White Paper

Tools for Assessing Learning Outcomes in Higher Education - Spring 2012

By:
Dr. Andrew L. Luna

Institutional Research, Planning, and Assessment

HISTORY in the MAKING

About the Author

Dr. Andrew L. Luna is Director of Institutional Research, Planning, and Assessment at the University of North Alabama. He has served over 26 years in higher education, with 15 of those years in institutional research. He has published research studies on many topics including salary studies, assessment, market research, and quality improvement. Dr. Luna received his Ph.D and M.A. degrees in higher education administration and his M.A. and B.A. degrees in journalism, all from the University of Alabama.

Table of Contents

INTRODUCTION

As the United States continues to hold higher education more accountable for that which students learn, assessment of learning outcomes becomes increasingly important to the development and maintenance of academic programs. While this assessment and accountability of higher education is relatively new, current practices indicate greater demands on colleges and universities in the future. During the 1950's and 1960's college enrollment was expanding with veterans seeking higher education, and women being afforded greater opportunities to go to college. During this growth surge, the value of a college education was assumed and institutions of higher education were, for the most part, not required to answer to the general public (Brubacher, 1982).

In the 1970's, however, higher education experienced several changes including financial hardships, an overall downturn in the national economy, as well as the increased diversity of college student populations. During this time, an anxious public raised concerns that college graduates may not have the skills and abilities needed for the workplace. In the 1980's, the economic profile of the US dramatically changed as traditional industry yielded to higher technological capabilities and the threat of global competition became more apparent. As higher education tried to adjust to these changes, federal and state governments increased their accountability of higher education while decreasing their financial support thus leaving colleges and universities to fight for an ever decreasing piece of the revenue pie (Garfield & Corcoran, 1986).

Oftentimes, revenue to colleges and universities has been tied directly to performance and learning outcomes, to which higher education has responded by seeking out the assistance of nationally normed assessment instruments, proprietary portfolio repositories, and assessment databases in order to demonstrate continuous improvement among its graduates. With the need for greater assessment in higher education, the private sector answered the call with a plethora of assessment instruments made for the taking - provided institutions had the money to do so. During the recent economic downturn, the federal government has opened the doors of alternatives to traditional assessment instruments, and institutions of higher education are increasingly looking to internal measure-

"During the 1950's and 1960's...the value of a college education was assumed and institutions of higher education were, for the most part, not required to answer to the general public."

ments, such as capstone courses and embedded assessments, to fulfill their assessment requirements (Alan, 2004).

The purpose of this paper is to discuss briefly the historical perspective of continuous quality improvement and to demonstrate the theory behind this methodology and how institutions of higher education may exploit traditional research and quality tools in order to assess and improve learning.

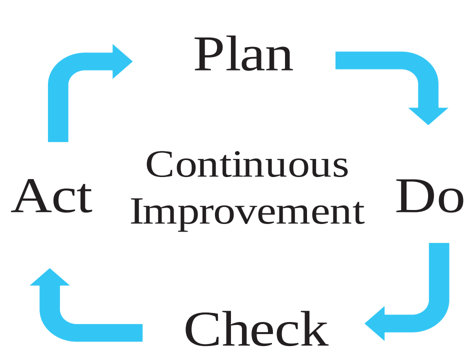
HISTORICAL PERSPECTIVE

The science of quality improvement is not new. While working for the Western Electric Company, a manufacturer of telephone hardware for Bell Telephone, Walter Shewhart developed advanced quality thinking techniques during the early 1920's. His two greatest achievements were the Plan, Do, Check, and Act (PDCA) cycle and his use of statistical process control charts. The advancements Shewhart made in quality measures were used throughout the manufacturing sector, adopted by the American Society for Testing and Materials (ASTM) in 1933, and carried over during World War II when the US government used his techniques to improve production (Shewhart, 1986).

The essence of Shewart's PDCA cycle is a four-fold process that is shown in **Figure 1**. The four steps to the Shewhart Cycle are as follows:

- **Plan** - Create a strategy as to what the institution wants to do and how it will measure success.
- **Do** - Follow the plan.
- **Check** - Assess the effectiveness of the current plan by looking at the success outcomes measures.
- **Act** - Make changes to the strategies to improve the measured outcomes (Shewhart, 1986).

Figure 1
Shewhart Cycle of Continuous Improvement



Unfortunately, after the war, US business and industry soon forgot about quality improvement as newer and easier business processes were developed. When the US supported efforts to rebuild Japan after the war, W. Edwards Deming took

"The advancements Shewhart made in quality measures were used throughout the manufacturing sector..."

Shewhart's quality concepts overseas and they were quickly adopted by Japanese business and industry.

In the 1980's, when US businesses were suffering significant losses due to Japanese imports, NBC News broadcasted the documentary "If Japan Can, Why Can't We." This was the introduction of Deming's Total Quality Management to the US. Soon many American businesses were eager to try the seemingly new practice, not realizing that the foundation of TQM rested on Shewhart's work in the US 60 years earlier (Breyfogle, 2003).

One company, IBM, adopted TQM in the early 1990's and realized great success with this practice. They also saw that the concepts of TQM could transcend the realms of business and industry and enter into the halls of academe. Soon, IBM offered substantial monetary rewards to those institutes of higher education that could successfully adopt TQM. The idea of applying TQM concepts to higher education interested many administrators. Just a few years beforehand, the seven major accrediting agencies changed their evaluation standards to place greater emphasis on quality improvement and institutional effectiveness.

Many in higher education quickly jumped on the TQM bandwagon without really understanding what TQM was or how it could be used effectively in higher education. Institutions hired TQM consultants – many of whom did not understand higher education – and together they tried to retrofit the business model of TQM into the non-business environment of colleges and universities.

TQM in higher education failed due to the following reasons:

- **Lack of measurement** – Probably the most important, yet least understood, component of TQM was measurement. After all, TQM's chief innovators including Deming, Juran, Ishikawa, and Shewhart were all statisticians. The basic premise behind TQM was that all processes are not perfect and vary to some degree. This variation is due to either random variation or an assignable cause or causes. Without measurement, there could be no TQM.
- **Not supported by upper administrators** – TQM's success depended on support and understanding of TQM

"Many jumped on the TQM bandwagon without really understanding what TQM was or how it could be used effectively in higher education."

by deans, directors, vice presidents, and the president. Without this support, many institutions could not commit the time or resources to successfully implement TQM.

- **Hidden agendas** – Anytime TQM was used in the same sentence with cost containment, consolidation, cutback, or attrition, what was being done was usually not TQM. While TQM could reduce costs, the main reason for implementing TQM was to improve quality.
- **Institutions not concentrating on the vital few problems** – Well intentioned institutions jumped into the TQM philosophy and immersed the entire operation into the program without first identifying those few problems which are vital to the success of that institution. The result is that the institution bit off much more than it could chew.
- **Higher education hostile toward innovation** – Because the concepts of TQM came from business and industry, many in higher education were skeptical about its application and/or usefulness to higher education. Without institutional support, TQM failed.
- **Failure of education to properly identify customers** – While a customer is usually defined as someone who purchases a good or service, in higher education this definition is much different. Tax payers and employers are some of the many customers in the higher education arena. Furthermore, the assumption that the student is a customer was an erroneous assumption at least half the time. While the student should be considered the customer in the areas of student and business-related services, the student is not the customer in the academic side of higher education. Using an industry analogy of the assembly line, the student should be actually considered the “raw” material who is transferred into a useable commodity for employers and the community at large.
- **Trainers and teams with limited knowledge** – When TQM was introduced into higher education, “experts” were everywhere, and many were unqualified in TQM, higher education, or both.
- **Expecting too much too soon** – Positive results with TQM took time. By concentrating on a quick fix, institutions became dissatisfied with the promises of quality.

“The basic premise behind TQM was that all processes are not perfect and vary to some degree. This variation is due to either random variation or an assignable cause or causes.”

From these reasons, there are two central themes surrounding TQM's failure in higher education. First, the more TQM was used, the more measurement was tossed aside in the process. Since measurement is centric to the TQM philosophy, it soon failed. Second, administrators lacked the knowledge as to how TQM could actually work on college and university campuses. Rather than use existing problems found in academe and use TQM to help solve them, implementers of TQM tried to force a business/industry model onto higher education and seek out problems that fit traditional TQM templates.

At the same time TQM was failing in higher education, its importance was significantly diminishing among the leaders in US business and industry due to many of the same reasons listed above. At this point, the CEO's of both Motorola and General Electric saw a need to keep the quality improvement philosophy of TQM, while bringing back the statistical measurement and scientific method components which were almost non-existent with many quality initiatives throughout the US. The end result was Six Sigma (Eckes, 2001).

Six Sigma is a statistical concept that represents how much variation there is in a process relative to customer specifications. The Sigma value is based on "defects per million opportunities where Six Sigma is equivalent to 3.4 DPMO. The core theme of Six Sigma is improving quality by increasing customer specifications and decreasing variability (Breyfogle, 2003).

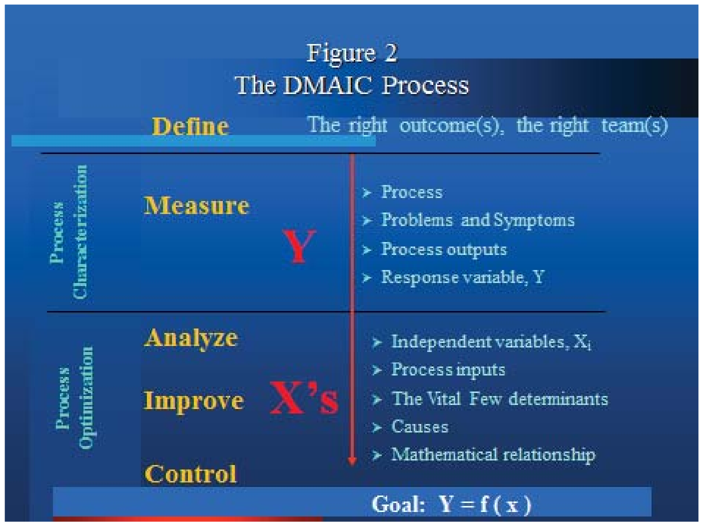
While the science behind Six Sigma is very complex, the process behind it is easier to understand and use than TQM. Much like Shewhart's Plan, Do, Check, and Act Cycle from the 1920's, the steps of Six Sigma follow a DMAIC (pronounced DEE-MAY-ICK) Cycle of **Define**, **Measure**, **Analyze**, **Improve**, and **Control**. Therefore, while some of the more difficult statistical tools of Six Sigma cannot be rationally used to solve academic problems within higher education, other useful tools can.

Figure 2 indicates the DMAIC process. Under the **DEFINE** portion of this cycle, a department chooses the most significant learning outcome for improvement. This can be done using a variety of methods, including statistical analysis, focus groups, outcomes assessment, and/or benchmarking. Once the outcome has been chosen, it is important that the measure defining this learning outcome be operationalized and **MEASURED** over a period of time. This measure becomes the dependent

"The core theme of Six Sigma is improving quality by increasing customer specifications and decreasing variability."

variable in the improvement process. It should be noted that more than one dependent variable may be used.

The next phase of the cycle concerns the **ANALYSIS** of key independent variables that factor into the variability of the dependent measure/s. During this phase, data are collected on the independent variables and the strength of the relationship between these variables and the dependent measure/s is tested. In the **IMPROVE** phase, the organization finds ways to improve and/or decrease variability of the independent measures which, in turn, will increase the quality and decrease the variability of the dependent outcome measure. The last phase of the Six Sigma process, **CONTROL**, stabilizes and sustains the changed independent measures in order to ensure continuous improvement. Within the academic environment, this last part involves changing curriculum, policies, and/or procedures while continuing to monitor the learning outcome/s.



Under the new reaffirmation criteria by the Southern Association of Colleges and Schools, Commission on Colleges, all institutions are required to submit a Quality Enhancement Plan that "describes a carefully designed and focused course of action that addresses a well-defined issue or issues directly related to improving student learning." Carefully looking at SACS's description of the QEP, one can easily find and distinguish each of the DMAIC steps of Six Sigma. Specifically, Six Sigma can be used to enhance the QEP by doing the following:

- **Define** – What is a significant area for improvement at a college or university? What is the current research regarding this area?
- **Measure** – What outcome measures define or operationalize improvement/success within this area (the Y measure)? How can this data be gathered?
- **Analyze** – What factors contribute to variation of Y (the X measures)? What are the attributes of these measures (descriptive statistics), and how strong are the relationships between these measures and the

Y outcome (inferential statistics). What is the cost of analysis?

- **Improve** – Which factors (X measures) have the greatest impact, and how can they be controlled?
- **Control** – How can the improvements to the X measures be sustained and stabilized so that they become a part of the process? What is the cost of sustaining and stabilization compared to the costs of not making the improvement (the costs of doing nothing)?

By using the DMAIC process in designing the QEP, an institution will be better suited to identify the most significant area for improvement, define a clear and scientific process for measuring and controlling improvement, and ensure that the improvements are maintained. Furthermore, the DMAIC process will aid the institution in developing a clearer and more concise report that focuses on prior research and benchmarking, the improvement process itself, the costs of improvement, as well as the costs and processes needed to maintain the improvements.

ASSESSMENT TOOLS

Since the foundation of the DMAIC method is rooted in the scientific method, it is not surprising that many of the statistical and quality tools used are equally suitable for assessing student learning outcomes in higher education. In general, these tools are designed to measure trends over a period of time, differences between and within groups, as well as relationships between variables.

It should be noted, however, that many quality tools, like control charts, are not suitable for measuring learning outcomes in higher education. These tools require long runs of 20 or more events. In higher education, events are usually defined as semesters or years. Therefore, a tool that requires 20 or more semesters/years to be analyzed is not appropriate for higher education.

Furthermore, it is important to note that two different types of data exist that should be equally utilized in higher education. The first type of data is longitudinal and is measured at specific points in time (i.e. semesters). While this type of data is optimal for determining trends, it can also be used to test other hypotheses as well. The other type of data is cross-sectional. This type of data is created by choosing representative samples from a larger population within the same period of time. For example, one could use scores from randomly selected sections from the total population of sections in English 101.

While neither exhaustive nor thorough, the examples below should give the reader an idea of how these statistical and quality tools may be used to improve learning in higher education and may facilitate further discussion of quality improvement within individual institutions.

Fishbone Diagram

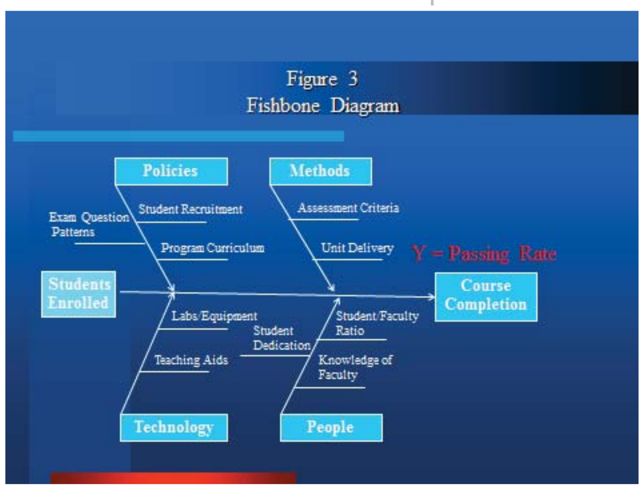
The Fishbone diagram is also known as the cause and effect diagram, the root cause analysis, and the Ishikawa diagram, named after its originator Kaoru Ishikawa, the Japanese quality pioneer, and the diagram got its name because looks similar to a fishbone. Essentially, the fishbone diagram relates causes and effects. It can be used to structure a brainstorming session where the major causes of a problem are placed under generic

"...the examples below should give the reader an idea of how these statistical and quality tools may be used to improve learning in higher education..."

headings such as Method, Equipment, People, Materials, Measurement, and Environment.

Figure 3 shows how a fishbone diagram was used to investigate ways for improving passing rates among basic engineering subjects (Abraham, Dereje, & Lim, 2011). The headings were labeled Policies, Procedures, Technology, and People. The faculty participated in a brainstorming session where they further delineated specific root causes within each of the four headings.

For instance, under the Policies heading of the diagram, the faculty found that the curriculum of the program, student recruitment, and exam question patterns were all policy factors that could lead to student failure or success. Likewise, under the People heading, faculty found that student/teacher ratios, pedagogical experience and knowledge of faculty, degree of motivation of faculty, and student/faculty interactions were all factors of student success or failure.



After these root causes were identified, further investigation was developed using more sophisticated techniques to determine which of these factors were significant contributors toward student failure.

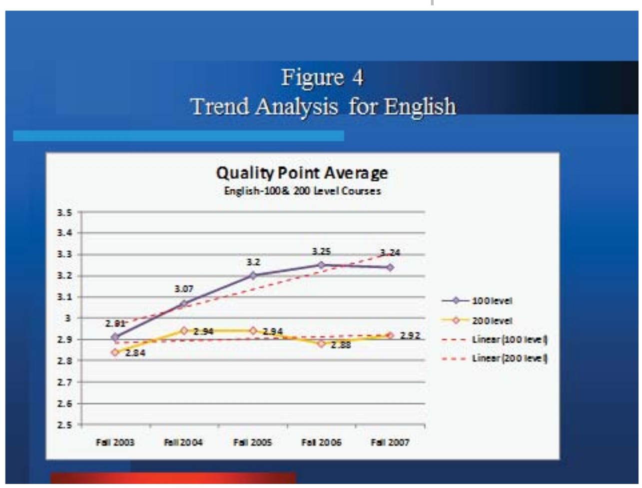
Trend Charts

Trend charts are also known as run charts, and are used to show trends in data over time. All processes vary, so single point measurements can be misleading. Displaying data over time increases understanding of the real performance of a process, particularly with regard to an established target or goal.

The example in **Figure 4**, a trend analysis was created based on the aggregate quality point average for all students who completed English 100 and 200 level courses at a four-year institution. The top line indicates scores for the 100-level courses while the bottom line indicates scores for the 200-level courses. Before this analysis was run, the English faculty made two assumptions:

1. That student earning higher grades in the 100-level courses for a given fall semester would also earn higher grades in the 200-level courses during the next fall.
2. That there was a direct relationship between performance at the 100-level and 200-level.

Clearly, this graph demonstrates that these assumptions may not be supported. After this analysis was run, further investigation was made. As it turned out, the faculty realized that the 100-level was focused more on writing, while the 200-level was focused more on reading. While reading and writing are compatible, they may not be comparable. Furthermore, it was found that students, who took the 200-level courses during those fall semesters, were not always the same students who took the 100-level courses the fall prior. Many students transferred to the institution from a local community college and, it was later determined after disaggregating Collegiate Assessment of Academic Profeciency (CAAP) scores, students who came from the community college had significantly lower scores than did those students who took the 100-level courses at the four-year institution.



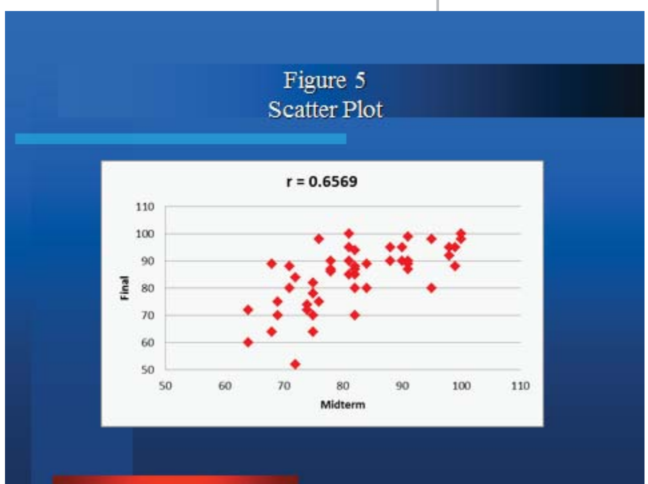
Scatter Plot

Scatter plots are similar to line graphs in that they use horizontal and vertical axes to plot data points. However, they have a very specific purpose. Scatter plots show how much one variable is affected by another. The relationship between two variables is called their **correlation**.

Scatter plots usually consist of a large body of data. The closer the data points come when plotted to making a straight line, the higher the correlation between the two variables, or the stronger the relationship.

If the data points make a straight line going from the origin out to high x- and y-values, then the variables are said to have a **positive correlation**. If the line goes from a high-value on the y-axis down to a high-value on the x-axis, the variables have a **negative correlation**.

A perfect positive correlation is given the value of 1. A perfect negative correlation is given the value of -1. If there is absolutely no correlation present the value given is 0. The closer the number is to 1 or -1, the stronger the correlation, or the stronger the relationship between the variables. The closer the number is to 0, the weaker the correlation. So something that moderately correlates in a positive direction might have a value of 0.67, whereas something with an extremely weak negative correlation might have the value -.21.

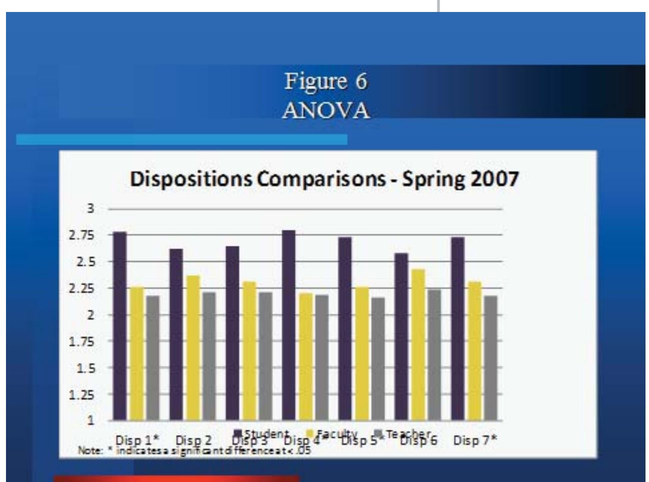


The example in **Figure 5** demonstrates a simple scatter plot. In it, students' midterm and final exam scores were plotted. The final grade (dependent measure) was plotted on the Y axis while the midterm grade was plotted on the X axis. The results show a moderate positive correlation of .65 indicating that midterm grades may be a reasonable predictor of final grades.

Analysis of Variance (ANOVA)

The Analysis of Variance is a technique that allows the comparison of two or more means to determine if significant differences exist between or among them. For example, the ANOVA could be used to determine if one class scored significantly higher than another class, or if one method of teaching resulted in significantly higher scores than another method. The ANOVA is used in a wide variety of applications and in varying degrees of complexity. The applications can range from a simple analysis of means from different independent groups using only one independent variable, to models that contain complex factorial (multiple independent variables) pre-test/post-test variations that are designed to test interactions between and among variables.

Figure 6 is an example of an ANOVA that was used to determine if significant differences existed between classroom teachers, student teachers, and their professors on the perception of competence with various dispositions/characteristics. Toward the midpoint of a student teacher's class assignment, questionnaires were given to the student teacher, classroom teacher, and professor to assess the student teacher's performance on key dispositions. After all assessments for each student teacher were gathered, mean scores were computed and the ANOVA was used to determine if significant differences existed.

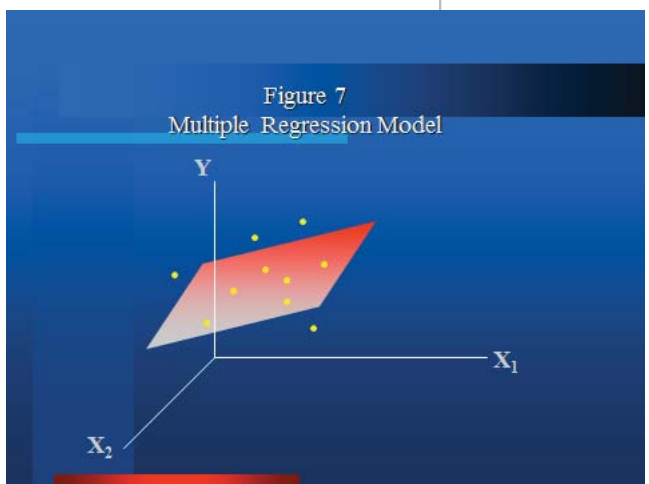


A bar graph of the aggregate scores by group clearly indicates that the student teachers perceived they had a higher level of competence than did either the classroom teacher or the professor. Furthermore, in most instances, the professor ranked the student teacher's level of competence higher than the classroom teacher. However, the ANOVA indicated that significant differences existed only between student teacher and both classroom teacher and professor, and only with dispositions 1, 4, and 7.

Regression Analysis

When an institution wants to determine how variables relate to each other, the regression analysis is used. Regression analysis is a statistical tool for the investigation of relationships between variables. Usually, the investigator seeks to ascertain the causal effect of one variable upon another—the effect of a college entrance test score to overall college success, for example, or the effect of changes in teaching methods upon the exam scores. To explore such issues, the investigator assembles data on the underlying variables of interest and employs regression to estimate the quantitative effect of the causal variables upon the variable that they influence. The investigator also typically assesses the “statistical significance” of the estimated relationships, that is, the degree of confidence that the true relationship is close to the estimated relationship.

As shown in **Figure 7**, the multiple regression analysis uses multiple independent variables to either predict or explain the variance of a single dependent variable. A multiple regression analysis was used to determine if cooperative learning (small group learning) improved student learning outcomes in economic instruction (Yamarik, 2010). The study applied cooperative learning to one section of an intermediate economics course, and taught another section using a traditional lecture format. According to the study, five learning outcomes were established and operationalized: interest, preparation, participation, attendance, and academic performance. Data for the interest, preparation, and participation variables came from a survey given to students. Data for attendance came from daily attendance records, while data for academic performance came from a combination of mid-term and final exams.



To test the effect of cooperative learning, the study estimated an empirical model where each learning outcome depended upon teaching pedagogy, demographic factors, economic knowledge, and other academic factors. The model determined that students taught by cooperative learning achieved greater academic performance in the form of higher exam scores.

Path Analysis

Path analysis is a mode of Structural Equation Modeling (SEM) involving assumptions about the direction of causal relationships among linked sequences and configurations of variables. SEM is a cross-sectional statistical modeling technique that has its origins in econometric analysis. SEM is a combination of factor analysis and multiple regression. The terms factor and variable refer to the same concept in statistics. Path analysis, a variation SEM, is essentially a type of multivariate procedure that allows an examination of independent variables and dependent variables.

Path analysis was used to determine if a newly created chemistry workshop was effective in raising student scores on the American Chemical Society's General Chemistry exam. Beforehand, it was determined that the institution's students were scoring significantly lower than the national average on the ACS exam. The department created a new workshop to coincide with the actual chemistry course, and required students to take it. It was hoped that the workshop would increase students' scores in both the class and the ACS exam.

To make this determination, a path analysis was constructed by running two regression models where the dependent variable were the final course grade and the ACS scores and the independent variables were SAT Math, high school GPA, classroom content, laboratory work, homework, and grades from the newly designed workshop. From these two regression equations, the path analysis (**Figure 8**) was constructed.

As shown, the relationship between the workshop grade and the final grade was moderate at 0.49 while the relationship between the workshop grade and the ACS exam was very low at 0.15. High school GPA was moderately related to the final course grade at 0.66, and SAT Math was moderately related to the ACS exam but not with the workshop grade.

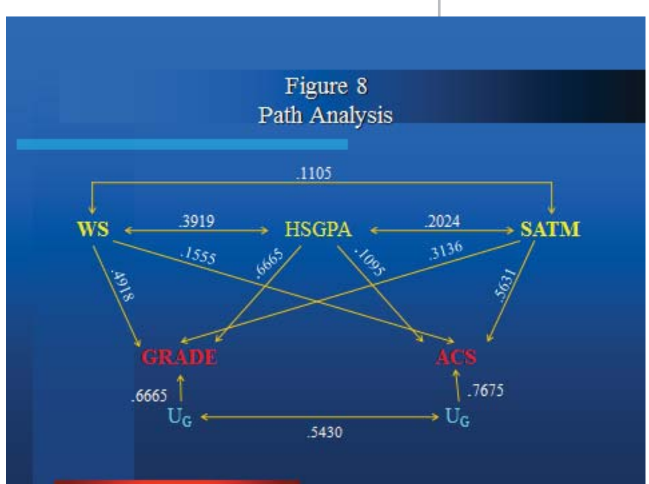


Figure 8
Path Analysis

It was determined by the faculty that, because the workshop grade and final grade were related, but that the workshop grade was not related to the ACS exam, that the content of the course and workshop needed to be reviewed and changed to better represent the skills needed for the ACS exam.

After a year, the path analysis was run again and the workshop grade was found to be highly related the ACS exam score. Furthermore, since the course was redeisigned, additional analysis found that the new course content was now more related to the ACS exam score.

CONCLUSION

As assessment demands increase and revenues decrease, institutions of higher education do not have to rely soley on expensive external assessment instruments. By utilizing the concepts of the scientific method as well as the processes of quality improvement, colleges and universities have the ability to create their own effective measures to assess student learning outcomes.

To do this, the institution must operationalize the outcome (dependent measure/s), create baseline measures from historical analysis, identify key independent variables that affect the outcome/s, and take measures to improve the independent variables in order to affect positive change in the learning outcome/s.

By practicing this type of in-house assessment over a period of time, an institution will be able to create important trend data, correlate its in-house measures with any external assessments used, and effectively improve student learning.

REFERENCES

Abraham, D.W., Dereje, E.W., & Lim, C.I. (2001). *Fishbone diagram approach for improving the passing rate for basic engineering subjects*. Paper presented at the Enhancing Learning: Teaching and Learning Conference.

Alan, M.J. (2004). *Assessing academic programs in higher education*. Bolton, MA: Anker Publishing Company

Breyfogle, F. W. III (2003) *Implementing six sigma*. New York: John Wiley & Sons.

Brubacher, J. (1982). *On the philosophy of higher education*. San Fransisco: Jossey-Bass.

Eckes, G. (2001). *The six sigma revolution*. New York: John Wiley & Sons.

Garfield, J. & Corcoran, M. (1986). *Assessment in American higher education*. Paper presented at the Annual Conference on Assessment In Higher Education.

Shewhart, W.A. (1986). *Statistical method from the viewpoint of quality control*. New York: Dover.

Yamarik, S. (2010). Does cooperative learning improve student learning outcomes? *The Journal of Economic Education*, 38 (3), 259-277.