

## 研究論文

# 李克式量表標示語之 心理量尺值研究\*

游琇婷\*\*

### 摘要

社會科學和心理學研究經常以問卷或量表方式收集受試者在各題目的反應資料。李克式量尺則為此類研究收集資料時問卷或是量表常用的方法，而量表標示語的心理量尺值在資料的分析和解讀上有重要意涵，但標示語所真實表徵的心理量尺值有不同的估計方法，也未有明確公認的結論。過往研究曾以連續區間法、簡化連續區間法、歸等測驗和自由反應測驗等方法探討量尺標示語由不同估計法所得之心理量尺值的差異。本研究採用配對比較評量、單一直接評量，以及同時排序評量等三種方法對同意、符合及頻率三類標示語進行量尺值估計。本研究針對各個度量方法所估計之心理量尺值進行分析比較，並探究量表標示語之間的相對關係、對稱性，以及語詞結構對量尺估計值的影響。文中並綜合討論三個量尺值估計法所得結果之差異及其影響，文末提出本研究結果於實徵應用研究的可能貢獻，並對實徵應用研究使用李克式量尺時在標示語選擇上的具體建議。

關鍵詞：李克氏量尺、量表標示語、心理量尺值、調查研究

\* 本研究獲科技部 MOST 108-2410-H-004-100 及 MOST 109-2410-H-004-075-MY2 計畫部分支持。本研究感謝政大心理系尤譯靈、蘇勤、陳伊柔、鄧祈允同學以及政大心理計量與量化方法研究室成員於資料收集及整理之協助

\*\* 政治大學心理學系／心智、大腦與學習研究中心。臺北市文山區指南路二段 64 號（國立政治大學心理學系）。電話：(02)2938-7384，E-mail: htayu@nccu.edu.tw。

## **Estimating Scale Values of Response Anchors in Likert-type Scales**

**Hsiu-Ting Yu\***

### **ABSTRACT**

Social science and psychological research often uses questionnaires or scales to collect data. Likert-type scales are the most commonly used scales in questionnaires or scales when collecting data. The psychological scale values of the scale labels/descriptors/anchors have important implications in the analysis and interpretation of data. Different methods and formats have been proposed for estimating the scale values of a Likert-type scale. Previous studies have used successive interval and simplified successive interval estimation methods. Studies have also used an open-response format and category-level assigning format to study the scale values. However, there is no consensus and still a lack of studies comparing different estimation methods. In this study, three estimation methods of label scale values are investigated: the paired-comparison method, direct rating method, and simultaneous ranking evaluation method. Three common types of scale labels are investigated in this study: agreement-, conformity-, and frequency-type. A review on the usages of Likert scale labels in literature was conducted, 24 agreement-type labels, 24 conformity-type labels, and 20 frequency-type labels were included in this study. There were 97 subjects (27 males and 70 females) who participated in this study. The mean age of the subjects is 20.68 years old with a SD of 1.54. The subjects took part in lab experiments with three tasks. In the direct rating task, subjects used a mouse to drag a marker on a horizontal bar to indicate the rating of a given label stimulus. In the pair-comparison task, subjects used the mouse to drag a marker on a horizontal bar to

---

\* Department of Psychology/Research Center for Mind, Brain & Learning, National Cheng-chi University. Email: htyu@nccu.edu.tw.

indicate the degree of difference of a given pair of labels. In the simultaneous ranking evaluation task, subjects were presented with all the labels on the top panel of the screen and the subjects' task was to drag these labels in any order they preferred to any location on the horizontal bar located in the center of the screen. Subjects could adjust the location of any label on the horizontal bar before they clicked the "confirm" button to submit their answer. Subjects' responses to the three tasks were recorded on a 0-1000 scale for further analysis. The collected data were analyzed and compared the psychological scale values estimated by the three estimation methods. The results suggested the equal space assumption between points on the Likert scale is in general not hold based on the estimated scale values of this study. The results also suggested substantial differences among the three estimation methods. Moreover, the estimated scale values of the selected labels were also not evenly distributed on the evaluation scale; more labels were loaded toward the extreme portion of the scale. The possible explanations for this observation are discussed in the discussion section. Since it is difficult to locate appropriate labels based on the estimated psychological scale values on a Likert-type scale, we suggested using only the "end-label" on a Likert scale in practice, for example, "completely disagree-completely agree" and "never-always" as the labels on a Likert-type scale. The limitations of the present study were presented in the discussion section as well as references for further reading on these topics. Other psychometric methods such as item-response theory (IRT) and Fuzzy Linguistic/numbers approach for dealing with Likert-type scale labels were also discussed in the concluding remarks.

**Keywords:** Likert-type scale, scale labels, psychological scale values, survey research

---

## 一、緒論

李克式量尺 (Likert-type scales) 為調查研究中廣為使用的收資收集的方法。此方法由 Likert 於 1932 提出，因其簡易直覺的特性受到研究者廣泛使用，尤其常使用於關於態度的測量。李克式量尺的使用通常為請受試者評量給定的陳述句，而受試者則在一有特定標示語的量尺上標示（勾選、圈選或劃記）適當的程度。李克式量尺也需由研究者依適合準則選定量尺點數以及量尺上的標示語 (labels, scale descriptors)。常見的李克式量尺之量尺點數型態有包含中間值的五點、七點或九點量尺，以及不包含中間值，如四點、六點、及八點量尺。研究者通常以研究情境或是適合包含中間值與否等考量，來選擇量尺點數和型態。

除了量尺點數和量尺的呈現方式考量外，使用李克式量尺另一個重要面向就是量尺標示語的選擇。標示語通常標示於量尺上方，其作用提供受試者評定的參考基準。而量尺的標示語和評定量尺的呈現方式也有不同的選擇：（1）每個量尺數值皆有所對應到的量尺標示語、（2）僅標示兩個極端值及其對應之量表數值、（3）僅標示兩個極端值及中間選項及對應之量表數值，以及（4）僅標示部分量表標示語和對應之分數。

研究者在分析以李克式量尺施測所得的量表分數或總分時，通常直接加總個別單一題目分數，這樣的計算是建立在測量尺度為等距量尺 (interval scale) 的前提假設之上。等距量尺的意義認定數值（分數）間相同的差距不會因為其絕對數值大小而不同；例如，數值標示 1-2 間的差距意涵和數值標示 4-5 間差距相同。然而，此一等距量尺的假

設在實徵研究中並不一定成立，資料很可能僅符合較寬鬆定義的順序量尺（ordinal scale）而不具等距量尺的特性。資料違反等距量尺的假設時，研究中常用的描述統計量（如平均值、總分）的使用和解讀都不具實質意義。分數的加減運算（例如：加總、平均）需在資料符合等距量尺的情況下才有意義，也才能比較受試者內或受試者間的分數差異。資料違反等距量尺的假設，使得能使用統計分析方法大為受限，經典的變異數分析、迴歸分析或是潛在變項模型等分析方法的使用都受到限制，可能需以進階的資料處理方式或採其他的統計方法做後續處理和分析。

為使李克式量尺的使用符合等距量尺的假設，李克式量尺中所使用的量尺標示語具重要角色。心理計量中的度量化方法（scaling method）之目的為估計每個刺激項在受試者心中的心理量尺的位置，為使李克式量尺的使用能符合等距量尺，需進一步瞭解量尺中標示語的心理量尺意義，以利後續研究使用標示語之參考。唯有準確估計每個量尺標示語的心理量尺值才能最佳使用此量尺所測得的分數或結果。如選用適當的量尺標示語以及量尺點數於評量尺規上，受試者在對題目作答時能有較符合等距量尺的尺規參照點，其結果分數也應會較符合等距量尺的特性。

過去研究中已就李克式量表點數數量影響（Matell and Jacoby 1972; Ooster 1989; Spector 1976）、量表點數與量表信效度之間關係（Masters 1974; Weng 2004）、作答格式（Newstead and Arnold 1989），和對此量尺在標示語使用及標示位置（Klockars and Yamagishi 1988; Low et al. 1991; Myers and Warner 1968; Wildt and Mazis 1978）進行探究。但上列多數並非針對中文的標示語進行研究。

目前對於中文標示語量尺值較完整的研究為翁麗禎（1998）對同

意度和頻率兩大類心理量尺值研究，其中同意度用語再細分成同意、符合及贊成詞三類。翁麗禎（1998）針對 41 個頻率詞、42 個同意度詞、67 個符合度詞及 35 個贊成度詞以 Edwardz（1957）提出的連續區間法（successive intervals, SI）及 Edwards 與 Gonzales（1993）的簡化連續區間法（simplified successive intervals, SSI）計算標示語的心理量尺值。其結果提供計算連續區間法（SI）、簡化連續區間法（SSI）、平均數和中位數於四類標示語心理量尺值；也以區辨離散值（discriminal dispersion）、標準差與四分位差三個統計量描述所包含標示語心理量尺值之離散程度。翁麗禎（1998）以連續區間法估計量尺值避免了等距區間法對於每個區段間距離相等的假設，求算各區值間的寬度以計算標示語的量尺值。連續區間法和簡化連續區間法的優點在於並未假設等距量尺的特性，但兩種方法仍保留資料為常態分配之假設，意即唯有資料本身呈現常態分布下才可以正確估計標示語的心理量尺值。因此，此二方法的量尺估計值仍需要實徵資料具有常態性分配的前提假設。然而資料違反常態性假設時，以對資料分配無特定假定配對比較法來估計量尺值較佳。過往研究雖建議以配對比較法所估算的心理量尺值作為不同量尺值估計方法的比較基準，目前仍然缺乏以配對比較法進行之中文心理量尺值估計研究。

以配對比較法估計標示語心理量尺值時，無需設立資料量尺等距和常態分配兩前提假設。配對比較法的進行方式通常為一次呈現兩個刺激給受測者，受測者評定兩者刺激之間的相異性（或相似性），整合所有配對的相異性評量可得出整體全部刺激兩兩配對的相異度距離矩陣，再由 Torgerson（1952）提出的典型度量化方法（classic scaling），也稱 Torgerson Scaling 或 Torgerson-Gower Scaling，對相異度距離矩陣進行分解，其後由所得之特徵值和特徵向量尋找特定數目維度下之

座標，再轉換為標示語的心理量尺值。配對比較法雖然不受區間等距及常態性假設所限制，但因該方法需將所有標示語兩兩配對評量，故需要評量的配對數隨著標示語數量增加而快速大幅上升。過去對於中文量尺標示語的研究也常因此一限制，而未以配對比較的方法來進行量尺值的估計（翁麗禎 1998）。

除以單一直接評量和兩兩配對評量的量尺值估計法外，本研究也增加同時排序評量的估計法。翁麗禎（1998）研究中所採用的連續區間法以及簡化連續區間法是由受試者將標示語分類（*sorting*）於十或十一個由低至高程度的類別中，因為類別是有程度的高低分別，這樣的分類也可視為排序（*ranking*）；本研究由電腦輔助施測，同時排序評量可以將所有同類的標示語同時呈現於螢幕上，受試者可以任何順序選取標示語於標準量尺上進行評量。因同時排序作業中受試者可同時考量標示語的位置及其相互相對位置，因此受試者在此作業下較其他兩個作業更有彈性，對於標示語的心理量尺估計可能異於其他兩種作業，加入這一個估計法能對應比較不同估計法的異同，以增進對於標示語心理量尺值的了解。

頻率類標示語在量尺使用與代表意涵和同意與符合類標示語很不同。相當大量的研究探討機率詞或是頻率詞的心理意涵（例如：Aarts and Dijksterhuis 1999; Blair and Burton 1987; Gaskell et al. 1994; Hakel 1968; Pepper and Prytulak 1974; Wallsten et al. 1986），但多數都以英文詞研究為主，中文機率詞的經典研究則是胡志偉、彭昭英、沈永正和楊金龍（1989）提出。胡志偉等人（1989）研究以「自由反應測驗」（Hakel 1968）與「歸等測驗」（Foley 1959）方式探討 29 個常用的機率詞在受試者心中所代表的機率。其研究結果認為機率詞於受試者心中的意義不同，也認為機率詞所代表的意義為一段機率區間而非一個

點，不同機率詞也會有區間部分重疊的現象；並建議以五個不同的機率詞（絕對、通常、或許、不太可能、絕不）來表示從 0 到 1 之間的五個機率區間以降低不確定性。然而胡志偉等人（1989）的研究中所選取的 29 個機率詞為較屬於日常口語溝通用語，例如：搞不好、很難說、似乎、大概、要不然等。本研究所設定的標示語需能適合於李克式量尺的研究和應用上，因此過往的研究結果仍有不足。

標示語量尺值的研究對於能適切使用李克式量尺所得結果具重要角色，雖已有部分中文研究論文探討此問題，但仍缺乏對不同的量尺值估計法間之系統比較。另外，以使用李克式量尺的量表中也常包含反向題的設計，而反向題在計分需以反向計分；李克式量尺中反向計分的影響已有文獻進行討論（例如：Józsa and Morgan 2017）但未有具體定論；本研究也在量尺對稱性上對所包含的三類型的量尺標示語加以分析和討論，其對稱性的分析能增加對於反向計分或是量尺標示語文字使用上影響的了解。本研究目的為針對實徵研究中李克式量尺三類常用的標示類別（同意、符合和頻率），以三種不同標示語的心理量尺值估計法進行探究也比較其差異；並由研究結果提出對於標示語於實徵研究上的使用作出建議。

## 二、研究方法

本研究以配對比較評量、單一直接評量和同時排序評量等三種作業對李克式量尺上的同意、符合及頻率類標示語，探討其心理量尺值，並比較不同方法下量尺估計值的差異。以下說明研究參與者、研究刺激材料及實驗進行流程。



## （一）研究參與者

參與本研究受試者共計 97 位，其性別分布為 27 名男性及 70 名女性。全部受試者皆為以中文母語使用者。而受試者年齡平均數為 20.68，標準差為 1.54。本研究通過國立政治大學倫理審查審核（NCCU-REC-201812-1084），受試者皆簽署知情同意書同意參與本實驗。

## （二）研究材料

本研究參考翁麗禎（1998）量尺值研究刺激材料，也另蒐集整理近十年（2008 年至 2017 年）《中華心理學刊》中論文使用的李克式量表中之量表標示語共 245 篇，其中有 92 篇論文使用李克氏量表作為研究資料之收集工具，共計有 279 個量表使用李克式量尺。其中有 54 個量表（19.4%）的量尺標示語資訊不清或未提供，而「符合詞」標示語計 55 量表（19.7%），「同意詞」標示語計 54 個量表（19.4%），「頻率詞」標示語計 42 個量表（15.1%）為前三大標示語類型。除上述三類之外的各類樣式量尺標示語的則有 74 個量表（26.5%）。其中以「贊成詞」作為標示語僅有同一篇論文的兩個量表。因此本研究僅包含「同意詞」、「符合詞」以及「頻率詞」三類標示語的量尺值研究，並未包含翁麗禎（1998）中的「贊成詞」一類。所收集的標示語整理為同意、符合和頻率三個類別及其使用頻率。在考量標示語之實際使用情況及頻率後，選定三類的標示語進行研究，分別為 24 個同意詞、24 個符合詞及 20 個頻率詞標示語。本研究包含之三類量表標示語及其編號整理可見於表 1 至表 3 中。

表 1 配對比較作業中的各配對相異程度平均值（右上角）與中位數（左下角）

同意詞類	完全不同意	強烈不同意	非常不同意	相當不同意	很不同意	大部分不同意	多半不同意	不同意	不太同意	部分不同意	有點不同意	少部分不同意	少部分同意	還算同意	有點同意	部分同意	同意	多半同意	大部分同意	相當同意	很同意	非常同意	強烈同意	完全同意
完全不同意	0	55	115	176	170	221	208	194	308	496	485	583	502	747	765	780	859	864	856	879	904	951	938	979
強烈不同意	31	0	71	144	126	205	330	249	272	527	449	565	640	763	808	783	861	875	836	902	950	900	939	965
非常不同意	44	36	0	73	145	215	431	227	441	469	443	569	628	830	768	709	806	900	907	851	932	993	918	979
相當不同意	121	118	48	0	134	218	151	255	367	549	389	493	480	717	739	674	790	866	735	850	924	856	905	936
很不同意	95	86	66	112	0	242	296	208	252	395	313	561	584	757	763	713	752	807	823	910	837	934	881	853
大部分不同意	171	126	149	148	175	0	157	238	254	293	459	630	352	699	664	567	770	829	847	773	823	806	882	847
多半不同意	171	256	364	117	204	105	0	278	210	465	406	580	264	672	687	512	668	791	768	822	804	796	828	868
不同意	129	177	134	182	190	161	208	0	276	315	284	427	433	665	724	641	773	879	805	867	880	867	848	913
不太同意	221	213	366	307	219	222	184	245	0	326	146	440	335	635	599	542	640	675	721	747	744	802	797	800
部分不同意	446	491	395	605	303	241	493	271	255	0	237	273	475	465	521	500	371	642	320	527	615	566	655	668
有點不同意	390	379	367	296	273	383	271	209	136	210	0	253	435	474	651	337	539	635	469	685	657	719	820	696
少部分不同意	565	514	661	440	560	691	623	363	438	224	221	0	741	311	359	409	363	418	244	298	475	301	555	413
少部分同意	559	672	627	618	644	294	160	430	322	512	395	713	0	334	194	449	532	494	665	646	583	719	663	603
還算同意	766	738	840	704	774	731	749	659	660	499	508	239	326	0	152	238	182	257	244	306	368	284	378	350
有點同意	768	809	801	737	749	733	715	736	648	565	690	294	182	105	0	218	215	328	332	336	335	413	358	411
部分同意	798	783	720	709	702	575	511	652	628	525	297	387	420	230	129	0	359	389	414	321	447	433	397	331
同意	859	878	821	825	800	819	726	788	693	285	631	267	535	143	185	309	0	223	183	266	219	250	347	239
多半同意	861	866	913	866	810	844	841	998	709	662	661	423	497	209	287	314	222	0	204	238	165	194	240	150
大部分同意	904	830	891	766	864	924	822	851	761	249	517	155	685	183	267	427	145	160	0	346	224	348	148	354
相當同意	868	957	879	871	919	802	843	884	799	639	717	262	759	244	260	263	173	146	211	0	108	99	192	176
很同意	930	965	964	954	944	874	837	935	755	668	664	554	615	313	306	467	170	151	175	87	0	132	115	126
非常同意	979	984	1,000	910	969	836	840	920	844	610	761	283	767	239	353	367	218	106	229	48	112	0	106	109
強烈同意	997	1,000	977	978	952	892	823	898	889	718	804	602	742	287	273	391	217	145	129	124	72	30	0	94
完全同意	1,000	994	999	961	920	885	867	970	777	751	751	364	717	271	323	314	194	122	363	110	113	63	10	0

表 1 配對比較作業中的各配對相異程度平均值 (右上角) 與中位數 (左下角) (續)

符合詞類	完全不符合	強烈不符合	非常不符合	相當不符合	大部分不符合	多半不符合	不符合	不太符合	部分不符合	有點不符合	少部分不符合	少部分符合	有點符合	部分符合	還算符合	符合	多半符合	大部分符合	很符合	相當符合	非常符合	強烈符合	完全符合	
完全不符合	0	72	55	163	196	222	305	184	300	519	519	433	603	756	772	762	829	790	817	945	853	941	970	978
強烈不符合	15	0	99	132	133	163	283	197	371	404	441	620	639	729	726	758	836	880	787	939	915	963	965	973
非常不符合	50	17	0	125	93	221	284	226	433	395	465	487	507	774	735	742	860	805	858	882	896	964	963	958
相當不符合	136	92	47	0	161	257	165	224	324	532	404	484	482	714	625	773	883	733	827	861	857	866	933	771
很不符合	154	107	74	109	0	223	248	241	283	425	420	464	608	739	743	653	877	741	812	798	928	906	912	864
大部分不符合	161	114	190	181	180	0	194	180	172	395	503	227	577	572	623	663	746	776	818	855	827	818	857	853
多半不符合	279	162	233	118	172	120	0	260	206	425	436	302	550	566	552	630	814	703	694	765	813	758	759	825
不符合	153	196	191	173	175	156	229	0	274	374	239	452	448	673	681	646	937	787	739	854	808	854	858	942
不太符合	237	279	348	261	215	179	189	244	0	362	148	320	426	537	561	540	730	665	688	765	768	802	882	799
部分不符合	523	349	326	659	341	313	341	344	282	0	240	459	327	460	544	354	592	462	411	627	573	596	703	669
有點不符合	569	406	415	362	364	583	372	234	134	252	0	433	239	536	359	477	600	495	473	707	656	772	729	644
少部分不符合	367	637	491	501	382	147	202	449	291	514	357	0	651	353	418	393	481	724	656	640	642	638	653	541
少部分符合	660	762	490	401	677	655	554	355	351	303	150	716	0	348	344	253	378	340	242	408	378	423	579	435
有點符合	761	725	773	734	714	608	651	673	581	463	525	310	275	0	170	106	385	303	347	327	373	390	435	494
部分符合	789	780	748	680	752	693	595	719	615	619	380	372	268	154	0	167	365	290	282	315	343	418	459	392
還算符合	803	804	796	837	679	714	694	669	573	265	510	361	164	60	141	0	247	167	219	394	361	383	332	364
符合	857	884	885	907	903	790	819	997	782	641	621	458	276	346	326	224	0	304	230	203	162	213	215	195
多半符合	846	889	835	779	758	772	770	784	733	416	562	751	250	247	261	100	261	0	139	303	341	226	339	316
大部分符合	856	832	901	863	861	869	769	787	740	405	514	737	107	291	232	168	189	53	0	186	267	311	171	330
很符合	966	966	929	895	855	872	804	879	776	648	740	667	362	282	247	366	164	211	129	0	185	101	145	148
相當符合	862	935	899	892	949	863	853	839	782	606	702	691	298	325	300	318	144	249	134	173	0	87	104	102
非常符合	971	983	996	905	931	849	776	876	822	682	750	732	309	304	381	382	169	135	191	83	32	0	54	88
強烈符合	993	1,000	983	953	961	899	823	892	902	736	789	699	670	383	431	296	172	322	103	113	58	34	0	134
完全符合	1,000	989	993	870	937	900	874	965	811	711	652	510	338	429	397	326	181	242	216	86	71	34	28	0

表 1 配對比較作業中的各配對相異程度平均值 (右上角) 與中位數 (左下角) (續)

頻率詞類	從未	從不	不會	完全沒有	從來沒有	幾乎沒有	很少	不常	偶爾	有時	時常	常常	經常	很常	大部分時間	大多時間	幾乎總是	一直	總是	每次都
從未	0	136	86	10	10	294	249	337	665	782	732	858	845	893	873	877	916	918	915	959
從不	1	0	73	30	10	215	160	465	725	714	879	821	858	861	873	874	903	944	914	966
不會	2	17	0	250	30	150	262	593	630	753	856	808	879	881	898	856	891	919	929	901
完全沒有	0	1	36	0	76	153	292	498	724	836	857	921	919	858	892	866	900	987	961	918
從來沒有	0	0	24	4	0	54	341	395	773	737	877	880	826	872	872	888	938	933	952	876
幾乎沒有	206	199	153	76	65	0	115	248	672	701	841	894	828	823	879	870	912	900	909	916
很少	160	191	149	179	201	90	0	156	316	487	733	839	768	788	832	801	890	810	858	865
不常	236	559	668	487	282	207	165	0	260	510	773	711	818	777	834	845	741	726	768	846
偶爾	665	730	698	764	737	722	228	205	0	217	391	384	333	511	569	484	573	596	447	691
有時	751	668	719	845	704	783	547	565	127	0	218	305	507	373	333	464	408	374	396	455
時常	752	872	852	877	887	867	790	775	367	208	0	56	131	223	236	111	223	227	230	343
常常	901	886	816	942	880	909	856	684	302	326	40	0	95	127	295	133	146	160	123	189
經常	857	854	858	935	892	860	745	886	242	617	58	66	0	88	179	122	254	161	119	217
很常	910	900	896	949	865	831	834	763	572	344	152	100	52	0	131	85	140	193	155	120
大部分時間	902	866	960	885	892	910	818	856	639	267	155	229	133	88	0	114	139	322	225	203
大多時間	933	915	846	890	903	913	781	850	536	525	109	112	83	93	12	0	56	294	255	192
幾乎總是	956	961	945	972	975	952	935	723	674	299	165	128	152	72	154	66	0	100	109	85
一直	991	964	955	998	958	924	935	725	647	334	162	113	181	172	211	192	48	0	87	138
總是	999	973	942	991	1,000	953	913	774	435	341	220	123	122	150	146	195	92	34	0	30
每次都	971	996	967	994	984	916	869	858	634	378	243	175	128	120	129	178	80	62	5	0

註：表中標示說明如下：數值<200 為灰字，201<400 為黑字，401<600 為淺灰底色黑字，601<800 為中灰底色黑字，801<1000 為深灰底色黑字。

表 2 單一直接評定之標示語量尺估計值

	Mean	SD	Med	Q		Mean	SD	Med	Q		Mean	SD	Med	Q
完全不同意	33.5	(145.3)	0	(10)	完全不符合	30.9	(127.4)	0	(6)	從未	30.5	(126.1)	0	(3)
強烈不同意	32.9	(117.5)	6	(26)	強烈不符合	41.9	(132.1)	5	(32)	從不	42.5	(134.4)	0	(8)
非常不同意	52.7	(149.5)	10	(50)	非常不符合	64.6	(148.5)	17	(79)	不會	43.8	(134.6)	0	(11)
相當不同意	116.2	(126.2)	90	(129)	相當不符合	120.1	(135.2)	104	(134)	完全沒有	49.7	(165.5)	0	(6)
很不同意	127.1	(113.8)	113	(99)	很不符合	141.9	(123.1)	124	(126)	從來沒有	63.3	(208.4)	0	(5)
大部分不同意	181.1	(141.5)	161	(121)	大部分不符合	159.8	(115.3)	139	(105)	幾乎沒有	100.6	(158.5)	62	(64)
多半不同意	217.1	(126.8)	196	(135)	多半不符合	213.1	(126.2)	195	(93)	很少	203.0	(167.6)	151	(129)
不同意	219.9	(141.4)	208	(126)	不符合	217.2	(137.7)	205	(150)	不常	261.1	(149.5)	240	(125)
不太同意	314.5	(109.7)	328	(160)	不太符合	311.4	(140.7)	322	(164)	偶爾	531.6	(203.9)	596	(257)
部分不同意	364.9	(130.9)	333	(144)	部分不符合	358.4	(138.3)	324	(122)	有時	650.0	(109.5)	651	(118)
有點不同意	371.3	(107.7)	373	(99)	有點不符合	379.2	(111.4)	372	(103)	時常	765.3	(120.9)	777	(117)
少部分不同意	551.0	(216.2)	458	(388)	少部分不符合	436.1	(197.5)	529	(356)	常常	793.1	(85.4)	801	(113)
少部分同意	456.7	(188.7)	545	(269)	少部分符合	523.4	(216.2)	449	(358)	經常	800.4	(103.5)	817	(109)
還算同意	629.5	(88.4)	635	(104)	有點符合	602.3	(109.3)	602	(89)	很常	813.2	(105.4)	828	(134)
有點同意	636.1	(102.0)	644	(102)	部分符合	613.5	(145.5)	649	(124)	大部分時間	826.6	(139.0)	857	(106)
部分同意	641.7	(121.8)	660	(117)	還算符合	636.5	(95.3)	629	(129)	大多時間	840.4	(96.3)	849	(102)
同意	749.6	(143.5)	780	(126)	符合	722.2	(150.3)	734	(167)	幾乎總是	884.9	(198.3)	951	(65)
多半同意	757.4	(121.1)	754	(140)	多半符合	748.7	(118.9)	764	(115)	一直	930.8	(95.1)	973	(112)
大部分同意	819.9	(103.3)	833	(128)	大部分符合	804.0	(120.2)	824	(138)	總是	940.8	(117.1)	985	(63)
相當同意	845.9	(116.3)	851	(129)	很符合	822.6	(125.8)	847	(113)	每次都	953.5	(87.5)	995	(48)
很同意	859.7	(86.2)	862	(95)	相當符合	838.5	(123.4)	854	(122)					
非常同意	934.7	(118.7)	970	(93)	非常符合	925.8	(124.4)	961	(92)					
強烈同意	973.1	(50.6)	993	(36)	強烈符合	941.4	(156.5)	987	(57)					
完全同意	973.3	(102.2)	1,000	(9)	完全符合	954.6	(171.8)	1,000	(6)					

表 3 同時排序評量之標示語量尺估計值

	Mean	SD	Median	IQR		Mean	SD	Median	IQR		Mean	SD	Median	IQR
完全不同意	19.2	(26.1)	5.2	(31.5)	完全不符合	17.8	(37.7)	3.9	(26.4)	從未	20.9	(44.3)	6.3	(18.7)
強烈不同意	26.2	(32.3)	16.9	(38.3)	強烈不符合	28.5	(42.7)	18.0	(38.7)	從不	26.6	(46.2)	7.9	(28.1)
非常不同意	56.1	(48.9)	45.3	(53.3)	非常不符合	63.2	(51.9)	54.4	(46.9)	不曾	30.6	(46.9)	9.7	(39.7)
相當不同意	139.4	(109.8)	113.2	(100.6)	相當不符合	127.6	(76.7)	123.7	(95.2)	完全沒有	27.4	(61.5)	3.7	(21.7)
很不同意	135.2	(65.6)	124.5	(74.5)	很不符合	137.7	(72.3)	132.2	(76.2)	從來沒有	29.4	(52.7)	5.3	(32.6)
大部分不同意	174.7	(86.5)	158.2	(109.9)	大部分不符合	173.8	(89.1)	158.2	(121.0)	幾乎沒有	96.8	(93.7)	65.2	(73.8)
多半不同意	218.9	(102.3)	207.0	(121.5)	多半不符合	222.7	(103.1)	224.3	(110.0)	很少	227.2	(136.7)	199.1	(168.0)
不同意	220.9	(98.2)	219.0	(114.3)	不符合	240.9	(121.5)	228.5	(123.6)	不常	249.4	(113.9)	241.8	(150.0)
不太同意	334.3	(100.8)	344.2	(153.0)	不太符合	337.7	(128.8)	333.5	(158.9)	偶爾	536.7	(180.3)	571.4	(219.0)
部分不同意	347.6	(119.5)	333.8	(103.3)	部分不符合	355.0	(141.4)	335.3	(122.2)	有時	612.1	(119.6)	616.9	(132.6)
有點不同意	391.9	(90.4)	404.9	(95.3)	有點不符合	414.7	(122.1)	415.3	(93.5)	時常	750.3	(114.1)	770.5	(156.0)
少部分不同意	475.9	(173.8)	428.2	(185.2)	少部分不符合	490.5	(198.6)	569.4	(314.9)	常常	787.6	(105.8)	785.1	(136.5)
少部分同意	501.9	(166.4)	548.9	(174.0)	少部分符合	517.2	(208.3)	433.8	(372.4)	經常	785.5	(108.0)	794.6	(150.0)
還算同意	647.3	(93.0)	637.4	(117.7)	有點符合	627.7	(87.4)	618.6	(102.5)	很常	821.7	(78.3)	826.7	(110.7)
有點同意	629.4	(76.6)	614.8	(97.0)	部分符合	667.5	(94.5)	672.1	(87.3)	大部分時間	833.3	(117.6)	849.9	(115.5)
部分同意	657.0	(98.1)	663.6	(100.2)	還算符合	641.8	(102.2)	630.4	(131.4)	大多時間	839.1	(106.5)	857.8	(105.1)
同意	777.3	(104.8)	785.6	(138.1)	符合	761.3	(123.2)	766.1	(130.7)	幾乎總是	926.5	(98.5)	956.5	(44.8)
多半同意	750.0	(117.7)	774.6	(121.9)	多半符合	765.8	(98.9)	776.0	(115.1)	一直	937.2	(82.2)	967.3	(89.2)
大部分同意	807.9	(144.9)	819.3	(135.7)	大部分符合	828.6	(74.1)	845.1	(102.3)	總是	961.3	(86.8)	993.2	(49.9)
相當同意	866.1	(82.9)	881.1	(107.5)	很符合	851.6	(86.2)	869.4	(93.9)	每次都	956.8	(80.6)	994.4	(45.3)
很同意	857.1	(69.9)	870.6	(91.2)	相當符合	881.2	(89.1)	907.9	(94.9)					
非常同意	942.4	(49.8)	949.3	(51.7)	非常符合	944.5	(63.3)	956.1	(48.4)					
強烈同意	976.1	(35.7)	988.7	(39.4)	強烈符合	958.8	(134.4)	983.4	(30.6)					
完全同意	981.9	(31.9)	998.8	(22.9)	完全符合	979.0	(98.7)	1,000.0	(21.4)					

### （三）研究流程

本研究於實驗室中以電腦施測，受試者在簽署同意研究知情同意書後進行。本研究共有三種評量作業：首先進行配對比較評量，其後進入單一直接評量階段，最後進行同時排序評量作業。三類量表標示語以隨機順序方式呈現，分別進行三種評量作業。每階段作業受試者皆先閱讀研究指導語以瞭解研究流程，正式施測前皆有練習嘗試階段以確保受試者熟悉作答規則和操作介面。為避免疲勞效應影響到資料品質，受試者在階段間皆有休息時間，待休息完畢後繼續研究流程。受試者平均花時 40-50 分鐘完成三階段的實驗作業。

#### 1. 配對比較評量

配對比較階段中三類的詞語分別以成對的方式配對。本研究包含的 24 個同意詞、24 個符合詞以及 20 個頻率詞共有 276、276 和 190 種配對組合。若對刺激詞呈現於左方和右方位置做平衡，三類詞語所有兩兩配對的刺激分別有 552、552 和 380 種組合。因受試者作答的時間限制，本研究將所有標示語的配對組合進行順序以及呈現位置的隨機組合隨機排序，將每類標示語配對組合分成八組，使每位受試者在配對比較階段所需評量的題數約為 186 題。如考量詞語呈現位置平衡共需 8 位受試者完成一組完整刺激的比較配對；如不考量刺激值左右呈現的位置，則需 4 位受試者完成一組完整配對比較。本研究所收集之資料每一組配對刺激之評量結果約為 24 位受試者的作答統計而得。

配對比較階段時，作答畫面的示意圖呈現於附錄 1 之圖示 1。兩個受評量的配對標示語組合呈現於螢幕中間部分的左右兩側；畫面下方為一個用於評量的橫軸並有一可移動游標。受試者則依該配對之相異性程度以滑鼠拖曳評分游標判定其相異度。每一作答畫面下方都有

指導語標明「軸的越右邊代表越相異」。受試者可多次調整指標位置，在確定評量後按鍵進到下一個標示語評量。本研究以 0-1000 的尺規紀錄受試者在各題作答時的指標位置。

## 2. 單一直接評量

同一類的標示語（24 個同意詞、24 個符合詞以及 20 個頻率詞）呈現順序以隨機方式呈現以平衡可能的順序效應。單一直接評量階段的作業畫面示意於附錄 1 之圖示 2，畫面上方為需要評量之單一量表標示語，下方為一水平橫軸，受試者以滑鼠拉動指標作答。水平橫軸並未有兩端點的文字標示，但於畫面下方標示「軸的越右邊代表同意（符合／出現頻率）程度愈高」。本研究以 0-1000 的尺規紀錄受試者在每一評量後確認的指標位置。附錄 2 整理每一作業施測時的指導語。

## 3. 同時排序評量

此階段作業為同時評量同一類的所有量表標示語。此階段的作業畫面的呈現與作答示意圖呈現於附錄 1 之圖示 3a、3b 和 3c。同一類詞的所有量尺標示語同時呈現於螢幕的上部其順序為隨機排列。受試者的作業為將上部的陳列之標示語以滑鼠左鍵選取並拖曳至下方的量尺橫軸欲評定之位置。受試者可以任意順序選擇標示語再將其拖曳至下方的量尺橫軸適當位置。已被經被拖曳至量尺橫軸的標示語也可以不限次數的以滑鼠長按左鍵調整其於量尺上的位置。此階段作業沒有時間限制，當受試者將同類所有標示語拖曳於量尺軸上並調整確認其位置後，受試者可按下方「完成作答」鍵完成這階段的作業。本研究以 0-1000 的尺規紀錄受試者確認的每一個標示語於橫軸上之位置。受試者分別對三類的標示語完成排序作業，而三類的標示語排序作業的順序以電腦隨機排定。



### 三、結果

#### （一）配對比較評量

本研究包含之三類詞共需要 552、552 與 380 個配對以完成所有可能的兩兩配對，此階段作業以電腦隨機選取組合及順序呈現，每個配對組合由多位受試者的反應同配對平均估計而得。在不考量左右兩邊的差異下，三類標示語每個配對差異性評量的平均數和中位數整理成列於表 1 中。表中右上三角形區域中數值為每配對相異性的平均數，左下三角形中數值為每配對差異的中位數。表中以每一個數值代表相對應的行與列標示語的平均數（右上三角形）或中位數（左下三角形），並以底色深淺標示不同程度的受試者在該配對的相異程度的評量高低。

同意詞類的配對比較結果大致與預期的趨勢型態相同，表 1 中靠近右上或是左下角的配對大致都有較高的差異程度。表 1 中相異程度的主要趨勢為大多數配對的平均相異程度偏高（數值較大），而只有在兩個標示語在相同的極端時，其相異程度才大幅下降（表 1 的左上和右下的配對組）。而幾個比較特異的趨勢則是「少部分同意—少部分不同意」和「有點不同意—有點同意」兩組配對不論是平均數或是中位數的相異程度都相對高；而「部分不同意—大部分不同意」、「部分不同意—同意」、「有點不同意—部分同意」、「少部分不同意—大部分同意」以及「少部分不同意—多半同意」五個配對組則比整體趨勢預期的相異程度要低，也就是受試者認為這幾個配對其隱含的程度較為相近。

在符合詞類，整體的相異程度的評量和同義詞類相近，但是在相

異程度的絕對數值上估計量平均較低，也就是在評量符合詞配對時，受試者認為符合詞的各組配對之差異性比同意詞類的標示語要低。而「少部分不符合—少部分符合」也和同意詞類一樣在相異程度上比預期的高。而「少部分不符合—多半符合」和「少部分不符合—大部分符合」則是僅在符合詞類中的配對中相異程度相對較高的配對。而和「部分不符合」配對的幾個符合詞標示語則比預期的相異程度要低，也就是「部分不符合」和其他幾個程度形容詞（如「還算」、「多半」、「大部分」和「非常」）間較不易區分兩者之間的差異。

而頻率詞類的標示語同樣在越靠近右上或是左下角的配對組被評量的差異程度越高。因頻率詞類每個配對受到受試者的評定次數差異較大，因此解讀評量頻率詞的差異時應以中位數結果為主以避免極端值的影響。頻率詞的配對中，「大部分時間—常常」以及「一直—大部分時間」兩個配對有比預期大的差異性。而另一個明顯的規律則是和「有時」配對的頻率詞時，除了和「經常」配對時差異性較大外，其他的配對都有較低的相異性。另外，「偶爾」標示語也和特定幾個頻率詞的相異性較低。這兩個詞都是位於中段程度的頻率詞，其主觀的頻率定義可能較為模糊。

為估計配對比較作業下標示語的量尺估計值，三類量尺標示語以評量相異程度的中位數<sup>1</sup>距離矩陣（表 1 中下三角形區域）以 Torgerson (1952) 提出之典型多向度度量法（classical multidimensional scaling, CMDS）進行估計，此方法也稱主座標分析（principal coordinate analysis; Gower 1966）分別分析。雖然多向度度量法常作為資料視覺化或資料潛在維度探索之工具，但 CMDS 為矩陣分解，也可作為潛在向度

---

1 因考量平均數可能較受極端值影響，故以中位數相異矩陣進行 CMDS 分析。

估計量之方法，因此研究主要找出潛在量尺值，也與其他兩個量尺估計法做比較，故選擇單一維度為目標維度數，而非一般 MDS 分析中以適配指標尋覓最佳維度數目之作法。所得之單一維度的座標解在保留刺激值間相對關係不變下，將估計座標值以等比例轉換，使數值介於 0-1000 範圍間。三類量表標示語以 CMDS 一維度分析所得的估計值結果呈現於圖 1。圖中顯示同意詞和符合詞的多項度量尺法估計值相當接近。尤其在低強度的標示語中一致性很高。在同意詞類，同樣出現了「少部分不同意」和「少部分同意」兩標示語和預期順序倒反的情況，但符合詞類則無此狀況。頻率詞類標示語則聚集在兩端，而僅

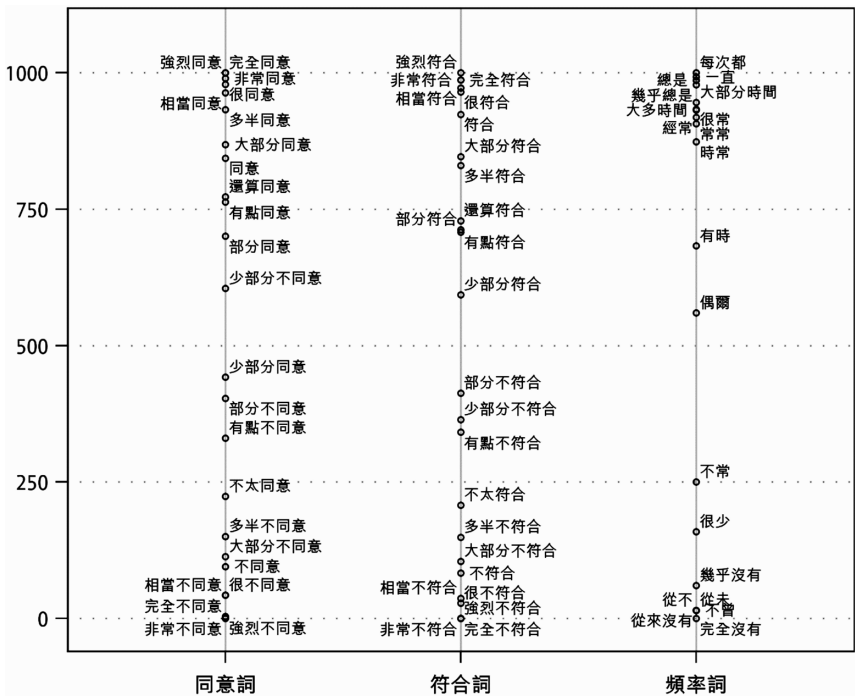


圖 1 配對比較作業以 CMDS 單一維度分析的三類標示語估計值

「有時」和「偶而」位於中段；而「不常」和「很少」位於中低段。

## （二）單一直接評量

由於直接評量作業是以電腦施測收集受試者於一線段上對於刺激標示語的反應，本研究以 0-1,000 之間數值紀錄受試者每一題的作答反應。表 2 為 97 位受試者在同意、符合和頻率三類標示語以直接單一評量作答的平均值和標準差以及中位數和四分位距。在同意詞類，計算每個標示語的偏度（skewness）發現越靠近兩個極端的標示語的偏態越大，而同意度愈低的標示語正偏的程度越高，同意度越高的標示語則越負偏。另外峰度（kurtosis）指標也顯示靠近兩個極端的標示語之估計值具高狹峰分配，而中間區段的標示語則為低闊分布型態。而符合詞類和頻率詞的標示語也呈現相似的分布型態。然而因為量尺受邊界值 0 和 1,000 的限制，因此以中位數估計之結果應較平均數估計更能反映受試者之心理量尺值。

單一直接評量各標示語所得的平均數和中位數的估計值呈現標示語量尺估計值於尺規中分布的並不均勻。同意詞類和符合詞類，位於四個等分尺規範圍（0-250、251-500、501-750、751-1,000）內的標示語數為 8、4、4 和 8 個，而頻率詞則更不平均（7、1、1 和 10 個）；換言之，所評量的標示語中，較低和較高程度的標示語數量較位於中段的標示語多出很多，因此不易在量尺上挑出等距量尺值的標示語使用於李克式量尺上。這樣的狀況在頻率詞標示語中尤其明顯，所選取的 20 個頻率詞中有一半在絕對評量數值上皆大於 750，而前五個頻率詞之中位數皆為 0，不能區別差異。整體而言，中段區域的標示語較前段和後段標示語有較大的變異程度（較大的變異數或四分位距）。這樣的結果可以解讀為受試者間對於兩端的標示語有較一致的評量，或可

以解讀為受試者對於較強（高／低）的標示語之心理意涵較為確切。

雖然在平均數上「少部分不同意」絕對評量平均數值（551.0）高於「少部分同意」絕對評量平均數（456.7）並不符合預期的方向性，但在中位數的評量上則符合預期的順序（458.0 和 545.0），可能為資料中極端評量值的影響。其他標示語所得的絕對評量平均數或中位數皆大致符合字面上的意涵和預期順序。「少部分不符合」和「少部分符合」則僅在絕對評量中位數值上有倒反，但絕對評量平均數的順序則和文字意涵相符。以上這些順序倒反的狀況也可能是因為這幾個標示語語意上有雙重否定（double negative）的狀況，受試者在作答時因語意不清而無法準確判斷。

檢驗同意和符合詞類的標示語中，有兩個標示語有相對較大的變異數和四分位距：「少部分不同意」和「少部分同意」以及「少部分不符合」和「少部分符合」。此外，「不太同意」和「不太符合」也具有較大的四分位距，這些詞有較大的評分者間變異，意即受試者間對於「少部分」或是「不太」兩個形容詞的評量有很大的個別差異。另外，同意詞類中有較大的個別差異，還有「部分不同意」和「多半同意」兩個標示語；而符合詞類則是「符合」和「不符合」在受試者間的評量上有較大的歧異。而在頻率詞類的標示語中，較大受試者間差異的標示語為「偶爾」、「很常」、「很少」與「不常」等。以上標示語有較大的變異數或四分位差，應可解讀為受試者在評量時對於其程度的認定較不一致，因此使用這些標示語於李克式量尺時，可能較無法準確的反映受試者真實的心理量尺值。

### （三）同時排序評量

同時排序作業時，每位受試者將同時呈現的所有同類標示語以拖

曳至橫軸的方式進行評量。每個受試者在每個標示語的評定的平均數和中位數及對應之標準差和四分位數結果整理於表 3。四分位距較大的同意詞類標示語依次為「少部分不同意」、「少部分同意」、「不太同意」、「同意」、「多半同意」和「多半不同意」。以相同標準的符合詞類標示語為「少部分符合」、「少部分不符合」、「不太符合」、「還算符合」、「符合」、「不符合」和「大部分不符合」。而「少部分符合」和「少部分不符合」的變異度更約為「少部分不同意」「少部分同意」的兩倍。因此，「少部分」形容或描述符合程度比以形容同意程度有更大的個別差異。

頻率詞的標示語在同時排序評量時位於中段的標示語（F07-F13）有較大的四分位距，顯示受試者對於中段的頻率標示語有較大受試者間歧異，越往兩極端則變異性降低。這可能是因為受試者在判斷時能以兩個極端（全有或全無）作為參照點；而離這兩個參照點越遠的標示語，受試者因無明顯參照點，所以所估計的結果歧異較大，也比較不容易一致。

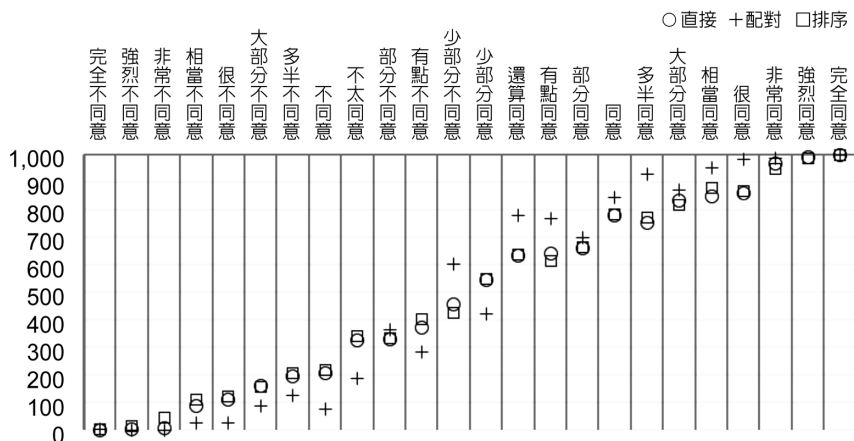
同時排序作業所估計的標示語心理量尺值在數值上有很不平均的分布。同意詞類和符合詞類位於四等分位距內的標示語數量同為 8、4、4、8，同樣在中段具有較少的標示語。而頻率詞四等分切分下各區間有標示語數為 8、1、1、10，分布更為不均勻，有一半的頻率詞標示語估計之平均數或中位數更高於 750。另外，所估計之量尺值上在越往極端時標示語的相對評量值間的差異越小，此結果和單一直接評量作業所得結果相近。而同時排序評量方法下所得的平均數和中位數的在同意詞類並沒有出現如直接單一評量結果時在「少部分不同意」和「少部分同意」順序倒反的狀況；在符合詞類的中位數評量上則如直接單一評量結果時在「少部分不符合」和「少部分符合」兩標示語

有順序倒反的狀況。這些順序倒反狀況也可能因為標示語雙重否定而造成的語意不清所致。研究者使用李克式量尺應該避免這類語意上可有雙重否定的標示語。

## 四、心理量尺值估計方法綜合比較

### (一) 量尺估計值比較

為易於比較本研究中三個作業所得估計之量尺值的相互關係，圖 2 將三個作業所得之估計值同時呈列，圖中以「+」為配對比較評量標示語之估計值「○」，標示單一直接評量之標示語估計值，而「△」為同時排序評量之標示語估計值。圖 2 由上而下子圖分別為同意、符合和頻率類的標示語的估計值於三個作業所得之量尺值。在所有類別中，直接單一評量和同時排序評量的量尺值相近，而以配對比較所得之量尺值則呈現較大差異。配對比較所得的標示語量尺估計值在正向詞時高於另兩個量尺估計法，而在負向詞時低於另兩個量尺估計法。



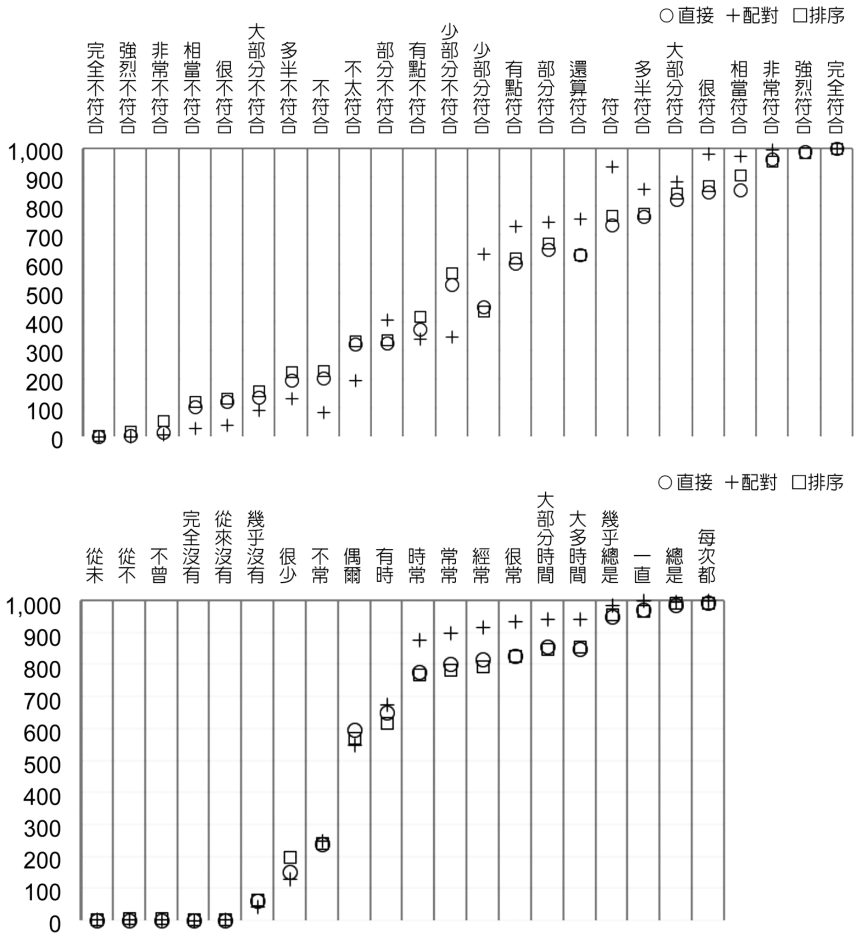


圖 2 不同量尺值估計法所得之估計值比較

配對比較所得之量尺值在同意和符合詞類和其他兩估計值在中段的差異較大；而頻率詞類標示語時則在中高頻率段有較大的差異。配對比較評量估計值都明顯高於其他兩個方法所得的量尺估計值。而頻率詞類前五個負向標示語估計量都接近底標，並無太大差異。



## （二）標示語對稱性分析

本研究亦對副詞相同的兩極標示語（例如：完全不同意—完全不同意、非常不同意—非常不同意）的意涵是否對稱進行分析。以 Lichtenstein 及 Newman（1967）的分析法將副詞相同的兩極標示語之中位數相加，而所得之和介於 990-1,010 之間的標示語則界定為高對稱性。以此法分析發現，同意詞的副詞對稱性較符合詞類的標示語高，其中同意詞類高對稱的副詞有：完全、強烈、大部分、部分、少部分；而符合詞類僅有「完全」與「強烈」兩副詞具有較高的對稱性。

頻率詞因語詞結構關係幾乎沒有副詞相同的兩極標示語，以同對稱性判定標準（兩個頻率詞標示語之中位數和介於 990-1,010 間）則發現，「每次都」和「從不、不曾、完全沒有」高對稱，「總是」和「完全沒有、從來沒有」高對稱；「一直」和「從來沒有」高對稱；以及「很少」和「常常、經常」高對稱。

## （三）估計值離散度分析

本研究以標準差（SD）和四分位距（IQR）兩個指標來衡量平均數和中位數估計的離散度。離散指標反映出受試者在同一標示語上的不一致程度。先前已就同一估計法中三類標示語的離散程度比較，現就不同估計方法間的差異進行比較。由「單一直接評量」（表 2）和「同時排序評量」（表 3）的離散程度發現，以「同時排序評量」的估計法在同意和符合類的標示語都具較小的離散程度，其差異又以標準差較四分位距來的大。這個結果應可以解釋為在「同時排序評量」的作業時，受試者因有同類的所有標示語的相對位置作為參照資訊，所以在最後評量結果反映出較有受試者間的一致性。而在「單一直接評

量」作業中單次僅評量一個量尺標示語，受試者間的變異也較大。在頻率詞類因在「單一直接評量」作業中，幾個低頻率的標示語在中位數評量時皆為 0（尺規的下界），不能區分出差異，然而在同時評量作業時因能同時參照所有標示語使其估計值可以拉開而區分出差異。對比「單一直接評量」的估計方法，「同時排序評量」在所得的量尺估計值有較高的受試者間一致性和較佳的標示語間的區辨力。翁麗禎（1998）結果發現各估計法在離散指標上的相關均偏低，但無充分資訊解釋其可能原因，本研究上述結果則可補充離散指標上的差異可能與作業方法有關，因同時排序評量同時具兩端參照點及所有待判定的標示語相對位置之資訊，而使得其心理量尺估計值比單一直接評量的作業的估計值有較小的離散度。

## 五、討論

翁麗禎（1998）標示語心理量尺的研究發現，中位數和平均數在連續區間法和簡化連續區間法所得的量尺估計值的相關極高，幾近於 1.00；其結果代表比較的估計方法之結果間有很強的線性關係，因而建議平均數和中位數等容易計算的量尺值估計法皆可加以應用。本研究因採電腦施測並以為精細的評量尺度（0-1,000）而非該研究中的 10 等級或 11 等級的級別，而本研究所比較的三種量尺值估計方法在中位數和平均數的差異是相對明顯的，雖然本研究與翁麗禎（1998）研究中共有的標示語其相對順序排序大致相同，但是就量尺值的估計值在絕對數值上還是相當不同。因研究方法之不同，中位數和平均數兩個集中量數的結果差異反映出資料仍受資料型態所影響，例如極端值、上下限等因素，因此在量尺值的估計上仍應注意選取適合的集中

量數於分析中。

本研究以三個不同的量尺值估計法皆發現各類標示語均未能平均分散於設定的量尺區間，尤其在配對比較評量下，量表標示語的量尺值更趨於兩端。整體而言，三個估計法在兩極端時的結果較為一致，而中間區段的標示語的差異相對大。考量各個標示語未能平均分布於量尺區間，也未能有較一致的等距區分的標示語，本研究建議在使用李克式量尺時可僅標示兩端之標示語以保持等距量尺之特性。除了分布不均勻的問題外，每個估計法所得的心理量尺值也都發現位於中間區段的標示語數量較少，此現象在頻率詞類尤其明顯，由於中段的詞類特別少，也同樣使我們建議不於李克式量尺標示中間區段標示語。統整本研究的結果，本研究具體建議在同意、符合和頻率類的李克式量尺兩端標示語為：完全不同意—完全同意、完全不符合—完全符合、從未—總是。

胡志偉等人（1989）的研究中所提出等分的中文機率詞由高至低（高、中高、中、中低、低）為「絕對」「通常」「或許」「不太可能」、與「絕不」。但本研究於頻率詞類的語詞並未包含這五個詞，不能直接加以比較，如以該研究考量的等分量尺和不重疊性原則，本研究結果建議的五個頻率詞為「總是一時常—偶而—不常—從未」。

本研究的樣本數稍嫌不足，和受試者的性別比例也不平均，是本研究的限制之一，因此對本研究的結論的概推性可能有所影響；另外在本研究結果是否可以由學生樣本推到異質性較大的群體我們也是相對保守，可以參考姚開屏、林茂榮與王榮德（2000）對於同質團體和異質團體在量表量尺語詞的研究討論和結果。本研究也未對標示語的使用情境（context）做設定，過去研究（Newstead and Collis 1987）發現，情境在頻率詞類的使用上會有影響，但中文研究中尚未有對此

效果進行系統性的研究，仍待進一步的研究來釐清。頻率詞類的情境影響已有非常多的研究加以探討，幾乎確定情境有非常一致和且相當強度的影響，但本研究在頻率類的量表標示語並未設定特定情境，因此在估計其心理量尺值時較同意和符合類的標示語都有更高的不一致性（較大的標準差和四分位距）。本研究也未對使用情境如何影響同意類和符合類標示語心理量尺的估計深入探討，也是本研究限制之一。另外，本研究以受試者評量每一個刺激的作答標示於電腦螢幕中一水平橫軸上的作答方法也和過去研究非常不同，雖然這樣的介面假定連續尺度並可以精細紀錄受試者的作答反應，但此一作答反應介面較類似於視覺類比量尺（visual analog scale, VAS），而這樣量尺和傳統李克式量尺適用的點數圈選、勾選或劃記的方法之間的具體關聯仍待其他相關研究繼續深入探討。

除本研究以三種不同作業估計的量尺標示語的量尺值外，也可以項目反應理論（Item Response Theory）模型來處理李克式量尺點數間可能不等距的疑慮。例如，研究者可以「等級反應模式」（Graded Response Model; Samejima 1969）或「部分計分模式」（Partial Credit Model; Masters 1982）直接估計量尺中點數的參數值，此方法對量尺不做點數間等距的假設。而如以本研究中第三種「同時排序」作業所取得的資料，也可以排序模型結合貝氏估計法來估計量尺值（例如：Yao and Böckenholt 1999）也是以直接估計的方式。此外，也有研究者提出以模糊語意或是模糊點數（Fuzzy number）來對於量尺值的估計（楊明慧 2002；Chen and Yao 2015）。

## 六、結語

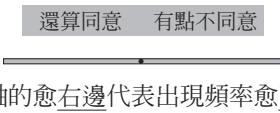
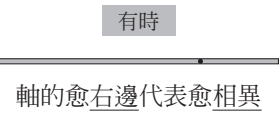
本研究以三種心理量尺值估計法對於在李克式量尺上使用同意、符合和頻率類的標示語進行研究。研究結果提供心理學研究中常用之量表標示語所表徵之潛在單一向度及其心理量尺值，並討論估計值在不同方法間的差異；此外，本研究也就對稱性和離散度兩個面向深入討論和解釋其可能原因。爲使標示語的標註能輔助李克式量尺爲等距量尺，本研究也根據分析結果針對標示語的選取和標示位置，提出只標示於兩端標示語的具體建議，例如：強烈不同意—強烈同意、完全不符合—完全符合、從未—總是，期本研究結果能提供未來研究者使用李克式量尺時之參考。

# 附錄 1 實驗畫面

圖示 1 配對比較評量

圖示 2 單一直接評量

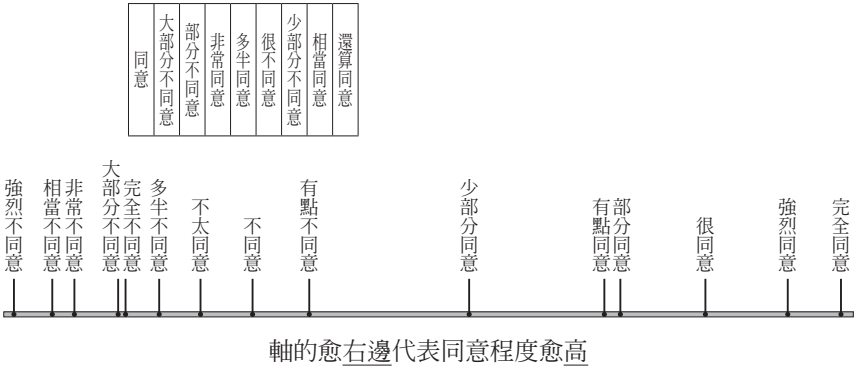
以下兩個量表標示語相異的程度為何？ 以下量表標示語出現頻率的程度為何？



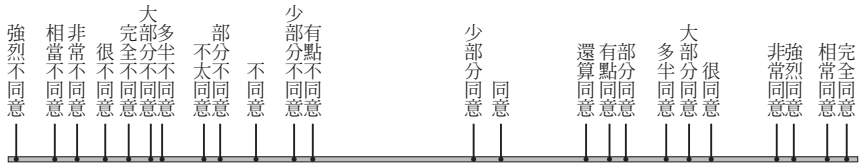
圖示 3a 同時排序評量實驗刺激

完全不同意	完全同意	同意	大部分不同意	大部分同意	非常不同意	部分不同意	不太同意	很同意	有點同意	少部分同意	非常同意	強烈同意	多半同意	不同意	很不同意	少部分不同意	有點不同意	相當不同意	還算同意	多半不同意	強烈不同意	部分同意	相當同意
-------	------	----	--------	-------	-------	-------	------	-----	------	-------	------	------	------	-----	------	--------	-------	-------	------	-------	-------	------	------

圖示 3b 同時排序評量操作示意



圖示 3c 同時排序評量完成示意



軸的愈右邊代表同意程度愈高

完成作答

## 附錄 2 指導語

### 配對比較評量

在您作答時，畫面上會隨機出現兩個同意詞，我們會詢問您認為這兩個量表標示語的「相異程度」，請您依據您的感受和想法作答。

- 點擊灰色軸後，作答用的「標示點」即會出現在軸上。
- 您可以使用滑鼠拖曳「標示點」，移動到軸上任何一個位置來回答。
- 軸的愈右邊代表「相異程度」愈高。

### 單一直接評量

在您作答時，畫面上將呈現一個頻率詞，我們會詢問您認為這個量表標示語「出現頻率」的高低，請您依據您的感受和想法作答。

- 點擊灰色軸後，作答用的「標示點」即會出現在軸上。
- 您可以使用滑鼠拖曳「標示點」，移動到軸上任何一個位置來回答。
- 軸的愈右邊代表「出現頻率」愈高。

### 同時排序評量

本階段畫面上將呈現 24 個同意詞以及一個灰色軸，請依據您認為各個量表標示語的「同意程度」高低相對關係，逐一將各標示語拖曳至灰色軸上。

- 請使用滑鼠左鍵拖曳各標示語到軸上。



- 您可以拖曳移動在軸上的標示語以調整相對關係。
- 請您將所有的標示語拖曳至灰色軸上。
- 軸的愈右邊代表「同意程度」愈高。

## 參考文獻

- 胡志偉、彭昭英、沈永正、楊金龍，1989，〈常用中文機率詞所代表的意義〉。《中華心理學刊》31(1): 1-6。(Hue, C.-W., C.-Y. J. Peng, Y.-C. Shen, and G.-L. Young, 1989, “Numerical Translation of 29 Chinese Probability Term.” *Chinese Journal of Psychology*, 31: 1-6.)
- 姚開屏、林茂榮、王榮德，2000，〈同質團體與異質團體在選擇量表量尺語詞上的比較研究〉。《中華心理學刊》41(2): 141-153。(Yao, G., M.-R. Lin, and J.-D. Wang, 2000, “A Comparative Study on Scale Descriptor Selection: Heterogeneous Group vs. Homogeneous Group.” *Chinese Journal of Psychology*, 41: 141-153.)
- 翁儷禎，1998，〈評定量表標示語之心理量尺值研究：頻率及同意度詞〉。《中華心理學刊》40: 73-86。(Weng, L.-J., 1998, “Scale Values of Anchor Labels in Chinese Rating Scales: Responses on Frequency and Agreement.” *Chinese Journal of Psychology*, 40: 73-86.)
- 楊明慧，2002，〈網路系統之模糊語意問卷設計與分析〉。臺中市：臺中師範學院教育測驗統計研究所碩士論文。(Yang, Ming-Hui, 2002, “Design and Analysis of the Fuzzy Linguistic Questionnaire Network System.” Unpublished master dissertation, Graduate Institute of Educational Information and Measurement, National Taichung University of Education, Taichung, Taiwan.)
- Aarts, H., and A. Dijksterhuis, 1999, “How Often did I Do It? Experienced Ease of Retrieval and Frequency Estimates of Past Behavior.” *Acta Psychologica* 103(1-2): 77-89. DOI: [https://doi.org/10.1016/S0001-6918\(99\)00035-9](https://doi.org/10.1016/S0001-6918(99)00035-9).
- Blair, E., and S. Burton, 1987, “Cognitive Processes Used by Survey Respondents to Answer Behavioral Frequency Questions.” *Journal of Consumer Research* 14(2): 280-288. DOI: <https://doi.org/10.1086/209112>.
- Chen, P.-Y., and G. Yao, 2015, “Measuring Quality of Life with Fuzzy Numbers: in the Perspectives of Reliability, Validity, Measurement Invariance, and Feasibility.” *Quality of Life Research* 24(4): 781-785. DOI: <http://www.jstor.org/stable/44849080>.
- Edwards, A. L., 1957, *Techniques of Attitude Scale Construction*. East Norwalk, CT, US: Appleton-Century-Crofts.
- Edwards, A. L., and R. Gonzalez, 1993, “Simplified Successive Intervals Scaling.” *Applied Psychological Measurement* 17: 21-27. DOI: <https://doi.org/10.1177/014662169301700106>.

- Foley, B. J., 1959, "The Expression of Certainty." *American Journal of Psychology* 72: 614-615. DOI: <https://doi.org/10.2307/1419512>.
- Gaskell, G. D., C. A. O'Muirheartaigh, and D. B. Wright, 1994, "Survey Questions about the Frequency of Vaguely Defined Events: The Effects of Response Alternative." *Public Opinion Quarterly* 58: 241-254. DOI: <https://doi.org/10.1086/269420>.
- Gower, J. C., 1966, "Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis." *Biometrika* 53: 325-338. DOI: <http://dx.doi.org/10.1093/biomet/53.3-4.325>.
- Hakel, M. D., 1968, "How Often is Often?" *American Psychologist* 23(7): 533-534. DOI: <https://doi.org/10.1037/h0037716>.
- Klockars, A., and M. Yamagishi, 1988, "The Influence of Labels and Positions in Rating Scales." *Journal of Educational Measurement* 25(2): 85-96. DOI: <https://doi.org/10.1111/j.1745-3984.1988.tb00294.x>.
- Józsa, K., and G. A. Morgan, 2017, "Reversed Items in Likert Scales: Filtering out Invalid Responders." *Journal of Psychological and Educational Research* 25(1): 7-25.
- Likert, R., 1932, "A Technique for the Measurement of Attitudes." *Archives of Psychology* 22(140): 55.
- Lichtenstein, S., and J. R. Newman, 1967, "Empirical Scaling of Common Verbal Phrases Associated with Numerical Probabilities." *Psychonomic Science* 9(10): 563-564. DOI: <https://doi.org/10.3758/BF03327890>.
- Low, G., I. Tasker, and L. Hong, 1991, "The Wording of Bipolar Attitude Scales in Chinese." *Educational Research* 33(2): 141-150. DOI: <https://doi.org/10.1080/0013188910330207>.
- Masters, E. R., 1974, "The Relationship between Number of Response Categories and Reliability of Likert-type Questionnaires." *Journal of Educational Measurement* 11(1): 49-53. DOI: <https://doi.org/10.1111/j.1745-3984.1974.tb00970.x>.
- Masters, G. N., 1982, "A Rasch Model for Partial Credit Scoring." *Psychometrika* 47: 149-174. DOI: <https://doi.org/10.1007/BF02296272>.
- Matell, M. S., and J. Jacoby, 1972, "Is There an Optimal Number of Alternatives for Likert-scale Items? Effects of Testing Time and Scale Properties." *Journal of Applied Psychology* 56(6): 506-509. DOI: <https://doi.org/10.1037/h0033601>.
- Myers, J. H., and W. G. Warner, 1968, "Semantic Properties of Selected Evaluation Adjectives." *Journal of Marketing Research* 5(4): 409-412. DOI: <https://doi.org/10.2307/3150266>.
- Newstead, S. E., and J. Arnold, 1989, "The Effect of Response Format on Ratings of Teaching." *Educational and Psychological Measurement* 49(1): 33-43. DOI: <https://doi.org/10.1177/0013164489491004>.

- Newstead, S. E., and J. M. Collis, 1987, "Context and the Interpretation of Quantifiers of Frequency." *Ergonomics* 30(10): 1447-1462. DOI: <https://doi.org/10.1080/00140138708966038>.
- Oaster, T. R. F., 1989, "Number of Alternatives Per Choice Point and Stability of Likert-type Scales." *Perceptual and Motor Skills* 68(2): 549-550. DOI: <https://doi.org/10.2466/pms.1989.68.2.549>.
- Pepper, S., and L. S. Prytulak, 1974, "Sometimes Frequently Means Seldom: Context Effects in the Interpretation of Quantitative Expressions." *Journal of Research in Personality* 8(1): 95-101. DOI: [https://doi.org/10.1016/0092-6566\(74\)90049-X](https://doi.org/10.1016/0092-6566(74)90049-X).
- Samejima, F., 1969, "A General Model for Free Response Data." *Psychometrika Monograph* No. 17.
- Spector, P. E., 1976, "Choosing Response Categories for Summated Rating Scales." *Journal of Applied Psychology* 61(3): 374-375. DOI: <https://doi.org/10.1037/0021-9010.61.3.374>.
- Torgerson, W. S., 1952, "Multidimensional Scaling: I. Theory and Method." *Psychometrika* 17: 401-419. DOI: <https://doi.org/10.1007/BF02288916>.
- Wallsten, T. S., D. V. Budescu, A. Rapoport, R. Zwick, and B. Forsyth, 1986, "Measuring the Vague Meanings of Probability Terms." *Journal of Experimental Psychology: General* 115(4): 348-365. DOI: <https://doi.org/10.1037/0096-3445.115.4.348>.
- Weng, L. J., 2004, "Impact of the Number of Response Categories and Anchor Labels on Coefficient Alpha and Test-retest Reliability." *Educational and Psychological Measurement* 64(6): 956-972. DOI: <https://doi.org/10.1177/0013164404268674>.
- Wildt, A. R., and M. B. Mazis, 1978, "Determinants of Scale Response: Label versus Position." *Journal of Marketing Research* 15(2): 261-267. DOI: <https://doi.org/10.2307/3151256>.
- Yao, G., and U. Böckenholt, 1999, "Bayesian Estimation of Thurstonian Ranking Models Based on the Gibbs Sampler." *British Journal of Mathematical and Statistical Psychology* 52: 79-92. DOI: <https://doi.org/10.1348/000711099158973>.