

5 調査問題の妥当性と信頼性（S－P表処理等による分析）

令和2年度高等学校入学者数学学力調査[A]、[B]について、S－P表処理等を基にして差異係数、信頼性係数、内容別平均正答率、正答率帯別問題数、正答率、注意係数、UL指数、問題間の相関等を考察したところ、次のような結果を得た。なお、データは、テスト[A]については参加19校から124名、テスト[B]については30校から376名を抽出して作成した。

[1] 問題全体について

表6

(1) 差異係数

差異係数とは、S、P両曲線のずれの程度を数量化したもので、生徒の理解と一連の学習内容がうまくみ合っているかを見るものである。差異係数は0から1までの値を

		(1) 差異係数		
テスト	年度	H30	H31	R2
テスト	[A]	0.341	0.221	0.309
テスト	[B]	0.281	0.286	0.296

とり、0.5より小さい値のとき生徒の理解と指導の密着性が高いとされている。簡単な確認テストのようなドリル演習型のテストではS曲線とP曲線の乖離は小さく、差異係数は小さくなる。実力テストのような多面にわたる総合的な問題ではS曲線とP曲線は大きく乖離して、差異係数は大きくなる。差異係数が0.5を超えたとき、指導内容に問題がなかったか、出題に問題がなかったか、学習者の理解やモチベーションは高かったかなどを検討する必要がある。今回のテストでは表6のように差異係数は0.3前後であり、出題にとりわけ大きな問題はないと考えられる。

(2) 信頼性係数（ケダー・リチャードソンの公式20による）

表7

信頼性係数とは、作成されたテスト問題が内容的に妥当で信頼できるものなのかを算出するものである。ここで言う信頼性とは、同一条件下で再度試験を実施しても同じ結果が出ると思われる安定性のことで、0から1までの値を

		(2) 信頼性係数		
テスト	年度	H30	H31	R2
テスト	[A]	0.870	0.914	0.893
テスト	[B]	0.831	0.867	0.840

とり、1に近いほど信頼性が高いとされている。今回のテストでは表7のように信頼性係数は0.85前後であり、信頼できる良好な問題であったことが分かる。

(3) 内容別平均正答率（ ）内の数字は問題数

表8

テスト 内容	年度	テスト[A]			テスト[B]		
		H30	H31	R2	H30	H31	R2
① 数と式		60.3%(10)	64.9%(10)	63.7%(9)	55.9%(7)	68.8%(7)	61.5%(7)
② 図形		36.7%(6)	33.8%(6)	34.4%(6)	24.8%(6)	40.0%(6)	51.7%(6)
③ 関数		34.6%(6)	32.1%(6)	35.2%(6)	39.5%(5)	46.5%(5)	48.0%(5)
④ 資料の活用		55.1%(3)	61.8%(3)	52.0%(4)	65.5%(4)	59.1%(4)	78.2%(4)

(4) 正答率帯別問題数

表9

テスト 正答率	年度	テスト[A]			テスト[B]		
		H30	H31	R2	H30	H31	R2
0.851以上		0	0	0	1	1	2
0.667～0.850		7	8	6	6	5	9
0.333～0.666		12	10	12	8	11	7
0.150～0.332		3	6	4	2	3	2
0.149以下		3	1	3	5	2	2

(5) 全体の正答率との相関別問題数

表10

テスト 相関	年度	テスト[A]			テスト[B]		
		H30	H31	R2	H30	H31	R2
0.70以上		0	1	0	0	0	0
0.60～0.69		4	12	7	0	4	1
0.50～0.59		7	5	10	6	4	9
0.40～0.49		9	5	2	11	10	9
0.30～0.39		2	2	6	2	3	2
0.29以下		1	0	0	0	1	1

[2] 検討を要する問題群

テストA、テストBの全ての問題について、②注意係数、③UL指数、④相関係数を算出した。表11は、三つの指標のうち一つでも基準値を満たさない問題を抽出し、基準を満たさない指標に注意マーク“×”を付け、①正答率が基準を満たす“I群”と、①正答率が基準を満たさない“II群”とに分け整理した表である。

②から④までの指標については、上位群と下位群の正答率の差が小さいときに基準値から外れる傾向にあり、正答率が非常に高い問題（正答率75%以上）と正答率が基準を満たさない（II群）問題のときに起こりやすく、これらの問題については、難易度に関して検討する必要がある。それ以外の問題で、②から④までの指標について検討を要する問題は2問あり、表11に※印で示した。

テストAの[1](9)は、選択形式の問題であったので、たまたま正解してしまう者がいて上位群と下位群の差が小さくなったことが原因である。テストBの1は、文字を含む分数の計算問題であるが、上位群であっても計算ミスをする者がおり、上位群と下位群の差が小さくなったことが原因である。

(×印は該当項目について検討を要する数値であることを示す)

表 11

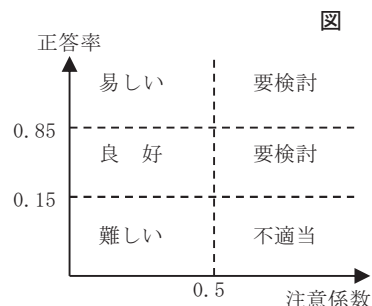
問 題	指 標 基準値	①正 答 率	②注意係数	③UL指数	④相関係数	
		>0.333	<0.500	>0.400	>0.400	
I	テストA	1	0.790	0.505×	0.358×	0.351×
		[1](9)※	0.556	0.560×	0.418	0.371×
		[1](14)	0.815	0.511×	0.358×	0.333×
		[2](1)	0.766	0.490	0.329×	0.376×
	テストB	1※	0.689	0.544×	0.392×	0.366×
		[1](9)	0.822	0.445	0.353×	0.402
		[1](10)	0.957	0.623×	0.078×	0.183×
		[2](1)	0.840	0.503×	0.284×	0.350×
II	テストA	[3](1)	0.331×	0.155	0.777	0.673
		[3](2)	0.298×	0.103	0.836	0.698
		[4](2)	0.145×	0.345	0.358×	0.407
		[5](1)	0.242×	0.298	0.508	0.512
		[5](2)	0.073×	0.183	0.209×	0.396×
		[6](1)	0.274×	0.175	0.657	0.628
	テストB	[6](2)	0.065×	0.250	0.209×	0.347×
		[3](2)	0.287×	0.204	0.676	0.577
		[4](2)	0.194×	0.190	0.520	0.526
		[5](2)	0.125×	0.171	0.382×	0.472
		[6](2)	0.120×	0.212	0.353×	0.443

(各項目の説明)

①正 答 率：各問題の正答率を示す。

$$\frac{\text{正答者数}}{\text{受検者数}}$$

②注意係数：S-P表において、ある問題の正誤の状況と全ての問題の正誤の状況を比較して、その関係性を数値化したものである。0.5より小さい方が適切な問題であるとされている。右図に示すように正答率と併せて検討するとよい。



③UL指数：
$$\frac{(\text{上位27\%の正答者数}) - (\text{下位27\%の正答者数})}{(\text{生徒27\%の人数})}$$

UL指数は上式で算出する。「上位27%の正答者数が多く、下位27%の正答者数が少ない」場合、UL指数は大きくなるが、「上位27%の正答者数が少なく、下位27%の正答者数が多い」場合、UL指数は小さくなる。UL指数が0.4より大きい方が適切な問題であるとされている。

④相関係数：生徒の得点合計とその問題の正解との相関を示す。基準値を0.4として大きい方が適切な問題であるとされている。