

「全民英檢」第一次優級測驗研究報告

一、前言

第一次「全民英語能力分級檢定測驗」(以下簡稱「全民英檢」)優級測驗已於九十三年二月七日在台北考區舉辦。本測驗為「全民英檢」五級測驗最高的一級,旨在檢測考生是否具有「接近受過高等教育之母語人士」的英語能力,其聽、說、讀、寫之分項能力指標如下:

聽:能聽懂各類主題及體裁的內容,理解程度與受過高等教育之母語人士相當。

說:能在各種不同場合以正確流利之英語表達看法;能適切引用文化知識及慣用語詞。

讀:能閱讀各類不同主題、體裁文章。閱讀速度及理解程度與受過高等教育之母語人士相當。

寫:能撰寫不同性質的文章,如企劃報告、專業/學術性摘要、論文、新聞報導及時事評論等。對於各類主題均能有效完整地闡述並作深入探討。

在題型設計方面,優級測驗為一整合聽、讀、說、寫四項能力的測驗,分兩節進行(表一)。第一節為整合式寫作測驗,考生須先看一段錄影節目並閱讀相關主題的文章,然後依據題目指示寫一篇 750 字左右的文章。第二節為整合式口說能力測驗,採面談方式進行,考生須依據前一節所寫的文章,作口頭報告及即席應答。兩項測驗各評分項目均達通過標準,始為通過。

表一 優級測驗題型

測驗項目	內容	長度	時間
第一節: 整合式寫作測驗	第一階段:聽力-錄影節目	10~15 分鐘	共 3 小時
	第二階段:閱讀-期刊文章	約 3,000 字	
	第三階段:主題寫作	約 750 字	
第二節: 整合式口說測驗	第一階段:準備報告要點	30 分鐘	約 50 分鐘
	第二階段:口頭報告	10 分鐘	
	第三階段:問答	約 10 分鐘	

第一次優級測驗應試考生共 43 人¹,平均年齡為 27 歲,其中男性 20 名,女性 23 名。依據九十三年三月四日揭曉的成績,43 位考生中有 2 位寫作及口說能力測驗均達通過標準,取得優級測驗合格證書,通過率為 5%。就單項測驗而言,3 位考生寫作能力測驗達通過標準,通過率為 7%;7 位考生口說能力測驗達通過標準,通過率為 16%。

本研究將依據第一次正式測驗結果檢視優級測驗之效度,分析內容包含:成績統計資料、考生作答語料,及試後訪談考生、寫作評分委員與口試主考官所蒐集的意見。分析結果就第一節測驗(寫作測驗)及第二節測驗(口說測驗)分別詳述如下。

¹ 本次測驗以試辦方式推出,同時為回饋各界對「全民英檢」的支持,採不收費方式辦理,考生須具下列任一種資格:1)「全民英檢」高級複試通過;2) TOEFL-CBT 達 280 分;3) IELTS 達 8.5 級分。

二、第一節測驗（整合式寫作測驗）

1. 測驗目標

評量考生是否能依據題目所設定之情境及要求，有效地運用聽、讀部分所提供的資訊，加以組織，並與個人意見整合，以書面方式深入、完整地探討一主題。

2. 測驗方式

測驗時間全長約三小時²，分三階段進行，考生可選擇以紙筆或電腦介面作答³：

第一階段：考生先看一段錄影節目，長度為 10~15 分鐘，節目只播出一遍，考生須在聆聽時作筆記。

第二階段：閱讀與錄影節目主題相關的文章，文章總長 3,000 字左右，考生可選擇從電腦螢幕上或試題本上讀取文章內容。

第三階段：依題目指示寫一篇約 750 字的作文，考生可選擇以紙筆或電腦介面作答。使用電腦作答者，須於考前參加電腦講習以熟悉操作。作答時，考生可使用電腦軟體中所提供的文字編輯功能，如剪貼、複製及數字計算等。電腦每兩分鐘則會自動存檔並備份。

寫作之前的聽力與閱讀活動旨在協助考生深入了解測驗主題內容及提供考生寫作時所需之背景資訊。錄影節目及閱讀文章皆擷取自電子及平面媒體之真實題材 (authentic materials)，俾以更精確地推估考生實際的語言運用能力。

3. 評分標準

優級寫作採用分項式評分量表（請參見附錄一），評分項目包括內容適切性 (Relevance and Adequacy)、組織與連貫性 (Organization and Coherence)、遣詞用字 (Lexical Use)、文法結構 (Grammatical Use)、寫作效果 (Effect on the Target Reader)、作者角色 (Writer's Role) 六項。每個分項皆針對該項的通過標準 (passing criteria) 加以敘述，而不細分成數級。考生需在每個分項表現均達通過標準，才視為通過。

4. 評分方式

本次共有四位評分委員參與閱卷。為求評分標準的一致與穩定，評分委員在正式閱卷前，須先熟悉測驗內容及評分量表。每份答卷分別由二位評分委員評閱，如兩閱結果不一致時，交由第三位評分委員複閱，最後成績則綜合三道評閱結果採多數決。

² 寫作測驗進行中考生可自由選擇休息至多 15 分鐘，但休息時間不另行補足。

³ 提供電腦作答選擇的原因，係考量本級考生可能經常需要大量書寫，習於使用文書軟體，若能提供一較符合其寫作習慣之測驗環境，將有助於考生表現真實的能力。本次測驗使用的文書軟體為 EmEditor。

5. 考生成績統計分析

5.1 所有考生之總通過率

參加本次優級測驗之 43 名考生中，3 名通過寫作測驗，通過率為 7%。其中 2 名為男性，1 名為女性，依性別分佈之通過率請參見表二。

表二 九十三年優級寫作測驗
成績分布情形 (依性別分佈之通過率)

類別		人數	寫作 通過	寫作 通過率	優級 通過	優級 通過率
性別	男性	20	2	10%	2	10%
	女性	23	1	4%	0	0
總計		43	3	7%	2	5%

5.2 通過考生背景

本次通過優級寫作測驗之三名考生集中在 21 至 29 歲年齡層，其中兩名為英語主修（含國內研究所畢業一名、國內大學在學生一名），另一名為非英語主修（國內大學畢業）。

5.3 分項平均通過率

表三為全體考生各分項通過人數及通過率，表四及表五則分別為依學歷及年齡層分析之各分項通過人數及通過率。以所有考生之分項成績而言，通過率最高的項目為「作者角色」，其次為「組織與連貫性」，而通過率最低的項目為「遣詞用字」。而就各類學歷考生表現分析，較值得注意的是國內大學英語主修學生在「組織與連貫性」及「作者角色」兩項的表現最為突出（詳請參見表四）。此外，以年齡層而言，25 至 29 歲的考生整體表現較佳（詳請參見表五）。

表三 九十三年優級寫作測驗
分項平均通過率

評分 項目 考生 總人數	內容適切性		組織與連貫性		遣詞用字		文法結構		寫作效果		作者角色		六項全過	
	通 過 人 數	通 過 率	通 過 人 數	通 過 率	通 過 人 數	通 過 率	通 過 人 數	通 過 率	通 過 人 數	通 過 率	通 過 人 數	通 過 率	通 過 人 數	通 過 率
43	11	26%	17	40%	7	16%	12	28%	10	23%	24	56%	3	7%

表四 九十三年優級寫作測驗
依學歷分佈之分項通過人數及通過率

評分項目 學歷	人數	內容適切性		組織與連貫性		遣詞用字		文法結構		寫作效果		作者角色		六項全過	
		通過人數	通過率	通過人數	通過率	通過人數	通過率	通過人數	通過率	通過人數	通過率	通過人數	通過率	通過人數	通過率
大學以下	6	2	33%	2	33%	1	17%	2	33%	1	17%	1	17%	0	0%
國內大學 -英語主修	9	3	33%	7	78%	1	11%	3	33%	3	33%	7	78%	1	11%
-非英語主修	13	1	8%	3	23%	2	15%	3	23%	2	15%	7	54%	1	8%
國內研究所	2	1	50%	1	50%	1	50%	1	50%	1	50%	1	50%	1	50%
國外大學	6	2	33%	2	33%	1	17%	1	17%	2	33%	4	67%	0	0%
國外研究所	6	2	33%	2	33%	1	17%	2	33%	1	17%	4	67%	0	0%

表五 九十三年優級寫作測驗
依年齡層分佈之分項通過人數及通過率

評分項目 年齡	人數	內容適切性		組織與連貫性		遣詞用字		文法結構		寫作效果		作者角色		六項全過	
		通過人數	通過率	通過人數	通過率	通過人數	通過率	通過人數	通過率	通過人數	通過率	通過人數	通過率	通過人數	通過率
19歲以下	6	2	33%	2	33%	1	17%	2	33%	1	17%	1	17%	0	0%
20-24歲	17	3	18%	7	41%	2	11%	5	29%	4	24%	9	53%	1	6%
25-29歲	8	2	25%	5	63%	2	25%	2	25%	2	25%	6	75%	2	25%
30-34歲	3	1	33%	1	33%	1	33%	1	33%	1	33%	2	67%	0	0%
35-39歲	4	0	0%	1	25%	1	25%	0	0%	1	25%	3	75%	0	0%
40-44歲	3	1	33%	0	0%	0	0%	1	33%	0	0%	1	33%	0	0%
45歲以上	2	2	100%	1	50%	0	0%	1	50%	1	50%	2	100%	0	0%

6. 考生作答表現分析

6.1 研究目的

依據「全民英檢」優級寫作測驗之設計，評分者須就考生作文在前述六個分項的表現進行評量，各評分項目均達通過標準，始達寫作測驗通過標準。本部分擬藉由比較通過與未通過考生在六個分項的作答表現，檢視其間是否有質或量的差異，並同時驗證通過與否分界點（cutting point）的適切性。

6.2 抽樣對象

由於通過寫作測驗的考生僅有三名（簡稱 P 組），為與未通過考生做對等的比較，故儘可能挑選在年齡、性別、與教育背景上相當的考生，共選出六名未通過考生做為研究對象。此六名考生再依其通過項目分為兩組：其中三名考生六個評分項目通過四項，距離通過標準較接近，歸為 F+ 組；另外三名考生六個評分項目通過兩項(含)以下，距離通過標準較遠，歸為 F- 組。九名抽樣考生皆為「全民英檢」高級複試通過，其相關資料摘要如表六：

表六 九十三年優級寫作測驗
抽樣考生基本資料

考生代碼	最後成績	性別	學歷	主修	平均年齡
A	通過 (P)	男	國內研究所	英語主修	25
B	通過 (P)	男	國內大學	非英語主修	
C	通過 (P)	女	國內大學	英語主修	
D	未通過 (F+)	男	國外大學	非英語主修	25
E	未通過 (F+)	男	國內大學	英語主修	
F	未通過 (F+)	女	國內大學	英語主修	
G	未通過 (F-)	女	國外研究所	非英語主修	29
H	未通過 (F-)	男	國內大學	英語主修	
I	未通過 (F-)	女	國內研究所	英語主修	

6.3 研究方法

本研究採取「定量」(quantitative)與「定性」(qualitative)研究方法兩者並用之設計。其中有關「遣詞用字」及「文法結構」兩項係採用「定量」研究方法，以檢測考生書面表達的字彙量 (range)、文法結構的複雜度 (complexity) 及正確性 (accuracy); 而有關「內容適切性」、「組織與連貫性」、「寫作效果」及「作者角色」四項則採「定性」研究方法，分析評分委員針對每份試卷所寫的評語 (assessor's comments)。主要研究項目及方法謹摘要如表七，詳細分析方法說明請參見 6.3.1 至 6.3.3。

表七 九十三年優級寫作測驗
研究項目及方法摘要

研究項目	研究方法
字彙量 (range)	<ul style="list-style-type: none"> ▪ 字數計算 ▪ 相異詞比例計算 (type-token ratio , 簡稱 TTR) ▪ 高字頻/中字頻/低字頻比例計算 (參考字表 : Collins COBUILD Word List)
文法句構的複雜度 (complexity)	♦ T-unit 總數與長度
文法句構的正確性 (accuracy)	♦ Error-free T-unit 總數與百分比
內容適切性、組織與連貫性、寫作效果、作者角色	♦ 分析評分委員針對每份試卷所寫的評語 (assessor's comments)

6.3.1 字彙量分析

本研究採用兩種方式分析字彙量，其一為比較每篇文章的相異詞比例 (type-token ratio)⁴，以了解通過與不通過考生使用的字彙豐富程度是否明顯不同。相異詞比例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。理論上，程度愈佳者，詞彙應相對較豐富，重複字相對較少，因此相異詞比例愈高；而字彙能力有限者，使用重複字表達意見的情形較明顯，變化較少，因此相異詞比例較低 (Huang 1999; Richards *et al.* 1992)。

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 Collins COBUILD 字表間的差異。Collins COBUILD 將搜集於 The Bank of English 的字彙，依出現頻率分為五級 (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%。根據 Collins COBUILD 統計 (Sinclair *et al.* 1995)，五級約有 680 字，四級約有 1,040 字，四、五級囊括了 75% 的英語使用字彙 (common grammar words and very frequent vocabulary items)。三級約有 1,580 字，此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、二級為高級字彙，其使用頻率較前三級為低，但仍非常重要。

依據「全民英檢」優級寫作分項能力說明：通過此級寫作測驗者，應具有足夠的字彙，針對於各類主題作有效完整地闡述及深入探討。因此通過優級寫作測驗的考生，Collins COBUILD 一、二級的高級字彙量應較未通過者多。

此外，在考生拼字錯誤處理方面，本研究參考 Kennedy & Thorp (2002) 研究 IELTS 寫作的做法，若拼字錯誤不影響文意解讀，則在分析前加以更正，若拼字錯誤嚴重，影響文意解讀，則不列入分析。

⁴ 相異詞比例的定義如下 (Biber *et al.* 1999: 52-53)：

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

6.3.2 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 (syntactic complexity of sentences)。本分析所採用的 T-unit 定義，係依據 Richards *et al.* 所編 *Dictionary of Language Teaching & Applied Linguistics* (Richards *et al.* 1992)：

"...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. For example, the sentence *After she had eaten, Kim went to bed* would be described as containing one T-unit....Compound sentences contain two or more T-units."

根據文獻記載 (Weigle 2002; Homburg 1984; Hunt 1983)，T-unit 長度及 error-free T-unit 的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.

6.3.3 定性分析

本研究「定性分析」部分採取「立即回想」(immediate retrospection) 的方式，請評分委員在評完每份試卷之後，針對考生在六個評分項目的表現寫下評語。其目的在了解寫作評分委員如何詮釋評分量表，以決定考生的作文是否達到通過標準，特別是在「內容適切性」、「組織與連貫性」、「寫作效果」及「作者角色」四項。為求得較完整的分析結果，除調閱九位抽樣考生作文並與評分委員評語比對外，亦參考評分委員對其他未抽樣試卷的評語。

6.4 結果與分析

6.4.1 相異詞比例統計分析結果（表八）

九位抽樣考生中，除考生 D 與 I 外⁵，其餘考生作文字數皆達 750 字長度要求，每篇作文平均長度為 789 字 (tokens)，平均相異字數 (types) 為 324 字，平均相異詞比例 (type-token ratio) 為 40.8。以三組考生詞彙運用變化 (variety) 差異而言，通過優級寫作測驗的三位考生 (P 組) 字彙使用似乎較有變化，相異詞比例平均值為 44.8，較 F+ 組之 42.22 稍高，較 F- 組之 35.3 則高出許多。

表八 九十三年優級寫作測驗
抽樣考生作文相異詞比例分析

分組	Types 平均值	Tokens 平均值	相異詞 比例平均值
通過 (P)	356	799	44.8
未通過 (F+)	318	756	42.2
未通過 (F-)	299	812	35.3
總平均	324	789	40.8

6.4.2 Collins COBUILD 字表比對結果（表九）

九位抽樣考生作文所使用的字彙，約有 94.5% 至 97% 落在 Collins COBUILD 的五級字表。以整體平均值而言，常用字彙 (三、四、五級) 佔 79.4%，高級字彙 (一、二級) 佔 16.3%⁶。而以各組考生而言，通過的考生 (P 組) 使用的高級字彙佔其用字的 18%，雖較 F- 組之 12.5% 高出許多，但較 F+ 組之 18.5% 稍低，此結果與原假設略有差距。

表九 九十三年優級寫作測驗
抽樣考生字彙量與 Collins COBUILD 字表比對結果

分組	Collins COBUILD 字表					Bands 1-5
	Band 5	Band 4	Band 3	Band 2	Band 1	
通過 (P)	46.7%	18.3%	12.7%	11%	7%	95.7%
	77.7%			18%		
未通過 (F+)	46.5%	17.5%	12%	12%	6.5%	94.5%
	76%			18.5%		
未通過 (F-)	47.5%	22%	15%	8.5%	4%	97%
	84.5%			12.5%		
總平均	46.9%	19.3%	13.2%	10.5%	5.8%	95.7%
	79.4%			16.3%		

⁵ 考生 D 與 I 作文字數分別為 659 與 697 字。

⁶ 依據衛世華、郭曉玲 (2003) 抽樣「全民英檢」高級考生作文所使用的字彙，約有 98% 落在 Collins COBUILD 的五級字表，其中約 88% 落在三至五級，10% 在一、二級。

經進一步比對 F+組三位抽樣考生之分項成績，發現考生 E 通過「用字遣詞」分項，其使用的高級字彙的比率較高，為 22%，相異詞比例亦達 46.31，顯示其字彙使用有變化，惟此考生作答部分離題，且未能清楚說明其個人立場，故未能通過「內容適切性」與「寫作效果」兩項。F+組的另兩位考生（D 與 F）未通過「遣詞用字」分項，其使用的高級字彙的比率分別為 16.5% 及 17%，則均較 P 組為低。

6.4.3 T-unit 統計分析結果（表十）

九位抽樣考生作文 T-unit 平均個數為 37.6，平均長度為 21.6 字。值得注意的是，考生成績表現並未與 T-unit 平均長度成正比，通過優級寫作測驗的三位考生（P 組）其 T-unit 平均長度反而較 F+與 F-組稍低。

表十 九十三年優級寫作測驗
抽樣考生作文 T-unit 數量與長度分析統計

考生代碼	整體寫作成績	文法結構分項成績	T-unit 個數	該組平均值	T-unit 平均長度 (字數)	該組平均值 (字數)
A	通過	通過	45	40	17.0	19.8
B		通過	38		22.7	
C		通過	37		19.7	
D	未通過 (F+)	未通過	22	35.3	30.1	23.2
E		通過	33		23.2	
F		通過	51		16.3	
G	未通過 (F-)	未通過	43	37.3	19.7	21.9
H		未通過	34		25.9	
I		未通過	35		16.3	
總平均			37.6		21.6	

進一步檢視每位抽樣考生作文 T-unit 的分析資料，發現考生句構使用呈現兩種傾向，而此傾向區分似乎與通過或不通過無關：

- 1) 九位考生中，有四位考生（A, C, F, G）傾向使用較短的 T-unit（平均長度在 16 至 20 字之間），而整篇文章 T-unit 的個數則偏多（37 至 51 個）。
- 2) 另一類考生（如 D, H）則有企圖以「較長句構」展現其寫作能力的傾向，T-unit 平均長度可達 25 至 30 字，惟如表十一之錯誤量分析顯示，其 error-free units 之百分比偏低，且未通過文法結構分項，顯示其雖然傾向於寫較長的複雜句，但正確性的掌握有待加強。以考生 D 為例，評分委員的評語指出其所寫的句子時常“unnatural and tediously long”，並有代名詞指涉不清及連接詞誤用的情形。

表十一 九十三年優級寫作測驗
抽樣考生 Error-free T-unit 數量與百分比統計分析

考生代碼	整體寫作 成績	文法結構 分項成績	Error-free T-unit 個數	該組 平均值	Error-free T-unit 百分比	該組 平均值
A	通過	通過	17	16	37.8%	40.1%
B		通過	16		42.1%	
C		通過	15		40.5%	
D	未通過 (F+)	未通過	4	11	18.1%	28.9%
E		通過	11		33.3%	
F		通過	18		35.3%	
G	未通過 (F-)	未通過	8	5.7	18.6%	14.9%
H		未通過	4		11.8%	
I		未通過	5		14.3%	
總平均			10.8		27.8%	

錯誤量分析方面,表十一呈現抽樣考生 error-free T-unit(即無文法結構錯誤的 T-unit)之數量與百分比統計結果。整體而言,總體成績表現越佳的考生,其 error-free T-unit 的百分比越高,顯示其錯誤量越低。通過優級寫作的三位考生 error-free T-unit 之百分比為 40.1%, F+組則為 28.9%, F-組最低,為 14.9%。此外,通過文法結構分項的考生(A, B, C, E, F)之 error-free T-unit 之平均百分比為 37.8%,遠高於未通過此項的考生(D, G, H, I)之平均百分比(15.7%)。

此外,經調閱考生作文分析錯誤發現,通過文法結構分項的考生,其錯誤多屬獨立字、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。未通過文法結構分項考生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,造成解讀困難。

6.4.4 定量分析總結

謹將上述定量分析之數據彙整如表十二。

表十二 九十三年優級寫作測驗
抽樣考生字彙量、句構複雜度及錯誤量分析結果摘要

分組	相異詞 比例 平均值	Collins COBUILD 五級字彙分布		T-unit 平均數目	T-unit 平均長度 (字數)	Error-free T-unit 百分比
		三/四/五級 百分比	一/二級 百分比			
通過 (P)	44.8	77.7%	18%	40	19.8	40.1%
未通過 (F+)	42.2	76%	18.5%	35.3	23.2	28.9%
未通過 (F-)	35.3	84.5%	12.5%	37.3	21.9	14.9%

整體而言，通過優級寫作測驗的考生（P 組）較未通過的考生（F+與 F-組）表現較佳部分有二：

- 1) 相異詞比例較高，顯示其字彙使用較少重複、較有變化；
- 2) Error-free T-unit 之百分比比較高，顯示其文法句構的正確性較佳。

然而，僅以字彙及文法句構分析結果詮釋考生寫作表現，仍有其不足之處。本研究將進一步以定性方法，分析評分委員針對每份試卷所寫的評語（assessor's comments），以了解評分委員如何判定考生在「內容適切性」、「組織與連貫性」、「寫作效果」及「作者角色」四方面的表現。

6.4.5 定性分析研究結果

優級寫作閱卷前，評分委員須先熟悉評分量表中各分項通過標準的敘述（請參見附錄一）。每份試卷評閱完畢後，評分委員須立即針對考生在六個評分項目的表現寫下評語。下文將針對「內容適切性」、「組織與連貫性」、「寫作效果」及「作者角色」四項，分析比對評分量表的敘述與評分委員的評語。

6.4.5.1 內容適切性（Relevance and Adequacy）

優級寫作評分量表中，「內容適切性」分項的通過標準如下：

- ◆ Entirely relevant
- ◆ All parts of the task are addressed.
- ◆ Personal opinions are appropriate to the task and effectively stated with an appropriate number of supporting points given based on the input.
- ◆ Minimum length of the task is mandatory.

本次寫作試題說明中，對寫作情境（situation）及寫作重點（task requirements）皆有清楚的界定，因此評分委員對於上述四項通過標準的詮釋相當一致。整體而言，

考生必須能針對探討的主題，正確清楚地闡述分析錄影節目及閱讀文章中呈現的主要論點 (main points)，然後提出自己的看法 (personal opinions)，並將此兩部分加以整合 (例如引述相關論點來支持自己的論述)。

而綜合評分委員的意見，考生寫作未能通過「內容適切性」分項的常見原因如下：

- 1) 本次測驗採用人文藝術類題材，部分考生因缺乏相關的文化知識，無法正確理解閱讀文章⁷中探討的抽象概念及隱含的思維脈絡，以致對相關論點的解釋錯誤或不完全；
- 2) 未能掌握「切題」的原則 (例如：個人意見部分隨興發揮，包含一些與寫作主題或聽/讀內容無關的細節)；
- 3) 在某一寫作重點上著墨過多，以致在其他寫作重點上發揮不足 (例如：超過二分之一的篇幅為個人意見，針對聽讀內容的闡述不足；或花過多的篇幅在說明聽讀內容，直到文末才說明自己的主張，未深入討論即下結論)。

6.4.5.2 組織與連貫性 (Organization and Coherence)

優級寫作評分量表中，「組織與連貫性」分項的通過標準如下：

- ◆ The text is logically organized throughout.
- ◆ Appropriate paragraphing
- ◆ Wide and appropriate use of linking devices

在上述三項通過標準中，評分委員多認為優級考生在基本的段落區分 (appropriate paragraphing) 方面較無問題，能清楚地顯示每個段落的目的與功能。

然而在維持論述的連貫性方面，許多考生的表現不甚理想，常見的問題如下：

- 1) 有些考生的作文雖然看似包含三部分的基本架構 (闡述錄影節目的論點、闡述閱讀文章的論點、提出自己的看法)，然而未能建立三部分共同的論述主軸，或無法有效運用引述、舉証等方式來維持三部分論點的邏輯關聯性。
- 2) 如前述提及，部分考生未能充分理解閱讀文章的內容，以至在闡述摘要的部分語句前後不連貫、組織零亂 (“Many points are listed but not logically connected.”)。

此外，在 linking devices 的使用方面，本次優級測驗考生多為通過高級測驗者，因此大致能掌握連接詞及轉折詞的用法。惟仍有評分委員指出考生誤用 (misuse) 或過度使用 (overuse) 連接詞及轉折詞的情形。此外，寫作能力佳的考生除能正確使用連接詞及轉折詞外，亦較善於使用同義詞等技巧建立上下文詞彙的連貫性 (lexical coherence)。

⁷ 本次試後問卷調查顯示，考生多數認為錄影節目難易適中，閱讀文章則偏難。

6.4.5.3 寫作效果 (Effect on the Target Reader)

優級寫作評分量表中，「寫作效果」分項的通過標準如下：

- ◆ Reader's interest is engaged and sustained throughout.
- ◆ Desired effects on the target reader are achieved.

多數評分委員認為「寫作效果」乃一整體性「語用」(pragmatics)層次的概念，前述提及的內容切題度、論述的組織與連貫性、字彙豐富程度、句構複雜度及錯誤量等均是形成「寫作效果」是否良好的重要因素。因此評分委員在此部分的評語多為針對考生寫作表現的整體印象描述，詮釋評分量表的方式與研究者預期不同。

除整體印象外，評分委員評語亦顯示下列影響寫作效果的因素：

- 1) 由於優級寫作涉及抽象概念的表達，本次許多考生因缺乏足夠程度的字彙，無法清楚地陳述想要表達的理念，造成解讀困難，影響寫作效果；
- 2) 條理分明、立論清楚的 introductory paragraph 對整體印象有加分效果，然而亦有考生雖然寫出不錯的 introductory paragraph 或 thesis statement，後續的申論 (elaboration) 及舉證 (support) 卻不足，留給讀者「虎頭蛇尾」的印象。

6.4.5.4 作者角色 (Writer's Role)

優級寫作評分量表中，「作者角色」分項的通過標準如下：

- ◆ Clear and appropriate use of register for the task.
- ◆ Purpose of writing is appropriately stated.

本次寫作試題明確要求考生須寫一短篇課後報告 (a short paper for an academic course)，對於寫作情境及目的均有清楚說明，因此評分委員對於上述兩項通過標準的詮釋相當一致。整體而言，考生必須清楚說明寫作目的，並使用符合學術情境的語言、文體 (register) 及論述方式。本次考生常見的問題為無法維持文體的一致性，文章中出現不當的「語體轉換」(style shift) 或正式與非正式用語交互出現的情況。此外，評分委員亦認為「作者角色」亦為整體性「語用」(pragmatics)層次的概念，為優級寫作評分的重要指標。

6.4.6 定性分析總結

整合上述定性分析結果，研究者發現本次優級考生寫作表現似與下列兩項因素有較密切的關聯：

- 1) 考生對閱讀文章的理解程度：如 6.4.5.1 (內容適切性) 與 6.4.5.2 (組織與連貫性) 之分析顯示，考生若未能掌握閱讀內容的思想脈絡，則無法正確引述相關

的論點來支持自己的論述或比較觀點的異同，因此在「內容適切性」與「組織與連貫性」的表現勢必受影響。

- 2) 考生表達抽象思維的能力：依據優級能力指標，通過考生應能以書面方式探討各類主題，包含抽象概念。如 6.4.5.3 (寫作效果) 之分析顯示，本次許多考生無法清楚地表達想法，以至影響寫作效果，推測可能係描述抽象概念的字彙不足所致。

此外，上述定性分析亦顯示：優級評分過程中，「寫作效果」與「作者角色」與其他評分項目應為 hierarchical 與 modeling 的關係，而非平行重疊關係。換言之，評分者首先檢視考生能否針對題目要求，使用符合作者身分的文體 (register)，將訊息傳達給指定的讀者，有效的達成寫作目的。組織方法、詞彙及語法結構則可視為達成寫作效果的工具。因此，優級寫作首要評分項目應為屬「語用」(pragmatics) 層次的「寫作效果」與「作者角色」兩項標準，其次才為屬「語意」(semantics) 與「語法」(syntax) 層次的「內容適切性」、「組織與連貫性」、「遣詞用字」與「文法結構」。

7. 問卷調查結果

7.1 考生意見調查結果

43 名應考考生中，測驗後，以電話訪談或郵寄問卷方式進行問卷調查，回收之有效樣本 20 份，佔出席考生之 46.5%。依問卷內容之類別，摘要問卷結果如下，問卷調查內容請參考附錄二：

7.1.1 測驗方式

- ◆ 測驗流程說明：受訪考生普遍認為測驗流程的說明相當清楚。
- ◆ 作答時間：七成的考生認為作答時間充足，三成的考生認為稍嫌不足。
- ◆ 試場安排及操作介面：優級寫作測驗考生可選擇以電腦或紙筆作答，本次共有 30 位考生選擇以電腦作答，13 位選擇以紙筆作答。選擇電腦作答的考生普遍認為本次提供的電腦設備新穎、文書軟體容易操作，且電腦作答速度較快、容易修改；但亦有部分考生表示坐位安排稍嫌擁擠、電腦桌上空間不足。

7.1.2 測驗內容取材

本次測驗採用人文藝術類題材。約半數的考生認為主題難度可接受，有創意、具趣味性及挑戰性；但亦有約半數的考生認為主題較抽象及專業、難度頗高，非文學或藝術主修者恐難以發揮。以聽力及閱讀內容而言，多數受訪考生認為錄影節目難易適中，閱讀文章則偏難。

7.2 寫作測驗評分委員意見調查結果

本次優級寫作測驗共有四位評分委員參與正式閱卷。研究人員除收集分析閱卷評語之外，並於閱卷完畢後訪談每位評分委員，了解其對測驗內容、評分方式、量表設計之看法，以下為四位評分委員之綜合意見。

7.2.1 寫作測驗內容

評分委員普遍認為錄影節目內容有趣，而閱讀文章難度頗高，具挑戰性，不熟諳人文藝術、社會學或哲學領域思維及語彙的考生恐不易理解。寫作題目方面，題型設計與一般寫作測驗相當不同，頗具創意；惟考生必須受過論文寫作訓練，同時具備相關領域的文化素養，才能有較佳的表現。

7.2.2 寫作評分量表

有評分委員認為優級寫作評分應以內容為主要考量，建議「遣詞用字」與「文法結構」兩項標準可合併為「語言使用」(Language Use)一項，以降低技術性細節(mechanics)在評分上的比重。

8. 結論與建議

本次優級寫作測驗「考生表現分析」及「試後意見調查」顯示下列結果及幾項值得後續研究的課題：

8.1 字彙量、句構複雜度與正確性

本研究使用「定量」方式比較通過與不通過考生在字彙量、句構複雜度與錯誤量的表現，結果顯示通過優級寫作測驗的考生字彙使用較有變化(相異詞比例較高)且句構正確性較佳(error-free T-unit 之百分比較高)。

然而，通過與不通過考生在高級字彙(Collins COBUILD 一、二級)使用比例及 T-unit 平均長度方面的表現並非通過與否的關鍵。此現象或與優級考生多為已通過高級測驗者，其能力值集中有關，抑或顯示以字彙及文法句構分析考生寫作表現，仍有其侷限性。囿於此次報考人數及符合比對條件的樣本較小，此部分仍需後續研究觀察。

8.2 內容適切性、組織與連貫性、寫作效果、作者角色

本研究「定性分析」部分採取「立即回想」(immediate retrospection)的方式收集評分委員的閱卷評語。其目的在了解評分委員如何詮釋評分量表，以決定考生的作文是否在「內容適切性」、「組織與連貫性」、「寫作效果」及「作者角色」四項達到通過標準。結果顯示除「寫作效果」一項外，評分委員對評分量表的詮釋大致如研究者所期。

8.3 寫作評分量表

有關評分委員提出將「遣詞用字」與「文法結構」兩項標準合併為「語言使用」(Language Use)一項的建議，由於優級寫作評分主要考量為「內容」，此建議未來應可採行，以降低技術性細節 (mechanics) 在評分上的比重。此外，依據評分委員的意見，「寫作效果」與「作者角色」乃較為整體性「語用」(pragmatics) 層次的概念，與其他評分項目應為 hierarchical 與 modeling 的關係，而非平行重疊關係。

綜合上述分析，優級寫作評分項目從整體 (語用) 到細部 (語意、語法) 層次順序應為：寫作效果、作者角色、內容適切性、組織與連貫性與語言使用。

8.4 寫作內容取材

本次測驗採用人文藝術類題材，主要考量為人文課題的接受面較廣⁸。定性分析結果顯示考生對閱讀文章的理解程度及表達抽象思維的能力為影響其寫作表現的重要因素。考生意見調查結果則顯示：約半數的考生認為主題難度可接受，但亦有約半數的考生認為主題較抽象及專業、難度偏高。

由於國內英語教育多以「語法詞彙正確」及「表達流利」為主要教學目標，文化學習多限於日常生活習俗及流行文化的介紹，在藝術涵養、哲學思維等方面之訓練著墨較少，以致部分考生由於缺乏相關的文化背景知識而無法深入理解閱讀文章，故雖能說寫流利，但表達內容的深度仍有加強的空間。建議未來可研究修訂優級測驗能力指標，將文化素養與評析能力納入，以幫助考生、教師及成績使用單位了解本級所要求的程度。

8.5 結語

本研究初步顯示，優級寫作通過與未通過考生在六個分項的作答表現的確存有質與量的差異，分項評分量表大致適切，細部則仍有修改空間。惟囿於此次報考人數及符合比對條件的樣本較小，抽樣文章有限，日後仍有賴更多資料的收集與驗證。

⁸Clapham(1996)以 IELTS 考生為對象，研究背景知識對閱讀理解表現的影響，其結果顯示：人文背景考生在閱讀科技、生物及醫學類文章時，理解程度較易受到背景知識的限制；而科技背景考生在閱讀人文、教育、藝術類題材時，背景知識的影響力則較小。

三、第二節測驗（整合式口說測驗）

1. 測驗方式

第二節測驗時間約五十分鐘，為一整合聽、讀、寫、說四種活動之測驗。首先，考生有三十分鐘的時間，根據第一節寫作測驗所看之錄影節目、所閱讀的文章、及自己所作的筆記及文章，準備報告要點。接下來考生進入口說試場面對主考作十分鐘的口頭報告。之後，考生再回答主考所提與主題相關的問題。主考同時須依據分項式評分量表給分（量表請參閱附錄三）。口說試場內，全程錄音、錄影。

2. 測驗題型

第二節測驗分兩部分，第一部分為考生的口頭報告(oral presentation)，共計十分鐘。第二部分為考生依據主考(interlocutor)所提問題，準備兩分半鐘後，回答兩分半鐘，共計十分鐘。本節的測驗重點在於檢測考生是否能靈活運用聽、讀、寫、說各種策略，自題目的內容中擷取重要資訊，加以組織摘要，有條理地將這些資訊及個人意見作口說的的表達。

3. 評分標準

本測驗採分項式評分，分項式評分量表之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

主考在正式測驗前一週，先行閱讀優級研究報告並評聽通過與未通過考生之錄音帶，以熟悉評分標準及測驗進行方式。測驗當天，主考於口說測驗開始前兩小時報到，閱讀試題、熟悉測驗流程並相互交換意見。原則上，考生必須六項分項成績均達標準才視為通過優級。由於優級為整合式測驗，為求審慎，若遇考生寫作測驗通過而口說測驗未通過，或是口說測驗通過而寫作測驗未通過，均進行二評；一、二評成績不一致者，亦須再經三評。另，由於優級題目內容專業，考生若僅未通過適切性該項，則均二評以求最後的確認。

5. 考生成績

5.1 考生表現

5.1.1 整體表現

43 名考生中，7 名通過口說測驗，佔總人數之 16%，其中僅 2 名考生通過優級測驗，佔總人數之 5%。

所有考生依性別分佈之通過人數及通過率請參見表十三。由表十三可知，男性考生之口說測驗及總通過率均高於女性考生，分別為 20% 比 13% 及 10% 比 0%（如陰影部分）。

表十三 九十三年優級口說測驗
依性別分佈之通過人數及通過率

類別		人數	人數 %	口說 通過	口說 通過率	優級 通過	優級 通過率
性別	男性	20	47%	4	20%	2	10%
	女性	23	53%	3	13%	0	0%

另外，所有考生依學歷分佈之各分項通過人數及通過率請參見表十四。

表十四 九十三年優級口說測驗
依學歷分佈之分項通過人數及通過率

評分項目 學歷	人數	發音 語調		內容 適切性		用字 遣詞		文法 結構		流利度		連貫性		六項均達 通過標準	
		通 過 人 數	通 過 率	通 過 人 數	通 過 率	通 過 人 數	通 過 率	通 過 人 數	通 過 率	通 過 人 數	通 過 率	通 過 人 數	通 過 率		
大學以下	6	6	100%	0	0%	3	50%	5	83%	5	83%	1	17%	0	0%
國內大學															
英語主修	9	8	89%	5	56%	5	56%	4	44%	5	56%	4	44%	2	22%
非英語主修	13	11	85%	2	15%	8	62%	8	62%	8	62%	7	54%	1	8%
國內研究所	2	2	100%	1	50%	2	100%	2	100%	2	100%	1	50%	1	50%
國外大學	6	6	100%	1	17%	3	50%	1	17%	3	50%	3	50%	1	17%
國外研究所	6	6	83%	3	50%	5	83%	3	50%	4	67%	3	50%	2	33%

由表十四可看出，六項全過之通過率最高者為國內研究所學歷考生，通過率為 50%（如陰影部分），不過樣本數太少（僅兩名），是否具代表性則有賴更多的考生參與方能確定；國外研究所學歷考生名列第二，通過率為 33%（如上斜線部分）；具有國內大學英語主修學歷考生名列第三，通過率為 22%（如直線部分）；國外大學

學歷考生第四，通過率為 17%（如橫線部分）；國內大學非英語主修學歷考生第五，通過率為 8%（如格線部分）。值得注意的是，大學以下學歷的考生無人通過，並且在適切性一項均未達通過標準（如下斜線部分）。經查，這六名考生中，五名為高中生，一名國中生，年齡均在 19 歲以下，並非本測驗所檢測的目標考生（亦即「英語能力接近受過高等教育之母語人士」）。據推測應為年紀過輕，歷練與學養尚不足之故。

考生依年齡層分佈之口說測驗各分項通過人數、通過率及總通過率請參見表十五。

表十五 九十三年優級口說測驗

依年齡分佈之分項通過人數、通過率及總通過率

評分項目 年齡	人數	發音語調		內容適切性		用字遣詞		文法結構		流利度		連貫性		六項均達通過標準	
		通過人數	通過率	通過人數	通過率	通過人數	通過率	通過人數	通過率	通過人數	通過率	通過人數	通過率	通過人數	通過率
19 歲以下	6	6	100%	0	0%	3	50%	5	83%	5	83%	1	17%	0	0%
20-24 歲	17	15	88%	2	12%	9	53%	10	59%	11	65%	8	47%	2	12%
25-29 歲	8	7	88%	5	63%	6	75%	3	38%	5	63%	4	50%	2	25%
30-34 歲	3	3	100%	2	67%	2	67%	1	33%	3	100%	1	33%	1	33%
35-39 歲	4	3	75%	1	25%	3	75%	1	25%	2	50%	1	25%	0	0%
40-44 歲	3	3	100%	1	33%	1	33%	1	33%	1	33%	2	67%	1	33%
45 歲以上	2	2	100%	1	50%	2	100%	2	100%	1	50%	2	100%	1	50%
所有考生	43	39	91%	12	28%	26	60%	23	53%	28	65%	19	44%	7	16%

就考生之六項分項成績而言，以發音語調之通過率最高，其他各項依序（由高至低）為流利度、用字遣詞、文法結構、連貫性、內容適切性（如陰影部分）。而考生於發音語調及流利度之通過率較其他四項佳，此結果與預試結果相符（請參閱全民英語能力分級檢定測驗優級測驗研究報告）。

另外，根據歷年「全民英檢」其他各級考生之表現，通常年齡層愈低考生的發音語調平均成績大致愈高，然在優級此一現象並不明顯（如斜線部分），推測是因優級考生已為語言能力金字塔頂端之考生，語言掌握已具相當的成熟度，其發音的精準正確性已不因年齡層不同而有明顯差異。

而各年齡層中，以 20 至 24 歲之考生人數最多（17 人），佔總人數之 40%。值得一提的是，45 歲以上考生，在六項評分標準中，其中四項通過率為 100%，總通過率也高達 50%，居於首位；惟該年齡層僅有 2 名考生應考，而其中一名通過應是造成通過率領先的主要原因。此外，19 歲以下的考生，在內容適切性及連貫性兩項的

通過率均最低，原因同之前所述，應是年紀過輕，歷練、學養尚不足之故。

5.2 個別表現

5.2.1 目的

藉由比較通過與未通過考生的語言表現，了解其間之差異，並同時驗證通過與否分界點 (cut-off point) 的適切性。研究題目如下：

- 考生在字彙的表現上，是否有差異
- 考生在語法句構的表現上，是否有差異
- 考生在流利度的表現上，是否有差異

5.2.2 對象

由於通過考生僅有七名，為與未通過考生進行對等的比較，故儘可能挑選年齡、性別、與教育背景相當的考生，另考慮到繕謄工作之耗時，僅選八名考生做為研究對象。八名考生之相關資料請參見表十六：

表十六 九十三年優級口說測驗
抽樣考生基本資料

		通過				未通過			
		年齡			教育程度	年齡			教育程度
		個人	代碼	平均		個人	代碼	平均	
性別	女	24	A	30	國內大學	22	E	29.5	國內大學
		43	B		國外研究所	39	F		國外研究所
	男	27	C		國內研究所	22	G		國內大學
		26	D		國內大學	35	H		國外大學

5.2.3 方法

相較於口說測驗第二部份的回答主考提問，考生於本節測驗之第一部分，須針對題目準備 30 分鐘後，再以 10 分鐘的時間表述所準備的內容，因此口語表現最為獨立 (independent)，而且發表的時間也最長、最完整。故將八名考生於第一部分之口頭報告謄錄成文字，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

是否因通過與否而有所不同。

- 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 LMC 軟體 (Chandler, Longman Group UK Ltd., 1989) 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) 是否因通過與否而有所不同。
- 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、平均每 AS-Unit 的字數及平均每 AS-Unit 的子句數。此外, 子句數量、平均每一子句的字數也列入計算, 以了解考生所用句構之複雜性 (complexity of sentence structure) 是否因通過與否而有所不同。

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。

- 語言能力愈佳者, 其語言使用的流暢性愈佳, 無須聽者費力解讀, 而據文獻記載 (Oliveira, 2002), $\frac{1}{4}$ 秒之無聲停頓 (silent pause) 即會引起聽者的注意, 輕微影響原屬流暢之溝通, 而停頓時間愈長, 流暢性愈差。因此將 0.5 秒 (含 0.5 秒) 以上之停頓列入計算, 以期了解考生言談中的停頓多寡是否亦因通過與否而有所不同。

5.2.4 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十七:

表十七 九十三年優級口說測驗
抽樣考生所用字彙量與級區分佈

口 說 成 績	考生代碼	Collins Cobuild Word List					總字數
		Band 5	Band 4	Band 3	Band 2	Band 1	
		高字頻		中字頻	低字頻		
		較常用字 ←			→	不常用字	
通 過	A	154	53	24	19	8	258
	B	174	58	55	40	21	348
	C	186	73	48	38	29	374
	D	223	62	42	27	30	384
	平均字數	184.25	61.50	42.25	31.00	22.00	341.00
	百分比	72%		12%	16%		
未 通 過	E	140	32	21	16	12	221
	F	182	53	57	32	24	348
	G	162	41	31	22	18	274
	H	219	75	40	35	26	395
	平均字數	175.75	50.25	37.25	26.25	20.00	309.50
	百分比	73%		12%	15%		

- 由表十七可以清楚看出，雖然在高字頻 (high frequency words) 中字頻及低字頻的百分比上，通過考生與未通過考生的差別不大，但是在平均總字數及在 Bands 1~5 的各別平均字數上，通過考生均較未通過考生來得多 (341 vs. 309.50 ; 22.00 vs. 20.00 ; 31.00 vs. 26.25 ; 42.25 vs. 37.25 ; 61.50 vs. 50.25 ; 184.25 vs. 175.75)。

- 經 LMC 軟體計算，考生第三部分作答之相異詞比例請參見表十八：

表十八 九十三年優級口說測驗
抽樣考生字彙相異詞比例計算結果

口說成績	考生代碼	Type/Token Ratio	Average Type/Token Ratio
通過	A	268/1053= .25	362/1220= .30
	B	362/1084= .33	
	C	421/1220= .34	
	D	397/1522= .26	
未通過	E	233/707= .32	330/1127= .29
	F	375/1182= .31	
	G	298/997= .29	
	H	416/1621= .25	

由表十八可看出通過與未通過考生，無論彼此之間或整體平均而言，其相異詞比例均無太大差異，其中相異詞比例最低者為.25，通過與未通過考生中各有一人(考生 A 與 H)，而相異詞比例最高者為.34，為通過考生(考生 C)，然而未通過考生之相異詞比例最高者亦有達.32 者(考生 E)。

推測應是本次參加優級測驗之考生均為高級測驗通過或托福 287 分以上者，語言程度類似，字彙量亦差不多，才會出現表六之結果。

進一步查看未通過考生的分項成績時，發現 E 與 F 考生在「用字遣詞」該項未達通過標準，且四位未通過考生在「內容適切性」該項均未達通過標準，因此能否正確使用字彙及表達的內容是否適切，亦同時影響考生的整體表現。

- 經統計,八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十九:

表十九 九十三年優級口說測驗
抽樣考生所用 AS-Unit 數量及平均字數

考生代碼	口說成績	AS-Unit 數量	子句數量	總字數	每一 AS-Unit 之平均字數	每一子句之平均字數	每 AS-Unit 之平均子句數	0.5 秒以上(含 0.5 秒)之停頓 (pause)數量
A	通過	47	98	1053	22	11	2	44
B	通過	53	98	1084	20	11	2	129
C	通過	54	121	1220	23	10	2	84
D	通過	53	105	1522	29	14	2	137
通過考生之平均值		52	106	1220	24	12	2.0	99
E	未通過	49	104	707	14	7	2	155
F	未通過	69	135	1182	17	9	2	187
G	未通過	55	125	997	18	8	2	162
H	未通過	70	177	1621	23	9	3	155
未通過考生之平均值		56	119	1127	18	8	2.1	165

表十九可發現,雖然通過考生之平均 AS-Unit 數及子句數都沒有未通過考生來得多(如斜線部分),但是就總字數、每一 AS-Unit 之平均字數及每一子句的平均字數而言,通過考生均多於未通過考生(如菱線部分),顯示通過考生使用之句子較長。就每一 AS-Unit 之平均子句數而言,通過與未通過考生差不多(如橫線部分),顯示考生使用句子的複雜程度相差不多。再就 0.5 秒(含 0.5 秒)以上的停頓數量而言,通過考生明顯少於未通過考生(如直線部分),顯示通過考生的整體語言表現較為流利。

6. 問卷調查結果

43 名應考考生中,測驗後,以電話訪談或郵寄問卷方式進行問卷調查,回收之有效樣本 20 份,佔出席考生之 47% (問卷調查內容請參閱附錄四)。

依問卷內容之類別,摘要問卷結果如下。

- 整體而言,所有填卷考生均認為本次口說測驗流程的說明清楚。
- 三分之一填卷考生認為本次口說測驗可以考出實際的英語口說能力。另有三分之二考生則認為因為口說和寫作採同一主題,如前項表現不好,連帶口說內容也不可能豐富,且題目太專業,不易發揮。

- 填卷考生大致上均肯定一對一面談之測驗方式。
- 除了兩位填卷考生外，餘均認為主考的說話速度適中。
- 僅兩位填卷考生認為第一部分(presentation)的準備時間(30 分鐘)太長或不足。
- 少數填卷考生認為第一部分(presentation)的作答時間(10 分鐘)過長。
- 大多數的填卷考生認為第二部分(questions and answers)每題作答之前需要 2 分 30 秒的準備時間。
- 大多數的填卷考生認為第二部分(questions and answers)每題的作答時間(2 分 30 秒)足夠。
- 有關測驗難度，第一、二部分均有考生選填較難，也有考生認為兩部分的難易度差不多。
- 有關考生表現，第一、二部分均有考生選填表現較佳，也有考生認為其在兩部分的表現差不多。
- 大多數填卷考生對口說測驗試場安排(例如桌椅、燈光)表示肯定。

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)。

7. 結論與建議

7.1 考生表現摘要如下：

- 字彙
通過考生使用較多的字彙，然在各頻率的字彙使用上則與未通過考生無明顯差異。
- 句構
通過與未通過考生之句構使用及句子的複雜度均差不多。
- 流利度
通過考生平均停頓(0.5 秒及以上)較少，語言流利度優於未通過考生。

7.2 建議

- 日後研究重點：
由本分析發現可看出，通過與未通過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及流利度是有些不同，可說明本測驗所使用的評分量表應屬適切。惟所採樣本有限，能否將此發現視為必然，則仍有賴日後更多資料的收集與驗證。
- 未來施測方向：
依據實際報考人數及國內英語學習現況預估，一般民眾自行報考優級測驗的需求不高，在測驗行政考量下，建議優級測驗宜改採不定期方式舉辦—若遇學校、機關、團體有特殊需求，就其需求擇期專案辦理，並委請相關領域的學者專家參與命題及評分。

附錄一：寫作評分量表

Score Focus	Pass
Relevance and Adequacy	<ul style="list-style-type: none"> ◆ Entirely relevant ◆ All parts of the task are addressed ◆ Personal opinions are appropriate to the task and effectively stated with an appropriate number of supporting points given based on the input ◆ Minimum length of the task is mandatory
Coherence and Organization	<ul style="list-style-type: none"> ◆ Text is logically organized throughout ◆ Appropriate paragraphing ◆ Wide and appropriate use of linking devices
Lexical Use <ul style="list-style-type: none"> ◆ Range ◆ Appropriateness 	<ul style="list-style-type: none"> ◆ Effective use of a wide range of vocabulary ◆ Vocabulary is used appropriately and errors are rare
Grammatical Use <ul style="list-style-type: none"> ◆ Range ◆ Accuracy 	<ul style="list-style-type: none"> ◆ Effective use of a wide range of structures ◆ Structures are used accurately and appropriately and errors are rare
Effect on the Target Reader	<ul style="list-style-type: none"> ◆ Reader's interest is engaged and sustained throughout ◆ Desired effects on the target reader are achieved
Writer's Role	<ul style="list-style-type: none"> ◆ Clear and appropriate use of register for the task ◆ Purpose of writing is appropriately stated

*Plagiarism: More than three consecutive words are copied from the input without appropriate quotation.

Non-ratable compositions

A composition is considered "non-ratable" when

- the length is shorter than 250 words, or
- the content is totally irrelevant, or
- the content is apparently a copy or memorization from other known materials, or
- the handwriting is completely unintelligible.

Non-ratable compositions will receive a final grade of "Fail" and grades for the subcategories will not be reported.

I. 寫作測驗內容部分

1. 整體而言，您覺得本次寫作測驗流程的說明是否清楚？
2. 您覺得本次寫作測驗作答的時間是否足夠？
3. 您對本次寫作測驗主題(topic)的接受度如何？
4. 您覺得本次測驗是否可以考出您實際的英語寫作能力？為什麼(不)？
- 5-1. 您覺得 Phase 1 的錄影節目難度如何？
A. 容易 B. 適中 C. 稍難 D. 很難
- 5-2. 您覺得 Phase 2 的閱讀文章難度如何？
A. 容易 B. 適中 C. 稍難 D. 很難
- 6-1. 您是否能從 Phase 1 的錄影節目中擷取足夠的資訊供寫作使用？
- 6-2. 您是否能從 Phase 2 的閱讀文章中擷取足夠的資訊供寫作使用？

II. 寫作測驗試場安排及操作介面部分

7. 您對本次寫作測驗試場安排(例如桌椅、燈光等)的感覺如何？
8. 您選擇使用電腦/紙筆作答的原因為何？
9. 如果可以重選，您會選擇使用電腦或紙筆作答？

選擇使用電腦作答者請回答第 10~13 題：

10. 您覺得使用電腦對您的作答表現有幫助嗎？
11. 您覺得本次測驗所提供的文書軟體是否容易操作？
12. 您覺得本次測驗前針對該文書軟體提供的說明是否清楚？
13. 您在 Phase 2 閱讀文章時，主要使用的是閱讀題本或電腦上 reading text 的檔案？

附錄三：口說測驗評分量表

Criteria	Description
Pronunciation •stress •rhythm •intonation •individual sounds	♦The candidate is able to produce entirely accurate utterances and sounds with appropriate stress, and natural rhythm, and intonation.
Relevance and Adequacy	♦The discourse is relevant to the task, and the contribution is more than adequate to complete the task.
Lexical Use •range •appropriateness	♦The candidate is able to use a wide range of vocabulary to effectively complete the tasks. Lexis is used appropriately. Errors are rare.
Grammatical Use •range •accuracy	♦The candidate is able to use a wide range of structures to complete each task accurately, appropriately and effectively. Errors are rare
Fluency	♦The candidate is able to keep communication flowing smoothly with minimal hesitation.
Coherence	♦The contribution is logically organized throughout a task.

The candidate has to pass all the six criteria in order to pass the superior speaking test.

III. 口說測驗部分

14. 整體而言，您覺得本次口說測驗流程的說明是否清楚？
15. 您覺得本次口說測驗是否可以考出您實際的英語口說能力？為什麼(不)？
16. 您對本次一對一面談的測驗方式感覺如何？
17. 您覺得主考的說話速度是否適中？
18. 您覺得第一部分(presentation)的準備時間(30 分鐘)是否足夠？
19. 您覺得第一部分(presentation)的作答時間(10 分鐘)是否足夠？
20. 您覺得第二部分(questions and answers)每題作答之前是否需要 2 分 30 秒的準備時間？為什麼(不)？
21. 您覺得第二部分(questions and answers)每題的作答時間(2 分 30 秒)是否足夠？
22. 本次口說測驗中哪一部分您覺得較為困難？為什麼？
23. 本次口說測驗中哪一部分您覺得您表現較好？為什麼？
24. 您對本次口說測驗試場安排(例如桌椅、燈光)的感覺如何？

參考書目

中文部分

- 吳若蕙, 吳怡芬, 衛世華, 郭曉玲, 廖秀育, 馬冬梅, 金韶 (2002)。全民英語能力分級檢定測驗優級測驗研究報告。台北: 語言訓練測驗中心。
- 衛世華, 郭曉玲 (2003)。「全民英檢」九十一年高級複試測驗。第十二屆中華民國英語文教學國際研討會論文集, 頁 599-616。台北: 中華民國英語文教師學會。
- 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)

英文部分

- Alderson, J. C. *et al.* (1995) *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.
- Alderson J. C. and Bachman, L. F. (eds.) (2004) *Assessing Speaking*. Cambridge: Cambridge University Press.
- Bachman, L. F. and Palmer, A. S. (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. and 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.
- Bygate, M., Skehan, P. and Swain, M. (eds.) (2001) *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing*. Harlow: Pearson Education.
- 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. and Taylor, L. (1997) Recent developments in IELTS. *ELT Journal*. 51 (4): 374-380.
- Chiang, G. S. and Dunkel, P. (1992) The effect of speech modification, prior knowledge, and listening proficiency on EFL lecture learning. *TESOL Quarterly*. 26: 345-374.
- Clapham, C. M. (1996) *The Development of IELTS: A Study of the Effect of Background Knowledge on Reading Comprehension*. Cambridge: Cambridge University Press.

- Foster, P., Tonkyn, A. and 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.
- Fulcher, G. (2003) *Testing Second Language Speaking*. Pearson Education Limited.
- Grabe, W. and Kaplan, R. (1996) *Theory & Practice of Writing*. London: Longman.
- Halliday, M. A. K. and 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. (1987) *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.
- Hughes, A. (1989) *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Hunt, K. (1965) *Grammatical Structures Written at Three Grade Levels*. Urbana, IL: National Council of Teachers of English.
- Hunt, K. (1983) Sentence combining and the teaching of writing. In Martlew, M (ed.) *The Psychology of Written Language: A Developmental Approach*. New York: J. Wiley. 99-125.
- Juan, E. U. and Silveira, J. C. P. (1998) A product-focused approach to text summarisation. *The Internet TESL Journal*.
<http://www.aitech.ac.jp/~iteslj/Articles/Juan-TextSummary.html>
- Kennedy, C. and Thorp, D. (2002) *A Corpus-Based Investigation of Linguistic Responses to an IELTS Academic Writing Task*. Unpublished manuscript. Cambridge ESOL.
- Knight, B. (1992) Assessing speaking skills: a workshop for teacher development. *ELT Journal*. 46 (3): 294-302.
- Lennon, P. (1990) Investigating fluency in EFL: a quantitative approach. *Language Learning*. 40 (3):387-417.
- Lee, J and McChesney. (2000). Discourse rating tasks: a teaching tool for developing sociocultural competence. *ELT Journal*. 54(2): 161-168.
- Lumley, T. and 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.
- Nation, I.S.P. (1990) *Teaching and Learning Vocabulary*. Boston, Massachusetts: Neinkle & Heinle Publishers.

- North, B. (1995) Scales of language proficiency. *Melbourne Papers in Language Testing*. 4 (2): 60-111
- North, B. (1995) Standardisation of continuous assessment grades. *Language Testing in the 1990s*. Macmillan Publishers Ltd.
- North, B. and 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.
- Oliveira, M. (2002) The role of pause occurrence and pause duration in the signaling of narrative structure. In Ranchod, E. and Mamede, N. J. (eds.) *Advance in Natural Language Processing, Third International Conference Proceedings*.
- O'Sullivan, B., Weir, C. J. and Saville, N. (2002) Using observation checklists to validate speaking-test tasks. *Language Testing*. 19 (1): 33-56.
- Pollitt, A. and Murray, N. (1996) What raters really pay attention to. *Studies in Language Testing*. 3. Cambridge: Cambridge University Press.
- Riazantseva, A. (2001) Second language proficiency and pausing. *Studies in Second Language Acquisition*. 23: 497-526.
- Richards, J., Platt, J. and Platt, H. (1992) *Longman Dictionary of Language Teaching & Applied Linguistics*. Hong Kong: Longman.
- Riggenbach, H. (ed.) (2000) *Perspectives on Fluency*. Ann Arbor: The University of Michigan Press.
- Shohamy, E., Reves, T. and Bejarano, Y. (1986) Introducing a new comprehensive test of oral proficiency. *ELT Journal*. 40 (3): 212-220.
- Sinclair, J., Fox, G., Bullon, S., and Manning, E. (1995) *Collins Cobuild English Dictionary*. London: HarperCollins Publishers Ltd.
- Skehan, P. and 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.
- Upshur, J. A. and 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. and 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".
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 and 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. (1993) An investigation of planning time and proficiency level on oral test discourse. *Melbourne Papers in Language Testing*. 2(1):61-71.
- Yorozuya, R. and Oller, J. (1980) Oral proficiency scales: construct validity and the halo effect. *Language Learning*. 30 (1): 135-153.
- Young, R. (1995) Conversational styles in language proficiency interviews. *Language Learning*. 45 (1): 3-42.
- Young, R. and Milanovic, M. (1992) Discourse variation in oral proficiency interviews. *Studies in Second Language Acquisition*. 14 (4): 403-424.
- Zellner, B. (1994) Pauses and the temporal structure of speech. In Keller, E. (ed.) *Fundamentals of Speech Synthesis and Speech Recognition*. Chichester: John Wiley.