

これらの過程には教師の経験やかんにもとづく判断が必要とされたり、予備テストの結果に根拠を求めるべきこともある。いずれにしても適切な項目を累積し、改善を加えていくことは評価を改善するうえで大切である。

(2) 総括的テストの用法

総括的評価は成績の評定に用いられるほかつぎのような用法が考えられる。

ア、後続する学習における成功の予測

総括的評価は、その後の関連する学習での成功を予測する機能がある。この予測は、後続する学習の方法・内容等が近似するものである場合に精度は高くなる。

イ、後続する学習における教授活動の開始点の決定

総括的テストの結果は、個々の児童・生徒や学級全体に対する教授活動をどこから始めたらよいかを決定するのに役立てることができる。このとき、得点のカテゴリーを十分理解していることがたいせつである。

ウ、児童・生徒へのフィードバック

学習の進歩の状況に関する情報を即時フィードバックしてやることは形成的評価の本質であるが、総括的評価もまたこのような利用ができる。総括的テストは概念間の関係の理解や種々の能力の習得状況というような総合的な成果を測定することができるので形成的評価とはちがった面で有益な情報を与えてやることのできる。

6. 評価問題についての技術的諸特性からの吟味

(1) 妥当性 (Validity)

あらかじめ測定しようと思っていたものを測定できたときそのテストは妥当性をもつという。妥当性は、内容的にかたよりのないか (内容的妥当性—Content validity), 概念の構成のしかたはよいか (概念的妥当性—Construct validity), 他の測定結果と一致するか (併存的妥当性—Concurrent validity) 結果から将来を予測できるか (予測的妥当性—Predictive validity) といった角度から吟味されるのが普通である。

(2) 客観性 (Objectivity)

採点者や評定者が異なっても同じ結果が得られるとき、客観性が高いといわれる。また、同じ答案を同じ採点者が2度採点したとき採点結果に変動が少なれば客観性は高いという。したがって客観性の程度は、採点者間、あるいは重複した採点結果間の相関などの測度が用いられる。客観性を高めるために、○×式や選択肢形式のテストを用いることがあるが、評価のねらいや内容によっては妥当性を欠いたり、信頼性をそこなう恐れのあることは周知の通りである。たとえば、ある項目の選択肢が  $n$  個、そこから  $r$  個選択させて正答となるように構成された問題で、みかけの正答率が  $P'$  %であれば、その項目の真の正答率  $P$  %はつぎの補正で得られる。

$$P = (nC_r P' - 100) / (nC_r - 1)$$

(3) 信頼性と信頼度係数

信頼性は測定の一貫性あるいは安定性の意味で用いら

れる。妥当性は低くても信頼性が高いということは考えられるがこの逆はあり得ない。信頼性を数量的に表示するためには信頼度係数が用いられる。これには、平行テスト法、再検査法、折半法などがあるが、ここでは実際的である内部的一致性による方法をあげてみる。

内部的一致性をもって信頼度を測るのに Kuder-Richardson の式が用いられる。いま  $n$  個の小問からなるテストで、1小問の得点が1か0の2種類しかないものとし、各小問の正答率を  $p_i$  とおけば信頼度係数  $r$  はつぎの式で与えられる。(  $S_x^2$  は全体の分散)

$$r = \left( \frac{k}{k-1} \right) \left( 1 - \frac{\sum_{i=1}^k p_i (1-p_i)}{S_x^2} \right)$$

いま、6個の小問からなるテストの信頼度係数を求めてみよう。正答率  $p_i$  はつぎの表のようにになっている。

問題番号	正答率 $p_i$	$p_i (1-p_i)$	$S_x^2$
1	57.6	0.244	3.452
2	(1) 28.9	0.205	
	(2) 19.8	0.159	
3	ア 33.5	0.335	
	イ 32.7	0.107	
4	40.1	0.240	

$$\sum_{i=1}^6 p_i (1-p_i) = 1.290$$

$$S_x^2 = \frac{1}{6} \sum X_i^2 - \left( \frac{\sum X_i}{6} \right)^2 = 3.452$$

上の数値を代入して  $r = 0.752$

信頼度係数が0.85未満のものは信頼性がやや低いとみるのが普通である。信頼性を高めるには、内部的一致性のある問題を付加してテストの長さを増す必要がある。

(4) テストの長さ信頼度係数

信頼度を高めるために付加すべき小問数をきめるには Spearman-Brown の一般式がある。信頼度係数  $r$  のテストを  $k$  倍の長さにしたときの信頼度係数  $R_k$  は

$$R_k = \frac{kr}{1+(k-1)r}$$

となる。たとえば、信頼度係数0.752のものを0.850以上にするには、上の式に代入して  $k = 1.867$  を得る。よってテストの長さをおよそ2倍以上にすればよい。

7. むすび

今日、教育には、豊かな人間性の回復や個性化への対応といった期待や要請が課されており、教育評価も受身的評価から主体的評価へ、一時的評価から、継続的評価へ、遅延的評価から即時的評価へ、平均的評価から個別性尊重の評価へ、結果の評価から過程の評価へ、再生的評価から思考力の評価へ etc、と変ぼうしてきている。このようなとき評価のあり方を吟味し改善を加えていくことは学習指導の最適化をはかるうえでももちろん、教育に対する期待や要請に応えるためにも極めてたいせつなことではないだろうか。