Alicias, Jr., Eduardo R.

Toward an Objective Evaluation of Teacher Performance: The Use of Variance Partitioning Analysis,
VPA

Education Policy Analysis Archives/Archivos Analíticos de Políticas Educativas, vol. 13, 2005, pp. 1-13

Arizona State University

Arizona, Estados Unidos

Toward an Objective Evaluation of Teacher Performance: The Use of Variance Partitioning Analysis, VPA[1]

Eduardo R. Alicias, Jr.
New England College
The Philippines

Abstract
Evaluation of teacher performance is usually done with the use of ratings made by students, peers, and principals or supervisors, and at times, self-ratings made by the teachers themselves. The trouble with this practice is that it is obviously subjective, and vulnerable to what Glass and Martinez call the "politics of teacher evaluation," as well as to professional incapacities of the raters. The value-added analysis (VAA) model is one attempt to make evaluation objective and evidenced-based. However, the VAA model—especially that of the Tennessee Value Added Assessment System (TVAAS) developed by Dr. William Sanders—appears flawed essentially because it posits the untenable assumption that the gain score of students (value added) is attributable only and only to the teacher(s), ignoring other significant explanators of student achievement like IQ and socio-economic status. Further, the use of the gain score (value-added) as a dependent variable appears hobbled with the validity threat called "statistical regression," as well as the

---

[1] Copyright Registration No. O 2004-037, February 26, 2004, The National Library, Manila, The Philippines

problem of isolating the conflated effects of two or more teachers. The proposed variance partitioning analysis (VPA) model seeks to partition the total variance of the dependent variable (post-test student achievement) into various portions representing: first, the effects attributable to the set of teacher factors; second, effects attributable to the set of control variables the most important of which are IQ of the student, his pretest score on that particular dependent variable, and some measures of his socio-economic status; and third, the unexplained effects/variance. It is not difficult to see that when the second and third quanta of variance are partitioned out of the total variance of the dependent variable, what remains is that attributable to the teacher. Two measures of teacher effect are hereby proposed: θ_1 for proportional teacher effect and θ_2 for direct teacher effect.

The Need for an Objective Teacher Evaluation

There is an obvious need for objective teacher evaluation. First, on equity considerations, there is a need to establish a direct link between teacher productivity and teacher compensation. Clearly, it should be the case that the more productive teachers should be paid more and/or should be given priority for promotion. Or, by the same token, the teaching laggards will have to be considered first for compulsory retraining or even dismissal, if the law allows. Second, on optimality grounds, there is a need on the part of the school administrator to deploy his teachers in the teaching of courses where they can demonstrate their utmost competencies—or, in the language of the economist, where their largest respective marginal productivities lie. For example, if it is shown that a mathematics teacher is more productive in the teaching of, say, algebra than in geometry, then the Pareto principle of optimality would dictate that said teacher, *ceteris paribus*, would have to teach algebra instead of geometry. Consequently, on the whole, the school will then tend to move to a higher level of productive optimality. And, third, it makes significant political sense that the school administration is perceived to be fair and impartial in the assignment and compensation of teachers. Obviously, this will minimize the occurrence of distractive uncooperativeness or at times even destructive resistance on the part of the teachers.

What's Wrong With the Traditional Practice of Teacher Evaluation?

Teacher evaluation is essentially and almost always done with the use of ratings made by students, peers, and principals or supervisors, and at times, self-ratings made by the teachers themselves. The trouble with this evaluation scheme is that it is obviously subjective and vulnerable to the quirks and frailties of the raters, not to mention their professional incapacities. For example, what sense can one make of a principal whose professional specialization is English and then observing and suggesting that the mathematics teacher handle, say, quadratic equations in this or that way? Or, what if the principal, for one reason or another, simply dislikes the teacher? Indeed, these occurrences, as well as those described in Glass & Martinez (1993).

The traditional practice is teacher-centered in that it uses ratings about traits and behavioral patterns of the teacher, rather than those about the students. Of course, students are the reasons for existence of schools and teachers; and therefore, whatever happens or not happens about them in the name and process of teaching should be the basis of measuring the effectiveness of said process.

Implicit in the traditional scheme are some global and commonly accepted but essentially unvalidated assumptions about teacher traits and behavior, such as: (1) teacher performance is a monotonic increasing function of educational attainment and/or professional seminars and in-service trainings undertaken, and (2) there is a standard teacher classroom behavior against which individual teacher behavior is measured.

In regard to the first, there is no convergence of evidence showing that the more highly educationally qualified teachers are more effective in the classroom. The second is simply heroic. What is that standard behavior? Does it make empirical sense—and, if so, in all or what subject areas? Who should define this behavioral pattern? The principal or a professional body? On what light would they do that? That of revelation (dogma) or that of science? By the way, what is the nature of teaching? Is it art or science? If it is art, then why can't we just leave the individual teacher alone to his own artistic devices? On the other hand, if it is science, where then is that unambiguous corpus of scientific knowledge that predicts with great probability that, say, this particular teacher behavior will produce this much of this type of student achievement within this length of time? In this age of post-Einsteinian relativity (e.g., supersymmetry and superstrings are on the horizon), does it make sense to consider the nature of teaching as deterministic as that?

In any event, the traditional practice also moves against the flow of professional autonomy of the teacher who is licensed by the state to practice his profession. This is anchored on the academic freedom of the teacher, which apparently is now a well-settled and universally accepted principle.

The Value-Added Analysis: Is This a Valid Attempt at Objectivity?

The value-added model of teacher evaluation seeks to isolate the additional learning (the value added) that is presumed to have occurred at the end of a teaching-learning cycle, say, at the end of a term or school year—and, by some mode of reasoning, attribute this increment to the teacher.

Ernest Pascarella (1986) described this model as attempting to separate the net effects of instruction from previous ability or simple maturation. He suggested ways to improve value-added assessment, namely: cross-sectional research design, methods of estimating the effect of a particular learning experience independent of students' pre-learning differences, multiple regression analysis, analysis of joint or redundant effects not directly attributable to instruction, and the development of causal models. (ERIC CD-ROM, 1985-1998). Likewise, he noted the possibility that not all students may benefit equally from the same experience. This is a reality that always happens in the classroom, and apparently this is something that the value-added model is unable to capture.

In 1998, Jill Berlin Slack and Edward P. St. John used a variant of the value-added model—the sequential analysis—to examine the association of specific factors to test score improvement. Among others, their findings showed the significant impact of age, gender, school environment, and curriculum and instruction on improvement.

However, their most consistent and significant finding is that "higher ability students were less likely to improve than lower ability students." They argued that this finding is consistent with the Accelerated Schools philosophy that "disadvantaged" students stand the most to gain from innovative teaching approaches. (ERIC CD-ROM, 1985-1998)

The *Tennessee Value Added Assessment System* (TVAAS) is arguably the most famous of such assessment systems. It was designed and operationalized by Dr. William Sanders, acknowledged by many as the value-added analysis guru. He claims that by carefully tracking student progress over time—with his "mixed-model" statistical methodology—he can gauge student academic performance—and the teacher effect on that performance—that is more accurate and fair than earlier measures.

Apparently, however, Dr. Sanders has not cared to publish a complete and detailed description of his model (other researchers are encouraging him to do so in a refereed journal), so we rely on how others describe his model and/or his own general description of his model. For example, Jeff Archer (1999) describes Sanders' approach as follows:

> While other researchers have spent years struggling to control for differences in students' backgrounds—such as family income and parents' educational levels—Mr. Sanders lets each student act as his or her own control. To do that he focuses on gains, instead of raw scores, so that each student's performance is compared not with that of similar students, but against his or her own past performance. The tool he uses is called mixed-model methodology. Though written into the Tennessee school code, its exact operation is nearly incomprehensible to a layperson. (1999 Editorial Projects in Education, Vol. 18, Number 34, pp 26-28)

It appears that Sanders attempted an improvement on traditional statistics, that is, an improvement on conventional trend or time-series analysis. Is it a multivariate analogue of a simple or even an interrupted time-series analysis? If so, how many time intervals are included in his model? Did he say more or less three years? Would that be adequate enough to yield a valid analysis?

Anyway, consider the following example as mentioned by Archer. If a teacher taught just one student for one year and that student made poor progress, then traditional statistics would predict that the teacher's next student would falter as well. However, on the other hand, Sanders' mixed-model would take that single result and predict that the next student would make gains that would only be slightly worse than the average for all the teachers' students.

Archer described the mixed-model as that involving a weighting of results based on how much information is available. He further described the statistical algorithm as a "magic" called "shrinkage estimation," and what it yields is called a Best Linear Unbiased Predictor (BLUP).

Sanders' BLUP—Magical, Mystical. We may grant that Sanders' BLUP is a magical algorithm—but, at least for now, its magic appears far too mystical to be clearly understood and appreciated by those directly concerned—the ordinary classroom teachers and/or school principals. Indeed, he owes it to the interested readers—the

scientific community at large—to publish a clear description of what is reported to be his claim as the "best" estimate of the teacher effect on student achievement.

In view of the absence of a clear description of how he mixed the ingredients, as it were, of his mixed-model, we now speculate and interpret Sanders' model as follows.

Let us ask some basic questions, but first let us lay down some basic premises consistent with what are reported to be such ingredients. Well, first, he tracks student progress over time, presumably to assemble a set of time-series data consisting of a finite temporal chain of discrete incremental values (value-added quantities). Second, the student acts as his own control—allegedly controlling for vital background factors such as family income and parents' educational levels (as reported by Archer). With the data thus assembled, let, say, *delta 10th* be the observed value-added of student X for the last (10th) stage in, say, a 10-stage learning cycle. Then the BLUP algorithm is applied, weighting into account as much relevant available information as possible. And bingo, the BLUP of the teacher effect on student X's achievement—presumably a portion of delta 10th—comes to the fore.

Now, what nonzero weights would Sanders assign to such relevant information as family income and parents' educational levels? Would the weights partake of the nature of Bayesian probabilities? Anyway, there appears to be the necessary implicit assumption that BLUP must always be less than delta 10th; otherwise, if he should assume equality, then the weights of all the other background factors would each be reduced to a nullity—contrary to an implied premise of his algorithm—and, of course, contrary to the weight of empirical evidence. At any rate, it would be much too unrealistic and utterly counterintuitive for him to posit that the whole of delta 10th is determined by just one and only one factor—the teacher.

Further, in light of available literature, why is there no explicit mention of two other factors which are probably more important, namely: the intelligence quotient (IQ) and the learning or cognitive state of student X at the beginning of the 10-stage cycle (pretest score). If these are included in the available "relevant information," then again, what would be their subjective weights? And by how much would delta 10th be further adjusted downward because of said weights? Does he also assume that family income, parents' educational levels, and the other available relevant information are invariant over time?

If the unit of analysis is a student or a cohort of students undiminished by attrition, it may be granted that IQ is invariant over time (although the basic question remains, what is the effect of that invariant IQ), but then what about the other relevant factors? For example, what happens if the parents' educational levels increase, say, at the end of the 8th stage of the cycle? Is the BLUP algorithm designed to handle such intertemporal variations of some relevant covariates? Also, what about the effect of maturation on the part of the student? In short, why did he purposely exclude factors like IQ, family income, and parents' educational attainment—or equivalently, assume the same to be constant—factors which time and again appear in the literature as significantly impinging upon student achievement?

In a related vein, Sherman Dorn (see Glass, 1995) mentions a number of problems afflicting the value-added assessment system, the most important of which apparently vitiates the gain score (value added) as a basis for statistical analysis. Dorn pursues the point as follows:

> "A gain score is a questionable basis for statistical analysis. Gain scores conflate the effects of two different teachers. VAA may seriously underestimate the effects of prior knowledge, social background, etc. Would the effects be different if you put in sex, race, economic class, perhaps a square of last year's scores, in the equation? I bet no one knows."

Glass, on the same occasion, further drives home the point as follows:

> "Now imagine—and it should be no strain on one's imagination to do so—that we have Teacher A and Teacher B and each has had the pretest (Sept) achievement status of their students impeccably measured. But A has a class with average IQ of 115 and B has a class average IQ 90. Let's suppose that A and B teach to the very limit of their abilities all year long and that, in the eyes of God, they are equally talented teachers. We would surely expect that A's students will achieve much more on the posttest (June) than B's. Anyone would assume so; indeed, we would be shocked if it were not so."

Indeed, there are many questions in need of answers, and there may be more that are needed to be asked, not to mention a number of other possible assumptions and a slew of other subjective weights (probabilities). There is small wonder then that according to Dorn "VAA is not an evaluation system accessible to teacher understanding." Indeed, the crucible where Sanders mixed the ingredients of his mixed-model recipe appears much too mystical and cryptic.

Are There Flaws and/or Weaknesses of a Value-Added Model? We may grant that Sanders had the foresight and wisdom to anticipate and adequately factor into his model all such aforementioned questions, assumptions, and subjective weights; but it is not clear from the available literature how he handled what apparently are some possible flaws and/or weaknesses of such a value-added model.

In fact, woefully, instead of meeting head-on the straightforward and pointed remarks of Dorn and Glass, he dished out arcane technical jargon and asserted tangential generalities as follows:

> We do not even calculate simple gains. For example, we use the whole observation vector for each child over all subjects and grades. This approach is superior to traditional multivariate approaches. As we apply these approaches in the context of the estimation of the teacher and school effects on the academic growth of populations of students, we take advantage of the prior knowledge of the distribution of the variance-covariance structure among populations of teachers, as well as the variance-covariance structure among students.

Regardless, **first**, the value-added model (using gain scores) appears flawed with the methodological threat to internal validity called "statistical regression." This is the phenomenon wherein larger incremental values (value-added quantities) are observed on administration of post-test among students with lower pretest scores than those with

higher pretest scores. Apparently, this was what Pascarella referred to when he noted, as stated earlier, the possibility that not all students may benefit equally from the same experience. Ironically and in fact, Sanders himself noticed this phenomenon which he called "shed patterns." Unfortunately, wittingly or unwittingly, he apparently chose to ignore the significance of this phenomenon. In this respect, Archer reported about Sanders' observation, thus:

> In many urban schools, he has noticed a pattern in which students with the lowest past performance make the greatest gains, but those who start with high scores make little headway. A graph of such gains against past performance creates a downward sloping line from left to right. He calls these "shed patterns." (1999 Editorial Projects in Education, Vol. 18, Number 34, pp 26-28)

This phenomenon probably underpinned the apprehension articulated by Tom Mooney, president of the Ohio Federation of Teachers, as reported by Willard and Oplinger (2003), about the use of a value-added model in the evaluation of teachers. It was reported that Mooney did not want to use a value-added approach to the evaluation of Ohio teachers because the same "could divide teachers as they try to avoid hard-to-educate children." If, in fact Sanders failed to account for this phenomenon in his algorithm, then indeed that weakness would eventually be translated into the behavior of teachers trying hard to avoid low IQ or slow-learning students. (Willard, Dennis J. and Doug Oplinger. "'Value-added analysis' credits school districts for progress a student makes," Columbus, Ohio: *Beacon Journal*; May 26, 2003).

This problem is essentially related to that posed by the likelihood that this phenomenon may yield the spurious conclusion that an observed incremental value is significant, when in fact it may just be a statistical artifact. The fundamental question is then asked: which is indicative of greater teacher productivity, the larger gains that may arise from the lower pretest scores or the smaller gains that may arise from the higher pretest scores? Or, by the same token, which is indicative of greater productivity: smaller increments arising from low-IQ students or larger increments arising from high-IQ students.

There are two possible answers to this question, depending on the underlying assumption. If the *unrealistic assumption* is made that the entire value-added (gain score) is caused by the teacher and only by the teacher (apparently this is Sanders' basic assumption but which we deem it much too heroic), then the larger gains arising from the lower pretest scores or those arising from high-IQ students would certainly be indicative of greater teacher productivity.

On the other hand, if the *realistic assumption* is made that only a portion of that value-added (gain score) is causally attributable to the teacher, then it is possible that—after isolating the effects of the relevant non-teacher factors, including but not limited to IQ and the pretest scores—the smaller gross gains arising from the higher pretest scores or those arising from low-IQ students would yield *proportional net gains*, possibly even *larger net gains*—thereby truthfully indicating greater teacher productivity.

To belabor the point, an analogous though exaggerated question is asked: which is more productive of energy, splitting a massive log into smaller pieces of firewood or splitting just one extremely smaller uranium atom?

The **second** weakness of a value-added approach—particularly that of Sanders—appears to be that it cannot get away from the constraints imposed by current theory and practice of teaching, especially the concomitant difficulty of measurement. This difficulty arises because, as herein above described by Archer, it involves a weighting procedure depending on "how much information is available." This means that the BLUP should be more valid if there is so much more information made available. This consequently begets the fundamental question: what are the bits and pieces of information that are available—and perhaps more importantly, what are those that should be made available that were not factored into the estimation of Sanders' BLUP. For instance, what are the other teacher factors that are embodied in the teacher and therefore necessarily inputted into the teaching-learning process that produce a desired incremental value but were not captured by the BLUP?

In view of the fact that there is a great and bewildering multiplicity and indeterminacy of both observable and unobservable teacher factors, then the task of identifying which to conceptualize and to measure, aside from the few that are being traditionally used in research, should in fact be attended with mind-boggling difficulties. The intangible and unobservable teacher traits like desires, motivations, attitudes, values, philosophies, fears, anxieties, emotions, affections, dreams, mania, native abilities, etc., etc., etc. are particularly difficult to handle. What about their behavioral patterns in the classroom? Their teaching styles and techniques? Their physical appearance and dressing style? What about their patterns of social interactions—professional, family, marital or even extra-marital relationships? What about their religious convictions? Indeed, what about an almost infinity of other teacher-related factors?

The **third** apparent weakness has something to do with practical considerations. In addition to what has been mentioned earlier on, Lynn Olson (1998) reported that many worry about the complexity of the statistical techniques used, thus making the value-added approach vulnerable to misunderstanding by the public at large, particularly the parents and the taxpayers. Olson elaborated that "public support is considered a crucial element of accountability efforts, and states and districts have long been criticized for using language and statistics that confuse, rather than enlighten." Likewise, she cited the ambivalent description of Carol Ascher, senior research scientist at the Institute for Education and Social Policy, New York University, about the features of a value-added approach to teacher evaluation, thus: "As an idea, it's very appealing. It feels very progressive. It feels fair. But the execution of it is so problematic." (Olson, Lynn. "A Question of Value," *Education Week on the Web*; May 13, 1998).

The Variance-Partitioning Analysis (VPA) Model: Conceptual And Methodological Advantages

It appears that the attempt at crafting an objective teacher evaluation scheme in the context of an input-oriented and/or teacher-centered framework is fraught with conceptual, methodological, and measurement difficulties. More importantly, the purported measure of teacher effect constructed therefrom appears untenable on closer scrutiny.

We instead propose a variance-partitioning analysis (VPA) model in the context of an output-oriented and student-centered evaluation model. The concept appears elegant in its simplicity. There is a pie, as it were, that represents the total variance of a set of achievement scores on a particular dependent variable or criterion. This pie is

then partitioned into various angular portions representing: first, the effects attributable to the set of teacher factors; second, effects attributable to the set of control variables the most important of which are IQ of the student, his pretest score on that particular dependent variable (criterion), and some measures of his socio-economic status like family income and/or the occupations and educational attainments of his parents; and third, the unexplained effects or unexplained variance.

This proposed model (VPA) takes into account the discomfort, at least, of Glass and Dorn about the non-inclusion of relevant and significant covariates like student IQ, pretest scores, and some socio-economic variables. On the occasion of the aforementioned Internet Discussion of the Tennessee Value-Added Assessment System (Glass, 1995), Dorn instructively asserts, thus:

> "…entering the prior years' scores as covariates solves the problem. Solving Glass' conundrum means that one assumes a linear relationship between first set of scores and second set of scores, but that's much more tenable than assuming an expected gain that's constant across the distribution of first sets of scores."

Further, it takes into account the suggestions made earlier by Pascarella that multivariate analysis of covariance be used on cross-sectional data. Likewise, it is not difficult to see that this VPA model should not be infirmed with the major flaws and/or weaknesses of the value-added approach as described earlier on, including that which induces teachers to avoid hard-to-teach or low-IQ students.

So, let evaluation be student-centered. Shift the focus of evaluation from teacher traits and behavior to what fundamentally matters—that is, student achievement. Pending the appearance and wide acceptance of a better alternative, let traditional cognitive achievement tests measure student achievement. And by making the class the unit of analysis, cross-sectional data on the component students are thereby available and to be used in a manner consistent with Pascarella's suggestion.

The beneficial effect of an output-oriented (achievement-oriented) evaluation model stems from the fact that the teacher is at liberty to use his individual creativity and artistry—on top of the indications of scientific knowledge—to do and to act in a manner consistent with what he thinks is best in the classroom. The teacher is thus empowered in the classroom as he must be. In this connection, Sanders and Horn (1995)—quite acceptably this time—state as follows:

> By focusing on outcomes rather than the process by which they are achieved, teachers and schools are free to use whatever methods prove practical in achieving student academic progress.

The VPA may be done in a way outlined hereunder as follows:

Consider the basic variance equation:
$$\sigma_Y^2 = \sigma_X^2 + \sigma_U^2; \qquad\qquad \text{(Equation 1)}$$

> where $\sigma_Y^2 = $ total variance of student achievement (post-test criterion scores);

$\sigma_X{}^2$ = explained variance attributable to the set of independent variables measured, specified, and included in the analysis; and

$\sigma_U{}^2$ = unexplained variance attributable to all the other variables not included in the analysis.

In light of experience and practical reality, $\sigma_X{}^2$ (explained variance) can be broken down into:

$\sigma_T{}^2$ = the variance attributable to all the known and unknown, as well as the observed and unobservable factors (traits) that are embodied in the teacher (teacher variance); and

$\sigma_C{}^2$ = the variance attributable to the control variables the most important of which are the student's intelligence quotient (IQ), the student's socio-economic status (mainly indicated by family income or alternatively by the proxy variables: parents' occupation and parents' educational attainment), and pre-learning achievement level (pretest score).

By definition, it is presented that $\theta_1 = \sigma_T{}^2 + \sigma_U{}^2$;   (Equation 2)

where θ_1 stands for "proportional teacher effect."

Thus, the basic variance equation (Equation 1) may now be rewritten as:

$$\theta_1 + \sigma_C{}^2 = \sigma_Y{}^2. \qquad\qquad\text{(Equation 3)}$$

By transposition, this reduces to:

$$\theta_1 = \sigma_Y{}^2 - \sigma_C{}^2, \qquad\qquad\text{(Equation 3.1)}$$

By definition, it is also presented that:

$$\theta_2 = \sigma_Y{}^2 - \sigma_C{}^2 - \sigma_U{}^2, \qquad\qquad\text{(Equation 3.2)}$$

where θ_2 stands for the estimated teacher variance or "direct teacher effect."

The basic properties of θ_1 or θ_2, which can be inferred from the aforementioned equations, which are listed as follows:
1. θ_1 appears to be inversely proportional to the magnitude of the variance attributable to the control variables ($\sigma_C{}^2$).

2. Holding constant the variance attributable to the control variables, θ_1 is always greater than θ_2; except in the limiting case where the unexplained variance is zero, where they are equal.
3. If the magnitude of the unexplained variance (σ_U^2) is held constant and if the variance explained by a given set of control variables (σ_C^2) is taken into account, θ_1 and/or θ_2 appears directly proportional to the portion of student achievement variance (σ_Y^2) that is causally attributable to the teacher.

θ_1 or θ_2 is a catch-all indicator showing the combined proportional or direct effect of any and all observable and unobservable traits of the teacher that have something to do with the teaching-learning process. This is something that an input-oriented model or a value-added model like that of Sanders apparently cannot do.

Now, the fundamental question is asked: can we validly use the θ values to compare the teaching performance of a teacher versus any other teacher? The answer is a cautious yes, depending upon the magnitudes of the unexplained variance across the units of analysis (classes). If the magnitude fluctuates wildly across the units of analysis (across teachers or classes), then its validity is probably impaired.

Thus, the validity of the VPA and that of the calculated θ's appear to be dependent upon the realism of the assumption that the magnitude of the unexplained variance remains essentially homogeneous across the units of analysis. Now, is this assumption realistic? Apparently the answer is yes.

Let us examine the characteristics of a typical school setting. Teachers or classes are usually grouped into various departments along disciplinal lines. Departmental examinations are usually administered, thus to a certain extent homogenizing the magnitude of the unexplained variance. Teachers and/or students are subjected to the same administrator, same set of departmental curricula, the same administrative policies and procedures, and the same departmental learning resources and facilities; thus further homogenizing said magnitude. Therefore, within a disciplinal department, it is quite realistic to erect the assumption that the magnitude of the unexplained variance remains constant across teachers and/or classes.

Quite understandably, realism is diminished if the analysis is extended beyond the boundaries of a disciplinal department and/or across course categories. Two difficulties emerge. First, there is the difficulty posed by the different criterion scores arising from different course categories. Second, there is the problem posed by the perceived differential difficulties in the teaching and learning of the various course categories. In regard to the first, standardizing the scores appears to be the only curative procedure. In regard to the second, subjective weights reflective of the perceived or felt differential difficulties of teaching or learning the various courses can be factored into the algorithm. Perhaps they can be used to make adjustments to the calculated θ values. However, careful thought is needed, since doing so is likely to invite the risk of contaminating the inherent objectivity of the algorithm. At any rate, the standardization of the criterion scores is also probably useful in this regard.

The specification of the algorithm according to the specific conditions of a peculiar institutional setting, the calculation of the θ values, the guidance of some underpinning principles, and the caveats that must be borne in mind are explained in detail in a separate technical document. The interested reader—particularly the

school administrator who is desirous to use the VPA—may contact this writer at his E-mail address: ealicias@mla.nsclub.net.

References

Archer, J. (1999).   1999 Editorial Projects in Education, Vol. 18, Number 34, pp 26-28.

Dorn, S. (1994 and 1995). "Discussion of the Tennessee Value Added Assessment System." EDPOLYAN, Internet LISTSERVE, Arizona State University College of Education.

Glass, G. V (1995). Discussion of the Tennessee Value Added Assessment System. Retrieved May 1, 2005 from http://epaa.asu.edu/epaa/v13n30/tvaas.html.

Glass, G. V & Martinez, B. A. (1993). Politics of Teacher Evaluation. Retrieved May 1, 2005 from http://glass.ed.asu.edu/gene/papers/pte.pdf.

Hebbeler, K. M. (1985). "Approach of the Preschool Evaluation Project to Measuring the Effectiveness of Early Intervention." ERIC Accession No. ED259485.

King, F. J. and M. D. Reblyer (1984). "Alternative Designs for Evaluating Computer-Based Instruction." *Journal of Instructional Development*; v7 n3 p23-29. ERIC Accession No. –EJ318720.

Olson, L. (1998). "A Question of Value," *Education Week on the Web*; May 13, 1998.

Pascarella, Ernest (1986). "Are Value-Added Analyses Valuable? ERIC CD-ROM, 1985-1998.

Pennsylvania Department of Education and Lancaster-Lebanon Intermediate Unit 13. *Value Added Assessment System (VAAS)*.

Sanders, W. L. and Horn, S. P. (1995). "Educational Assessment Reassessed: The Usefulness of Standardized and Alternative Measures of Student Achievement as Indicators for the Assessment of Educational Outcomes." *Education Policy Analysis Archives*, Vol. 3, No. 6, March 3, 1995; ISSN 1068-2341.

Slack, J. B. and St. John, E. P. (1998). "A Model for Measuring Math Achievement Test Performance: A Longitudinal Analysis of Non-Transient Learners Engaged in a Restructuring Effort." ERIC CD-ROM, 1985-1998.

Willard, D. J. and Oplinger, D. (2003). "'Value-added analysis' credits school districts for progress a student makes," Columbus, Ohio: *Beacon Journal*; May 26, 2003.

About the Author

Eduardo R. Alicias, Jr.
Vice President for Academic Affairs
New England College
Quezon City, The Philippines

Professorial Lecturer (Graduate Program)
Bulacan State University, Malolos City &
Philippine State College of Aeronautics
Villamor Airbase, Pasay City

Email: ealicias@mla.nsclub.net

Doctor of Education, Educational Administration, 1981, University of the Philippines, Diliman, Quezon City, The Philippines. Diploma in Education (economics of education and education in developing countries), 1985-1986, University of London, UK, (British Council Fellow). Former Editor and Founder, *The SEARCH Journal*, ISSN 0117-6439, 1989 – 1998.

Education Policy Analysis Archives   **http://epaa.asu.edu**

Editor: Sherman Dorn, University of South Florida

Archivos Analíticos de Políticas Educativas