ISBN: 978-989-8533-18-0

AN ALTERNATIVE APPROACH TO TEST ANALYSIS AND INTERPRETATION

J. C. Powell
Better Schooling Systems

ABSTRACT

This reflection paper challenges current test scoring practices on the grounds that most wrong-answer selections are thoughtful not random, presenting research supporting this proposition. An alternative test scoring system is presented, described and its outcomes discussed. This new scoring system increases the number of variables considered, reduces the mesh of the analytical screen and provided considerable more information to inform teaching.

KEYWORDS

Alternative scoring procedures; Cognitive maturity scales; Conformity *vs.* diversity. New statistical procedure for test analysis; Score validity; Test profile interpretation; Qualitative *vs.* quantitative testing;

1. INTRODUCTION

This paper takes a critical look at current multiple-choice test-scoring practices. Consider Dr. Friedrich A. von Hayek's (1974) Nobel Prize acceptance lecture "The Pretense of Knowledge."

There is much reason to be apprehensive about the long run dangers created in a much wider field [than economics] by the uncritical acceptance of assertions which have the appearance of being scientific [because] what looks superficially to be like the most scientific procedures are often the most unscientific … Beyond this, there are definite limits to what science can achieve [arising] from our inability to quantify some important variables.

This paper quantifies these intractable variables in a novel manner.

2. THE "RIGHT-WRONG" SCORING PERSPECTIVE (LEARNING HYPOTHESIS I)

We currently presume that multiple-choice test scores estimate what students know as a cumulative proportion of an epistemological domain. The items are a representative sampling of course content. The frequencies of "right" answers are assumed proportional to the examinees' "knowledge." Options are often designed to reflect common procedural errors or common misconceptions (*e.g*. concept inventories; Halloun, I. & Hestenes, D. L. (1985)).

Most "wrong" answers are randomly selected and unrelated to the "right" answers (making scoring into a dichotomous process.) The frequency of "right" answers is necessary and sufficient information base about examinees' subject-matter knowledge.

An Editorial in *Science* (Coffey & Alberts, 2013) suggests that:

> "The [Common Core] contains a vast number of core disciplinary ideas and subideas. Current measurements and approaches do not allow these performances to be assessed easily…"

In this approach we can expect a number of supporting observations. The most important of these are:

1. The distribution of scores will resemble a typical *normal bivariate distribution* (Frame 1).
2. The changes of answer selections from one administration to the next (whether the same test or a parallel one) will be from any "wrong" answer to the "right" one (Frame 2).

3. That changes of answer and direction of learning increase will progress in equivalent directions (*the equivalency assumption*; Frame 3).

Current practice presumes there can be only one reasonable "right" answer for a question. Learning requires the exchange of "wrong" answers for "right" ones (Frame 1). Answering is a dichotomy whose distribution of total frequencies resembles the upper limit of the binomial expansion $(a + b)^n$ as $n \rightarrow \infty$. This is the *normal bivariate distribution* (Frame 2). The only meaningful changes are those in the total-correct scores. (Frame 3).

3. "MOST ANSWER SELECTIONS ARE THOUGHTFUL" (LEARNING HYPOTHESIS II)

Education suffers non-measurability as much as any other discipline. We can compare live data with the hypothetical patterns postulated in H_I. Which theory stands the test of oservation. If the results from live data show patterns different from the expectations for H_I, it must be rejected in favor of H_{II}?

4. WHAT DO MULTIPLE-CHOICE TEST ACTUALLY MEASURE?

Extending over more than half a century, interviewing students about their "wrong" non-random answer selection distributions (Frame 4), using thousands of students, a broad age range and more than one continent and more than one subject matter area showed *item interpretation* to be involved with the following observations:

1. Teaching interpretation skills causes huge increases in students' learning motivation (Powell, 2010a).
2. Using written selection explanations with adults, Powell (1968), showed explanations often predicted selections,
3. Testing 550 students (with Gorham, 1956) from the third through the eighth grade, showed (Frame 4; Powell, 1977) "wrong" answer subsets systematically ordered perfectly with chronological age (CA; Frame 5). More than half these correlations are *conjoint* (double-headed arrows; Luce & Tukey, 1964).
4. Giving this test twice (October-March) to 2,000+ students (third grade through the end of high school) supported this developmental sequence. We defined students who followed the observed developmental sequence as showing increasing *cognitive maturity* (CM). This study showed four developmental pathways, refuting the dichotomous data assumption and displaying the dynamics of learning (Frame 6):
    a. A Piaget-like (1953) normal development (about 40% of the students),
    b. Declining CM (about 30% of the students),
    c. Students stalled at the *literal* (Concrete Operations) *thinking* level at a preadolescent level interpretation of questions (about 20% of the students) and
    d. A mature age multidimensional expansion of thinking skills beyond 2-value logic. Shown by shifting from a "right" answer to a "wrong" one (the remaining 10% of the students).
5. Using a representative sample (52) of these about 3,000 students, we drew developmental profiles of their CM changes (Frame 7), which shows a student whose CM declined while total-correct score increased and a student with the opposite pattern of change. These two contrary-to-expectations results are sufficient to refute H_I. Powell and Powell showed that the 10% can be increased to at least 40%.
6. Using 16,000+ students from years 4, 6 and 8 in India and two tests, one in science and one in mathematics, Powell, Bernauer and Agnihorti (2012) showed that item interpretation was the key ingredient for answer selection with different tests. It showed developmental sequences, procedural, information style and cultural bases for answer selection of "wrong" answers (No Frame).
7. Using the multidimensional transition to identify high order changes among students Powell and Powell (2012) showed that when teaching interpretation skills replaces transmitting information a six-fold increase in multidimensional thinking among college undergraduates occurs (No frame).

8. From a representative sample of the data from study 4, Powell, (2013) showed (Frame 8), the changes in *cognitive maturity* (CM) are unrelated to changes in total correct score and that the latter score misclassified more than two thirds of the students from the corresponding results drawn from the same test. (Compare this result from live data with the anticipated result from random "wrong" answering given in Frame 3). Current scoring practice is *invalid!*

5. CONCLUSIONS AND IMPLICATIONS

When including all answer-selection variables, multiple-choice tests measure *interpretation skills* independently of *knowledge* of the subject matter content of the questions except that fluency with the technical language of the item is partially involved. These skills are not dichotomous. Instead they can be expanded into logics of levels higher than 2 and this shift is observable using the *Thurs ()*. Adding the "wrong" answer variables to the mix makes these data multinomial, not binomial, invalidating current test-scoring practice at the expense of two-thirds to three-fourths of the data available from each student from a single administration of any test and more than this when the tests are administered more than once to expose the learning dynamics. When "wrong" answers are omitted from the scoring process, this loss of data destroys the integrity of the matrix.

6. THE FIX

The algorithm for calculating the *Thurs ()* [*þ*] (Frame 9), that bypasses linear dependency and detects non-linear data structure, is now in public domain for private use. The interpretation strategies found from interviews data from the 1977 study and the data samples from the 1992 and the 2013 studies may be requested from the author. The author will consult with any interested party on application of the *þ* or educational effective improvement strategies.

REFERENCES

Coffey, J. and Alberts, B. Improving (2013). Educational Standards, *Science 399*, p. 489

Gorham, D. R. (1957). *Proverbs Test*. Missoula, MT: Psychological Test Specialists

Halloun, I. & Hestenes, D. L. (1985). Common sense about motion. *Am. J. of Physics, 53*, (*11*) 1056 – 1065

Hayek, F. A. von. (1970) The Pretense of Knowledge. Nobel Prize Lecture.

Luce, R. D. & Tukey, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology, 1* (1), 1-27.

Piaget, J. (1953). *Logic and Psychology.* Manchester, UK: Manchester University Press

Powell, J. C. (1968). The interpretation of wrong answers from a multiple-choice test. *Educational and Psychological Measurement, 28*, 403-412.

Powell, J. C. (1977). The developmental sequence of cognition as revealed by wrong answers. *Alberta Journal of Educational Research. 23* (1) 43 – 51.

Powell, J. C., (2010a). *Making Peasants into Kings.* Bloomington, IN: Author House; ISBN: 1-978-4490-0634-1.

Powell, J. C. (2010b). Testing as feedback to inform teaching. In J. Michael Spector, *et al. Learning and Instruction in the Digital Age: Making a Difference through Cognitive Approaches*. New York: Springer. Chapter 3.

Powell, J. C. (2012). Are We Assessing Students Using the System Noise from Our Data Matrices? Paper presented to the 77th Meeting of the

Powell, J. C & Shklov, N. (1992). Obtaining information about learners' thinking strategies from wrong answers on multiple-choice tests. *Educational and Psychological Measurement, 52*, 847-865.

Powell, J. C. & Powell, V. J. H. (2010). Do profoundly informed students choose wrong answers, lowering their scores? Presentation to the *Psychometric Society*; Athens, GA.

Powell, J. C., Bernauer, J. and Agnihor, V. (2012). Learning from Incorrect Answers on Multiple Choice Tests: Implications for a New Teaching Chap. 16 in *Towards Learning and Instruction in Web 3.0. Advances in Cognitive and Educational Psychology.*

Support materials: Available on request from jpowell@tir.com, Phone: +1-412-595-7190.

1. Table of strategies identified.; 2. Profile data for 52 students.; 3. The *Thurs* (*þ*) program, in the process of being reimplemented for flexibility of private application, is available upon request.

Sample lesson plans can be found, along with an extensive discussion of the delivery model can be found at the website www.better-schooling-systems.org