The Shift from Modeling Observations to Applying Theory:

Some Timely Points About Measuring Latent Traits

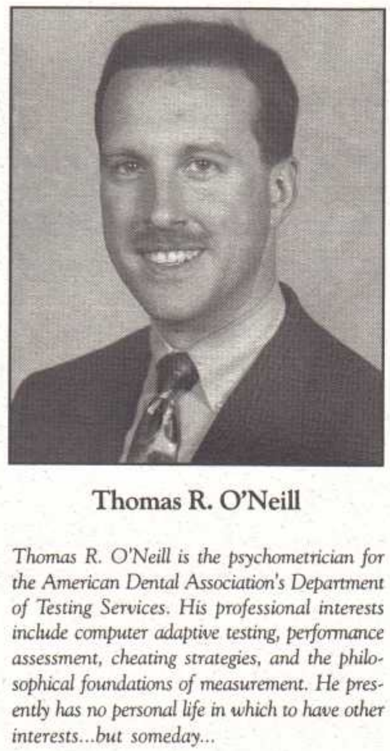
Thomas R. O'Neill

American Dental Association

MEASUREMENT MUSINGS

Tremendous progress has been made in the physical sciences in the last 500 years and the rate of change has been increasing, especially over the last 100 years. The discovery that some traits are transmitted genetically has led to the genetic engineering of fruits and vegetables, the cloning of mammals, and the promise of successful genetically engineered solutions to medical prob-lems. Military technology has been changed not only by the invention of the mass-produced rifle, but also by the radio, micro-chips, satellites, airplanes, and missiles which can deliver explosives or non-conventional weapons (chemical, biological, or nuclear). Medical technology has been changed not only by the invention of antibiotics, anesthesia and the development of a germ theory of disease, but also by dialysis technology, replacement joints and the development of sophisticated surgical technologies (i.e. micro, orthoscopic, laser, etc.). Human organs can be replaced with organs from other people or sometimes from other animals. Computer technology changes so quickly that businesses usually expect top-of-the-line technology to be obsolete in less than five years.

In contrast to the rapid advancement in "hard science technology", social technology has experienced almost no real advances in 100 years. In social science, one never finds a well-developed theory that (1) describes a phenomenon, (2) identifies its predisposing or precipitating conditions, (3) explains the mechanism through which the process works, and (4) permits the prediction and control of the phenomenon. Even Freud's famous psycho-dynamic theories fail. Although his theories distinctly describe the phenomenon and a mechanism through which predisposing factors become manifested as the phenomenon, it accounts for all unexpected observations by attributing them to "defense mechanisms", such as, repression, displacement, projection, and reaction formation. Although the theory is an excellent framework in which to understand events, it does not lend itself to verification or refutation. Theories regarding cognition, motivation, affect, and all



Thomas R. O'Neill

Thomas R. O'Neill is the psychometrician for the American Dental Association's Department of Testing Services. His professional interests include computer adaptive testing, performance assessment, cheating strategies, and the philosophical foundations of measurement. He presently has no personal life in which to have other interests...but someday...

Email: oneillt@ada.org

other important social science topics fail to adequately address these four issues. In the absence of powerful ways to verify and refute theories, researchers are left to assess the theories intuitively permitting them only to form opinions rather than empirical conclusions about the theory. As a result, every theory has proponents and opponents, which effectively thwarts any type of universal consensus. The acceptance of competing theories as all being equally good has distanced social scientists from the process of theory building. The distinction between theory building and modeling data has become blurred. If social science is to achieve the same status as physics, then the distinction must be made clear, and social scientists must shift toward building and applying theories.

A theory is a coherent set of principles that is used to explain a wide range of related observations. The quality of a theory is judged by the range of facts that it explains and the precision with which it explains them. Although explanations of past events are comforting, the value of a theory lies in the accuracy of its prediction of future events. When an observation contradicts a well-established theory, the researcher usually suspects an error in the data collection or analysis before disputing the theory because of the substantial accumulation of evidence already supporting the theory.

Modeling data is a very different enterprise. In the absence of a strong theory, researchers often collect data that they believe to be related to their topic. Assuming that truth can be found in the data, the researcher tries to find the most parsimonious mathematical representation that will recreate the observed data reasonably well. To see if these predictors will be applicable to future situations, the model must be cross-validated using a second sample. However, even when a model permits very accurate predictions across a wide variety of samples, it is still not a theory until the model can be meaningfully understood. Although models require the predictor variables to be operationally defined, their meaning may be ambiguous. Models may include variables that are correlated with the outcome, but are not conceptually part of the construct. For example, suppose that socio-economic status (SES) is moderately correlated with math ability. Although SES could be useful in making imprecise predictions about a person's math ability, it would be very difficult to coherently incorporate it into a theory of what math ability is. Variables that can't be discussed coherently in terms of the construct cannot be included in a theory. The essential difference is data modeling permits the selected observations to dominate the researcher's intentions, but in a well established theory, the researcher's intentions dominate the observations.

This difference has not always been clear to observers of physical phenomenon either. Barnett (1998) describes the historical development of the concept of time, as well as, how human needs, pre-existing concepts, and available technology influenced that concept. She provides many examples of the confusion and tension between modeling observations and applying a theory. Social scientists committed to advancing their respective fields would do well to understand how these issues have been resolved in the physical sciences. Although Barnett (1998) never addresses social science directly, the issues she highlights are quite pertinent. The following three paragraphs are a very abbreviated summary of Barnett's book with regard to some issues that are relevant to these tensions.

Primitive sundials were used to divide the day into segments, but not necessarily segments of equal size. Circa 1500 B.C. some sundials marked the calibrations for the morning and evening hours farther apart than for those hours near noon to adjust for the uneven increases in the length of the shadow cast throughout the day. This sundial produced 12 approximately equal daylight hours. However night was still a single unit of "non-day" and summer hours were longer than winter hours. Observations of the sun's position defined both the current time and the length of the hours. In the 1580s, Galileo noted that the swing of a pendulum is amazingly regular (it varies according to the length of the pendulum, not it's weight or the horizontal force applied to it). In 1657, Christiaan Huygens used this principle to build the first gravity-based pendulum clock which lost only about one second every two and a half hours. For short periods of time, this clock produced hours that were of the same duration regardless of the time of year and could work through the night. This clock produced more stable time than did observing the sun's position. Time was no longer tied to the relative position of the sun! But not entirely. In the long run these clocks tended to slow down and lose time due to friction and other factors. To rectify this, pendulum clocks had to be reset occasionally according to the only standard that was relatively stable over long periods of time, the sun and stars. The mechanical clock was not without controversy. Some people objected that it did not adequately model the position of the sun in the sky. Had the clock makers possessed the technology to accomplish such a feat, they probably would have, in effect, destroying the equal hours that they had just created. In towns, these pendulum clocks were installed in towers which permitted the town's activities to be coordinated using "local time". Methods to minimize the amount of friction in the mechanism extended the amount of time a clock could go without recalibration (resetting the time), but these improvements were limited by a precision ceiling of one second every 250 days. However, that ceiling would soon be removed.

Pierre Curie's discovery that quartz crystals vibrate at a very stable frequency when pressure (or electric current) is applied to them, led W. A. Marrison of Bell Laboratories to create the first quartz crystal clock in 1928. This clock was accurate to about one second every nine years. Today quartz crystal wristwatches are still quite popular. Despite their utility, quartz crystals are not the perfect solution. In addition to the imperfections inherently found in the crystals, the vibrations themselves cause some wear on the crystal which in turn changes the frequency with which it vibrates. Greater precision could be achieved if regularity was a property of the substance rather than form. This became possible with the new atomic theory and quantum mechanics. Atoms seem to function as a miniature solar system in which there is no friction. Using these ideas, atomic (cesium-133) clocks have been devised that are accurate to approximately one second every 10 million years.

Despite these improvements in precision, the original concepts of year and day as based upon the earth's orbit and rotation have not been vanquished. People find these models easy to understand and easy to relate to the experience of time. Although the production of stable hours, minutes, seconds, nano-seconds, etc. is better accomplished by observing more regular and controllable occurrences of nature (i.e. pendulum swings, crystal vibrations, etc.), the count of those occurrences are then incorporated back into an abstracted framework based upon the original concept. The idea of a mean solar day recognizes that the rotation of the earth is not constant. With the invention of atomic clocks that are precise to one second in 3 million years, it seemed silly to use the mean solar day as the standard from which seconds were derived. Rather than define a second as 1/86,400 (1/24x60x60) of a mean solar day, the 13th General Conference of Weights and Measures redefined a second as 9,192,631,770 oscillations between two specific energy levels of a cesium 133 atom under highly specified conditions. This, in effect, redefines a solar year as 86,400 "atomic" seconds rather than vice-versa.

The regularity of the sun's position relative to the earth's was replaced by the regularity of gravity's effect on a pendulum, which was replaced by the regularity of the vibrations of a quartz crystal, which, in turn, was replaced by the regularity of an electron's orbit around the nucleus of an atom. The discovery of finer gradations of regularity in nature permits humanity to extend the concept of time.

As these advances have occurred, the notion of time has become clearer. Time is certainly an abstraction created by man to make the world more understandable, but is the primary purpose to predict certain types of events or is it to create a framework to understand the events. When the framework and outcome agree, there is no conflict, but when there is a discrepancy, which one should dominate? If the purpose of time is to accurately predict the position of celestial bodies relative to a particular point on a rotating planet that orbits a star, then the failure of an equal interval measurement system to predict those positions indicates that adjustments should be made to the model. Furthermore, these adjustments should be made even if it degrades the interval quality of the model. This approach would be popular in pre-electrical societies whose concern is the amount of useable time (daylight) remaining before nightfall. The disadvantage is that it would be acceptable and probably necessary to have a different model for every point on the planet and for every day of the year for which you wanted a prediction. As a result, time would be very specific to location, which in turn would make coordinating operations over any distance quite imprecise.

However, if time is intended as a theoretical framework to make sense out of events, then having a stable equal interval framework is important. Rotational and orbital anomalies can then be regarded as merely imperfections in the cosmic machine rather than a shortcoming of the framework. In the quest to harness time, chronometry specialists have done two things. First, they have sought out ways to increase the regularity of the phenomenon that they observe to make their measurement system more stable. Second, they have investigated those observations that seem to depart from what the theory predicts to find why the observation was anomalous. The theory is only modified when the source of the anomaly is conceptually part on the construct of time. If the source of the anomaly is unrelated to the construct of time, then ways to remove its influence are sought out.

If social science is to experience the same rapid advancement as the physical sciences, then social scientists must improve their instruments and clarify the constructs implied by those instruments. Social scientists must free their ideas about the construct from the particular observations (modeling) and permit the theory to dominate. The lessons for social scientists are twofold. First, seek out methods that will permit finer and more stable regularities. Search for social science pendulums, quartz crystals, and cesium atoms. Second, do not attempt to incorporate the influences of extraneous forces into your theoretical framework. Control them! When creating a social science clock, seek to reduce the friction in the mechanism, control the temperature of the pendulum, and stay alert for other sources of error.

Barnett, J. E. (1998). Time's Pendulum: The quest to capture time from sundials to atomic clocks. New York: Plenum Publishing.