

中數值約.55。從所得數值判斷本測驗難度適中。

常態化難度指數 Δ ，介於8.5~15.9之間，平均12.3。

二系列相關 r 值，介於.35~.85之間。由 r 看出，除少數較容易的題目外，各試題皆和總分有顯著關係($P < .01$)。全測驗 r 的中數值約.62。

鑑別力指數 D 介於.33~.89之間，平均值約.59，屬於鑑別度良好的情況(陳英豪、吳裕益，民71)。

伍、信度與效度

一、信度

為考驗本測驗的穩定性和內部一致性量數，採重測法、折半法、庫李法及測量標準誤等方式驗證測驗之信度。信度樣本見表5-1。

(一)重測信度

本測驗的重測信度，以間隔四個星期，前、後所測得的分數，用統計Pearson r 求取相關係數。重測信度係數見表5-2。

(二)折半信度

折半信度是以一次測驗結果，分成兩半求其相關。分成兩半的方法很多，其中最常用的有兩種：一是依照隨機方法將題目分為兩半，另一種是依照奇數題和偶數題分為兩半。本測驗採後者，再用皮爾遜積差相關方法，求得受試者兩半分數的相關，最後以斯布公式(Spearman-Brown Formula)加以校正。見表5-2。

年所提出估計信度的一種方法。是依據受試者對所有試題的反應，分析題目間的一致性(interitem consistency)，以確定測驗中的題目是否測量相同的特質。本測驗依庫李21號公式計算出測驗的信度。見表5-1。

(四)測量標準誤

重複測驗同一個人相當多次，而不會改變所測量的特質，實際上是不可能的。因此，測量標準誤通常是從團體的資料，加以估計。本測驗根據所得的折半信度係數算出測量標準誤。測驗的信度愈高，測量標準誤愈小，從測量標準誤的大小，可以推知測驗的信度及解釋個別分數與誤差的關係。見表5-2。

表5-1 信度樣本明細表

年 級	學 校	男	女	合 計
四	台中市大鵬國小	2 2	1 5	3 7
	台中縣龍泉國小	1 9	2 5	4 4
	彰化縣田中國小	3 1	2 0	5 1
	南投縣草屯國小	2 6	2 0	4 6

註：重測、折半、庫李等信度與測量標準誤取樣人數皆相同。

表5-2 本測驗之信度及測量標準誤

年 級	人 數	重測信度	折半信度	庫李信度	測量標準誤
四	1 7 8	· 8 4	· 9 0	· 8 2	2 · 5 8

註：1. 以上各信度係數皆達.01顯著水準。
2. 折半信度係經斯布(Spearman-Brown)公式校正。
3. 測量標準誤根據重測信度係數估計。

二、效度

效度是指測驗分數的正確性，易言之，就是指一個測驗能夠測量到它所想要測量的特質的程度(郭生玉，民74)。本測驗分別就內容效度(content validity)、效標關聯效度(criterion related validity)之同時效度(concurrent validity)，及構念效度(construct validity)，來說明效度考驗之結果。

(一)內容效度

內容效度是指測驗內容的代表性或取樣的適切性，但是內容效度尚無法以一種簡單的統計方法來評估試題的抽樣是否適當，因此沒有一種數量的表示方式。本研究敦請數學系教授、測驗編製專家、資深國小數學科教師數人就本測驗之題項，配合國小數學課程標準及本測驗之命題計畫、雙向細目分析表考驗本測驗之內容效度、表面效度。

研究者經各專家之指正及接受各專家寶貴意見，曾做多次修改，經各專家認為內容效度、表面效度頗佳，方才決定預試。

(二)同時效度

同時效度是指測驗分數與實施測驗同一個時間所取得的效標之間的相關，目的在使用測驗分數估計個人在效標方面的目前實際表現。本測驗主要以在校數學科成績作為效標，並參考「國民小學數學能力發展測驗」之結果，進行效標關聯效度之考驗。測驗之同時效度，達 .01顯著水準，因此可說明學童數學成就的高低。結果見表5-3、5-4。

表5-3 本測驗四年級之同時效度

年級	學 校	與學校數學科成績之相關		與數學能力發展測驗之相	
		N	r	N	r
四	台中市大鵬國小	1 8 7	· 7 1		
	台中縣龍泉國小	1 5 9	· 6 3		
	台中縣龍井國小	4 4	· 6 2		
	彰化縣田中國小	2 4 5	· 7 4		
	南投縣草屯國小	2 1 1	· 6 7		
	台中縣大雅國小			4 9	· 8 1

註：表中各相關皆達.01顯著水準。

(三)構念效度

構念效度是指測驗能夠測量到理論上的構念或特質的程度，也就是指測驗分數能夠依據某種心理學的理论構念加以解釋的程度。

本測驗之構念效度採下列五種方式考驗之：

1. 內部一致性分析。
2. 本測驗與其他測驗分數之相關研究。
3. 數學能力發展之比較研究。
4. 男女性別差異之比較研究。
5. 因素分析法。

1. 內部一致性分析

內部一致性分析(internal consistency)主要是以測驗

本身的總分為效標，本測驗以下列二種方式來分析：

(1)各題項和總分之相關

比較高、低分組的受試者在各題中答對的百分比，高分組答對的比率皆高於低分組答對的比率，全測驗平均鑑別度達.53(見表4-1)。

(2)內部相關及分測驗和總分之相關

本測驗根據常模樣本所得資料建立內部相關及分測驗和總分之相關，各交互相關矩陣見表5-4。

表5-4 本測驗之交互相關矩陣

	測驗一	測驗二	測驗三	全測驗
測驗一	—	.6538	.5568	.8385
測驗二	.6538	—	.6018	.8652
測驗三	.5568	.6018	—	.8664
全測驗	.8385	.8652	.8664	—

註：表中各相關係數皆達.01顯著水準

2. 本測驗分數與其他測驗分數之相關研究

新編測驗和已公認有效的測驗之相關，可用來做為考量新編測驗相關特質的依據。本測驗採與羅桑非語文智力測驗、黑白瑞文氏測驗、國民小學系列學業性向測驗等之相關，來做為考量新編測驗相關特質的依據。結果見表5-5、5-6。

表5-5 本測驗與羅桑非語文智力測驗之相關

年 級	學 校	與羅桑非語文智力測驗之相關	
		N	r
四	台中市大鵬國小	1 8 7	· 7 1
	台中縣龍泉國小	1 5 9	· 6 7
	台中縣龍井國小	4 4	· 6 8
	彰化縣田中國小	2 4 5	· 7 3
	南投縣草屯國小	2 1 1	· 7 0

表5-6 本測驗與系列學業性向及黑白瑞文氏測驗之相關

年級	學 校	系列學業性向測驗		黑白瑞文氏測驗	
		N	r	N	r
四	台中市光復國小	43	· 77		
	高雄市永清國小			40	· 64

註：表中各相關皆達.01顯著水準。

3. 數學能力發展之研究

依據文獻探討所得的理論，數學能力會隨著年級或年齡的增高，而能力也隨之增加。因此較高年級兒童在數學測驗上的得分也應當高於較低年級兒童在測驗上的得分。研究者分別算出三年級和四年級兒童在各分測驗及全測驗得分之平均數和標準差，用 t 檢定來考驗各年級平均數的差異是否達到顯著性，考驗結果見表5-7。從表中可看出，四年級兒童在分測驗或全測驗的得分均優於三年級的兒童。本研究自彰化市民生國小三、四年級中各隨機抽取一班做為樣本。

表5-7 本測驗年級差異之顯著性考驗

測 驗	年 級	平 均 數	標 準 差	t 值
測 驗 一	三年級	7.02	3.12	2.24**
	四年級	7.82	2.78	
測 驗 二	三年級	7.92	3.33	1.87**
	四年級	8.21	2.81	
測 驗 三	三年級	7.12	3.41	1.24**
	四年級	7.42	3.32	
全 測 驗	三年級	22.07	8.57	2.12**
	四年級	23.45	8.02	

註：1.**p < .01

2.一年級每組 N =45

二年級每組 N =45

4. 男女性別差異之比較

為進一步探討相同年級，不同性別兒童的數學能力發展是否有所差異，研究者以常模之樣本為分析對象，考驗性別差異在本測驗得分的情形，結果見表5-8。

表5-8 性別差異之顯著性考驗

測 驗	性 別	平 均 數	標 準 差	t 值
測 驗 一	男	8.09	2.89	.37
	女	8.02	2.81	
測 驗 二	男	7.91	3.05	.37
	女	7.98	2.96	
測 驗 三	男	7.12	3.79	.56
	女	6.99	3.65	
全 測 驗	男	23.12	8.32	.25
	女	22.99	8.12	

註：男生 N =548

女生 N =508

5. 因素分析(factor analysis)

本測驗採用常模樣本1043人，先求出各分測驗之相關矩陣(見表5-4)，再以此作為因素分析之資料，因素分析結果見表5-9。

表5-9 各分測驗共同因素矩陣

分 測 驗	因 素 1	共 同 性
測 驗 一	.85984	.73932
測 驗 二	.88000	.77440
測 驗 三	.83393	.69544
因素變異量		2.20915
佔總變異量之百分比		73.6

由以上因素分析之結果，可知本測驗在各分測驗中有一重要的共同因素，然各年級分測驗中尚有部份無法以共同因素解釋者，稱為獨特性(Uniqueness)。各分測驗的獨特性尚包含該測驗獨有的變異量，稱為特殊性(Specificity)，和誤差變異量。共同變異量與特殊變異量之和稱為信度(reliability)(黃光國，民72)。