1987

Measuring the Incremental Learning Achieved with Computer Enhanced Instruction

Larry N. Bitner
University of Richmond

Gail B. Wright
University of Richmond

Recommended Citation

Bitner, Larry N. and Gail B. Wright. "Measuring the Incremental Learning Achieved with Computer Enhanced Instruction." E.C.R.S.B. 87-8. Robins School of Business White Paper Series. University of Richmond, Richmond, Virginia.

MEASURING THE INCREMENTAL LEARNING

ACHIEVED WITH COMPUTER ENHANCED INSTRUCTION

Larry N. Bitner
Gail B. Wright

1987-8

MEASURING THE INCREMENTAL LEARNING
ACHIEVED WITH COMPUTER ENHANCED INSTRUCTION

Abstract

There has been increasing pressure by the AACSB on accounting educators to include the use of computers as an instructional tool in the 1980s with the advent of separate accreditation for accounting programs. In response, researchers have sought to study student attitudes and report on software available for classroom uses. From the educational perspective, however, the most important type of research would be that which would evaluate the impact of the computer on the level of learning achieved by the student or would indicate sensitivity to computer enhanced instruction (CEI). The focus of this paper is to suggest a research design to evaluate the computer's contribution to achievement in Accounting education in a model which will also examine the potential effects of confounding variables.

Introduction

Accounting educators have long recognized the need for students to learn
about computers. The American Accounting Association (AAA) Committee on
Accounting Instruction and Electronic Data Processing expressed this need in
1957. It was followed by other AAA committees of the 1960's also charged
with evaluating the role of computers in accounting education (AAA, 1970).
Currently, the American Academy of Colleges and Schools of Business (AACSB)
applies pressure through its accreditation standards to hasten computer
interactive learning processes. General business requirements for accreditation
anticipate sufficient computer hardware, software and support resources for
teaching and research. Additionally, separate requirements for accounting
accreditation specify that "Students shall receive instruction in the design,
use, control and audit of computerized information systems. Students are
expected to use the computer in accounting courses." (AACSB, p. 44). ·

The importance of computers in education is also acknowledged by accounting
practitioners who want to hire graduates with computer exposure. The use of
computers in audit and other client services demands that accountants seize
the opportunity presented by their familiarity of this tool in order to be
successful (Toth, 1985). The business community in general, represented by
a recent survey of controllers (Randall,1985), indicated that that portion
of the pool of new employees with no computer experience (31%) coupled with
another 21% of new employees with inadequate experience presented a weakness
of the present college-level education which needs to be overcome.

The question posed for study here focuses on the question of measuring
the potential educational impact of the use of computers in accounting
education. While information about the usage of computers--both hardware

and software--and attitudinal surveys are useful, a fundamental question is whether computers used to enhance instruction also improve the learning process and provide a positive force for academic achievement. To date, conclusive evidence of the computer's effect does not exist. Thus, a model should be developed which will more adequately measure the differential impact, if any, of computers used as an instructional tool in contrast with conventional instructional methods.

Literature Review

Recent research on the use of computers and computer software in accounting education can be divided into four categories: methods of computer integration, surveys of software, surveys of attitudes and measurement of differential educational effects. The first group of projects reports ways in which computers can be integrated into the accounting curriculum [Armitage and Boritz (1986) and Helmi (1986)]. The second category contains publications which surveyed the preprogrammed software available to educators [McKell and Stocks (1986)]. The third group of studies reported the attitudes of professors and students toward the use of computers in accounting education [Borthick and Clark (1986)]. The fourth category contains reports of experiments which attempt to measure the effects of computer integration on students' grades [Friedman (1981), Oglesbee, Bitner and Wright (1987)]. It is research on the impact of computer integration, or computer enhanced instruction (CEI), where the computer becomes an instructional tool, that is of great importance to the accounting educator.

Given the problem of testing for the differential effect of the computer as an instructional tool compared with traditional instructional methods, the primary research question may be stated as follows: Does the integration

of computer enhanced instruction (CEI) contribute to increased learning of accounting concepts? Several subsidiary questions may also be addressed. Three of the more important of these are stated as:

1. Does the instructor make a difference in the level of learning achieved by students exposed to CEI?

2. Does the learning potential or ability level of students affect the level of learning achieved with CEI?

3. Do established patterns of academic achievement make a difference in the level of learning achieved by students receiving CEI?

These would, of course, be appropriate questions in testing for differential effects of any innovative instructional technique.

Research Design

In order to address the above research questions, the investigator must select a surrogate variable for achieved learning and a research design that minimizes threats to the internal and external validity of the study. Variables that could be used as surrogates for achieved learning include the gain (loss) in scores between a pretest and a posttest and in other cases, the posttest score alone. Since a differential learning effect is being tested, the design should involve at least one pair of classes with a single instructor: an experimental or treatment (CEI) class and a control class which receives conventional instruction.

The selection of a design that minimizes threats to the internal validity of the study is complicated by the classroom environment in which the study would be conducted. In most cases, it is reasonable to expect that the experimental data will be collected in a college classroom where those students

participating in the test are chosen as a function of a college registration process. The important consideration is that there exists an implicit self-selection bias in a college registration process. Consequently, it is not possible to accomplish a true random selection of control and treatment subjects.

The inability to test random samples in this type of research is a critical issue. Given nonrandom (or nonequivalent) classes, the design chosen must permit a test which can distinguish between differential effects due to the treatment and those due to selection differences. A general research design suggested to test for the differential effects of CEI is demonstrated in Figure 1. The design is one of a pretest/posttest for nonequivalent groups. The pretest must be some measure which is related to the posttest measure. The researcher may choose to use either single or multiple pretest measures. Typical pretest measures that have been suggested include a student's SAT score, grade point average, or an examination score on a test which determines prior knowledge of course content. For advanced level accounting courses, grades in previous accounting courses may serve as a pretest.

FIGURE 1.

GENERAL RESEARCH DESIGN FOR
PRETEST/POSTTEST WITH NONEQUIVALENT GROUPS

	CEI Experimental Group (E)	Control Group (C)
Pretest measure	X_E	X_C
Posttest measure	Y_E	Y_C

Statistical Tests

Two basic statistical techniques are presented here as appropriate to test for any incremental benefits of CEI. These are matched pair testing and analysis of covariance (ANCOVA). Following is a discussion of the procedures for applying these techniques and the problems associated with each test.

Matched Pairs

Matched pair testing is an extreme case of analysis of variance (ANOVA) with blocking. For this case, each matched pair becomes one block and the statistical test reduces to a difference of means. Given the suggested experimental design, the experimenter has available two choices for a test statistic: these are the use of a gain score (G) or a posttest score only (Y). Use of a gain score presumes that an initial examination is given at the beginning of the experiment to test for pre-existing knowledge. The gain score then becomes the difference between the posttest and pretest scores. The difference of means test statistic compares the mean gain score of the experimental group with that of the control group to determine if any difference present is significant. The optimal situation is one where a standardized, validated instrument is available for a pretest instrument. When the posttest score is used as the test statistic, the researcher examines the difference in mean posttest scores for the experimental and control groups to determine whether the difference is significant.

Given the selection differences for nonequivalent groups, however small, the researcher is probably more interested in differential changes within groups (group gain scores) rather than differences between each group's posttest scores. Thus, gain score analysis with matched pairs is suggested as more
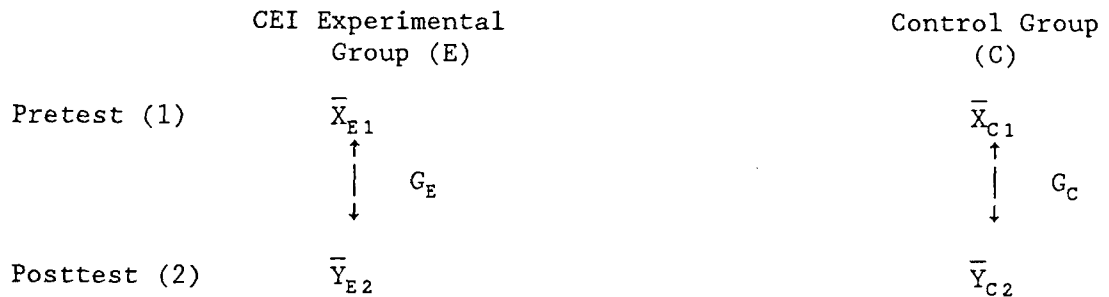
appropriate than posttest score analysis. The research design for gain score analysis is demonstrated in Figure 2. The mean test scores for each group are identified as $\overline{X}_{E1}$, $\overline{Y}_{E2}$, $\overline{X}_{C1}$ and $\overline{Y}_{C2}$. Finally, the mean gain score (G) is computed as follows:

$$G_E = \overline{Y}_{E2} - \overline{X}_{E1} \text{ and}$$

$$G_C = \overline{Y}_{C2} - \overline{X}_{C1}$$

FIGURE 2.

RESEARCH DESIGN FOR MATCHED PAIRS WITH
GAIN SCORE TESTING

	CEI Experimental Group (E)	Control Group (C)
Pretest (1)	$\overline{X}_{E1}$	$\overline{X}_{C1}$
	$\uparrow$ G_E $\downarrow$	$\uparrow$ G_C $\downarrow$
Posttest (2)	$\overline{Y}_{E2}$	$\overline{Y}_{C2}$

With this research design, hypotheses may now be developed and statistically tested as a means of answering the research questions previously set forth. The first hypothesis which addresses the primary research question is stated as:

H_0: Students receiving CEI demonstrate no statistically significant difference in learning accounting concepts from those receiving conventional instruction as measured by mean gain scores.

Notationally,

$$H_0: G_E = G_C$$

This hypothesis is tested using a difference of means. However, in order to increase the power of the test, students in the experimental group

are individually matched with students in the control group on the basis of some characteristic(s) which is expected to have a relationship with the treatment effect. Matching increases the power of the test to measure the effect of the treatment by reducing the standard error of the mean of gain scores. As a result, the test becomes more sensitive and diminishes the likelihood of accepting the null hypothesis when it is false.[1] Appendix A provides a statistical outline for the application of matched pair testing.

If the primary research hypothesis (no difference) is rejected, subsidiary hypotheses should be tested as they apply. In the case of multiple participating instructors, it is essential to be able to isolate an instructor effect if it is present. Thus, the first subsidiary question inquires whether the individual results are consistent with those of the overall analysis. Sections of students should be chosen so that control group sections can be paired with experimental group sections under the same instructor. The hypothesis to be tested is as follows:

H_0: There is no statistically significant difference in the performance of students receiving CEI from those receiving conventional instruction that can be attributed to the instructor.

Notationally,

H_0: $G_{Ei} = G_{Ci}$ for all i = 1,2,...m where m = the number of participating instructors.

The statistical test chosen to resolve this question should present evidence of any instructor's subsample of students which differ from those of the entire sample. Since there is no particular need to test for differences in learning (gain scores) between instructors at this point, a test for

_____

[1]For a more detailed discussion of matched pair testing see J.P. Guilford, <u>Fundamental Statistics in Psychology and Education</u>. New York: McGraw Hill, 1965, pp. 194-197.

difference of means again serves as an appropriate test.[2] Thus, each instructor's results are tested using the same procedure outlined for the total sample.

The second suggested subsidiary question ponders whether the students' ability level affects the learning to be achieved with CEI. Assuming no instructor differences appear, the total sample of matched pairs of students may be ranked by a measure of ability such as SAT scores. The top half would represent a high ability group and the lower half, the low ability group. A difference of means test may then be applied to the high and low ability groups separately. Results of the treatment effect would be observed on students subdivided by ability in order to make recommendations for instructional methods as may be found appropriate. The hypothesis to be tested here is stated as:

H_0: There is no statistically significant difference in the performance of students receiving CEI from those receiving conventional instruction that can be attributed to differences in learning ability.

Notationally,

H_0: $G_{Ej} = G_{Cj}$ for all $j = 1, 2, \ldots p$, where $p =$ the number of student ability levels tested

If teacher differences do appear in the results, ability level differences must be tested separately by instructor. If instructor differences appear or should the researcher desire to test achievement sensitivity with more than two student ability levels (e.g., high, medium and low) then ANOVA rather than simply a difference of means test would appropriately be applied.

---

[2]It may be noted that, when ANOVA with blocking is taken to the extreme of matched pairs, no test for interaction of blocks and treatment can be made.

The third subsidiary question involves the use of matching variables which measure prior achievement patterns. These might include GPA and/or grades in prior accounting courses. Testing this question requires an analytical procedure similar to that for the second question.

ANCOVA

ANCOVA tests for treatment differences by a matching procedure similar to that described in the preceding section. ANCOVA, however, compares the students statistically as opposed to the judgmental process utilized to select the criterion characteristic with matched pairs. ANCOVA is a variation of ANOVA which tests the significance of any difference between posttest score means after adjusting for initial group differences as measured by pertinent independent variables (pretest measures). Pretest measures represent variables determined to have a potentially significant effect on the posttest measure. Since these variables are not expected to have a constant effect across all sample subjects, ANCOVA adjusts the data to control for any initial differences. These pretest measures are covariates and, in order to be useful, must be correlated with the posttest score at a minimum level of 0.4. The research design to test for CEI treatment differences for the single covariate case is illustrated in Figure 3. The hypothesis to be tested may be stated as:

> H_0: Students receiving CEI demonstrate no statistically significant difference in learning accounting concepts from those receiving conventional instruction measured by posttest scores.

This is, of course, the same null hypothesis stated for the matched pairs test, however measured by a different test statistic: the adjusted difference in posttest scores between groups. (See Appendix B for a notational statement of H_0 and H_1.)

FIGURE 3.

RESEARCH DESIGN FOR ANCOVA TESTING
OF NONEQUIVALENT GROUPS WITH PRETEST/POSTTEST

	CEI Experimental Group (E)	Control Group (C)
Pretest measure (covariate)	X_E	X_C
Posttest measure	Y_E	Y_C

For the single covariate case used by Friedman, the first step in the ANCOVA process is to develop a regression equation with the posttest score (Y) as the dependent variable and a pretest measure (X) as the independent variable or covariate for each group.[3] A key assumption is that the regression lines for the experimental and the control group are both linear and parallel. If a treatment difference exists, then, a vertical displacement between the two regression lines will exist. A statistically significant difference would suggest that, given the same pretest measure(s), one group will outperform the other, i.e., the treatment effect is significant on the learning achieved.

Observation of the experimental data may indicate that the regressions are either nonlinear and/or nonparallel. The ANCOVA model can easily be adapted to either or both of these situations. A real interpretation problem for the nonparallel case, however, is to determine at what point to measure the treatment difference since the two regression lines will not be equally displaced over all values of X. A check should be made to determine whether results change as the point of measurement is arbitrarily changed. If results

---

[3]A more in depth presentation of using ANCOVA with nonequivalent group designs is presented in T. D. Cook and D. T. Campbell, Quasi-Experimental Design and Analysis Issues for Field Settings. Boston, Houghton Mifflin Company, 1979, pp. 153-175.

do change, stratifying the data to test for sensitivity with high or low groups as discussed previously under matched pairs may resolve the dilemma.

The ANCOVA model may be further expanded to use multiple covariates (pretest measures). Employing multiple covariates can increase the power of the test, assuming that the additional covariates are related to the posttest. SAT scores or GPA may be used in addition to pretest examination scores or grades in prior courses. For example, the researcher may feel that performance on the posttest will be affected by general ability levels (measured by SAT scores), by prior knowledge of course material (measured by a pre-experimental examination), or by prior achievement patterns (GPA or grade in prerequisite course(s)). Additional covariates which meet minimum correlation standards (e.g., $r > 0.4$) increase the power of the test by further controlling for the presence of initial differences between groups so that the significance of the treatment effect is appropriately measured.

As with any multiple regression model, multicollinearity checks should be made. The subsidiary questions are indirectly tested by introducing covariates into the analysis. If covariates which meet minimum correlation standards enter the model and affect the level of significance in the statistical results, it can then be determined whether differences in the posttest measure result from the treatment effect or arose as the result of initial differences between groups. Treatment sensitivity to student ability level or prior achievement patterns can be tested by stratifying the data (e.g., by SAT scores) and employing dummy variables in the ANCOVA model to represent each stratum. Appendix B provides specific details of the ANCOVA model used in testing for treatment differences.

Comparison of Matched Pairs with ANCOVA

The choice of the appropriate test for differences is difficult as it is not clear, for the nonequivalent case arising in class test situations, which method offers more precision. Any choice must obviously consider conditions unique to the researcher; however, some points for consideration are offered here.

If the researcher is either unwilling or unable to specify the shape of the regression of the posttest score on the pretest, he may also be unwilling to assume parallel slopes or identify a point of measurement such as the posttest score. In this situation, matched pair testing using gain scores as the test statistic may be more suitable than ANCOVA. Since the matched pair test does not require any specification of the shape of the regression of the posttest on the pretest, at least one potential source of bias is eliminated. On the other hand, if the regressions are linear and parallel and, in addition, the posttest score is used as the test statistic, then the expected value of the treatment effect for matched pairs converges on that for ANCOVA. Further, in the extreme case when all pretest scores are perfectly matched, there will be no difference in the treatment effects as measured by the two methods.

Another point to be considered is the potential for bias in the matching procedure when students for whom no match can be found are omitted. Further, when multiple pretest measures exist, i.e., SAT, GPA, grades in prior related courses, etc., it may be difficult to identify a single criterion variable or to establish a priority for multiple criteria for matching pairs of students. Thus, matching of students in the control group with those in the experimental group becomes a matter of subjective judgment.

For the nonequivalent case then (where perfect matches do not occur), the superior test is that which can best adjust for initial differences without a subjective choice in the criterion variable. Matching focuses on a comparison of within group changes (gain score) whereas ANCOVA focuses on between group changes after controlling for initial differences. If the linear and parallel assumption can be satisfied and if the covariate(s) used in ANCOVA were able to control for all initial differences, ANCOVA would be the clear choice.

A final point worth considering is ease of application. All other things equal, ANCOVA is clearly a simpler and less subjective test to apply. Unfortunately, few of these other things ever seem to be equal during actual testing. As usual then, no clear choice surfaces and all the above considerations must be weighed in making a minimally arbitrary choice. It may be noted that the early researchers in this area appear to favor ANCOVA testing.

Concluding Remarks

The intent of this paper is to provide a robust research design that will stimulate research to evaluate student achievement under new educational methods. Of particular concern is the ability to measure and interpret the effect that CEI used for instructional purposes has on the level of achievement of accounting students. The contribution of CEI beyond the aspects of technical training should be measured since integration into the curriculum is expected by accreditation committees and the business community alike.

To date, studies in accounting attempting to measure the effects of computer interactive learning, have provided inconclusive if not conflicting results. Some confusion could be eliminated if a rigorous research design were employed. If indeed, CEI is to be required as an integral part of accounting education, educators should be made aware of any benefits which

may exist beyond technical training. Conversely, if CEI does not aid in the teaching of accounting concepts, educators and the AACSB must determine whether technical training in computers is an appropriate goal in itself of the college accounting curriculum.

<u>Appendix A</u>:

Testing for Differences in Mean Gain Scores
Of Matched Pairs

A first step in the test requires a computation of the standard error
of the mean of gains (σ_{MG}) for each group E and C. Algebraically, for E
this may be stated as

$$\sigma_{MGE} = (\sigma_{GE}/\sqrt{N - 1})(\sqrt{1 - r_{12}^2})(\sqrt{1 - (N/N_p)})$$

where $\quad$ σ_{GE} $\quad$ = sample standard deviation of gain scores in the experimental
group;

$\quad\quad\quad$ N $\quad$ = number of students in the experimental group ($N = N_E = N_C$
for matched pairs);

$\quad\quad\quad$ N_p $\quad$ = total number of students enrolled in accounting during the
test semester;

and $\quad\quad$ r_{12} $\quad$ = sample coefficient of correlation of the pretest and posttest
scores in the experimental group.

The term ($\sqrt{1 - r_{12}^2}$) is needed to adjust the standard deviation for the cor-
relation between the pretest and posttest scores. The final term,
($\sqrt{1 - (N/N_p)}$) is a correction factor for a finite population.
Similarly, for the control group,

$$\sigma_{MGC} = (\sigma_{GC}/\sqrt{N - 1}) (\sqrt{1 - r_{12}^2})(\sqrt{1 - (N/N_p)}).$$

Once these standard errors are determined, the standard error of the
difference in means gains (σ_{DMG}) is computed as

$$\sigma_{DMG} = \sqrt{\sigma_{MGE}^2 + \sigma_{MGC}^2 - 2r_{EC}\sigma_{MGE}\sigma_{MGC}}$$

where $\quad\quad$ r_{EC} = coefficient of correlation between the gain scores of the
matched pairs of students.

The significance of the difference in mean gain scores ($G_E - G_C$) may be
determined by converting it to a z value as follows:

$$z = (G_E - G_C)/\sigma_{DMG}.$$

Given that, a priori, there is no reason to expect that one group will perform better than the other, a two tailed test is appropriately used here. Of course, a paired t-test may be used if class sizes are small.

Appendix B:

Testing for Treatment Differences
With ANCOVA

The regression model utilized by ANCOVA with a single covariate (the pretest such as grade of the preceding semester as used by Friedman) may be stated as follows;

$$Y_{ij} = \mu + \alpha_i + \beta (X_{ij} - \overline{X}) + \epsilon_{ij}$$

where Y_{ij} = the posttest score for the jth (j = 1,2,...,n)

individual in the ith group (i = 1 or 2, e.g., E or C)

μ = the population mean posttest score

α_i = the treatment effect

β = the coefficient of within group regression of posttest on

pretest

X_{ij} = the pretest score

$\overline{X}$ = the overall mean pretest score

and ϵ_{ij} = the error term.

The estimate of the treatment effect then becomes

$$\hat{\alpha}_E - \hat{\alpha}_C = (\overline{Y}_E - \overline{Y}_C) - \hat{\beta}(\overline{X}_E - \overline{X}_C).$$

The null hypothesis to be tested is that of no difference in the treatment.

Notationally,

$$H_0 : \alpha_E = \alpha_C$$

The general model for the k covariate case is

$$Y_{ij} = \mu + \alpha_i + \beta_1 (X_{1ij} - \bar{X}_1) + \beta_2 (X_{2ij} - \bar{X}_2) + \ldots + \beta_k (X_{kij} - \bar{X}_k) + \epsilon_{ij}.$$

The estimate of the treatment effect for this case becomes

$$\hat{\alpha}_E - \hat{\alpha}_C = (\bar{Y}_E - \bar{Y}_C) - \hat{\beta}_1 (\bar{X}_{1E} - \bar{X}_{1C}) - \hat{\beta}_2 (\bar{X}_{2E} - \bar{X}_{2C}) - \ldots - \hat{\beta}_k (\bar{X}_{kE} - \bar{X}_{kC}).$$

REFERENCES

AAA Committee on the Role of the Computer in Accounting Education. (1970) Report of the 1968-69 Committee on the Role of the Computer in Accounting Education, The Accounting Review, 45, 29-43.

American Assembly of Collegiate Schools of Business. (1985-86) Accreditation Council Policies Procedures and Standards. St. Louis: AACSB.

Armitage, H. M. and Boritz, J. E. (1986) Integrating Computers into the Accounting Curriculum, Issues in Accounting Education, 1, 86-101.

Borthick, A. Faye and Clark, Ronald L. (1986) The Role of Productive Thinking in Affecting Student Learning with Microcomputers in Accounting Education, The Accounting Review, 61, 143-157.

Cook, T. D. and Campbell, D. T. (1979) Quasi-Experimental Design and Analysis Issues for Field Settings. Boston: Houghton Mifflin Company.

Friedman, Mark E. (1981) The Effect on Achievement of Using the Computer as a Problem-Solving Tool in the Intermediate Accounting Course, The Accounting Review, 56, 137-143.

Guilford, J. P. (1965) Fundamental Statistics in Psychology and Education. New York: McGraw Hill.

Helmi, Hedhat A. (1986) Integrating Computers into the Accounting Curriculum, Issues in Accounting Education, 1, 102-111.

McKell, Lynn J. and Stocks, Kevin D. (1986) An Evaluation of Computerized Accounting Practice Sets, Journal of Accounting Education, 4, 177-190.

Oglesbee, Tom W., Bitner, Larry L., and Wright, Gail B. (1987) Measurement of Incremental Benefits in Computer Enhanced Instruction: A Field Study. (In review).

Randall, Robert F., Editor. (1985) Data Sheet, Management Accounting, 67, 16.

Toth, Edward J. (1985) Changes Affecting Accounting Education, The Ohio CPA Journal, 23-25.

AUTHORS

Gail B. Wright, Assistant Professor, Department of Accounting, University of Richmond, Richmond, VA 23173. Degrees: A.B., College of William and Mary; MS, Virginia Commonwealth University; DBA, George Washington University. Specialization: Accounting Educational Performance.

Larry N. Bitner, Assistant Professor, Department of Accounting, University of Richmond, Richmond, VA 23173. Degrees: BSME, MBA, West Virginia University; DBA, George Washington University. Specialization: Accounting Performance Assessment.