9%)
- D、 非常困難 ( 1%)

9、 我認為 Task 2 的寫作題目

- A、 非常容易發揮 (20%)
- B、 容易發揮 (60%)
- C、 稍難發揮 (17%)
- D、 非常難發揮 ( 2%)

10、 我認為 Task 2 的作答時間

- A、 稍長 ( 6%)
- B、 適中 (70%)
- C、 稍短 (21%)
- D、 太短 ( 3%)

11、 我認為 Task 2 可否測出我的英語寫作能力?

- A、 可以 (23%)
- B、 大致可以 (53%)
- C、 勉強可以 (20%)
- D、 無法測出 ( 4%)
- 未填 ( 1%)



附錄三：口說測驗整體式評分量表

1	2	3 (Pass)	4	5
<ul style="list-style-type: none"> <li>◆ Utterances are frequently unintelligible and incomprehensible.</li> <li>◆ Contributions are frequently irrelevant, and inadequate.</li> <li>◆ The range, accuracy and appropriateness of language used is inadequate.</li> <li>◆ The candidate doesn't speak with ease.</li> <li>◆ The candidate doesn't speak coherently in extended turns.</li> </ul>		<ul style="list-style-type: none"> <li>◆ Utterances are intelligible and comprehensible.</li> <li>◆ Contributions are relevant, and adequate.</li> <li>◆ The range, accuracy and appropriateness of language used is adequate.</li> <li>◆ The candidate generally speaks with ease.</li> <li>◆ The candidate speaks coherently in extended turns.</li> </ul>		<ul style="list-style-type: none"> <li>◆ The candidate has no problems with either sounds or utterances.</li> <li>◆ Contributions are both relevant and more than adequate.</li> <li>◆ The range, accuracy and appropriateness of language used is more than adequate.</li> <li>◆ The candidate speaks fluently with minimal hesitations.</li> <li>◆ Speech is very well-organized in extended turns.</li> </ul>



附錄四：口說測驗分項式評分量表

Criteria \ Score	1	2	3 (Pass)	4	5
<b>Pronunciation</b> <ul style="list-style-type: none"> <li>•stress</li> <li>•rhythm</li> <li>•intonation</li> <li>•individual sounds</li> </ul>	<ul style="list-style-type: none"> <li>♦Inaccuracies in individual sounds sometimes prevent communication.</li> <li>♦Inappropriate use of stress, rhythm, and intonation sometimes prevents communication.</li> </ul>		<ul style="list-style-type: none"> <li>♦Comprehensible utterances with easily recognizable sounds, appropriate stress, rhythm, and intonation.</li> <li>♦Inaccuracies and inappropriateness may sometimes occur, but these never prevent communication.</li> </ul>		<ul style="list-style-type: none"> <li>♦The candidate is able to produce entirely accurate utterances and sounds with appropriate stress, and natural rhythm, and intonation.</li> </ul>
<b>Relevance and Adequacy</b>	<ul style="list-style-type: none"> <li>♦The discourse lacks relevance to the task, and the contribution is inadequate.</li> </ul>		<ul style="list-style-type: none"> <li>♦The discourse is almost entirely relevant to the task, and the contribution is adequate to complete the task.</li> </ul>		<ul style="list-style-type: none"> <li>♦The discourse is relevant to the task, and the contribution is more than adequate to complete the task.</li> </ul>
<b>Lexical Use</b> <ul style="list-style-type: none"> <li>•range</li> <li>•appropriateness</li> </ul>	<ul style="list-style-type: none"> <li>♦The range of vocabulary is inadequate to complete the tasks.</li> <li>♦Vocabulary is frequently used inappropriately.</li> </ul>		<ul style="list-style-type: none"> <li>♦The range of vocabulary is adequate to complete the tasks.</li> <li>♦There may be some inappropriate use of vocabulary.</li> </ul>		<ul style="list-style-type: none"> <li>♦The candidate is able to use a wide range of vocabulary to effectively complete the tasks. Lexis is used appropriately. Errors are rare.</li> </ul>
<b>Grammatical Use</b> <ul style="list-style-type: none"> <li>•range</li> <li>•accuracy</li> </ul>	<ul style="list-style-type: none"> <li>♦Too limited a range of structures for task completion.</li> <li>♦Utterances are frequently used inaccurately.</li> </ul>		<ul style="list-style-type: none"> <li>♦An adequate range of structures to complete the tasks.</li> <li>♦There may be some inaccurate structures.</li> </ul>		<ul style="list-style-type: none"> <li>♦The candidate is able to use a wide range of structures to complete each task accurately, appropriately and effectively. Errors are rare</li> </ul>
<b>Fluency</b>	<ul style="list-style-type: none"> <li>♦Unnecessary hesitation interferes with communication and puts stress on the listener.</li> </ul>		<ul style="list-style-type: none"> <li>♦The candidate keeps communication flowing; however, hesitations which do not affect communication may sometimes occur.</li> </ul>		<ul style="list-style-type: none"> <li>♦The candidate is able to keep communication flowing smoothly with minimal hesitation.</li> </ul>
<b>Coherence</b>	<ul style="list-style-type: none"> <li>♦The contribution lacks logical organization.</li> </ul>		<ul style="list-style-type: none"> <li>♦The contribution is logically organized in general.</li> </ul>		<ul style="list-style-type: none"> <li>♦The contribution is logically organized throughout the task.</li> </ul>

Passing score: 3 or above in every category.



## 附錄五：口說測驗考生意見調查問卷結果

91/08/24 & 25 全民英語能力分級檢定測驗高級測驗  
口說能力測驗考生意見調查問卷結果(N=191)

### 測驗方式

1、 今天的口說測驗方式（面談），我	口說通過%	口說未通過%
A、 很喜歡 (15%)	15%	14%
B、 喜歡 (61%)	56%	64%
C、 不喜歡 (22%)	27%	19%
D、 很不喜歡(3%)	1%	3%

2、 今天口說測驗的試場安排（例如桌椅、燈光等），我	口說通過%	口說未通過%
A、 很喜歡 (19%)	23%	17%
B、 喜歡 (76%)	70%	79%
C、 不喜歡 (5%)	7%	3%
D、 很不喜歡(1%)	0%	1%

3、 今天口說測驗的錄音、錄影設備是否影響我，使我表現變差	口說通過%	口說未通過%
A、 是(有影響) (21%)	30%	16%
B、 否(沒有影響)(79%)	70%	84%

4、 我認為和我同一組考生的表現會影響我，使我表現變差	口說通過%	口說未通過%
A、 會(有影響) (24%)	21%	26%
B、 不會(沒有影響)(76%)	79%	74%

5、 我認為主考官的態度	口說通過%	口說未通過%
A、 很友善(35%)	35%	36%
B、 友善(56%)	52%	58%
C、 不友善(9%)	13%	7%
D、 很不友善(0%)	0	0

### 測驗各部分的時間

6、 我認為第二部分（Part II, Information Exchange）的作答時間	口說通過%	口說未通過%
A、 很充裕(14%)	10%	16%
B、 還算充裕(45%)	46%	45%
C、 稍嫌不足(34%)	39%	31%
D、 不足(7%)	4%	9%

7、 我認為第三部分（Part III, Presentation）的作答時間	口說通過%	口說未通過%
A、 很充裕(26%)	31%	22%
B、 還算充裕(53%)	46%	57%
C、 稍嫌不足(17%)	18%	17%
D、 不足(4%)	4%	4%



### 測驗表現

8、	整體而言，我認為今天口說測驗的題目	口說通過%	口說未通過%
A、	很容易發揮(8%)	13%	5%
B、	還算容易發揮 (66%)	57%	72%
C、	不易發揮 (22%)	26%	20%
D、	很難發揮(4%)	4%	3%
9、	我認為今天口說測驗的第一部分 ( Part I, Warm-Up Introduction )	口說通過%	口說未通過%
A、	很容易(25%)	30%	21%
B、	還算容易 (68%)	61%	71%
C、	稍嫌困難(7%)	9%	6%
D、	很困難(1%)	0%	2%
10、	我認為今天口說測驗的第二部分 ( Part II, Information Exchange )	口說通過%	口說未通過%
A、	很容易(24%)	30%	20%
B、	還算容易 (65%)	61%	67%
C、	稍嫌困難(11%)	9%	12%
D、	很困難(1%)	0%	1%
11、	我認為今天口說測驗的第三部分 ( Part III, Presentation )	口說通過%	口說未通過%
A、	很容易(9%)	16%	5%
B、	還算容易 (57%)	63%	54%
C、	稍嫌困難(32%)	21%	39%
D、	很困難(2%)	0%	2%
12、	我認為今天的口說測驗能否反映我的實際英語口說能力	口說通過%	口說未通過%
A、	絕對可以 (12%)	14%	10%
B、	還算可以 (54%)	51%	56%
C、	勉強可以(27%)	24%	28%
D、	不可以(7%)	10%	6%
13、	我今天複試測驗的作答內容可提供語言訓練測驗中心作日後研究使用	口說通過%	口說未通過%
A、	同意 (77%)	83%	74%
B、	不同意(21%)	17%	23%
	未填(2%)	0%	3%



## 參考書目

- 語言訓練測驗中心. 2002. 《全民英語能力分級檢定測驗高級測驗研究報告》台北: 語言訓練測驗中心.
- Alderson, J. C. et al. 1995. *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.
- Bachman, L. & Palmer, A. 1981. 'The Construct Validation of the FSI Oral Interview' *Language Learning* 31:67-86
- Bachman, L. F. 1990. *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Backman, L. F. and Palmer, A. S. 1996. *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. 1999. *Longman Grammar of Spoken and Written English*. England: Pearson Education Limited.
- Brown, J. 1996. *Testing in Language Programs*. New Jersey: Prentice Hall Regents
- Chalhoub-Deville, M. 1996. 'Performance assessment and the components of the oral construct across different tests and rater groups' *Studies in Language Testing* 3. Cambridge: Cambridge University Press
- Chapelle, C. et al. 1997. *Communicative Language Proficiency: Definition and Implications for TOEFL 2000*. (TOEFL Monograph Series, MS-10) Princeton, NJ: ETS.
- Charge, N. & Taylor, L. 1997. 'Recent developments in IELTS' *ELT Journal* 51 (4): 374-380
- Chiang, G. S. & Dunkel, P. 1992. 'The Effect of Speech Modification, Prior Knowledge, and Listening Proficiency on EFL Lecture Hearing.' *TESOL Quarterly*. 26: 345-374.
- Collin's COBUILD. 2001. *English Dictionary for Advanced Learners, Major New Edition*. Glasgow: HarperCollins Publishers.
- Elley, Wm B., I.H. Barham, H. Lamb, & M. Wyllie. 1979. *The Role of Grammar in a Secondary School Curriculum*. NCTE: Wellington, New Zealand: Council for Educational Research, 1978.
- Foster, P., Tonkyn, A. & Wigglesworth, G. 2000. 'Measuring Spoken Language: A Unit for All Reasons' *Applied Linguistics* 21 (3): 354-375
- Fulcher, G. 1987. 'Tests of oral performance: the need for data-based criteria' *ELT Journal* 41(4): 287-291
- Fulcher, G. 1996. 'Testing tasks: issues in task design and the group oral' *Language Testing* 13 (1): 22-51
- Halliday, M. A. K. & Hasan, R. 1976. *Cohesion in English*. London: Longman.
- Hamp-Lyons, L. (ed.). 1991. *Assessing Second Language Writing in Academic Contexts*. Norwood, New Jersey: Ablex Publishing Corporation.
- Henning, G. 198. *A guide to language testing*. Cambridge, Mass.: Newbury House.
- Homburg T., J. 1984. Holistic Evaluation of ESL Compositions: Can It Be Validated Objectively? *TESOL Quarterly* 18 (1): 87~107
- Huang, R. J. 1999. Using Different Measurements to Evaluate Vocabulary Size of Chinese Written Language. *Journal of Education and Psychology* 22 (1): 113~144. Taipei, Taiwan: National Chengchi University Publisher. (Article written in Chinese)
- Hughes, A. 1989. *Testing for language teachers*. Cambridge: Cambridge University Press.



- Hunt K.** 1965. Grammatical Structures Written at Three Grade Levels. *Research Report No. 3*. Urbana, Ill.:NCTE, 1965. First published in 1964 as Differences in grammatical structures written at three grade levels, the structures to be analyzed by transformational methods. Tallahassee: Florida State University, Project 1998, Cooperative Research Program, Office of Education, U.S. Department of Health, Education and Welfare.
- Juan, E. U. and Silveira, Beaudin, J. C. P.** 1998. 'A Product-Focused Approach to Text Summarisation.' *The Internet TESL Journal* 4:1 [available on-line]  
<http://www.aitech.ac.jp/~iteslj/Articles/Juan-TextSummary.html>
- Knight, B.** 1992. 'Assessing speaking skills: a workshop for teacher development' *ELT Journal* 46 (3): 294-302
- Loban, W.** 1976. Language Development: Kindergarten through Grade Twelve. Urbana, Ill.: NCTE, 1976.
- Lee, J & McChesney.** 2000. 'Discourse rating tasks: a teaching tool for developing sociocultural competence' *ELT Journal* 54(2): 161-168
- Lumley, T. & McNamara, T.F.** 1995. 'Rater characteristics and rater bias: implications for training' *Language Testing* 12 (1): 54-71
- Matthews, M.** 1990. 'The measurement of productive skills: doubts concerning the assessment criteria of certain public examinations' *ELT Journal* 44(2): 117-121
- McNamara, T. F.** 1996. *Measuring Second Language Performance*. New York: Addison Wesley Longman Limited.
- McNamara, T. F.** 1997. 'Interaction in second language performance assessment: Whose performance?' *Applied Linguistics* 18 (4):446-466
- Mellon, J. C.** 1969. Transformational Sentence Combining: A Method for Enhancing the Development of Syntactic Fluency in English Composition. *NCTE Research Report No. 10*. 1969.
- Nation, I.S.P.** 1990. *Teaching and Learning Vocabulary*. Boston, Massachusetts: Neale & Heinle Publishers.
- North, B.** 1995. 'Scales of Language Proficiency' *Melbourne Papers in Language Testing* 4 (2)
- North, B.** 1995. 'Standardisation of Continuous Assessment Grades' *Language Testing in the 1990s*. Macmillan Publishers Ltd.
- North, B. & Schneider, G.** 1998. 'Scaling descriptors for language proficiency scales' *Language Testing* 15(2):217-262
- Nunn, R.** 2000. 'Designing rating scales for small-group interaction' *ELT Journal* 54 (2): 169-178
- O'Donnell, R. C., Griffin, W.J., & Norris, R.C.** 1967. Syntax of Kindergarten and Elementary School Children: A Transformational Analysis. *Research Report No. 8*. Champaign, Ill.: NCTE, 1967.
- O'Hare, F.** 1971. Sentence Combining: Improving Student Writing Without Formal Grammar Instruction. *NCTE Research Report No. 15*, 1973.
- O'Sullivan, Barry, Weir, Cyril J. & Saville, Nick.** 2002. 'Using Observation Checklists to Validate Speaking-test Tasks' *Language Testing*. 19 (1): 33-56.
- Pollitt, A. & Murray, N.** 1996. 'What raters really pay attention to' *Studies in Language Testing* 3. Cambridge: Cambridge University Press
- Richards J., J. Platt and H.Weber.** 1985. *Longman Dictionary of Applied Linguistics*. Hong Kong: Longman
- Shohamy, E., Reves, T. & Bejarano, Y.** 1986. 'Introducing a new comprehensive test of oral proficiency' *ELT Journal* 40 (3): 212-220



- Sinclair, J., Fox, G., Bullon, S., & Manning, E.** 1995. *Collins Cobuild English Dictionary*. London: HarperCollins Publishers Ltd.
- Skehan, P. & Foster, P.** 1997. 'Task type and task processing conditions as influences on foreign language performance' *Language Testing Research* 1(3): 185-211.
- Swain, M.** 1985. 'Large-scale communicative language testing: A case study' in Y. P. Lee, A. C. Y. Fok, R. Loard, and G. Low (eds.): *New Directions in Language Testing*. Oxford: Pergamon Press: 36-46
- Upshur, J. A. & Turner, C. E.** 1999. 'Systematic effects in the rating of second-language speaking ability: test method and learner discourse' *Language Testing* 16 (1): 82-112.
- Urquhart, A. H. & Weir, C. J.** 1998. *Reading in a Second Language: Process, Product and Practice*. Addison Wesley Longman Higher Education.
- Vavra E.** 2002. Definitions of the "T-Unit". From Pennsylvania College of Technology homepage [http://nweb.pct.edu/homepage/staff/evavra/ED498/Essay009\\_Def\\_TUnit.htm](http://nweb.pct.edu/homepage/staff/evavra/ED498/Essay009_Def_TUnit.htm).
- Weigle S. C.** 2002. *Assessing Writing*. Cambridge, UK: Cambridge University Press
- Weir, C. J.** 1990. *Communicative Language Testing*. Prentice Hall.
- Weir, C. J.** 1993. *Understanding & Developing Language Tests*. Prentice Hall.
- Weir, C. J. et al.** 1997. *The Test for English Majors (TEM) Validation Study*. Shanghai Foreign Language Education Press.
- Wigglesworth, G.** 1997. 'An Investigation of planning time and proficiency level on oral test discourse' *Language Testing* 14 (1):85-106.
- Yorozuya, R. & Oller, J.** 1980. 'Oral Proficiency Scales: Construct Validity and the Halo Effect' *Language Learning* 30 (1): 135-153
- Young, R. & Milanovic, M.** 1992. 'Discourse variation in oral proficiency interviews' *Studies in Second Language Acquisition* 14 (4): 403-424