

# 考試院第 12 屆第 2 次會議考選部重要業務報告

民國 103 年 9 月 11 日

## 壹、考選行政

### 申論題評閱差異性分析－以實務界專家與學者評閱結果比較

#### 一、背景說明

各項國家考試應試科目申論題之閱卷委員，係經各該考試典試委員長遴提各專業領域學者專家，並經考試院同意後聘任，閱卷委員依參考答案及專業判斷評閱試卷後，給予評分。

相同應試科目（或同一題）評閱以同一閱卷委員為原則，若有多位委員評閱則分為單閱及平行兩閱兩種情形，在單閱情形下，當同一應試科目試卷數量過多時，分由多位委員評閱，惟每份試卷或每題仍僅 1 位閱卷委員評閱；平行兩閱情形則是每份試卷或每題均由 2 位閱卷委員同時評閱，兩閱分數差距如超過題分之三分之一則啟動第三閱，以三閱後較接近之兩閱分數平均為成績。

為維護評閱結果之公平性，閱卷過程中，除了維持個別委員給分寬嚴一致外，委員間之給分寬嚴亦需一致。為了解由實務界專家或學者擔任閱卷委員時給分之寬嚴是否有顯著差異，本研究以同一應試科目由多位閱卷委員評閱，分析比較委員間評閱分數之差異，以作為提昇閱卷品質之參考。

目前(103 年)計有 7 項考試全部類科採線上閱卷，其中司法官及律師考試第二試專業科目實施線上平行兩閱；另建築師考試「建築計畫與設計」、高考三級建築工程類科之「建築設計」與景觀類科之「景觀與都市設計」以及地方特考建築工程類科之「建築設計」實施紙本平行兩閱（參閱表一）。

表一、103 年考試採線上閱卷以及實施平行兩閱情形

項次	考試名稱	類科 (應試科目)	線上單 人閱卷	平行兩閱	
				線上	紙本
1	103 年第一次專門職業及技術人員高等考試醫師中醫師考試分階段考試、中醫師、營養師、心理師、護理師、社會工作師考試	各類科	V		
2	103 年專門職業及技術人員高等考試驗船師、第一次食品技師考試、高等暨普通考試消防設備人員考試、普通考試地政士、專責報關人員、保險代理人保險經紀人及保險公證人考試	各類科	V		
3	103 年第二次專門職業及技術人員高等考試醫師中醫師考試分階段考試、中醫師、營養師、心理師、護理師、社會工作師考試 103 年專門職業及技術人員高等考試法醫師、語言治療師、聽力師、牙體技術師考試	各類科	V		
4	103 年公務人員特種考試外交領事人員及外交行政人員考試、103 年公務人員特種考試國際經濟商務人員考試、103 年公務人員特種考試民航人員考試、103 年公務人員特種考試原住民族考試	各類科	V		
5	103 年公務人員高等考試一、二級考試	各類科	V		
6	103 年公務人員特種考試司法官考試(第二試)	司法官 (專業科目)		V	
7	103 年專門職業及技術人員高等考試律師考試(第二試)	律師 (專業科目)		V	
8	103 年公務人員高等考試三級考試暨普通考試	建築工程 (建築設計) 景觀 (景觀與都市設計)			V
9	103 年專門職業及技術人員高等考試建築師考試	建築師 (建築計畫與設計)			V
10	103 年特種考試地方政府公務人員考試	建築工程 (建築設計)			V

## 二、資料與分析方法

為了解實務界專家與學者評分寬嚴之差異性，本報告共擇定四項應試科目作為分析資料依據，分別是（一）102年一般警察特考四等考試行政警察類科之「刑法概要」；（二）102年司法官考試「民法與民事訴訟法」第一題；（三）102年律師考試「民法及民事訴訟法」第五題；（四）102年建築師考試「建築計畫與設計」。

上述各項應試科目，均有兩位以上閱卷委員評閱，依其評閱方式分為兩類進行統計分析：（參閱表二）

- （一）單閱：102年一般警察特考四等考試行政警察類科之「刑法概要」，由於該應試科目共有 8,222 份試卷，分由 11 位閱卷委員各自評閱 7 百餘份試卷（每卷四題）。11 位閱卷委員中有 7 位實務專家與 4 位學者，依評閱結果進行實務專家與學者兩組評閱分數之變異數分析。
- （二）平行兩閱：102 年司法官考試「民法與民事訴訟法」第一題試卷 1,997 份、102 年律師考試「民法及民事訴訟法」第五題試卷 2,687 份、102 年建築師考試「建築計畫與設計」試卷 2,835 份，均採平行兩閱方式，且兩閱委員分別為實務專家與學者，依評閱結果進行實務專家與學者兩閱評閱分數之相關分析及一致性分析。

表二、兩位以上閱卷委員評閱方式及分析項目

評閱方式 統計分析	單閱 (每份試卷或每題僅 1 位閱卷委員評閱)	平行兩閱 (每份試卷或每題均由 2 位閱卷委員評閱)
一、分析資料 (應試科目)	102 年一般警察特考四等考試行政警察類科之「刑法概要」	1. 102 年司法官考試「民法與民事訴訟法」第一題 2. 102 年律師考試「民法及民事訴訟法」第五題 3. 102 年建築師考試「建築計畫與設計」

二、分析方法	1. 各項統計指標 2. 實務專家與學者兩組評閱之變異數分析	1. 各項統計指標 2. 實務專家與學者兩組評閱之相關分析及一致性分析
三、分析項目	1. 多位委員間評閱分數之差異性 2. 兩組委員間評閱分數之差異性	1. 平行兩組委員間評閱之相關性 2. 平行兩組委員間評閱之一致性

### 三、統計資料分析

(一) 102 年一般警察特考四等考試行政警察類科之「刑法概要」(單閱、配分 100 分、8,222 份試卷，參閱附表 1)

1. 委員間評閱差異：本應試科目分由 11 位閱卷委員各評閱 7 百餘份，各委員評閱結果之平均成績最高自 48.86 分至最低 28.41 分不等，落差達 20 分(參閱圖 1)；又各委員評閱結果之高分群(60 分及以上)比例最高 37.45% 至最低 1.72% 不等，差距亦大，依各項統計指標顯示委員間評閱結果存有差異性，惟因各委員評閱之卷包地區不同，或因委員個人之見解不同或寬嚴標準不同或因應考人程度不同所致。
2. 實務專家與學者兩組間評閱差異：依 11 位閱卷委員之背景分為實務專家及學者兩組，實務專家組平均成績 34.8 分，學者組 38.6 分，實務專家組評閱平均分數雖低於學者組，惟 7 位實務專家評閱結果(參閱圖 1 之藍線)，與 4 位學者評閱結果(參閱圖 1 之紅線)高低互見；由變異數分析結果發現實務專家與學者兩組之組間變異小於組內變異，意即實務專家與學者兩組評閱結果並無顯著差異。

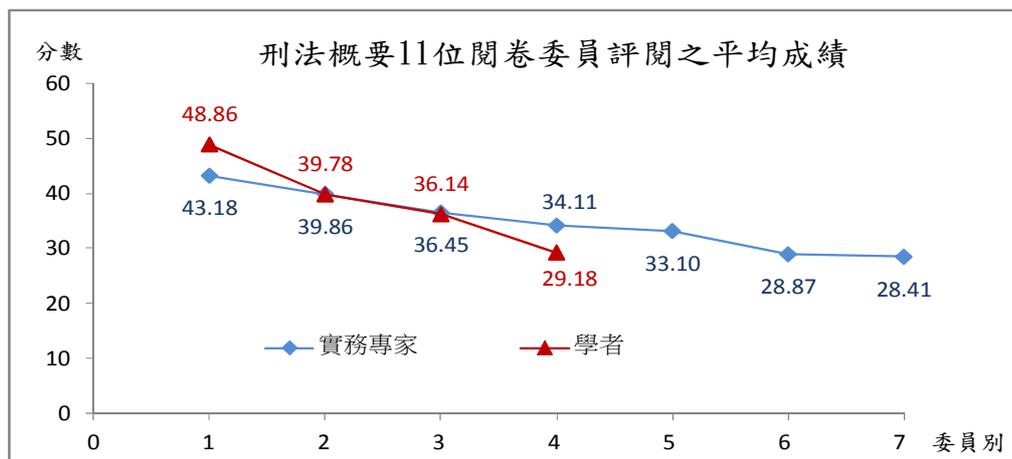


圖 1、102 年一般警察特考四等考試「刑法概要」成績圖

(二)102 年司法官考試「民法與民事訴訟法」第一題(平行兩閱、配分 45 分、1,997 份試卷，參閱附表 2)

1. **兩閱寬嚴標準**：本題平行兩閱之 2 位委員給分離散程度差異不大，意即個人對整體應考人程度區別之標準尚為一致，委員 1 (實務專家) 評閱之平均成績 20.35 分，略高於委員 2 (學者) 之平均成績 18.63 分，惟差異不大，依各項統計指標顯示兩位委員間並無顯著寬嚴不一情形 (參閱圖 2-1)。
2. **兩閱相關性**：從本題兩閱分數之線性相關程度觀察，相關係數達 0.7731，為高度正相關，表示 2 位委員給高分或低分方向一致，即兩位委員之見解趨於一致。
3. **兩閱一致性**：以同份試卷給分情形統計，兩閱分數差距均在 15 分以內 (題分三分之一內，參閱圖 2-2 綠線範圍)，無啟動第三閱情形。將 1,997 份試卷兩閱分數差距，再以 5 分 (15 分除以 3) 為級距，兩閱分數相同 252 人 (占 12.62%、完全一致)、相差 1 至 5 分 1,493 人 (占 74.76%)、相差 6 至 10 分 239 人 (占 13.97%)、相差 11 至 15 分僅 21 人 (占 0.65%)。由於申論題給分範圍大，委員對同份試卷給分不同實屬常態，2 位委員評閱之所有試卷八成五以上分

數差距均在 5 分內，顯示兩閱評閱結果一致性程度高。

4. 綜上，兩閱評閱結果無顯著差異，實務專家或學者背景寬嚴不一情形甚微，屬合理範圍。

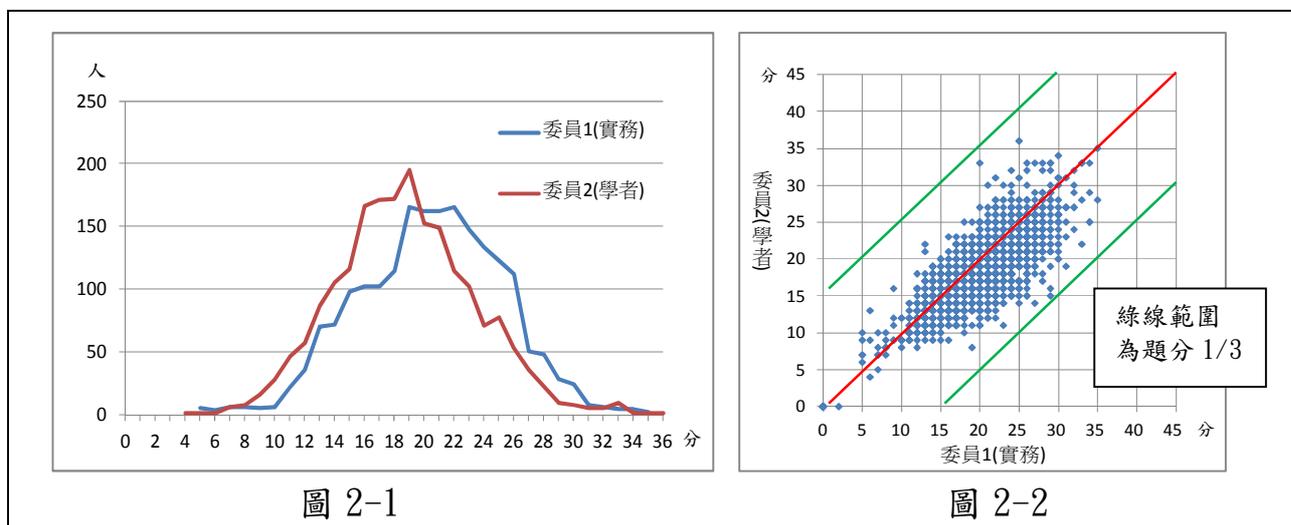


圖 2、102 年司法官考試「民法與民事訴訟法」第一題成績圖

(三)102 年律師考試「民法及民事訴訟法」第五題(平行兩閱、配分 35 分、2,687 份試卷，參閱附表 3)

1. 兩閱寬嚴標準：本題 2 位委員評閱之平均成績及標準差(給分離散程度)度均接近，意即個人對整體應考人程度區別之標準尚為一致，2 位委員間寬嚴標準接近 (參閱圖 3-1)。
2. 兩閱相關性：從本題兩閱分數之線性相關程度觀察，相關係數 0.6431，為中度正相關，顯示 2 位委員之見解稍有差異。
3. 兩閱一致性：以同份試卷給分情形統計，兩閱分數差距均在 12 分以內 (題分三分之一內，參閱圖 2-2 綠線範圍)，無啟動第三閱情形。將 2,687 份試卷兩閱成績差距，再以 4 分(12 分除以 3)為級距，兩閱分數相同 275 人 (占 10.23%、完全一致)、相差 1 至 4 分有 1,255 人 (占 46.71%)、相差 5 至 8 分有

919 人 (占 34.20%)、相差 9 至 12 分有 238 人 (占 8.86%)。2 位委員評閱之所有試卷近六成分數差距在 4 分內，超過四成分數差距在 5 至 12 分，顯示 2 位委員見解稍有不同。

4. 綜上，2 位閱卷委員存在見解不同造成給分差異，此即平行兩閱為降低試卷評閱受主觀因素影響之效果，但各自評閱之寬嚴標準接近 (參閱圖 3-1)，顯示實務專家或學者背景寬嚴不一情形甚微，屬合理範圍。

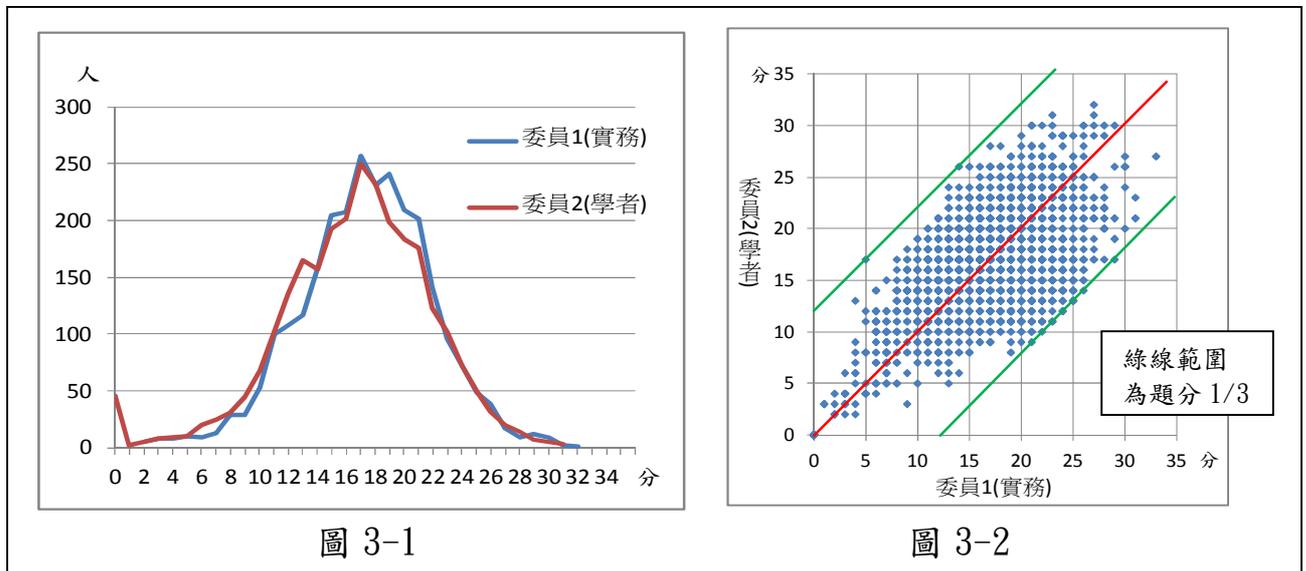


圖 3、102 年律師考試「民法及民事訴訟法」第五題成績圖

(四)102 年建築師考試「建築計畫與設計」(平行兩閱、配分 100 分、2,835 份試卷，參閱附件四)

1. 兩閱寬嚴標準：本應試科目平行兩閱之兩位委員評閱結果平均成績差異不大、標準差(給分離散程度)接近，又由於建築師考試為科別及格制，各科目以 60 分及格標準，兩閱 60 分及以上(及格線以上)各為 369 份及 370 份，顯示 2 位委員對整體應考人程度區別之標準一致，且寬嚴標準接近(參閱圖 4-1)。
2. 兩閱相關性：從兩閱分數之線性相關程度觀察，相關係數 0.8782，為高度正相關，顯示 2 位委員之見解

一致程度高。

3. 兩閱一致性：以同份試卷給分情形統計，同時評閱為60分(含)以上以及同時給60分以下高達2,830份試卷，僅有5份試卷不同，2位委員評閱結果之一致性高（參閱圖4-2）。
4. 綜上，兩位閱卷委員寬嚴標準接近、見解一致，實務專家與學者背景並無影響其寬嚴程度。

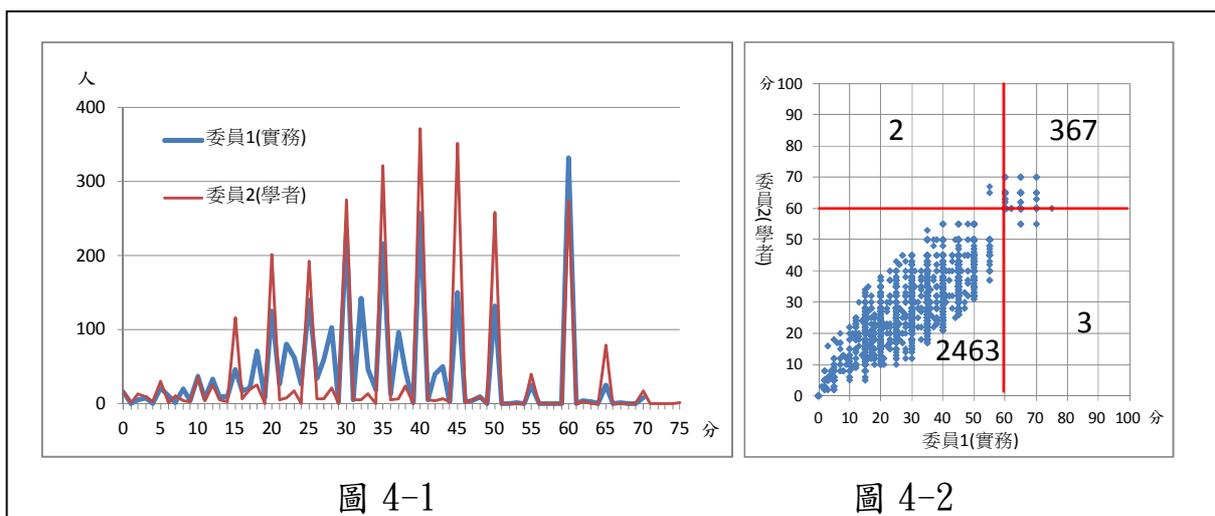


圖 4、102 年建築師考試「建築計畫與設計」成績圖

#### 四、綜合結語

依本報告四項應試科目評閱結果資料分析，不論是單閱之應試科目由多位委員評閱，或是平行兩閱之應試科目，委員間之評閱結果存在分數差異，其差異來源主要為委員間之評閱寬嚴不一、見解不同（主觀因素），或是應考人程度不同等因素，但委員身分別並非影響分數差異之主要因素，即由實務界專家與學者擔任閱卷委員，並無評閱寬嚴顯著不一情形。

為提昇閱卷品質，未來精進方向可考量以下幾點：

- (一) 考量調整分派試卷方式：單閱科目多位委員間之評閱分數差異來源有可能來自於委員間之評閱寬嚴不一、見解不同，或是評閱範圍應考人程度不同等因

素。依目前紙本閱卷作業，委員評閱範圍採依考區順序卷包分配，致無法排除不同考區應考人程度是否影響委員評閱分數差異之原因，未來當有多位委員評閱時，可考量採隨機分派試卷方式以排除應考人程度不同因素。

(二) **賡續強化兩閱溝通機制**：平行兩閱間分數差異有可能來自兩閱委員之評閱寬嚴不一或見解不同，依目前平行兩閱科目評閱，均經兩閱委員於評閱會議、試評及閱卷過程中意見溝通，以達評閱標準及寬嚴一致，並藉由兩閱分數超過題分三分之一啟動第三閱機制，以降低兩閱見解不同之分數落差。因此，依本報告平行兩閱評閱結果亦顯示，兩閱評閱之寬嚴程度差不多，兩閱分數之相關性與一致性程度均高，表示考試結果之信度高。

(三) **賡續增加研究範圍，以利更正確推論**：由於本報告僅分析四項應試科目資料，尚無法推論本部辦理之所有科目情形。國家考試評閱工作首重公平性，為使評閱標準一致，避免考試成績受個別委員評閱寬嚴不一因素影響，允宜針對平行兩閱及多位閱卷委員之應試科目，於評閱會議時即請委員討論訂定相同之給分標準，於閱卷期間請召集人抽閱，並建立查核機制，如有特殊差異俾及時反映改善。

附表 1、102 年一般警察特考四等考試行政警察類別之「刑法概要」  
統計表

1-1. 各項統計指標 (配分 100 分)

委員	份數	平均 成績	成績 標準差	及格線(60分)以上		最高 成績
				人數	%	
委員 1(實務)	748	43.18	18.41	170	22.73%	73
委員 2(實務)	724	39.86	15.79	78	10.77%	76
委員 3(實務)	733	36.45	18.79	84	11.46%	80
委員 4(實務)	747	34.11	16.96	49	6.56%	77
委員 5(實務)	741	33.10	18.53	61	8.23%	75
委員 6(實務)	755	28.87	13.67	13	1.72%	68
委員 7(實務)	769	28.41	19.99	58	7.54%	82
委員 8(學者)	761	48.86	22.54	285	37.45%	92
委員 9(學者)	753	39.78	16.49	64	8.50%	77
委員 10(學者)	767	36.14	17.84	70	9.13%	74
委員 11(學者)	724	29.18	16.09	21	2.90%	78

1-2. 變異數分析

	實務專家組	學者組	合計
樣本數	7	4	11
平均數	34.8543	38.4900	36.1764

ANOVA TABLE

變異來源	平方和	自由度	均方和	F 檢定	P 值
兩組委員間	33.6470	1	33.6470	0.7984	0.3948
組內委員間	379.2926	9	42.1436		
合計	412.9396	10			

附表 2、102 年司法官考試「民法與民事訴訟法第一題」統計表

2-1. 各項統計指標(題分 45 分)

委員	份數	平均 成績	成績 標準差	及格線(26 分)以上		最高 成績	偏態 係數
				人數	%		
委員 1(實務)	1997	20.35	4.97	173	8.66%	35	-0.39
委員 2(學者)	1997	18.63	4.78	95	4.76%	36	-0.01

2-2. 檢定及相關係數

檢定項目	平均成績	成績標準差	相關係數
檢定方法	Z-test	F-test	0.7731
檢定統計量	1.553878	1.039749	
P-value	0.060107	0.191988	

2-3. 一致性

兩閱分數差距	試卷份數	%
0 分	252	12.62%
1-5 分	1493	74.76%
6-10 分	239	13.97%
11-15 分	21	0.65%
合計	1997	100.00%

附表 3、102 年律師考試「民法及民事訴訟法第五題」統計表

3-1. 各項統計指標(題分 35 分)

委員	份數	平均 成績	成績 標準差	及格線(21 分)以上		最高 成績	偏態 係數
				人數	%		
委員 1(實務)	2687	17.10	5.03	305	11.35%	32	-0.65
委員 2(學者)	2687	16.65	5.19	305	11.35%	33	-0.52

3-2. 檢定及相關係數

檢定項目	平均成績	成績標準差	相關係數
檢定方法	Z-test	F-test	0.6431
檢定統計量	0.368703	0.969171	
P-value	0.356175	0.208582	

3-3. 一致性

兩閱分數差距	試卷份數	%
0 分	275	10.23%
1-4 分	1255	46.71%
5-8 分	919	34.20%
9-12 分	238	8.86%
合計	2687	100.00%

附表 4、102 年建築師考試「建築計畫與設計」統計表

4-1. 各項統計指標(配分 100 分)

委員	份數	平均 成績	成績 標準差	及格線(60分)以上		最高 成績	偏態 係數	
				人數	%			
總分	委員 1 (實務)	2835	34.82	14.33	369	13.02%	70	0.23
	委員 2 (學者)	2835	37.04	14.94	370	13.05%	75	-0.09

4-2. 檢定及相關係數

檢定項目	平均成績	成績標準差	相關係數
檢定方法	Z-test	F-test	0.8782
檢定統計量	-0.90812	0.95917	
P-value	0.18191	0.13361	

4-3. 一致性

試卷份數		委員 2(學者)		
		不及格	及格	合計
委員 1 (實務)	及格	2	367	369
	不及格	2463	3	2466
	合計	2465	370	2835