TEACHER EFFECTIVENESS AND COMPUTER ASSESSMENT OF READING

Relating Value Added and Learning Information System Data

W. L. Sanders
University of Tennessee
Value-Added Research and Assessment Center

&

K. J. Topping
University of Dundee, Scotland

Technical Report

University of Tennessee Value Added Research and Assessment Center
Knoxville, TN, USA

September 1999

TABLE OF CONTENTS

TEACHER EFFECTIVENESS AND COMPUTER ASSESSMENT OF READING

Relating Value Added and Learning Information System Data

1 INTRODUCTION

The Tennessee Value-Added Assessment System (TVAAS) has for several years used the largest longitudinally merged data base of student achievement data in the U.S.A. to generate estimates of school system, school, and teacher effects on indicators of student learning in a number of subjects, including reading comprehension. The Accelerated Reader (AR) (Advantage Learning Systems, 1993), a commercially available system for monitoring student reading practice, captures important data when students read a diversity of books during the school year. Students involved in the AR program may read a book from over 22,000 titles on the AR list, then take a multiple-choice comprehension test on the book at the computer, which scores the test and keeps records.

By merging the data from these two unique databases, the opportunity to examine the relationship between measures of student reading practice and improvement in reading comprehension as measured from TVAAS could provide special insights into factors that either accelerate or impede reading comprehension improvement. This is of special interest in Tennessee because, even though the statewide 8th grade means for math, science and language arts have been trending upward slowly since 1991, the overall reading comprehension scores have been slowly trending downward. Thus, this study examines two research questions: what are the relationships between independently obtained TVAAS value added measures of teacher effectiveness and AR measures of quality and quantity of reading practice, and what light does this shed on factors at the level of classroom management causative in improving reading comprehension?

2 THE TENNESSEE VALUE ADDED SYSTEM (OVERVIEW)

The Tennessee Value Added Assessment System (TVAAS) was developed to provide unbiased estimates of the influences that school systems, schools, and teachers have on the academic gains of students in a number of subjects, including reading comprehension (Sanders, Saxton, & Horn, 1997). To support this process, the largest longitudinally merged database of student achievement data in the U.S. has been constructed. Applying the statistical methodology to this database via software developed to handle the size and scope of this process, indicators of system, school and teacher effectiveness to the educational community in Tennessee have been provided for the past several years.

The state of Tennessee has 138 school systems. All students in grades 2 through 8 are tested annually, using the Tennessee Comprehensive Assessment

Program (TCAP). This includes norm-referenced and curriculum-referenced items and covers five areas: reading, language arts, math, science and social studies. It utilizes pattern (Item Response Theory) scoring, and was chosen by the state as providing a high degree of match with the state curriculum. Legislation mandated the deployment of TVAAS in 1992. The Act also requires the use of fresh, non-redundant, equivalent tests each year. Test items for each year are a new sample drawn from a large bank of equivalent items, in which a minimum of 70% must be items new in relation to the year before, making it difficult to "teach to the test."

TCAP was normed on a national (U.S.) sample of children in 1988, and Williams (1989) noted in a review of customized standardized tests that TCAP has proper statistical characteristics of reliability, adequate floors and ceilings, and articulation across test levels. Bock and Wolfe (1996) also found the psychometric properties of TCAP to be adequate. Additionally, the consistency of teacher effects from year to year suggests that the test-retest reliability of TCAP is satisfactory. In general, results for Tennessee are not significantly different from national averages. Obviously, test scores cannot reflect the totality of a student's learning experience or progress, but they can provide useful indicators. Equally, it can be asserted that no tests are culture free, and that TCAP scores are affected by the socio-economic background of students, but it has been demonstrated that the value added analysis of TCAP scores from TVAAS has eliminated these biases.

TVAAS inputs student scale scores derived from the norm-referenced component of the TCAP into a statistical process of multivariate, longitudinal analysis. TVAAS is based upon statistical mixed-model methodology (McLean, Sanders, & Stroup, 1991; Sanders, Saxton, & Horn, 1997), which seeks to provide the best linear unbiased estimates and predictors of the influence of school systems, schools and teachers upon student learning rates. TVAAS enables a massive multivariate, longitudinal analysis of student data over time. By so doing, the estimates of school, system and teacher effects have been shown to be virtually uncorrelated with socio-economic factors and measures of prior levels of student achievement. These results have mitigated the concern that has been expressed that dozens of exogenous variables must be included in the analyses to insure fair and equitable assessment.

The TVAAS model does not use gains as the dependent variable, but fits the entire observational vector. The system, or school, or teacher effects are simultaneously estimated considering the variance-covariance structure of the data. Using an appropriate set of estimable functions, these effects are back mapped into mean gains for reporting purposes. Thus, gains from different baselines are not directly compared, and the starting point of the students is irrelevant so long as the testing regime provides sufficient scale elongation to allow measurement of progress for the lowest and highest achieving students.

Unlike more traditional statistical methods, TVAAS methodology enables a multivariate, longitudinal analysis regardless of the degree of missing information for

each student. Students miss tests, change schools, move districts and go to other states, leaving fractured records. Nevertheless, inbuilt mixed model equations (Sanders, Saxton, & Horn, 1997) allow the unbiased use of incomplete longitudinal data. This is a major advantage over traditional statistical approaches which require that all students have complete information over all years. With such approaches, the analyst is caught in the dilemma of using fewer years of data in order to have more student records, resulting in a loss in sensitivity of the estimates, or more years of data with fewer students with the severe risk of obtaining severely biased estimates. The TVAAS methodology eliminates this problem.

The total number of equations used for all the school systems in the State is very large indeed, running into tens of thousands, each re-solved iteratively for every new assessment cycle. The general form of the underpinning mixed-model equations are given in mathematical form in Sanders and Horn (1994, 1995a, 1995b) and Sanders, Saxton and Horn (1997). The possibility of a teacher getting an "undesirable" estimate due to some quirks in the data, (e.g. too few observations or an unlucky collection of negative "errors of measurement") has been virtually eliminated by the TVAAS process, which incorporates shrinkage estimation, the layered model at the teacher level, and the use of all available information for each student. The robustness of the system has been demonstrated using computer simulations to evaluate worst-case scenarios.

For a more thorough discussion of the technical aspects of TVAAS, see Sanders, Saxton, & Horn (1997).

3 MEASURING READING PRACTICE WITH ACCELERATED READER

Recent developments in computer assisted assessment of reading comprehension seek to enable more frequent and detailed assessment in less time with greater consistency, and empower the teacher and student by yielding more detailed formative feedback to these key players. The Accelerated Reader (AR) is a type of free-standing, computer assisted, curriculum based assessment, voluntarily self-administered by students, and specifically intended to have strong formative effects on subsequent learning. It is designed to enable curriculum based assessment of reading comprehension of real ("trade") books by students, and analyze and summarize the results. It aims to help teachers promote and manage effective reading practice.

Students using the program read a book from over 22,000 titles on the AR list, then take a multiple-choice comprehension test on the book at the computer, which scores the test and keeps records. Each book has a maximum point value according to its length and difficulty. When the student self-tests, the computer awards points up to this maximum, according to the number of correct test responses. Students select their own preferred books and read at their own pace. Teachers may choose to allow students to test on books read to and with them, as

well as those read independently and silently, especially in the case of early or delayed readers. The default condition is that students can test on a book only once.

The quantitative feedback from the points system is intended to raise meta-cognitive awareness and motivate students to read more, longer and harder books. The computer also provides the teacher with an automatically updated analysis of scores for individuals or whole classes, which indicates average percent correct, difficulty of books read, points earned, and other diagnostic information. Computer generated "At Risk" reports enable the teacher to guide each student's reading practice. The Accelerated Reader system assigns a point value to each book based on the number of words in the book and its reading difficulty, using a formula based on the well-known Flesch-Kincaid readability index (Flesch, 1968, 1974; Chall & Dale, 1995). The formula considers the number of syllables in words and sentence complexity. Thus:

$$\text{AR points} = (10 + \text{reading level}) \times \frac{\text{words in book}}{100{,}000}$$

Once the student has chosen and read a book, he or she goes to the computer and takes a multiple-choice comprehension test on the content of the book. The computer scores the test, awards the student points based on test results, and keeps a complete record of results. For a book valued at 10 AR points, like Black Beauty, a student would receive 10 points for a score of 100 percent, nine points for 90 percent, and so on. However, the student must score at least 60 percent on the test to earn any points.

AR tests are created by specialists who read the book and draft an item bank of comprehension questions relating to story grammar and text structure. Tests may have 5, 10 or 20 items, depending on the length and difficulty of the book. Correct and incorrect (distracter) answers are carefully constructed. Then during two consecutive editing phases, adult reader-evaluators use the draft version on computer, considering consistency of vocabulary, plot detail, chronological order of events, the extent to which questions might cue certain answers, the balance between the multiple choices presented, and other issues. This leads to the discarding or revision of the less discriminatory and any ambiguous or unreliable items. Each question is re-checked against the book at each stage and checks made to ensure a child could not pass the test without reading the book. The tests are then separately reviewed by quality assurance professionals and the senior editor. Thus the psychometric properties of the AR comprehension tests are likely to be considerably more stable than those made up on the wing by creative teachers.

As the student tests on more and more books, the AR system enables close monitoring of their general reading performance level and the level of book currently chosen. It is recommended by the software designers that co-ordinating teachers target a test success rate of 85% correct as optimal for students, with either

independent or supported reading. (Percent correct is discussed at length later in this study.) Automatically generated "at-risk" reports alert the teacher to students who are not reading effectively, within their zone of proximal development. Clearly, the AR tests are brief, primarily assess literal comprehension rather than idiosyncratic reader inferences or other more complex responses (which might be culturally specific), and do not pretend to measure all relevant reading behaviors in school or elsewhere. Indeed, AR questions are deliberately restricted to those which demonstrate adequate reliability.

4 METHODOLOGY

4.1 Sampling Framework

The sampling framework for the TCAP data entered into TVAAS has been described above. All students in grades 2-8 take the tests in five subject areas near the end of each academic year, and TVAAS compensates for missing values. The current study used TVAAS data on value added at the individual teacher level in the school year 1996-7.

The sampling framework for the AR data was more complex. It was known from purchase records that about half of the schools in the state had acquired the AR software. On telephone enquiry, some schools indicated that although they had purchased AR, they had not yet implemented it, and these were excluded from all analyses. Accordingly, 379 schools were mailed a circular request to supply student level AR data, with the incentive of entry in a prize draw for two pairs of airline tickets to anywhere in the U.S. Of these, 313 schools (82.6%) responded, usually submitting data electronically on disc. This proportion was considered high enough to assure a fair degree of representativeness. Data on approximately 80,000 students was thus obtained.

In addition to the book tests provided within AR, it is possible for teachers to devise and insert their own tests for books of particular local interest. However, as the psychometric properties of such “home made” tests were completely unknown, results on them were excluded from the analysis.

Some books on which students had tested were marked as texts which had been read **to** the student, rather than read independently by the student. These were also excluded from subsequent analyses.

4.2 Data Merging

Given the different time scales of the data, the dates of origin of the AR data were matched with those of the TVAAS data. The number of students for each grade for whom AR data and TVAAS data were successfully merged is given in Table 1.

The total number of successfully merged individual student records was 62,739. *Inter alia*, this table suggests that beyond the third grade, students increasingly tend to read books below their actual grade level (also see Table 6). The number of books read declines through the grades, while the mean point value per book continues to rise quite steeply, suggesting that students have a stronger tendency to read longer books than to read more difficult books.

Table 1: Number of Students For Whom Data Successfully Merged

Grade	Students (n)	Mean number books read	Mean book reading level	Mean book Point value
Second	8573	40.92	2.47	0.52
Third	10359	34.88	3.04	0.96
Fourth	9773	25.21	3.70	1.94
Fifth	10761	22.67	4.19	2.83
Sixth	8906	16.03	4.84	4.14
Seventh	7356	12.66	5.50	5.55
Eighth	7011	11.71	5.72	6.29

4.3 Variables

As there is considerable potential for confusion from similar variable labels, note that in what follows,

Grade Level = Student's actual Grade Level in school

Grade Equivalent = Grade Equivalent level of student's tested reading score

Book Level = Assessed Readability Level of the book

Also note that some of the analyses were conducted at the level of the individual student, some at the level of the class teacher.

TVAAS Variables

TCAP Scale Scores

TCAP Gains-obtained by subtracting previous scale score from current scale score

Gender

Ethnicity

School System

School

Teacher

Teacher Effectiveness - TVAAS current teacher effect per individual teacher.

Student Reading Attainment - obtained by first averaging current and previous TVAAS reading scale score for each individual student, then converting into grade equivalent as indicated by the TCAP manual.

Class Attainment on average (RSSAVG) - obtained by first averaging current and previous TVAAS scale scores on all 5 elements of TCAP for each individual student, then obtaining the class average. Class Attainment (RSSAVG) was highly correlated with mean student reading attainment.

AR Variables

Books Read - the mean number of books read individually and tested on within the AR system by students during the study period (NOBOOKSI).

Mean Book Level of Books Read - the mean readability level of books read individually and tested on within the AR system by students within the study period (READLEVI).

Mean AR Point Value of Books Read - (PNT_VAL).

The mean number of books read (NOBOOKSI) and mean point value of books read (PNT-VAL) were supplanted in later analyses by a combined Reading Volume variable.

Reading Volume - the amount of engagement in reading by an individual student (RLWORDSI) - obtained by converting the mean number of Books Read by each individual student (NOBOOKSI) and the mean Point Value of the books read (PNT_VAL) (not the points actually earned on test) into an estimate of the total number of words read. (The manufacturers of the software supply coefficients to convert Point Value to words where book difficulty level is known). This is expressed on a log scale due to the severe degree of skewness in this variable.

Percent Correct - successful comprehension as measured by average percentage correct on AR tests on books read individually. Initially entered as raw data (PCTCOR). Later obtained by converting the average percent correct score on AR tests on books read individually by each student into a logit scale (RPCTCORI).

Composite TVAAS/AR Variables

Challenge on average in books read (RTPCHAL) - obtained by subtracting each student's grade equivalent for reading on TVAAS from the average book level of the books read by the student on the AR system (so low negative scores indicate low challenge and high positive scores indicate high challenge). The class means were grouped into top, middle and bottom thirds (levels 0, 1, 2).

Other VARIABLES

Reading Renaissance (RR) Training for teachers in implementation of AR. Some teachers had attended a one day seminar, a smaller number a 3-day seminar. The latter were too small a group to analyze separately, so the two groups were aggregated. Teachers trained very recently were excluded from the analysis on the grounds that the training had not had sufficient time to influence their practice and be reflected in their AR data.

Model Classroom status award, indicating specified implementation quality criteria met. This award was based on submitted AR data from their class, with some teacher self-report on a structured checklist. The qualification criteria were that according to the computer generated At Risk Report, less than 10% of students were at risk, average percent correct was between 85% and 92%, median points earned were at least 80% of expectations on goal-setting charts, and Reading Renaissance principles were in place (particularly with respect to amount of reading time, intervention in response to At Risk Reports, and other elements of implementation integrity).

5 RESULTS

5.1 Initial Analyses

Initial analyses explored the relationship between value added residual means and mean number of Books Read (NOBOOKSI), mean Book Level of books read (READLEVI), and Percent Correct (PCTCOR). The residuals were obtained from a very simple initial model:

> Predicted current TCAP reading score = linear function of (each student's previous year's TCAP reading, math and language arts scores).

Observe that these residuals were at the individual student level and no consideration was given as to the classroom teacher.

5.2 Mean Number of Books Read

The following set of tables presents simple means in only one dimension with no attempt made to consider the correlation that exists among the various book variables.

Blank cells contained too few subjects to be worth recording. At third grade, a consistent trend of rising value added with increased number of books read is evident. This is also true of fourth grade, except it takes fewer books to make a positive difference. Fifth grade shows some similarity, except for the additional striking negative impact of very large numbers of books. This pattern is sustained in the sixth grade, except the negative effects are weaker and the positive effects stronger.

Table 2: Mean Number of Books Read and Value Added Residual Means

Grade:	Third		Fourth		Fifth		Sixth		Seventh	
Books	n	mean	n	mean	n	mean	n	mean	n	mean
2	1791	-4.34	2040	-1.91	2091	-2.42	2467	-0.71	1861	-1.28
4	842	-1.57	950	1.06	938	-1.59	1042	-1.08	1109	0.93
8	1191	-1.17	1334	0.47	1466	-0.32	1156	-0.02	1140	0.04
16	1560	1.93	1551	2.29	1957	0.77	1455	1.33	1306	0.05
32	1641	0.89	1464	1.27	1930	2.14	1365	2.38	981	0.11
64	1389	3.09	1048	2.47	1002	1.46	494	1.94	249	0.73
128	742	3.90	432	0.85	304	-1.58	91	-1.44		
256	186	8.55	50	6.78	21	-6.57				

Perhaps reading a large number of very easy books has positive effects in grade 3 but negative effects if continued into grade 5 and beyond. Seventh grade gives scattered results, which are however approximately in line with the third grade results. There is an indication of positive value added commencing when books read get into double figures, especially in grades 3, 5 and 6. The number of books read declines through the grades, very probably reflecting a tendency in older students to read fewer but longer and harder books.

5.3 Mean Book Level of Books Read

Results are summarized in Table 3. At third grade, added value generally rises steadily with increased difficulty of books read. However, reading very easy books and very hard books appears to reduce value (although n is very small for this latter). Without knowledge of the inter-relationship between reading ability and difficulty of books read with the individual student as the unit of analysis, as with all of these one dimensional tables, interpretation should be cautious.

In fourth, fifth and sixth grade the general pattern of added value rising with book difficulty is more consistent, with book level more than 0.5 to 1.0 grades below the actual grade of the students showing negative effects. Seventh grade shows a similar pattern, although slightly more erratic. The largest value added tends to be where the mean book level is above the student's actual grade.

Table 3: Mean Book Reading Level and Value Added Residual Means

Grade:	Third		Fourth		Fifth		Sixth		Seventh	
Level	n	mean	n	mean	n	mean	n	mean	n	mean
1.5	120	-9.23								
2.0	685	-6.81	177	-3.01	149	-9.52				
2.5	2063	1.09	918	-3.53	486	-5.65				
3.0	2711	1.61	1445	-1.95	921	-4.14	620	-4.65	239	-7.10
3.5	1706	1.79	1713	0.29	1406	-1.35	653	-1.50	247	-4.74
4.0	626	3.13	1467	3.32	1525	-0.08	807	-0.71	388	-5.06
4.5	230	5.38	1007	4.16	1474	1.73	1005	-0.35	647	-2.50
5.0	99	4.73	577	5.29	1219	3.38	1170	1.25	805	-0.78
5.5	53	-3.40	273	5.62	754	4.68	1062	2.29	957	1.39
6.0	32	-0.67	205	3.69	349	4.58	792	3.35	1000	0.00
6.5					162	5.30	490	2.79	773	3.55
7.0					104	1.25	156	6.50	476	2.65
7.5							83	6.41	264	3.96
8.0							66	5.16	104	2.23
8.5									37	1.17
9.0									34	17.73

5.4 Mean Percent Correct

Results are summarized in table 4. The positive relationship between percent correct and value added is consistent across all grades. The recommendation given in the implementation training for AR customers is to sustain average percent correct above 85%. The data in Table 4 strongly support this recommendation in all grades, as this is clearly the point at which positive value added emerges. Alarmingly, however, more than half of the students in the study were operating below this level suggesting that many teachers were not using the data to monitor student performance as recommended. However, more than a third were operating above 90% correct in all grades.

Table 4: Mean Percent Correct and Value Added Residual Means

	Grade 3		Grade 4		Grade 5		Grade 6		Grade 7	
% correct	n	mean	n	mean	n	mean	n	mean	n	mean
45	97	-20.27	90	- 9.48	129	- 6.16	170	- 4.54	199	- 7.51
50	93	-15.09	109	- 7.19	121	-10.49	161	-3.54	170	- 4.16
55	82	-16.21	122	-15.69	146	- 5.38	163	-3.75	183	- 3.80
60	232	- 8.52	242	- 5.82	321	- 6.68	353	- 2.42	345	- 2.10
65	313	- 9.86	381	- 3.65	403	- 3.83	351	- 1.67	377	- 2.34
70	565	- 6.34	577	- 1.39	675	- 3.69	593	- 0.63	564	- 1.06
75	774	- 4.86	728	- 1.78	865	- 1.25	706	-1.83	592	- 2.05
80	1205	0.43	1106	0.70	1251	- 0.92	939	- 0.76	788	- 0.41
85	1552	1.25	1421	1.16	1528	0.94	1025	- 0.02	821	0.28
90	1947	5.58	1630	4.02	1664	2.77	1273	3.00	1005	3.09
95	1550	6.97	1435	5.59	1542	4.99	1231	5.37	975	3.48

5.5 Influence of Overall Student Ability

This relationship was then examined by ability of student, dividing students into six groups of ascending prior achievement on TCAP reading, language arts and math scores combined. In all grades (3-7), the lowest ability group had the lowest percentage correct (72-74) and the highest ability group the highest (89-92). However, the numbers of students at these extremes were relatively small, and the modal percent correct was just above 80% (range 77-83).

As would be expected, the mean book reading level of books read was strongly related to student ability. However, although students in grades 3 and 4 on average read books around their grade level, beyond this average book level increasingly fell behind grade level, as also indicated in Table 1.

The numbers of books read declined steadily through the grades (doubtless at least partially reflecting older students' reading fewer, longer and harder books), as also indicated in Table 2. However, within grades across ability groupings no consistent pattern was evident: grade 3 showed a tendency for the number of books read to rise with ability, while in grades 6 and 7 the number of books read fell with higher ability.

Considering value added residuals across ability groupings, the highest residuals were consistently found for students within the middle ability groups (which of course included the largest number of students).

Curiously, mean residuals were not consistent across grades, grades 4 and 6 having particularly high mean residuals (0.73 in both), while grade 5 had the lowest. This might reflect the structure of schools in this state, where sixth grade is often the initial grade of middle school. Perhaps in grade five there is some tendency for students to be focused more on traditional curriculum subjects before transfer, creating something of a plateau in literature based reading.

As was mentioned earlier, since all of the book variables are correlated with each other, the previous tables should be interpreted carefully and are included more for description than for rigorous interpretation. The next set of analyses include all of these variables simultaneously in an attempt to parcel the various relationships.

5.6 Relationship Between Simple Value-Added Residuals And Book Variables

The residuals were taken as the dependent variable and related to the various book variables via a traditional multiple regression approach. Tests of fixed effects within the analysis of variance were conducted on Type I sequential sum of squares and Type III partial sum of squares. Hypotheses tested from Type I SSS are model order dependent. Thus, if the first term in the model is 'significant', what is really being tested is this factor and its partial confounding with all other factors

which could be related to the dependent variable. Hypotheses tested from Type III PSS are the partial effects remaining on the dependent variable after accounting for the partial confounding with all the other factors in the model.

Table 5 shows that the largest number of significant effects are evident in the Third Grade. Mean Book Level of books read shows a significant effect in all grades except the sixth. Percent Correct shows a significant effect in grades 3, 4 and 7. The number of Books Read shows a significant effect only in grades 3 and 6. The interaction of Books Read and Percent Correct shows significant effects in the third, fourth and sixth grades.

Table 5: Tests of Hypotheses of the Partial Relationship of Book Variables on Value Added Residuals: Type III F & (Probability)

	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7
Books Read NOBOOKSI	**15.77** (0.0011)	3.69 (0.0547)	0.13 (0.7162)	**5.81** (0.0160)	0.30 (0.5833)
Book Level READLEVI	**20.20** (0.0011)	**4.51** (0.0337)	**14.43** (0.0011)	0.05 (0.8265)	**3.92** (0.0477)
Percent Correct PCTCORI	**19.46** (0.0011)	**5.98** (0.0145)	1.29 (0.2564)	1.32 (0.2511)	**4.48** (0.0343)
Books Read x % Correct NOBOOKSI x PCTCORI	**19.63** (0.0011)	**5.25** (0.0220)	0.69 (0.4067)	**8.49** (0.0036)	0.55 (0.4570)
Book Level x % Correct READLEVI x PCTCORI	**4.11** (0.0427)	0.05 (0.8168)	0.16 (0.6924)	2.86 (0.0908)	0.51 (0.4751)
DDF (NDF=1)	8317	7774	8541	6896	5963

(**emboldened** where $p<0.05$)

5.7 Influence of Student Reading Ability

Results are summarized in Table 6. In third grade, the average student is reading at or above their grade level on average. There is a long and fairly flat upward tail to the distribution, so that even the most able readers are not reading above grade 3.8 level on average. In fourth grade, the average student is reading below their grade level on average, even when their tested reading ability is above grade level. This tendency is even more evident in grade 5, where only students with eleventh grade tested reading ability are reading grade 5 books on average. Grades 6 and 7 show this tendency increasing further, with no students at any level of tested reading ability reading books at grade level on average. In both these grades, students of tested reading ability equal to their actual grade level appear to be reading on average books of a readability level one to one and a half grade levels below their ability.

Table 6 also indicates that the trend for average Book Level to regress as higher actual grade levels are attained, which was noted in Table 1, appears to apply to students of all tested reading ability grade equivalents after third grade. The analysis was differentiated by gender, but there were no significant differences between males and females.

Books read were flagged as fiction or non-fiction, enabling an analysis of the mean book level of fiction versus non-fiction books, differentiated by student grade level, grade equivalent, gender and ethnicity. In fact, there was little evidence of any significant differences in any respect, the modal difference between fiction and non-fiction book levels being 0.1 grade. In the few cases where there was a difference of any size, more girls than boys figured, and fiction was more likely to be at a higher level, but there was no gender interaction in this respect.

Table 6: Student Grade Equivalent in Reading & Book Level (n)

Student Grade Equivalent	Mean Book Level				
	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7
2	2.6 (581)	2.9 (186)	2.9 (197)	3.4 (41)	3.4 (14)
2.5	2.8 (1248)	3.1 (451)	3.2 (306)	3.4 (135)	3.7 (73)
3	2.9 (1012)	3.2 (472)	3.3 (285)	3.8 (130)	4.0 (70)
3.5	3.0 (1003)	3.3 (645)	3.7 (422)	3.8 (179)	4.4 (109)
4	3.0 (863)	3.4 (651)	3.7 (533)	4.1 (255)	4.3 (120)
4.5	3.1 (692)	3.6 (781)	3.9 (581)	4.2 (365)	4.7 (175)
5	3.2 (749)	3.7 (963)	4.0 (914)	4.3 (512)	4.6 (274)
5.5	3.3 (491)	3.9 (942)	4.1 (909)	4.6 (576)	4.9 (341)
6	3.4 (420)	4.0 (745)	4.3 (984)	4.7 (686)	5.2 (458)
6.5	3.4 (286)	4.1 (532)	4.4 (793)	4.9 (725)	5.3 (530)
7	3.5 (148)	4.2 (261)	4.5 (448)	4.9 (440)	5.5 (326)
7.5	3.8 (113)	4.1 (163)	4.6 (314)	5.2 (342)	5.4 (284)
8	3.6 (75)	4.3 (146)	4.6 (282)	5.2 (345)	5.6 (322)
8.5	3.6 (75)	4.2 (169)	4. 7(289)	5.2 (423)	5.8 (457)
9	3.5 (92)	4.3 (153)	4.8 (301)	5.2 (433)	5.8 (457)
9.5	3.8 (75)	4.4 (185)	4.8 (288)	5.3 (369)	5.9 (468)
10	3.6 (33)	4.5 (60)	4.9 (130)	5.6 (166)	6.0 (214)
10.5	3.6 (21)	4.5 (27)	4.9 (80)	5.5 (106)	6.0 (150)
11	3.8 (34)	4.5 (69)	5.0 (156)	5.6 (191)	6.2 (269)
Typical (modal) S.D.	0.7	0.8	0.9	1.0	1.1

5.9 Relationship between TVAAS Teacher Effects and Book Variables

For the next stage of analysis, more sophisticated metrics and analyses were employed. The previous analyses were at the student level and the "value added" was obtained from the residuals of a simple prediction model. In this stage, the formal TVAAS teacher effects (obtained from the multivariate, longitudinal analysis as referenced previously) were used as the dependent variable and class means of the book variables were used as the predictors. Each of the predictor variables was grouped into thirds then included in the model as classification variables.

The specific model employed was: TVAAS Teacher Effects is a linear function of the classification variables representing Reading Volume, Percent Correct, Challenge and their respective interactions.

The results of the analysis of variance are presented in Table 7 (see next page). All tests of hypotheses were based upon partial sum of squares (SAS Type III). Note that in these tables of analyses at the class teacher level, N refers to number of teachers. From Table 7, it appears that Reading Volume (RLWORDSI), Log Percent Correct (RPCTCORI) and Challenge (RTPCHAL) are significantly related to teacher effectiveness in almost all cases in grades 3 to 6. Additionally, there is some evidence of significant interaction between Reading Volume (RLWORDSI) and Log Percent Correct (RPCTCORI).

Tests of hypotheses for the interaction combinations Reading Volume (RLWORDSI) x Challenge (RTPCHAL), Log Percent Correct (RPCTCORI) x Challenge (RTPCHAL), and Reading Volume (RLWORDSI) x Log Percent Correct (RPCTCORI) x Challenge (RTPCHAL) were also conducted, but did not reach statistical significance in any grade (with one exception), and so are not reported here. The exception was Log Percent Correct (RPCTCORI) x Challenge (RTPCHAL) in sixth grade, for which SS = 57.93, F= 4.60, p=0.0013.

Table 7: Tests of Hypotheses of Impact of Book Variables on TVAAS Teacher Effects: Type III SS, {F} and (Probability)

	Grade 3	Grade 4	Grade 5	Grade 6
Reading Volume RLWORDSI	92.27 **{4.93}** (0.0075)	92.76 **{10.55}** (0.0001)	108.14 **{9.91}** (0.0001)	33.30 **{5.29}** (0.0056)
Log Percent Correct RPCTCORI	90.29 **{4.82}** (0.0083)	21.11 {2.40} (0.0915)	69.47 **{6.37}** (0.0019)	29.50 **{4.68}** (0.0100)
Challenge RTPCHAL	215.92 **{11.54}** (0.0001)	108.84 **{12.38}** (0.0001)	45.96 **{4.21}** (0.0154)	25.71 **{4.08}** (0.0180)
Reading Volume x % Correct RLWORDSI x RPCTCORI	128.78 **{3.44}** (0.0086)	23.66 {1.35} (0.2516)	25.71 {1.18} (0.3196)	50.03 **{3.97}** (0.0038)

(**emboldened** where $p<0.05$)

5.10 Effects of Reading Volume, Challenge and Log Percent Correct on TVAAS Teacher Effects

Table 8 shows that Reading Volume (RLWORDSI) and Percent Correct (RPCTCORI) are positively related to teacher effectiveness in grades 3 through 6, very consistently in grades 3 and 4. Figure 1 (see following page) demonstrates this graphically.

However, Challenge (RTPCHAL) is consistently **negatively** related to teacher effectiveness. In other words, teachers whose students read at high volume and obtain high AR test scores (indicating successful comprehension) add most value, but teachers who have students reading books which provide **low** challenge also tend to add value. At first sight this last finding is somewhat surprising, and seems to run counter to the previous literature and to previous findings in this study based on analyses at the student rather than teacher level and on other variables (see Tables 1, 3 & 6).

Table 8: Teacher Effectiveness (L. S. Means) for Levels of Reading Volume, Percent Correct & Challenge

	Third grade			Fourth grade			Fifth grade			Sixth grade		
	Reading Volume RL WORDSI	**Log % Correct** RPCT CORI	**Chall-enge** RTP CHAL	**Reading Volume** RL WORDS	**Log % Correct** RPCT CORI	**Chall-enge** RTP CHAL	**Reading Volume** RL WORDS	**Log % Correct** RPCT CORI	**Chall-enge** RTP CHAL	**Reading Volume** RL WORDS	**Log % Correct** RPCT CORI	**Chall-enge** RTP CHAL
Low	-0.30	-0.23	**0.94**	**-0.41**	-0.14	**0.81**	**-0.57**	**-0.42**	**0.47**	**-0.67**	-0.18	0.42
Medium	0.05	-0.04	0.18	0.12	0.20	0.17	**0.59**	**0.43**	0.37	-0.06	**-0.52**	-0.19
High	**0.72**	**0.75**	**-0.65**	**0.78**	**0.42**	**-0.49**	**0.50**	**0.52**	-0.31	0.44	0.41	**-0.52**
N	699			593			524			322		

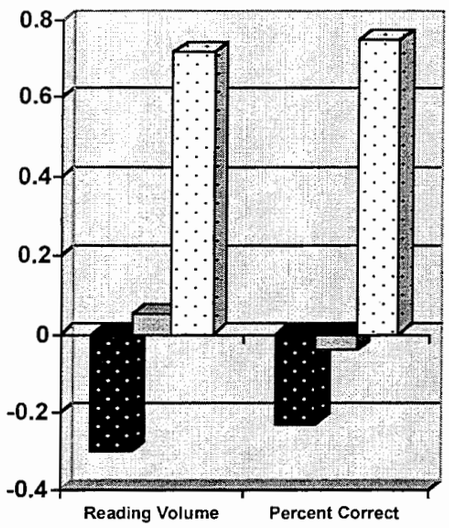
(**emboldened** where $p<0.05$ for difference from zero)

(N.B. An effect of zero implies that a teacher's estimate was equal to the district average)

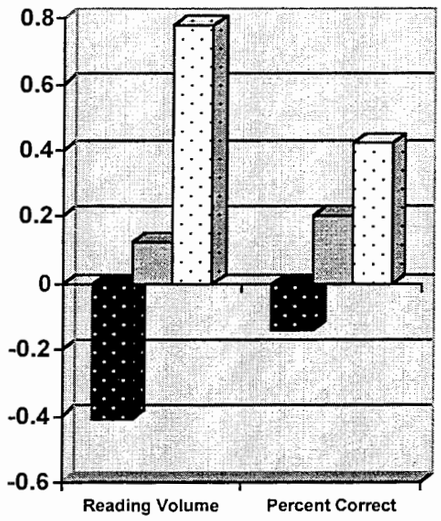
Figure 1: Teacher Effectiveness (Least Squares Means), by Category of Reading Volume and Percent Correct

■ Low ▨ Medium □ High

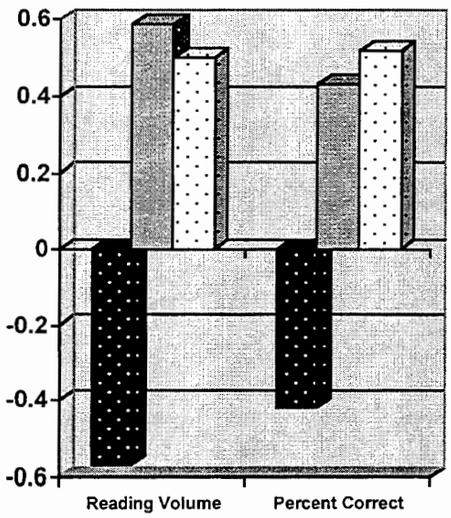
THIRD GRADE



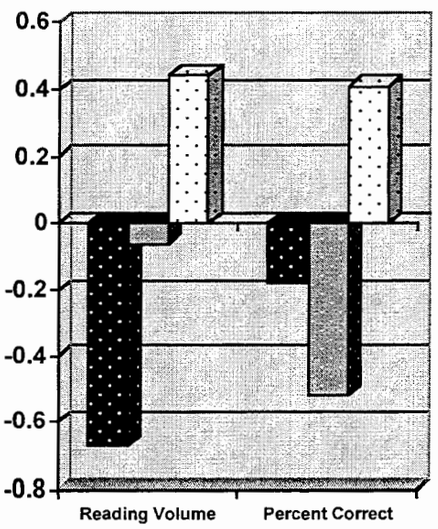
FOURTH GRADE



FIFTH GRADE



SIXTH GRADE



It might be assumed that the finding is mediated by interactions between other variables, but this was not evident from the Type III SS effects (other than for the single exceptional case Log Percent Correct (RPCTCORI) x Challenge (RTPCHAL) in sixth grade).

5.11 Interactions between Reading Volume, Challenge & Log Percent Correct

The interactions between variables were then explored in relation to the Low, Medium and High categories. As the ensuing tables were very complex, a summary is given below.

In third grade, where Reading Volume (RLWORDSI) was Low, increasing Log Percent Correct (RPCTCORI) from Low to High increased teacher effects from minus 0.55 to 0.04. When Reading Volume (RLWORDSI) was Medium, increasing Log Percent Correct (RPCTCORI) increased teacher effects from minus 0.84 to 1.38. When Reading Volume (RLWORDSI) was High, increasing Log Percent Correct (RPCTCORI) increased teacher effects from 0.69 to 1.24.

In fourth grade, where Reading Volume (RLWORDSI) was Low, increasing Log Percent Correct (RPCTCORI) from Low to High increased teacher effects from minus 0.57 to +0.11. When Reading Volume (RLWORDSI) was Medium, increasing Log Percent Correct (RPCTCORI) increased teacher effects from minus 0.28 to .39. When Reading Volume (RLWORDSI) was High, increasing Log Percent Correct (RPCTCORI) increased teacher effects from 0.44 to 0.76.

In fifth grade, where Reading Volume (RLWORDSI) was Low, increasing Log Percent Correct (RPCTCORI) from Low to High increased teacher effects from minus 1.37 to minus 0.39. When Reading Volume (RLWORDSI) was Medium, increasing Log Percent Correct (RPCTCORI) increased teacher effects from 0.39 to 0.78. When Reading Volume (RLWORDSI) was High, increasing Log Percent Correct (RPCTCORI) increased teacher effects from minus 0.30 to 1.17.

In sixth grade, where Reading Volume (RLWORDSI) was Low, increasing Log Percent Correct (RPCTCORI) from Low to High decreased teacher effects from minus 0.07 to 0.31. When Reading Volume (RLWORDSI) was Medium, increasing Log Percent Correct (RPCTCORI) increased teacher effects from minus 0.14 to 0.83. When Reading Volume (RLWORDSI) was High, increasing Log Percent Correct (RPCTCORI) increased teacher effects from minus 0.33 to 0.73.

It seems clear that increasing percent correct is associated with increased teacher effectiveness **at all levels of volume of reading**.

In third grade, where Log Percent Correct (RPCTCORI) was Low, increasing Challenge (RTPCHAL) from Low to High decreased teacher effects from 0.80 to minus 1.53. Where Log Percent Correct (RPCTCORI) was Medium, increasing

Challenge (RTPCHAL) from Low to High decreased teacher effects from 0.61 to minus 0.69. Where Log Percent Correct (RPCTCORI) was High, increasing Challenge (RTPCHAL) from Low to High decreased teacher effects from 1.41 to 0.26.

In fourth grade, where Log Percent Correct (RPCTCORI) was Low, increasing Challenge (RTPCHAL) from Low to High decreased teacher effects from 0.77 to minus 0.75. Where Log Percent Correct (RPCTCORI) was Medium, increasing Challenge (RTPCHAL) from Low to High decreased teacher effects from 1.01 to minus 0.65. Where Log Percent Correct (RPCTCORI) was High, increasing Challenge (RTPCHAL) from Low to High decreased teacher effects from 0.64 to minus 0.07.

In fifth grade, where Log Percent Correct (RPCTCORI) was Low, increasing Challenge (RTPCHAL) from Low to High decreased teacher effects from minus 0.03 to minus 0.90. Where Log Percent Correct (RPCTCORI) was Medium, increasing Challenge (RTPCHAL) from Low to High decreased teacher effects from 0.77 to minus 0.29. Where Log Percent Correct (RPCTCORI) was High, increasing Challenge (RTPCHAL) from Low to High decreased teacher effects from 0.66 to 0.26.

In sixth grade, where Log Percent Correct (RPCTCORI) was Low, increasing Challenge (RTPCHAL) from Low to High decreased teacher effects from minus 0.15 to minus 0.17. Where Log Percent Correct (RPCTCORI) was Medium, increasing Challenge (RTPCHAL) from Low to High decreased teacher effects from 0.72 to minus 2.10. Where Log Percent Correct (RPCTCORI) was High, increasing Challenge (RTPCHAL) from Low to High decreased teacher effects from 0.69 to 0.36.

In this analysis, it seems clear that increasing reading challenge is associated with **decreased** teacher effectiveness **at all levels of percent correct**. Even when Log Percent Correct (RPCTCORI) was in the High category (conforming to the training recommendation for minimum percent correct), the negative effect of increasing challenge was still evident (although for third, fifth and sixth grade students this meant the difference between a large positive teacher effect and a smaller positive teacher effect).

Comparable findings were found for all levels of Reading Volume. Even when Reading Volume (RLWORDSI) was in the High category (conforming to the training recommendation), the negative effect of increasing challenge was still evident (although for third, fifth and sixth grade students this meant the difference between a large positive teacher effect and a smaller positive teacher effect). Additionally, analyses considering the interaction Log Percent Correct (RPCTCORI) x Reading Volume (RLWORDSI) X Challenge (RTPCHAL) show no consistent influences on the above.

5.12 Impact of Training with AR

Advantage Learning, the developers of AR, offers teachers both one and three day training sessions on their recommended use of the AR product. Records were supplied of teachers who had attended these sessions. However, the numbers of teachers who had attended the three day session were rather small and were excluded from the subsequent analysis. To ensure that sufficient time was allowed for the influence of the one day seminar to be realized on student gain, only those teachers who attended prior to October 1, 1996, were included, given that the teacher effect data was for the Spring 1997 testing period. If any district sent at least one teacher to the seminar, all the other teachers in the district were included in the non-attending ("control") group, irrespective of whether they used AR or not.

The analysis showed trained teachers had significantly higher gains over 1995-7 than control teachers in grades 3 ($t=5.06$, $p<0.05$) and 5 ($t=3.33$, $p<0.05$), definitely not in grade 4 ($t=0.56$) (where student gains in both groups were negative) and not in 6 ($t=0.62$) (although this difference was in the right direction). Numbers in grades 7 and 8 were too small to permit useful analysis.

5.13 Influence of Model Classroom Status

The TVAAS teacher level gain in reading was related to whether their classroom had never been certified as a model classroom, had been certified once, or certified and subsequently re-certified (see Table 9).

Numbers of model classrooms were very small throughout. In fourth and fifth grade, model classrooms showed higher effectiveness than non-model classrooms, and re-certified model classrooms higher effectiveness than those certified once only, but this trend reached statistical significance only in fourth grade for the comparison between all model and non-model classrooms ($t=4.20$, $p<0.01$). In third grade, the difference between single and re-certified model classrooms reached statistical significance, but the non-model classroom result lay between these.

Table 9: Model Classroom (MC) Status & Mean Reading Gain, [s.d.], (n)

	Not a Model Classroom	Single Certified MC	Re-certified MC
Third Grade	37.8 [6.0] (2446)	35.1 [6.4] (18)	40.9 [7.7] (8)
Fourth Grade	19.7 [4.1] (2075)	22.3 [2.5] (11)	24.0 [3.0] (13)
Fifth Grade	18.6 [4.2] (1756)	19.6 [3.2] (14)	19.8 [0.5] (5)

6 DISCUSSION

Perhaps the most counter intuitive finding from this study is the consistent and strong negative relationship between the Challenge variable and the TVAAS teacher effects. The TVAAS teacher effects are a robust measure of the improvement that students make considering their previous academic history. In other studies these effects have been shown to be related to factors thought to impact positively on growth in academic achievement, just as is the case with Percent Correct and Reading Volume. Thus, the negative relationship observed between Challenge and teacher effects might not be due to the measure of teacher effectiveness, but rather due to other circumstances or variables.

The following are offered as hypothetical explanations (although derived from background anecdotal evidence):

- Some of the most effective teachers may be using AR as merely a supplement to their reading instruction. It must be remembered that the AR relates only to "real" (usually library) books, not to school basal reading schemes or textbooks. Some teachers might actually be highly effective in improving reading comprehension, but focus their efforts in this regard largely upon basals and/or textbooks, leaving students to read "real books merely for enjoyment, irrespective of challenge.

- In some classrooms teachers may have good reading gains but their students are not participating to any great extent in the program, or not testing on all the books they read.

- Perhaps there are constraints of limited stocks of AR books at appropriate levels.

- Perhaps the practice of some librarians of grouping books by grade levels prompts restricted self-challenge in students.

- Perhaps where "reading across the curriculum" policies are in place, there are few opportunities for dedicated reading time. Perhaps even if there are such opportunities, the reading is not monitored. Where little in-school time is devoted to reading practice, perhaps the strong readers read in their own time and the other students try to pass tests on books they have not read, resulting in high apparent challenge. Thus the negative challenge findings might in many schools emanate from a context of improper implementation. Certainly the variation in the percent correct statistic between schools is not reassuring in this respect.

7 CONCLUSIONS

7.1 Analyses of merged records for very large numbers of students (62,739) in Tennessee indicated that beyond third grade, students tended to read fewer but longer books, but increasingly tended to read books of a level of difficulty below their actual grade level (Table 1).

7.2 Value added rises with increased numbers of books read by students, except where a very large number of very easy books are read in the fifth grade and beyond (Table 2).

7.3 Value added rises steadily with increasing percent correct on AR reading comprehension tests, but only becomes positive at the level of 80% correct in grades 3 and 4, and 85% in grades 5 to 7 (Table 4). Alarmingly, more than half the students in the study were operating below this level.

7.4 There was some evidence that operation at low percent correct was particularly likely to be true of students of low overall ability. This implies that teachers were not generating or not responding to AR At-Risk Reports.

7.6 The most significant effects were in the third grade (Table 5). Mean book level showed a statistically significant effect for four out of five grades, and percent correct for three out of five. The interaction between mean book level and percent correct was significant for three out of five grades.

7.7 The average book level regressed after third grade irrespective of student tested reading ability, until in sixth and seventh grade no student of any reading ability level read books which were on average at their actual grade level (Table 6).

7.8 The difference between highly effective and highly ineffective teachers was most striking in the lower grades, especially for students of lower ability (section 8.8). Highly effective teaching of the most able students was rare.

7.9 The remaining conclusions stem from the teacher level analysis, using different variables. Reading Volume, Log Percent Correct, and Reading Challenge

showed a statistically significant relationship to value added in almost all cases in grades three to six (Table 7).

7.10 Reading Volume and Log Percent Correct were consistently positively related to value added. However, Reading Challenge was consistently negatively related to value added (Table 8). This was evident at all grades from 3 to 6.

7.11 Conclusion 7.10 was true at all levels of percent correct, reading volume, and student ability (section 8.11).

7.12 Teachers completing one-day RR Training were significantly more effective than control teachers in grades 3 and 5, more effective (but not significantly) in grade 6, but not more effective in grade 4.

7.13 In fourth and fifth grade, even though model classrooms showed higher effectiveness empirically than non-model classrooms, the differences reached statistical significance only in fourth grade. In third grade there was no consistent difference.

Thus both volume of reading done by students and their success in reading comprehension (as measured by percent correct on AR tests) have a positive impact on teacher effectiveness in terms of value added to student achievement as measured by TVAAS. When both are high, the positive effect is likely to be maximized.

The issue of Reading Challenge is much more complex. In the teacher level analysis, reading challenge was found to be negatively related to teacher effectiveness, irrespective of volume of reading, percent correct or student ability level. The teacher level analysis data suggest that encouraging more difficult books or emphasizing challenge is unlikely to have a positive effect, and is very likely to have an increasingly negative effect if percent correct is allowed to fall below a high figure. The teacher level analysis also suggests it is much more important for teachers to ensure a high volume of reading and a high percentage correct, than to encourage high challenge.

Both increased reading volume and high percentage correct are key aspects of the AR training, which this study thus validates. There is also direct evidence of the impact of AR training on teacher effectiveness in some grades (but not in others). Similarly, there is direct evidence of a relationship between model classroom status and effectiveness in some grades (but not in others).

This study does not permit definitive statements about effective teacher behaviors, since these were not directly measured, and students might engage in effective reading behaviors irrespective of the behavior of their teachers, including directly utilizing the feedback they receive from the AR software to self-manage their own reading more effectively. However, the findings of the study are strongly

suggestive of some very important parameters of reading behavior, which teachers should seek to enhance by whatever means best fits their personal context. The current study supports the view that higher levels of reading practice can yield higher reading achievement, but only when the reading practice is characterized by a high percent correct (i.e., is successful). It also strengthens the evidence regarding directionality of causation.

8 ACTION IMPLICATIONS

8.1 For Future Research

A fruitful area for further investigation is to explore whether students choose individual books **consistently** within their own zone of proximal development (which is a recommendation of the AR training, and supported by other studies) or tend to choose books that indeed are easy for them to read and comprehend. If the later is true, then it suggests that directed reading activity and more teacher intervention may be necessary to improve reading comprehension.

8.2 For Future Practice

Teachers should

Have students read as much as possible,

- But not expect positive effects to be discernible until 10 or more books have been read in the AR system,
- And guide students away from reading a large number of very easy books beyond fifth grade,

Monitor student progress carefully,

- Keep percent correct at 85% or higher,
- Generate and study AR At-Risk Reports,
- Intervene when indicated to achieve the above goals,
- Especially for low ability students,
- And probably especially for high ability students.

Increase challenge level slowly and gradually.

- Monitor reading challenge carefully to ensure that an excess of challenge does not begin to depress percent correct.

- Inform students of the above, by way of formative feedback

These findings on the variable effectiveness of teacher management of quantity and quality of reading practice have evident implications for local, state and national guidelines on literacy teaching which aim to raise teacher effectiveness and standards of achievement. **Locating teacher-assisting software in classrooms clearly does not guarantee it will be used intelligently, reinforcing the view that information technology is not a replacement for the professional teacher, but a tool to potentially enhance the effectiveness of teachers.** Appropriate quality and quantity of training and support for teachers is needed if implementation integrity is to be sustained at the level necessary to raise student attainment. In parallel, future re-engineering of software might elaborate error analysis feedback, more firmly scaffolding meta-cognitive reflection by both student and teacher, requiring either or both to consider possible causes of poor performance on current testing and the action consequently needed.

REFERENCES

Advantage Learning Systems (1993). **The Accelerated Reader** (Computer Program). Wisconsin Rapids, WI: ALS. (http://www.advlearn.com).

Bock, R. D. & Wolfe, R. (1996). Audit and review of the Tennessee Value-Added Assessment System (TVAAS). In R. D. Bock, F. Wolfe, & T. H. Fisher (Eds.), **A review and analysis of the Tennessee Value-Added Assessment System** (pp. 69-72). Nashville, TN: Comptroller of the Treasury, Office of Education Accountability.

Chall, J., & Dale, E. (1995). **Readability revisited.** Cambridge MA: Brookline Books.

Flesch, R. (1968). A readability formula that saves time. **Journal of Reading, 11**, 513-516.

Flesch, R. (1974). **The art of readable writing with the Flesch readability formula.** New York: Harper & Row.

McLean, R. A., Sanders, W. L., & Stroup, W. W. (1991). A unified approach to mixed linear models. **The American Statistician, 45**, 54-64.

Sanders, W. L. & Horn, S. P. (1994). The Tennessee Value-Added Assessment System (TVAAS): Mixed-model methodology in educational assessment. **Journal of Personnel Evaluation in Education, 8**, 299-311.

Sanders, W. L. & Horn, S. P. (1995a). The Tennessee Value-Added Assessment System (TVAAS): Mixed-model methodology in educational assessment. In A. J. Shrinkfield & D. Stufflebeam (Eds.) **Teacher evaluation: A guide to effective practice** (pp. 337-350). Boston, MA: Kluwer.

Sanders, W. L. & Horn, S. P. (1995b). 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, 3**(6): Available: http://info.asu.edu/asu-cwis/epaa/welcome/html, March 3, 1995.

Sanders, W. L., Saxton, A. M., & Horn, S. P. (1997). The Tennessee Value-Added System (TVAAS): A quantitative, outcomes-based approach to educational assessment. In J. Millman (Ed.), **Grading teachers, grading schools**. Thousand Oaks, CA: Corwin Press.

Williams, P.L. (1989). **Using customized standardized tests.** Washington, DC: Office of Educational Research and Improvement, US Department of Education. Eric Document Reproduction Service No. ED 314 429.