Adding It Up:

IMPROVED OUTCOMES AND ECONOMIC DEVELOPMENT

William P. Fisher, Jr., Ph.D.
Associate Professor
Department of Biometry & Genetics,
Louisiana State University Medical College, New Orleans

How can Rasch's probabilistic conjoint measurement models contribute to economic development? By deliberately extending into health care and education the crucial role that measurement has historically played in commerce. As anyone who has ever shopped for groceries knows, without fair units of measurement, there is no basis for economic activity. Unfortunately, health care and education lack fair units of measurement for many of their most important outcomes. The lack of stable outcome measures — common currencies for the exchange of quantitative value — in health care and education explains a lot about why it is so hard to know exactly what is obtained per dollar spent in either area, as well as why what is obtained per dollar varies so much across providers.

Functional assessments, test scores, consumer satisfaction surveys, and health status surveys all commonly produce units of measurement that literally do not add up. New technologies exist for correcting that situation and thus for enhancing the economic impact of health care and education.

In a nutshell, Rasch measurement is playing a crucial role in the founding of a new kind of organization, a developmentally-attuned organization that learns from the cumulative experience of its members. These organizations deliberately and scientifically measure learning and experiment with it, becoming organizations that coordinate their own evolution with the evolution of their members and partner organizations. This literal "co-ordination" will be the impact of units of measurement that do not vary in their size or order (by more than a known amount of error) depending on which brand instrument they are read off, where or when they are used, or by whom. With these instruments in systematic use throughout these organizations, even spontaneously occurring consistent variation in clinical or educational outcome measures will support better understanding and improved treatment and teaching effectiveness, quite apart from controlled experimentation.

The key is to put useful and meaningful information into the hands of those responsible for outcomes, the providers and consumers, teachers and students, in a form designed to be as developmentally well-targeted, and so intellectually accessible, as possible. The history of science shows over and over that widely recognized, interpretable, consistent, and reproducible effects and phenomena are essential to new understanding. The historian of science who coined the phrase "big science" once remarked that "thermodynamics owes much more to the steam engine than ever the steam engine owed to thermodynamics," and that "the chemical revolution resulted much more from the technique of the electric battery than from the careful measurements or new theories of Lavoisier" (Price 1986, pp. 240, 248). Computer adaptive testing and health assessment will be to medical education and health care what the steam engine was to thermodynamics and the electric battery was to chemistry because these technologies will provide universal access to consistently reproducible and interpretable quantitative information.

Those who are among the first to understand this will have the lead in making the new techniques pay off. The payoff is going to come in the form of evidence that supports or contradicts the effectiveness or efficiency of treatments, teaching, policies, or skill levels. Simply by placing fair and universally-recognizable units of measurement in the hands of people on the front lines of health care and education, many of those people will immediately take responsibility for the outcomes of their practices in ways that they never could before. When they are readily able to see the effect of variations in their treatments on outcomes, they will come to understand what they did not understand before. When they can compare the results of their interventions with the results of the clinic down the street, the practitioner across the hall, and the group across the country, they will either take new pride in their work or want to know how to do better.

And the comparisons, the reflection, and the decision to take action will not be a cumbersome, time-consuming, expensive process of data gathering and analysis. Instead the data system will already be in place. The relevant measures will all be expressed in a common quality assessed and monitored unit. In the same way that steam engines and electric batteries

permitted observations of thermodynamic and chemical effects to be routinized and generalized, so too will calibrated measures of educational and medical outcomes permit the routine and general observation of teaching and treatment effects. Only when such observations can be expressed in common units of measurement will they be able to play a significant role in the larger conversations of professionals actively engaged in learning and sharing their learning. Richer community lives for those impacted by measurement can be effected only when ways of sharing richer common unities of meaning are provided. Rasch measurement practitioners are discovering, inventing, and creating those unities.

Some Historical Background

As Ben Wright likes to point out, many historical documents, including the Bible and the Magna Carta, specify particular units of measurement as a standard in order to promote a common currency for the exchange of value. Looking back over the course of time we see that empires and political alliances bring about large economic communities that share much in the way of measurement standards, and that political fragmentation is associated with wide variation in measuring units. Before the French revolution, every town in Western Europe had its own system of weights and measures; Napoleon adopted the revolutionaries' metric system as a means of unifying the empire, with the effect of stimulating trade across a wide region.

Today, health care and education are like Western Europe before the French Revolution. There are wide regional variations in treatments, outcomes, and costs. Outcome measurement has been identified as a potentially useful means of overcoming some of the unwanted variation and of making treatment effects comparable. Existing quality of life and health status measures fall far short, however, of being equivalent to Napoleon's unified metric system.

Flaws in Current Practice and New Ideas for the Future

To date, virtually all efforts aimed at measuring outcomes have merely added more levels to the Tower of Babel as the proliferation of new instruments has brought along with it a proliferation of new units of measurement. Each different instrument has its own particular questions and its own rating response format, meaning that the sum of the ratings means something different for each instrument. One survey has 10 items and 3 response categories, for a 10-30 score range, and another one, intended to measure exactly the same thing, has 20 items and 6 categories, for a 20-120 score range. Scores from the two instruments plainly do not correspond. And even if complex statistics were used to establish correspondence, the two instruments still would not be shown to measure the same thing; the scores would still be nonlinear, nonadditive, and ordinal instead of linear, additive, and interval; the measures would not be accompanied with error and data quality estimates; and every respondent would have to answer every question on the instrument of choice for even the appearance of comparability.

What if, however, the two instruments had been calibrated to measure in one quantitative metric? What if the two instruments were used in two different hospitals in two different clinics seeing the same kind of clients? Using the new measurement technologies, given sufficient data quality, outcomes could be compared across the two clinics even if clients routinely skipped questions or if the instruments were in fact adaptively administered, so that clients were asked only those questions relevant to their condition.

Similar scenarios involving the comparison of test results across universities, course sections, or from year to year could be imagined in the educational arena. When universal access to universally interpretable and comparable educational measurement information is available, a new economy of educational effectiveness studies will be created.

Our goals ought to be 1) that each variable measured by means of client self-report satisfaction and quality of life surveys, clinician-administered functional or performance assessments, or psychological and educational tests is calibrated to a reference standard, 2) that the quantitative units of the majority of instruments in use for measuring each of those variables be traceable to that standard, and 3) that metrological systems be put in place for monitoring the quality of the instruments and the measures. Only when these goals are achieved will there be a basis for trading in and banking on a common currency of health and educational value. Only when these generalized measuring units are brought into the clinics and the classrooms and put into the hands of the care providers and consumers, and the teachers and the students, will people on the front lines of health care and education have the information they need to take responsibility for the outcomes of their efforts.

Suggested Plan

In the first five years of this plan, organizations interested in advancing a broad-based measurement agenda should publicly establish themselves as being aware that new management efficiencies could be provoked by the creation of unified systems of measurement for educational, psychological, and health care outcome variables. These organizations should make all of their faculty, students, staff, customers, etc. aware of several points, directing the early adopters to publicly available bibliographic resources:

1) that units of physical measurement (meters, grams, volts, ohms, degrees centigrade, etc.) do not exist in nature but are the result of ongoing
   a) experimental research establishing a convergence of results across samples, labs, instruments, and other variables according to strict mathematical data re-

quirements, and

b) efforts deliberately aimed at creating and maintaining measuring units as the common currency for the exchange of quantitative value;

2) that the only reason why there are no unified metrics for psychosocial variables is that no one has set out to create them with the right tools (perhaps because of the heavy computational burdens and lack of accountability demands);
3) that long-established measurement theory and data analysis techniques are available to help create unified metrics;
4) that applications of this theory and these techniques are establishing the expected experimental convergence of results across samples, labs, instruments, etc.;
5) that increased economic pressure and accountability in health care and education demand easily understood and comparable outcome measures, measures that can come into being only when sufficient attention is paid to instrument design and user training;
6) that it is no longer necessary to force people to adapt to the needs of measurement technologies, as it is now possible to adapt the measurement technology to the needs of people;
7) that examinees and survey respondents need not answer one single set of questions to be meaningfully measured in a common quantitative unit;
8) that measures can therefore be made comparable across classes, years, clinics, cultures, etc., even when tests and surveys are not identical, opening up vast new possibilities for understanding variation in learning and health;
9) that organizational growth of the kind we envision will most readily occur in a learning environment that recognizes that intellectual development does not stop at any age, but that adults can progress through as many marked transformational stages of development between 21 and 70 as they did between birth and 21; and
10) that part of the general mission of educational and many other kinds of organizations must be to provide an environment that supports continuing intellectual development to those they serve, including their own employees, and to take differences in reasoning attributable to developmental variations into account in educational, clinical, and managerial decision-making.

In the second five years, the paradigm-shifting organizations should establish themselves as the world leaders in the calibration and use of unified metrics in human resource management, health care, and medical education. They should set an example for the world to follow in its use of unified measures, linking up with other universities, hospitals, employers, schools, and government agencies locally, regionally, and globally to form the networks through which unified metrics for psychosocial variables will be created and maintained. Someday there will be a meaningful, useful, and quality-assessed quantitative metric for each psychosocial variable (clinical competence; physical function; consumer satisfaction; knowledge of anatomy, spelling or mathematics; environmentally sound behaviors; etc.) we're interested in, just as we have for each physical variable (meters, grams, volts, etc.). Organizations employing Rasch measurement practitioners could, should, and are playing a big role in making that happen.

The second five years should focus 1) on disseminating calibrated measuring instruments traceable to universally recognized reference standards; 2) educating faculty, students, clinicians, patients, employees, employers, etc., on instrument use, with demonstrations of each of the above 10 points; 3) on technical aspects of instrument design and calibration; and 4) on the needed information systems.

Overall effectiveness of the measurement program could be evaluated by surveying participants as to how much they are learning about their practices now, and surveying them again periodically as instruments come on line.

Summary and Conclusions

The economic need for common units of measurement greatly predates science. Economic development is impossible unless we can estimate amounts of value in a way that does not depend on the particulars of the measurement process, such as who is using which brand instrument where and when. Recent advances in measurement theory and in computerized information technologies support the emergence of new kinds of learning organizations capable of deliberately evolving in the direction of enhanced efficiency and effectiveness. Rasch measurement practitioners have a unique opportunity to help shape the organizations that will create new health care and educational economies. Carpe diem!

Reference

Price, D. J. d. S. (1986). Little science, big science...and beyond. New York: Columbia University Press.

William P. Fisher, Jr. was formerly Senior Research Scientist for Program Evaluation at Marianjoy Rehabilitation Hospital & Clinics in Wheaton, IL, serving on the Management Team, and on the Clinical Programs and Quality Assessment & Improvement Committees. After completing the University of Chicago's Social Sciences Divisional Master's degree in 1984, William was a Spencer Foundation Dissertation Fellow, earning a Ph.D. in Chicago's Department of Education in 1988, concentrating in Measurement, Evaluation, and Statistical Analysis (MESA). Dr. Fisher is still a MESA Research Associate, is on the Editorial Board of the Journal of Outcome Measurement, and on the Advisory Board of the Institute for Objective Measurement. He is professionally active in diverse organizations. Current tasks include designing and implementing an outcome measurement system for Louisiana's statewide public hospital system; consulting on the Social Security Administration's Disability Process Redesign Project; and drafting scale-free health status measurement standards for the ASTM E31 Committee on Medical Informatics.