de Brún Memorial Lecture
Measurement and Learning

Eugene Galanter[1]
Children's Progress

Experimental measurement in psychology commonly refers either to psychophysics or psychometrics. The metrology behind psychophysics presumes that numbers—invariant to some plausible transforms—can represent psychological experience, and in this way display experience as functionally related to physical metrics. So, for example, experienced "loudness" can be systematically related to acoustical power. The success of this enterprise can be appreciated by the application of such metrics to various palliatives or other practical ends, viz: eyeglasses or hearing aids.

These methods give information about the psychological functions of single individuals. They can be made normative by the usual statistical procedures. The scatter of psychophysical data represent the "noisiness" of neurological processes that translate the physics into psychology. Theories of this neural driving often concern the ways that neurological systems, themselves composed of huge populations of synchronous and asynchronous neural activity, preserve highly reliable functional relations between physics and psychology. The main point of psychophysics is that non-physical psychological events and experience can be understood as quantitative representations of physically measured events.

Psychometrics on the other hand measures abstract, and purely psychological properties, capacities, or achievements that have no observable physical counterpart. These numbers applied to individuals, e.g., skill in reading, ability to calculate correctly, psychomotor skills, political attitudes, all derive from early attempts to metricize purely mental processes. These metrologies were driven by the practical need to improve education. The foundation of such metrics, drawn in part from psychophysics, was first made quantitative by the empirical data reported by Binet & Simon (1909), and then validated by Karl Pearson's derivation of the correlation coefficient (Huber, 2004). Spearman's invention of factor analysis (1904), and Thurstone's (1927) scaling methods extended the technology. Psychometrically, an individual is denominated by his position on the abscissa of a distribution of like individuals. Differences between and among individuals are reflected by differences in standard deviation units.

These statistical conceptions, all designed to measure individual psychological quantities: intelligence, motor skill, mental competence, educational skill etc. have served as efficient *management* tools for the past century (Furr & Bacharach, 2007). Their implementation has made many practical inroads to improve teaching and learning by teacher selection, comparative children's performance etc. However, they all fail to advance educational practice in the way psychophysics has enhanced perceptual practice, because theories of learning have not provided solutions to educational problems. Indeed, these very psychometric educational measures have signaled to educational management continuing declines in the schooling of children.

[1] I am indebted to Scott Parker for his critical and technical help in the preparation of this report.

One may reasonably conjecture that part of the schooling problem resides in the failure of theories of learning to guide instruction. We address this issue by noting first that learning theories divide into two major classes (Galanter, 1960); stimulus–response chaining, mostly supported by animal experiments, and cognitive heirarchical structures, the formulation of plans to organize and constrain behavior. In various contexts both theories appear plausible. Indeed S-R theories are commonly associated with memorization and rote practice, whereas heirarchical structures are these days praised for their participation in the “cognitive revolution.”

We may be able to resolve some of these problems by turning from theories of how children learn to theories of human development, and then to some relevant neuropsychological imaging research. How a child grows and develops physically and psychologically has occupied many experimentalists, and has led to a body of data that must be addressed by the questions we raise. There is some agreement that in the early life of children, psychic growth divides conveniently into four segments that we name here proto-lingual (birth to age 3 plus), associative structural (age 3 to 6), hierarchical organizations and cortical pruning (age 5 to 10), intellectual-organizational (8 plus to puberty and beyond). (Stuart Kauffman)

These stages overlap, and often run in parallel. The central operative and practical question is how to take advantage of this structural model. There have been several views about such attempts, but we may summarize them as Piagetian and Vygotskyian. Piaget (1972) believed that structural change during early development imposed limits on what a child could address intellectually at various stages. Vygotsky (1986) believed that any child had a range of possible intellectual capacities, a S<>R view, that could be aided or adjusted by external coaching (sometimes referred to as scaffolding, or hinting). Skinner called this “operant conditioning” to distinguish it from Pavlov’s model.

Turn first to the earliest stages, when the child begins to build his or her linguistic capabilities. Betty Hart and Todd Risley, mounted an eight year effort to record the verbal activities in homes of infants through the first three years of life. The words spoken by the parent’s to their child, and later the child’s words were recorded for one hour every month. The collaborating parents were divided into three empirically distinguishable groups, based mostly on socio-economic criteria. I have replaced these nominative classes by calling such parents talkative, midlevel, and taciturn, independent of their social class. The records of those words by parents, siblings, and child were transferred from tape records to computer, and I show you first a sample of the 30,000 pages of print generated by those folks. The next figures are cumulative data similar to Skinnerian records, that let us distinguish clearly between the three groups.

Much of the talk by parents at this level is used most often to control child behaviors. H & R distinguish between positive and negative control expressions, e.g., “If you don’t eat your veggies, I am going to spank you” vs. “Let’s eat our veggies so we can go out and play.” The data suggest real differences among the parent groups.

H & R summarized the observed differences by converting parent and child talk into a normal distribution and correlated these linguistic interactions with the children into their 4 year old Wechsler-Bellevue IQs.

At this developmental level, I believe these children were forming simple chain-linked associations by contiguity and reinforcement, which changed over time and experience into cognitive heirarchical structures. That such internal reorganization was occurring may be reflected by recent fMRI observations that early brain scans show “pruning,” a reduction in the mass of neuronal associative tissue that occurs in children between 3 and 7 years of age. Kids brains get smaller as they get smarter.

In 2003 my daughter and I received a patent on a new way to assess children's intellectual development, based in part on Vygotskyian principles (Galanter, E. & Galanter, M. 1999); in particular the errors a child makes give more information about the child's mind, than do the correct answers. However, until the existence of broadband Internet and computers in schools, high-speed analysis and records of these interactions were not feasible. The consequence of this central idea leads to a new learning model.

Look at our example based on error analysis. Our technology allows us to make the next posed question contingent on the answer to the previous one. The next question is a Vygotskyian hint. The hint may lead to a correct answer, or give the teacher advice on how best practice suggests appropriate remediation. The structure of such a "test" becomes a tree, leading not merely to a "score," but also to teaching modules to address the problem the child has encountered.

The ramifications of this technology lead directly to a new model of teaching. This tripartite structure requires an assessment technology that is ongoing. The child may pass through a technically equivalent structure, and thereby reveal to the teacher the need for further remediation, or the behavior may suggest new concepts, all designed to promote the development of organizational hierarchies, although they may have originally developed as simple associative ("rote" to use a hateful term) structures. This requirement changes the administration of assessments from a linear sequence of questions, to a hierarchical structure with next questions constrained by last response.

This model removes "testing" from the classroom. The role of assessment is separated from the psychometric role of the summative test in school management. This separation of function retains the viable quantitative features of psychometrics while at the same time offering a quantitative role in teaching children, rather than the often punitive fear-inducing properties of the *test*. In scores of teacher-parent and parent only "round table" discussions, the most common concern about early education was the need for tests. Many parents reported actual nights of fear and trembling by their children; for which their only recourse was to acknowledge the justifiable basis of the fear, and consoling the children with their love and concern.

By offering the teacher suggested helpful remediation, we may take advantage of the corpus of educational research in language arts and mathematics that serve as the foundation for early schooling; age 4/5 to age 10, and so spanning the transformation of associative structures into hierarchical models that transcend the exemplars of the assessment. Not only have we changed assessment from a reporting tool, it now becomes part of the skill-set of the teacher for advancing the rate of learning among her charges.

Whereas we have minimized the role of the correct answer in learning, we cannot deprive management of the capacity to examine the competence of children for which she or he is ultimately responsible. Because the technology captures every keystroke of every child in every classroom, the responsible entity that oversees and supervises the school or schools, whether in a classroom, a school division or indeed in the national school network, can access summary data that is critical to his or her role.

Because the entire system is Internet based, special needs for special children can be identified, and every teacher, every appropriate school principal, educational specialist or special technologist, can call up with a few mouse clicks, summary data for entire the classroom, the school, the regions, indeed for entire nations.

The composite information for a particular classroom can be reconfigured for the appropriate needs of the moment. That capacity also ensures the timely attention to problems that may go unnoticed, while the causal issues for the children may establish inappropriate behavior in the face of their intellectual pursuits.

Questions and critical issues about the vitality of a new metric procedure must show that the new methods reflect older and other ways to address the same problem. We turn then again to the single most respected metrology for this purpose, the Pearson product moment correlation coefficient. This calculation permits a quantitative comparison between the old and the new. We do this by converting the metric performance of children based on the Children's Progress methods that are highly qualitative, by dividing possible end points on the vertical axis of the lattice into a scale ranging from 1 to 4. We use these numbers to reflect our qualitative description of performance as "below expectations" "approaching expectations" "at expectations" and "above expectations. Because our hierarchical lattice is administered at least 3 times during a school year, we can calculate an average over the four point scale. That average serves as our criterion measure. To assess the validity of the this new technology I selected an old style standardized grade three (year 9) assessment used by the New York State Department of Education. The ordinates of the figure shows the Children's Progress scores for 575 children collected over the first three years of school (age 6 to 9). The abscissa is the standardized score from the NYSA for these children. The value of Pearson's *r*'s are +0.68 for language arts and +0.70 for mathematical skills, absorbing about 50% of the variance of the scores on the State test accounted for by the Children's Progress assessment; a value way beyond most validity coefficients when the variables are not comparable in content or structure.

In collaboration with Dr. Mark Elliott and Dr. Stanislava Antonijevic we have been conducting pilot studies using the American version of CPAA in the Irish National Schools. Over the next few slides I am going to show you data that illustrates performance of Irish children in literacy and math.

These data are averaged over several schools. You can see that performance of Irish children is 'at expectations' in literacy and math which shows a high standard of education in Ireland.

In addition, we already know that there are some discrepancies in the Irish and US curriculum. This is most pronounced in 'measurement' because the American system uses imperial measures while Irish system uses metrics; this was shown in the assessment by an increase in errors, which were subsequently improved after a hint. The children did not fail to do the task, rather, they managed to learn how to solve the problem. Apart from showing that Ireland has bright students, this indicates that children can be taught by the "assessment." These results demonstrate the universality of human natures and the ability of children regardless of background or culture to acquire the foundations of a universally viable education.

References

Binet, A. & Simon. (1909) *Instructions: 30 Segment Version* pp.1 – 44.

Cliff, N. (1992) Abstract Measurement Theory and the Revolution that Never Happened. *Psychological Science,* 3, No.3, 186-190.

Furr, R. M. & V. R. Bacharach, (2007) *Psychometrics, An Introduction,* Thousand Oaks, CA:Sage Publications

Galanter, E. (1960) Two Models of a student, *Teachers College Record,* 61, 3.

Galanter, E. & Miller, G. A.(1960) Some Comments on Stochastic Models and Psychological Theories, Chapter 19 in *Mathematical Methods V. R. the Social Sciences* (Arrow, K., Karlin, S., and Suppes, P.).

Terrace, H. S. (1963) Discrimination Learning with and without "errors." *Journal of the Experimental Analysis of Behavior,* 6, 1.

Galanter, E. &, Sachtler, W. (2002) Color Vision Testing on Uncontrolled Displays, *Society for Informatipn Display*, Vol **33** No. 2.

Galanter, E. & Galanter, M. (2003) *Adaptive Evaluation Methods and Adaptive Evaluation Apparatus,* U. S. Patent No. 6,511,326 B1.

Huber, Peter. J. (2004). *Robust Statistics*. New York:Wiley.

Krantz, D.H.; Luce, R. ; Suppes, P. & Tversky, A. (1971). *Foundations of Measurement, Vol. I: Additive and polynomial representations*. New York:Academic Press.

Neisser, U. (1967) *Cognitive Psychology,* New York:Appleton-Century-Crofts.

Piaget, J. (1972) *The Psychology of the Child.* New York:Basic Books.

Spearman, C. (1904) "General Intelligence" objectively determined and measured, *American Journal of Psychology*, 15, 201-292.

Stevens, S. S. (1968), Measurement, Statistics, and the Schemapiric View, *Science,* 161, 3844, 849-856.

Thurstone, L. L. (1927) A Law of Comparative Judgment, *Psychological Review,* 43, 273-286.

Vygotsky, L. (1986) *Thought and Language.* Boston:MIT Press.