Refining Measurement
by

Eric C. Haughton*
(8 November 1983)

Precision Teachers offer some of the most accurate, sensitive and valid formative data available in the human service area. The use of B.F. Skinner's frequency as a basic datum and of O.R. Lindsley's monitoring of successive frequencies of thousands of human behaviors account for Precision Teaching's accuracy, sensitivity and validity.

Some of us in teaching or research run into confrontations with our historical measurement precedents. Test styles have been so rigid for the past 40 years that new approaches, however clear and operational, are often placed on the defensive. My own experiences as an early field advisor and supervisor causes me to be sensitive and practiced in discussing this interesting and complex area. In the past few years I've been developing an overhead and handout to attempt to explore this labyrinthian, nether region.

Contrasts

Repeated measurement forms a cornerstone of the foundation of information different from traditional or commercial testings. Testing attempts to relate group data to static individual data. Actuarial data (such as insurance and testing companies use) cannot forecast an individual's outcomes. Precision Teachers monitor performance on the standard celeration chart to observe and to forecast change. As we measure individual and program concerns, we compare the individual's charts to group or other reference data. Commercial tests cover the waterfront on a wide spectrum of topics to meet market demands and administrative needs-- not the needs of individuals. Both measuring systems attend to the two major quantities: Quantity 1 is temporal (calendar and interval); Quantity 2 is the content of performance. Frequency is our constant unit of information, while standardized tests inherently obscure frequency data.

Both systems work to insure the accuracy of data. Precision Teachers break performance into significant packages to explore and meet aims such as corrects, skips, and learning opportunities. Traditional testing generally relates only to accuracy, while obscuring accuracy into the artifacts of percentile and stanine rankings.

Measuring - from Testing to Monitoring

Commercial/Personal. Standardized tests often commit to multiple-choice and machine scoring formats for economic considerations. These formats can be intimidating and distracting to both behavers and managers. Precision Teachers strive for a fully informed team of learner and instructor, involving usual and relevant behaviors, high comfort and trust levels. Data on my personal pinpoints are for, and belong to, me personally.

Minifeedback/Maxifeedback. In the worst testing situations, teachers do not learn the results. In the best, they receive them, but often find them difficult to interpret, let alone to use them to improve student learning. Behavers who chart regularly receive maximum, immediate feedback while managing their own projects. They operate as self-managers and resource seekers.

Average/Proficient. Standardized tests relate performance to the mean of peers. On any standardized test, by definition, half must fall below the norm. Suppose you are in the first grade, and the mean peer-norm is 50 words correct per minute on oral reading. Is 50 words per minute competent, fluent, proficient? No. Or suppose the mean peer-norm is 250. Is 249 words per minute incompetent, stumbling, in need of improvement? No.

We Precision Teachers use different frames of reference depending on the behaver's interests, desires, and needs. We may ask for a personal aim at the start-- "better than I was." We may use some peer data. In the final analysis, we owe it to each behaver to determine levels that will insure retention, endurance, and application of his/her learning. Many of us recognize this topic as deserving immediate study. Since we are a "Nation at Risk", we

need to determine and implement education based on substantial proficiency levels. Few decision guidelines exist. What performance levels do you use when deciding on a new phase? How much is enough?

Fail/Support. If you cannot answer items on standardized tests, you fail the item. One of the classic IQ items is "What is Mars?" If Robin answers, "Candy bar," she fails. In Precision Teaching we structure monitoring to support Robin through changes. There is no failure in Precision Teaching. ("Feel better, Robin?")

Snapshot/Continuous. Testing, even pre and post-testing, is a one-shot event. In Precision Teaching we check performance repeatedly based on regular calendar cycles.

Now/Forecasts. Not navigators, weather folks, physicians, physicists, behavers, managers, not you, and certainly not I, can forecast from a single observation. Therefore snapshot, commercial tests offer at best only a static hint about a person's strengths and needs. Since we don't know the rate of change, we can't estimate the intensity of intervention required, or if any is required. Ten day screening data improves people at x1.3 M/m/week. We Precision Teachers have learned not to project a flat line from an initial frequency, an unfortunate and incorrect assumption in current testing and statistical approaches. Slope is one of our big power pieces for understanding measurement and individuals.

Unrelated/Relevant. U.S. federal law 94-142 requires that measurement relates to behaver's program and goals. Good-bye I.Q.! Adios, much traditional diagnostic and labeling testing! Au revoir, heterogeneous test sections! Hello, valid, reliable, usually homogeneous items with SENSITIVITY! We must give more consideration to the fact that our data are sensitive

Prompt/Produce. Prompted, test-taking behavior in multiple-choice format differs from normal performance ecology. Surely you know several anecdotes about people who have guessed their way to "success" on prompted tests. Monitored performance is similar to real-life production, often requiring multiple, compound channel sites. This is in marked contrast to commercial testing's slavish use of See/Select.

Monoview/Multiview. Traditional tests report some aspect of the quality of effort, then translate the statistically adjusted score, or artifact, into a meaningless grade-level statement. Does anyone know what 4.2 in math means? What exactly does that child know? What is "equivalent to grade 10 level of reading"? Can that person read any chemistry book, any novel? Our data set includes categories of performance (correct, legible, requires improvement, learning opportunities, skips, etc.), as well as presenting the rate of change through the family of Standard Celeration Charts. Changing the rate of change is our goal as we strive to maximize performance gains for each person.

Validity/Valid. Tomes have been written to justify the use of remotely chosen items presented on commercial tests. (If you wish to study this topic from an historical perspective, check the history of "operational definitions.") Our data are valid because when we measure a topic, we measure our area of programming and of concern. This approach allows us to verify empirically and continuously our data in each setting, on each project.

Metaphor/Relation. Perhaps one day society will deem it unprofessional or unethical (or both) to translate raw data into the unknown. We do not know what age 2.6 on a Denver (or any preschool test) means. We are unable to interpret what mental age of 6.9 means. We cannot program for a child who "scores" 8.2 on the language section of the California Test of Basic Skills or the Iowa Test of Basic Skills. On the other hand, directly quantified performance of a specific topic monitored over time, aids everyone's understanding. We Precision Teachers require clear awareness of relationships between events and performance.**

Afterward

About 20 years ago, Og Lindsley presented ideas about the deficits of standardized testing. Maybe it was in a course, perhaps at a local or national conference, maybe in a marathon rap session in some motel room in North America. He pointed out that we were in the process of standardizing the information format and flow relating to people, that we would gain significantly from our implementation of frequency monitoring using standard celeration

---

** We have the tools to emerge fully from the alchemy of education and psychology. The question is how do we improve our measurement and strategies, and thus our communication to the rest of the world about this unique and superior system of measurement. (an editorial comment of Abigail B. Calkin, Consulting Editor.)

charts. Traditional testing has worked strenuously to structure procedures, instructions, page format, administrative minutiae, as well as to attempt to determine appropriate content and sequence.

Overconcern and testing biases applied to inappropriate areas of classroom and research efforts contribute to weakening people. Certainly one of the leading causes of the failure of the education system is its lack of effective monitoring of student performance. If we assume that this lack is one of the lesser causes of education's failure, then we shall never know our error and therefore, shall be unable to correct it.

Historically and presently, performance levels have been seldom observed or recorded prior to the efforts of Precision Teachers. Our expectations are challenging. We support the behaver thoroughly while delighting in his/her gains. We are humble in the realization of the magnitude of our task and the potential gains to individuals and to the communities associated with maximizing personal development.

References

Haughton, E. (1972). Aims, growing and sharing. In J.B. Jordan and L. S. Robbins (Eds.), *Let's try doing some thing else kind of thing*. Arlington, VA: Council for Exceptional Children.

Johnston, J. M. & Pennypacker, H. S. (1980). *Strategies and Tactics in Human Behavioral Research*. Hillsdale, NJ: L. Erlbaum and Associates.

Kunzelmann, H. P. (1980). The learning metric. In *Human diversity and the assessment of intellectual development*. Media Association, Inc.

Lindsley, O. R. (1964). Direct measurement and prosthesis of retarded behavior. *Journal of Education*, 147, 62--81.

Peters, T. J. & Waterman, Jr., R. H. (1982). *In search of excellence*. New York: Harper and Row.

Skinner, B. F. (1953). *The science of human behavior*. New York: Macmillan.

Stevens, S. S. (1960). *The measure of man*. New York: Whitney Library of Design.

Abigail B. Calkin

Dr. Eric Haughton was one of the pioneers of Applied Behavior Analysis and Precision Teaching and also served as a contributing editor to this journal from its inception in 1980. At the time of his death in 1985, Eric was a faculty member at Loyalist College, Belleville, Ontario, Canada.