DOCUMENT RESUME

ED 191 884 TM 800 508

AUTHOR Carlson, Jerry S.: Wiedl, Karl Heinz
TITLE Dynamic Assessment: An Approach Toward Reducing Test Bias.
PUB DATE May 80
NOTE 20p.: Paper presented at the Annual Meeting of the Western Psychological Association (Honolulu, HI, May 5-9, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Black Students: Cognitive Style: *Conceptual Tempo: Culture Fair Tests: Elementary Education: *Ethnic Groups: Feedback: Grade 2: Grade 4: Hispanic Americans: *Intelligence Differences: *Intelligence Tests: Learning Problems: Nonverbal Tests: Personality Assessment: *Test Bias: Verbal Communication: White Students
IDENTIFIERS *Dynamic Assessment: Piaget Tests: Raven Progressive Matrices: *Tailored Testing

ABSTRACT

Through dynamic testing (the notion that tailored testing can be extended to the use of a learning oriented approach to assessment), analysis were made of how motivational, personality, and cognitive style factors interact with assessment approaches to yield performance data. Testing procedures involving simple feedback, elaborated feedback, and subject verbalization were applied to the Raven Coloured Progressive Matrices (CPM), a test of nonverbal intelligence, and selected Piaget measures, and were tested in 434 second- and fourth-grade children from the general population. Applicability of the approach was demonstrated for populations of children between 5 and 10 years of age, and of "normal" intelligence (IQ), of learning handicapped, low IQ, and from Anglo, Black, and Hispanic-American ethnic groups. Analysis was made of the interactive effects of testing condition with the cognitive style variable, impulsivity-reflectivity. Evidence was presented that compensatory effects can be obtained for impulsive children and children of ethnic minorities utilizing the procedures of verbalization and elaborated feedback. The indication that compensatory effects can be obtained in assessing between ethnic group differences may be useful in reducing test bias, but larger samples with control over factors such as socio-economic status are necessary before reaching firm conclusions. (Author/RL)

# Dynamic Assessment: An Approach Toward Reducing Test Bias

Jerry S. Carlson

University of California, Riverside

and

Karl Heinz Wiedl

University of Mainz

Dynamic Assessment: An Approach Toward Reducing Test Bias


## Abstract

The bases of the dynamic assessment approach are presented. Testing procedures involving simple feedback, elaborated feedback, and subject verbalization are applied to the Raven Matrices and selected Piaget measures. Applicability of the approach is demonstrated for several different populations of children. These include children of "normal" intelligence, low IQ children, and children from different ethnic/racial groups. Analysis is made of the interactive effects of testing condition with the cognitive style variable, impulsivity-reflectivity. Evidence is presented for compensatory effects for impulsive children and children of ethnic minorities under verbalization and elaborated feedback conditions. The results are discussed in terms of reduction of bias in testing.

Dynamic Assessment: An Approach Toward Reducing Test Bias

A significant concern of our era is related to the conceptualization of what constitutes appropriate assessment of cognitive and perceptual processes. The debate has focused on key terms such as "test bias" or "test fairness." Several different approaches to this issue have been taken. These range from the classical test fairness concept of equal prediction elaborated by Cleary (1968), Cleary, Humphreys, Kendrick, and Wesman (1975), and Eichorn and Bass (1971), to criteria involving group membership and cultural variables (Darlington, 1971), and to the incorporation of sociological data in arriving at performance level (Mercer & Lewis, 1975).

Recently, Flaugher (1978) has cautioned that test bias can result from many factors and influences and that focus on just one aspect may be myopic and ignore the breadth of the issues involved. He did suggest, however, that the development of "tailored testing" would perhaps provide a solution to one very important factor: bias related to the atmosphere in which the test is administered. The notion of tailored testing can be extended to the use of a learning oriented approach to assessment (Resnick, 1979). Through such an approach (referred to here as dynamic testing), analyses can be made of how motivational, personality, and cognitive style factors interact with assessment approaches to yield performance data. This allows for analysis of bias between and within ethnic and/or cultural groups and clarification of individual and group differences in terms of psychological and information-processing variables. In addition,

dynamic assessment provides a basis for modifications in the classroom to incorporate learning characteristics of the child in the strategy and/or teaching methodology used.

The rationale for the use of dynamic testing, and empirical results of selected studies validating the approach for various groups of children will be outlined.

Bases of Dynamic Testing

An assumption often made in assessment of cognitive functioning is that test scores obtained in the traditional, standardized manner serve as adequate estimates of the cognitive competence of the person or group of persons tested. The validity of the assumption is based on the notion that performance is a veridical measure of competence. This suggests that intra-individual variations in information processing, resulting primarily from intellectual and/or personality factors, play relatively unimportant roles in test outcome. The relationship between subject- and task-independent variables is ignored (cf. Detterman, 1979).

The static test approach and the assumptions behind it has been challenged by several researchers (Brown and French, 1979; Schmidt, 1971; Guthke, 1977; Haywood, in press). It is argued that testing procedures must be developed and applied which adequately take intra-individual variability into account and reduce the potential discrepancy between measured cognitive functioning and cognitive competence (Bortner and Birch, 1970). This calls for a dynamic testing approach, where modifications in testing procedures can be analyzed as

they affect performance levels and interact with sources of individual variation.

In this approach, variations in testing procedure can be built directly into the test situation itself or implemented through training outside of the actual testing. The latter requires subsequent assessment of the effects of the training on a transfer test. We favor the former, incorporating the modifications in the procedures into the testing situation, for pragmatic and methodological reasons. Pragmatically, it is convenient and directly applicable for use by psychological practitioners. Methodologically, it avoids problems related to the measurement of change (Cronbach and Furby, 1970).

Testing Conditions Leading to Increased Levels of Performance

First, several types of commonly used variables were isolated which can be incorporated into testing procedures and be independently manipulated (Carlson and Wiedl, 1979).

The testing procedures isolated involve different techniques of optimizing cognitive functioning. Thus, controlled comparison between types of approaches often used, but heretofore not systematically analyzed in the dynamic testing approach, was made. The dependent measure was the Raven Coloured Progressive Matrices (CPM), a test of nonverbal intelligence (Raven, 1965). The CPM was chosen as it is considered to be a relatively "culture fair" test and a good measure of general intelligence. The subjects tested were 434 second- and fourth-grade chidlren from the general population. The testing conditions (C) employed were:

$C_1$ Standard Instruction, according to the procedures outlined by Raven.

$C_2$ Verbalization During and After Solution, which requires the child to describe the main stimulus pattern prior to searching for the correct distractor and, after a particular distractor is chosen, to explain why he made the choice he did.

$C_3$ Verbalization After Solution, which involves the child describing the reasons for his choice after the choice is made.

$C_4$ Simple Feedback, where the child is informed only if he is correct or incorrect.

$C_5$ Elaborated Feedback, which involves, in addition to simple feedback, an elaboration by the experimenter of the reasons why the chosen distractor was correct or incorrect. The principles involved in the task were pointed out.

$C_6$ Elaborated Feedback plus Verbalization During and After Solution, which is a combination of Conditions two and five. That is, it involves the child's verbalization of the pattern to be completed, followed by solution and the child's explanation for the reasons for solution, and elaborated feedback by the experimenter informing the child of the correctness of his response and explaining the principles involved in the task.

In all cases, the child's initial response, regardless of condition, was scored.

For the second-grade children, Conditions 2, 5, and 6 led to higher levels of performance than were found for Conditions 1, 3, and 4. The most effective conditions were 5 and 6. For the fourth-grade children, Conditions 2 and 6 led to improved performance. A trend in this direction was noted under Condition 5. Thus, three conditions for younger children and two procedures for older children were found which appeared useful for improving test scores, i.e., bringing performance closer to actual competence.

Beyond ascertaining the main effects of the testing conditions on total score CPM performance, closer analysis of the results was

obtained by analyzing the interactive effects of performance with differentiations which the CPM allows in administration and scoring. The test can be administered in either the normal booklet form or in a puzzle, broad-form version. In the former, covert proceses must be used by the subject as he searches for solution. In the latter, overt trial-and-error procedures can be used. Differentiations in the test were distinguished through previous factor analysis (Wieul and Carlson, 1976). The factors isolated were: simple pattern completion, pattern completion through closure, and reasoning by analogy. (See Carroll and Maxwell, 1979 and Carlson and Jensen, 1980.)

Results of analysis of these differentiations showed that the puzzle version led to higher levels of performance for the second-grade, but not the fourth-grade children. The salient item groups were simple pattern completion and pattern completion through closure. The effects of the optimizing testing conditions could be traced mainly to those items requiring abstract thinking.

On the basis of these results, two types of studies were conducted. In one type, the goal was to replicate and extend the above findings by investigating whether or not the salient testing conditions have the effect of optimizing performance for children from special populations. In the other type, in addition to extending general applicability, the goal was to assess the compensatory effect which the procedures could have in reducing differences between particular groups of children.

Children with Learning Difficulties

In a first study involving learning handicapped children (Carlson and Wiedl, 1978), the same testing paradigm as described for the first study was used. In addition, however, the test was readministered under $C_1$ to all subjects after a two week interval. The sample consisted of 108 special school chidlren, mean age 10.5 years, mean WISC-IQ 70.8. Main effect differences were found for testing condition and repeated testing. Conditions 4, 5, and 6 led to higher performance than found under the other conditions. No testing condition by repeated testing interaction was found. The results are very similar to those obtained with children of average IQ, adding further evidence for the effectiveness of the salient conditions.

In addition to the main effect found for testing condition, a main effect due to test form (booklet-puzzle) was detected, with performance on the puzzle-form being higher than on the booklet-form. Further, it was evident that for the puzzle version of the test Conditions 5 and 6, but no Condition 2, were effective.

Localization of the effects to the subparts of the CPM revealed that testing conditions affected performance on the reasoning items only. Conditions 5 and 6 proved to be the most salient conditions. The effect of the board-form could be localized to pattern completion and gestalt closure items.

The most important results described in the preceding sections are summarized in Table 1.

---

Insert Table 1 about here

---

## Table 1

Summary of Salient Testing Conditions with Different Groups of Subjects

| Groups of Subjects | Testing Conditions: Differentiated $C_2$ Verbalization | Elaborated $C_5$ Feedback | Verbalization $C_2$ + $C_5$ + Feedback | Board-Form |
|---|---|---|---|---|
| | Item Requirements ++ | | | |
| | T p c r | T p c r | T p c r | T p c r |
| Regular School | | | | |
| Children 2nd grade | o o | ● ● | ● ● | ● ● ● |
| 4th grade | ● ● | | ● ● | |
| Children with Learning Difficulties (avg. 4th grade) | | o o | ● ● | ● ● ● |

\+ o = slight
● = marked

++ T = total test score
p = pattern completion items
c = closure items
r = reasoning items

Personality and Cognitive Style Variables

The above presentation of the dynamic assessment approach has been concerned primarily with the validation of the dynamic assessment procedure in terms of effects on mean performance scores. Beyond this, however, it is important to investigate how and to what extent learning characteristics (personality and cognitive style) of the child interact with dynamic testing approaches.

Several studies have stressed the interaction of learning and personality characteristics. Two will be mentioned here that relate to one particular personality (style) characteristic: impulsivity-reflectivity. Bush and Dweck (1975) have concluded that task demands influence strategies children use who display reflective information processing while impulsive children seem to be unaware of the strategies they should match with the demand of the task. In other words, they don't adapt strategies to task requirements. In a study relating the dynamic assessment approach to impulsive children, Wiedl (1980), hypothesized that the performance of impulsives could be made equal to that of reflectives by modified, more appropriate testing procedures. The subjects were 150 second-grade children, average age 7.6 years. The Matching Familiar Figures test (MFF) was administered to all children to assess degree of impulsivity-reflectivity. Based on their MFF performance the subjects were divided into two groups, impulsives or reflectives. The classification was obtained by medtansplit on the error score of the test. The CPM was administered under testing conditions 1, 2, 5, and 6.

The results showed that impulsives scored lowest on the CPM under the standard testing condition ($C_1$). No differences between the impulsives and reflectives were detected under Conditions 2 and 5. That is, significant gains in performance were noted for the impulsives, but not for the reflectives when conditions of verbalization ($C_2$) and elaborated feedback ($C_5$) were employed. Under Condition 6, no compensatory effect was detected for the impulsive children. In addition to the MFF error score classification, children were also categorized as impulsives or reflectives by a median-split on latency to response. For this classification, no compensatory effects attributable to testing condition were detected in matrices performance.

Compensation with Regard to Ethnic/Racial Differences: Anglo-American, Black-American, Hispanic-American Children.

The potential compensatory effects of the dynamic testing approach in assessing cognitive capabilities of three different ethnic groups, Anglo, Black, and Hispanic-American, was studied (Dillon and Carlson, 1978). Several investigations have been carried out comparing performance of these groups on tests like the Raven matrices and various Piaget tests. Most results indicate that Black children tend to perform approximately one standard deviation below Anglos, with Hispanic-American performance half-way between. It was hypothesized that the testing conditions shown to be salient in the previous studies would compensate for factors which may contribute to the relatively low Black and Hispanic-American performance.

The tests used were Matrices and Order of Appearance taken from a recently developed Piaget battery (Winkelmann, 1975). The testing procedures involved three conditions: Standard ($C_1$), Verbalization During and After Solution ($C_2$), and Verbalization During and After Solution plus Elaborated Feedback ($C_6$).

The tasks were administered to 189 children, representing three American ethnic groups: Anglo, Black, and Hispanic-American. The subjects ranged from five to ten years of age. Randomization and assignment to testing condition was carried out for three age levels: 5-6, 7-8, and 9-10 years. This resulted in a 3x3x3 randomized block design, with three levels of test administration, three age levels, and three ethnic group classifications. There were seven subjects in each cell.

Since results for the Matrices and Order of Appearance tests turned out to be essentially the same, only those for the Matrices will be given. They are summarized in Table 2. Inspection of the

---

Insert Table 2 about here

---

mean standard deviations reveals a marked decline in differences in performance between the groups. The differences ranged from approximately two-thirds of a standard deviation between Anglo and Black performance under Condition 1, to essentially zero under Condition 6. Although this could be attributed to a ceiling effect for the older group (means of just over 6 from an 8-item test), such an interpretation is not plausible for the younger and middle groups (means

Table 2

Means and Standard Deviations of Matrices for Three Age Groups, Ethnic Groups and Testing Conditions (C).

| | age: 5-6 | | | age: 7-8 | | | age: 9-10 | | |
|---|---|---|---|---|---|---|---|---|---|
| | $C_1$ | $C_2$ | $C_6$ | $C_1$ | $C_2$ | $C_6$ | $C_1$ | $C_2$ | $C_6$ |
| **Anglo** | | | | | | | | | |
| Mean | 1.29 | .71 | 2.43 | 3.43 | 2.43 | 4.29 | 4.14 | 6.00 | 6.29 |
| SD | .95 | 1.11 | 2.23 | 2.37 | 1.51 | .95 | 2.80 | 1.73 | 1.60 |
| **Mex-Am** | | | | | | | | | |
| Mean | .57 | 1.71 | 2.43 | 2.14 | 2.71 | 4.00 | 4.29 | 5.71 | 6.14 |
| SD | .98 | 1.70 | 2.23 | .90 | 1.50 | 1.41 | 2.29 | 1.38 | 1.07 |
| **Black** | | | | | | | | | |
| Mean | .57 | .86 | 2.00 | 1.43 | 2.71 | 4.14 | 2.71 | 5.72 | 6.00 |
| SD | .79 | .85 | 1.83 | 1.40 | 1.60 | .90 | 1.38 | 1.25 | 1.00 |

— difference not significant

___ difference significant

around 2 and 4 respectively). It is interesting that for all age groupings, regardless of ethnic classification, significant increases in performance were found under Condition 6 as contrasted with Condition 1. In addition, the verbalization condition ($C_2$) was effective only for the older age groups, again, regardless of ethnic classification. Apparently, for these groups the appropriate verbal mediation strategies are available.

Summary

The approach of applying optimizing testing procedures to the normal population as well as to special groups of children was demonstrated. Restrictions and differentiations in the effects of testing procedures, test format, and task demands were found. The applicability of certain testing procedures for purposes of assessing intellectual performance of normal and special school (learning handicapped, low IQ) children was shown. This indicates that specific forms of performance optimization can lead to more accurate, thus fairer, assessment of intellectual functioning. It was demonstrated that specific dysfunctional individual characteristics such as impulsivity can be compensated for under certain testing conditions. This underscores the necessity of critically analyzing how variations in testing procedure can interact with certain individual differences. The implication here is that the practitioner must be aware of these interactions and under which circumstances performance may be increased for some children, while decreased for others.

The indication that compensatory effects can be obtained in assessing between ethnic group differences is of practical and

theoretical significance. Extension of the approach by employing other measures of Level II intelligence (Jensen, 1973) to larger samples with control over factors such as socio-economic status is necessary before firm conclusions can be reached, however.

## References

Brown, A., & French, L. The zone of potential development: implications for intelligence testing in the year 2000. Intelligence, 1979, 3, 255-273.

Carlson, J. S. & Jensen, C. M. The factorial structure of the Raven Coloured Progressive Matrices: A reanalysis. Journal of Educational and Psychological Measurement, 1980, in press.

Carlson, J. S., & Wiedl, K. H. The use of Testing-the-Limits procedures in the assessment of intellectual capabilities in children with learning difficulties. American Journal of Mental Deficiency, 1978, 82, 559-564.

Carlson, J. S., & Wiedl, K. H. Towards a differential testing approach: Testing-the-Limits employing the Raven Matrices. Intelligence, 1979, 4, 323-344.

Carroll, J., & Maxwell, S. Individual differences in cognitive abilities. In M. Rosenzweig & L. Porter (Eds.), Annual Review of Psychology. Palo Alto: Annual Reviews, Inc., 1979, pp. 603-640.

Cleary, T. A. Test bias: prediction of grades of negro and white students in integrated colleges. Journal of Educational Measurement, 1968, 5, 115-124.

Cleary, T. A., Humphreys, L., Kendrick, S., & Wesman, A. Educational uses of tests with disadvantaged students. American Psychologist, 1975, 30, 15-41.

Cronbach, L. & Furby, L. How should we measure "change"--or should we? Psychological Bulletin, 1970, 74, 68-80.

Darlington, R. Another look at "cultural fairness." Journal of Educational Measurement, 1971, 8, 71-82.

Detterman, D. A job half done: the road to intelligence testing in the year 2000. Intelligence, 1979, 3, 295-306.

Dillon, R., & Carlson, J. Testing for competence in three ethnic groups. Educational and Psychological Measurement, 1978, 38, 437-443.

Eichhorn, H., & Bass, A. Methodological considerations relevant to discrimination in employment testing. Psychological Bulletin, 1971, 75, 261-269.

Flaugher, R. L. The many definitions of test bias. American Psychologist, 1978, 33, 671-679.

Guthke, J. Zur Diagnostik der intellecktuellen Lernfahigkeit. Stuttgart: Klett, 1977.

Haywood, H. C. Alternatives to normative assessment. In P. Mittler (Ed.), Research to Practice in Mental Retardation. Baltimore: University Press, in press.

Raven, J. C. Guide to Using the Coloured Progressive Matrices, Sets A, Ab, B. Dumfries, Scotland: Grieve & Sons, 1965.

Resnick, L. The future of IQ testing in education. Intelligence, 1979, 3, 241-253.

Schmidt, L. Testing the limits in Leistungsverhalten: Moglichkeiten und Grenzen. In Duhm, E. (Ed.), Praxis der klinischen Psychologie. Band II. Gottingen: Hogrefe, 1971, pp. 9-29.

Wiedl, K. H. Kompensatorische Interventionen im Rahmen intelligenz-diagnostischer Untersuchung bei kognitiv impulsiven Kindern. Klinische Psychologie: Forschung und Praxis, 1980, in press.

Wiedl, K. H. & Carlson, J. S. Dynamisches Testen bei Lernbehinderten mit dem Farbigen Martizentest von Raven. Heilpadagogische Forschung, in press.

Winkelmann, W. Test zur Erfassung kognitiver Operationen. Braunschweig: Westerman, 1975.