DOCUMENT RESUME

ED 054 849 PS 004 842

AUTHOR Reid, J. Christopher

TITLE A Critique of Film Tests of Visual Cognition and Memory.

PUB DATE Feb 71

NOTE 6p.; Paper presented at the American Educational Research Association Convention, New York, New York, February, 1971

EDRS PRICE MF-$0.65 HC-$3.29

DESCRIPTORS *Films; *Memory; Testing; *Testing Problems; *Test Reliability; Test Validity; *Visual Perception; Visual Stimuli

ABSTRACT

This paper discusses some methodological, theoretical and physiological issues to be considered in the use of film tests in research relating to visual cognition and memory. One trouble spot in methodology is the use of a square data matrix with an insufficient number of observations in proportion to variables. A second methodological problem arises from a failure to use the best tools available to determine reliabilities of experimental tests. Another concern is the theoretical problems involved in film test research. Although test authors may define a trait, the designing of a test to measure that trait is not always successful. The final concern relates to the problems of human perception. Care needs to be taken that the observer can see a stimulus nonperipherally in the experimental situation. In short, while the findings of film tests studies remain as interesting as ever, their role in discrimination and prediction will increase only after their practical utility is established. (WY)

# A CRITIQUE OF FILM TESTS OF VISUAL COGNITION AND MEMORY

Presented at the American Educational Research Association Convention
February 1971

J. Christopher Reid
University of Missouri - Columbia

My colleagues have discussed their research and experiences with film tests of cognition and memory. They have shown that film tests may tap unexplored cells of Guilford's Structure of Intellect, that film tests are practical to administer, and that traits measurable by film tests and perhaps only by film tests can be replicated.

In this symposium my position is that of the caboose, and the job of the man in the caboose is to look ahead at the freight that has gone before, watch out for hot boxes--potential trouble spots, and to stop the train momentarily, if need be, to attend to the hot boxes.

In the journey thus far I have noticed three types of hot boxes that deserve attention by the engineers.

The first type of hot box is caused by methodological problems in analyzing the data. The work by Seibert and Snow (1965) on film testing, affectionately called the "Green Report" because of its green cover, used a data matrix with 100 subjects and 96 variables. A square data matrix capitalizes on chance to a great extent. To obtain data on 96 variables you should have nearly a thousand observations. We generally prefer to have several times as many observations as variables. Although this point has been made before, the fact that it is sometimes forgotten indicates that it needs to be made again.

The second film testing report, called the "Yellow Report" (Seibert, Reid, and Snow, 1967) and the McDaniel-Kephart report have a somewhat more comfortable ratio of observations to variables: between 4 and 6 to 1.

---

The preparation of this paper was supported by a grant from the Markle Foundation.

A second methodological problem arises from a failure to use the best tools available to determine reliabilities of experimental tests. Cronbach's alpha serves as an upper estimate of equivalent-form reliability, and it is easy to do by machine. Reliability can also be computed by analysis of variance procedures. Despite the relative ease with which reliability estimates may be obtained, only one third of the 96 experimental variables in the "Green Report" have Kuder-Richardson estimates. For the remaining 65 variables, the authors give the communality as the estimate of reliability. One could argue that communalities are lower-bound and thus conservative estimates of reliability. True. However, the communalities that were reported are only estimates of communalities, and so we have the situation of trying to estimate a parameter by a statistic that is two generations removed. It is like trying to guess what your daughter will look like by looking at her grandmother. Why disturb her grandmother when her mother is right beside you? The second film testing report, the Yellow Report, does show progress since it reports K-R 20's for 17 of 23 experimental tests. The general Kuder-Richardson is, of course, a special case of Cronbach's alpha. The Yellow Report also has one test-retest reliability (p.24). A film test called Short Term Color Memory I has a test-retest reliability of .82, quite respectable for an experimental test. This figure is surprisingly high when you consider how it was obtained. Normally, in the determination of a test-retest reliability, subjects are given a test, then perform some unrelated task, and then given the test again. In this study of film tests, subjects were given this test, then had about five hours of similar tests on which to practice, and then were given the retest. That the retest reliability was as high as .82 despite all this practice deserves more emphasis than it received in the report.

Three of the reliability estimates for the 10 McDaniel-Kephart film tests are too low to be acceptable and most could be improved. The authors are aware of this and suggest ways of improvement. McDaniel and Kephart also conducted item analyses, but they used another achievement variable to determine high and low groups rather than test total score. The use of an outside variable in partitioning subjects for an item analysis has certain problems and I would suggest that their final report dispense with that procedure and instead use the total score of that particular test when analyzing.

Some interesting questions arise upon examining the film test data in a

multitrait, multimethod (Campbell and Fiske, 1959) perspective. All film investigators were conscious of correlations of film tests with other standardized or experimental cognitive tests. McDaniel (1971) listed 91 significant correlations of the ten film tests with subscores of 9 standardized achievement and ability tests, and pointed out that these correlations were acceptably low. The highest of these 91 correlations was only .55 (Temporal Memory Span with Gates Reading speed); one fourth of these correlations were in the 40's and the remainder were usually in the 30's. Assuming the 10 film tests are reliable, they are not measuring what the standardized achievement tests are measuring. Their shared common variance is 25% at the largest and typically is 10%. If the film tests were extremely easy, then their smaller range and/or the skewed distribution will have shrunk their correlations with these standardized criteria having a different--presumably symmetric--distribution. If the correlations remain small after test revision, and the McDaniel-Kephart film tests prove reliable, then the next steps would be the seeking of instruments that film tests correlate negatively with and the designing of other (non-film) methods to measure and support the hypothesized traits.

Let me close this section with a point of commendation about the Yellow Report. Chester Harris (1967) pointed out that if a data matrix is factored by two or three methods rather than by one method, greater faith can be placed in those factors common to all methods. The Yellow Report used three different methods of "factor" analysis and noted that all three solutions were nearly the same.

So much for methodological problems. The second hot box concerns theoretical problems. A strength of all the reports has been the authors' efforts to define a trait, and then to design tests to measure that trait. But the designing of a test to measure a trait is not always successful. Suppose you design a test for memory for figural transformations. If subjects actually process the information pertaining to the test as memory for figural units then the test measures memory for figural units, and not memory for figural transformations. A factor analysis may not offer wholly decisive results, since other tests for figural transformations may be complex and also contain some figural units

variance. As an example: the time-space translation test presents two colored pegs that move across a colored checkerboard. Do subjects remember the movement of the pegs or do they simply remember the final position of the pegs? The criterion task only demands that subjects remember the final position of the pegs. Does the test measure time-space translation or does it measure simple positional memory? This whole issue of validation can be directed at several of the film tests reported. McDaniel and Kephart hypothesized that their ten tests would describe four factors. A Kaiser image analysis (Kaiser, 1963: Reid, 1968) and a components analysis (little jiffy) on their ten film tests both resulted in only two "factors," not four, and both factors are complex. Although their four hypothesized factors have an appealing theoretical rationale, the implementation of their four factors into film tests has not been satisfactory so far.

Let me give two more examples of possible theoretical problems before proceeding to the third and last hot box.

Both the Green and Yello Report describe a "Fleishmanesque" analysis. Three of their film tests had items that were presented for various lengths of time. The authors found that subjects used a differing proportion of abilities to process information appearing for different lengths of time. Some abilities were not used at all at some time intervals. Unfortunately, the findings in the Green Report were not wholly replicated in the Yellow Report. Further, no satisfactory hypothesis has yet been generated to explain the differences that occur across intervals as short as tenths of a second. Two things need to be done: first, we need to know what proportion of differences across replications are due error variance and what are due to true variance. Second, when the amount of true variance is known, a hypothesis should be generated and tested to explain the variation in true variance across stimuli differing in time interval.

One of J. J. Gibson's (1954) more interesting ideas was the interaction between vision and kinesthesia. In a film testing situation, no interaction can occur, since the subject's behavior is being sampled at a constant (motionless) state. My last theoretical query is: To what extent can the results from the film test studies be generalized to the normal state of an observer in motion?

I turn now to the third and last hot box, the hot box of physiological

problems of human perception.

Care needs to be taken that the observer can see the stimulus. Both the Green and Yellow reports indicate that viewing distance has a significant negative loading on the serial integration factor. The serial integration factor should be replicated in an experimental situation where viewing distance is not a critical factor.

With slow-moving stimuli, subjects' eyes move rapidly enough so that detail is not lost in peripheral vision. With short, tachistoscopic stimuli, subjects' eyes may not be able to move fast enough to encompass all the stimuli by non-peripheral vision. If peripheral vision is required then as stimulus time intervals grow shorter, subjects will not obtain information for color and form even though they may still obtain information requiring visual acuity. For example, subjects looking at a tachistoscopic array of colors might not see all the colors, whereas they might be able to discern a similar array of broken circles. It is not enough to say that the subject is seated 6 feet from the screen. The experimenter must give assurance that the angle of view is small enough that all the stimulus can be seen non-peripherally, or must demonstrate that the subject obtains all the essential information by eye-movement.

Horizontal angle of the viewer from the screen is a third consideration. The viewer off to the far right or left of the screen sees a different and distorted image. Apparently, however, viewers have been seated within horizontal angle tolerances.

The findings of the film tests studies remain as interesting as ever. More attention should be given to their role in discrimination and prediction, of which the McDaniel-Kephart studies are a start. I would like to prophesy that unless a practical utility can be found for the film tests, they will gather dust and be sold for scrap. The train of film tests has had a few hot boxes, but nothing insuperable. I trust that the train will reach a useful destination, and not just go down the track, becoming smaller and smaller and finally disappear into the sunset.

## BIBLIOGRAPHY


Campbell, D. T., and Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 1959, 56, 81-105.

Gibson, J. J. The visual perception of objective motion and subjective movement. Psychological Review, 1954, 61, 304-314.

Harris, C. W. On factors and factor scores. Psychometrika, 1967, 32, 363-380.

Kaiser, H. F. Image analysis. In C. W. Harris (Ed.) Problems in measuring change. Madison: University of Wisconsin Press, 1963.

McDaniel, E. New film tests of visual perception. Paper read at the American Educational Research Association convention, New York City, February, 1971.

Reid, J. C. Kaiser image analysis. Behavioral Science, 1968, 13, 259.

Seibert, W. F., Reid, J. C., and Snow, R. E. Studies in cine-psychometry II: Continued factoring of audio and visual cognition and memory. Lafayette, Indiana: Purdue University Audio Visual Center, 1967. (USOE Grant No. 7-24-0280-257.) [The "Yellow Report."]

Seibert, W. F., and Snow, R. E. Studies in cine-psychometry I: Preliminary factor analysis of visual cognition and memory. Lafayette, Indiana: Purdue University Audio Visual Center, 1965. (USOE Grant No. 7-12-0280-184.) [The "Green Report."]