

Alignment as a Predictor of Student Achievement Gains

John L. Smithson

Ana Cristina Collares

The analyses presented below follow from the theoretical premise, widely accepted as a truism among educators, that students perform better on tests that measure content they have had the opportunity to learn. It follows that a measure of a students' opportunity to learn assessed and/or standards-based content should contribute to explaining variation in student achievement gains. Such a measure would have the added value of providing a diagnostic tool with which to determine where insufficient opportunity to learn subject-matter content is contributing to poor student performance. By identifying areas of low alignment, and using descriptive results of content descriptions of both the intended and enacted curriculum, teachers are able to target instructional content in order to increase alignment. This increased alignment, in turn is expected to lead to increased gains in student achievement scores. This paper examines the extent to which alignment between instruction and state benchmarks (essentially the test specifications for one state's elementary reading assessments) is able to explain or predict achievement gains for a large sample of student achievement data and teacher self-reports of instructional content.

The procedures and instruments used to generate alignment indices reported here were developed by Porter and Smithson while conducting research on the impact of key education policies regarding instructional practice, and the resulting effect of this on student achievement (Porter & Smithson, 2001). In the Upgrading Mathematics study lead by Porter & Gamoran, a strong positive relationship between opportunity to learn content (alignment) and student achievement gains across a range of high school mathematics courses was found (Gamoran, et.al. 1997). The analyses conducted and reported below attempt to determine if similar results can be established for elementary reading, using results from state standardized reading tests and teacher self-reports of instructional content.

Both analyses use a similar methodology, involving content analysis of documents (i.e. standards and assessments), combined with teacher self-reports of instructional content (using an annual survey), and student assessment scores, combined with key control variables using hierarchical linear models designed to examine the effects of alignment on student achievement.

Methodological similarities aside, the data set utilized here is distinct from the original study in important ways, and these should be noted. The original study focused on high school mathematics, consisting primarily of ninth and tenth grade students. The analyses reported here focus on reading instruction and standards for elementary students. The earlier study used an assessment instrument designed especially for the study, with pre- and post- testing using a common assessment to measure learning gains. The results reported here are drawn from student scores on a statewide assessment, using different assessment instruments to measure achievement at different points in time. Further, the earlier study drew from a relatively small sample of classrooms, spanning diverse levels of student achievement, course types and school locations (i.e., urban, suburban, rural).

By contrast, results reported here are drawn from a relatively large sample of students, but with an over-representation of low achieving students from urban schools and/or economically disadvantaged neighborhoods. These characteristics combine to make the current data set a more challenging sample from which to demonstrate the predictive capacity of alignment measures.

Central to the alignment methodology employed here is a systematic content language used to generate descriptions of content, whether those descriptions are embedded in standards, assessments or based on teachers' self-reports of content coverage (Blank, et.al., 2001). These content descriptions, whether collected through teacher self-reports (for instructional content) or document analysis (for standards content) utilize a common taxonomy or language to describe and quantify content coverage.

The content language is based on a two dimensional construct that defines content as the intersection of topic and expectations for student performance (sometimes referred to as cognitive demand). The English, language arts and reading taxonomy consists of 139 topics organized into 14 content areas and crossed with five categories of expectations for student performance. This yields a two-dimensional matrix with 695 cells (139 * 5), each cell representing a unique content description. Thus a description of a teacher's instructional content can be quantified by placing values in each of the relevant cells of the matrix, representing the proportion of time a teacher reports covering one or another unique content description. Similarly, content descriptions of a standard, benchmark, assessment, or other curriculum document can be quantified using the taxonomy, yielding another content matrix with cells containing values indicating the relative emphasis of text (for standards) or score points (for assessments) embedded in the document. The relationship between any two sets of content descriptions (e.g., instruction compared to benchmarks) can be quantified by making cell-by-cell comparisons between the two sets of quantified descriptions (see Porter & Smithson, 2001). This relationship between content descriptions is referred to as alignment for the purposes of this discussion.

Challenges to the Quality of Data

In trying to demonstrate the predictive capacity of our alignment measure (or index); thereby providing supporting evidence for the validity of the procedures employed, we are faced with certain challenges to the quality of data. These deserve mention. If the results of our analyses do not support the assertion that alignment (read opportunity to learn) helps to predict, or explain variation in student achievement gains, any one or combination of the following could be a factor in explaining the negative outcome. On the other hand, if the results are positive, affirming the assertion that opportunity to learn (as measured by our alignment index) can predict student achievement gains, it will be in large part despite these challenges.

Clearly, the adequacy of the language used to describe content, and the procedures employed to collect those descriptions play an important role in determining the predictive capacity of any resulting alignment measure. The English, Language Arts & Reading (ELAR) survey is a relatively new content language, having been developed in

2003 and first employed on a large-scale basis in 2004. The ELAR language was substantively revised in 2006, based upon feedback from the field (consisting of teachers and content analysts).

The analyses conducted for this report are based on the original ELAR version, and thus the content language may not have provided the optimal descriptive measures upon which to calculate alignment. Subsequent alignment analyses will help to determine the extent to which improvements in the content language will contribute to stronger effects.

Adequacy of the content language aside, it has been suggested for some time that teachers receive an orientation and training on the surveys of enacted curriculum (SEC) prior to survey administration. The orientation is designed to motivate teachers to take the survey conscientiously, to assure them that their results will remain confidential, and to inform them of the varied ways that SEC data can serve teachers. In some cases teacher orientations include additional training on the concept of cognitive demand, one of the two dimensions used to describe content. Absence of, or inadequate teacher orientations are likely to yield data less reflective of instructional practice. The 2005 administration of the SEC was the first large-scale attempt at SEC data collection in Ohio, and many lessons were learned from that effort. It is generally accepted by those involved in the survey administration that Ohio teachers completing the SEC in 2005 had varied levels of, and quality in, the orientation and training provided prior to survey administration (Osthoff & Karatzas, 2005). Thus inadequate training of teachers in completion of the SEC could also account for weak effects. The 2006 SEC administration in Ohio was substantially improved, both in terms of the amount and quality of SEC orientations delivered to the participating teachers (Osthoff, 2006). Subsequent alignment analyses will help to determine whether improved quality of teacher reports will contribute to the level and strength of alignment in predicting student achievement gains.

Just as teachers' reports of content coverage can vary in the quality of data provided, the content analysis procedures themselves can vary in quality from one team of analysts and workshop to another. While the training process is largely consistent across groups, variation in the level of content expertise and experience of analysts using the procedures does exist across teams of analysts. That said, results for the benchmarks used in these analyses show good inter-rater reliability, and utilized the recommended four person teams to conduct the content analyses. (For a more in-depth examination into inter-rater reliability and generalizability of the content analysis procedures (see Porter & Zeidner, 2007).

The final factor to be noted that may have negatively impacted these analyses, concerns the sample of teachers and students for whom data was available. The teacher data was collected in 2005, and was drawn almost exclusively from schools that had been identified by the state as not meeting adequate yearly progress (AYP), yielding a sample of teacher and student data skewed toward low achieving, high poverty schools. In addition, prior achievement (2004) data for students was available for only about one third of the sample for which 2005 achievement data was available, limiting substantially

the extent to which prior achievement could be controlled for. This is particularly unfortunate for alignment measures, which represent a value-added measure, indicating opportunity to learn for a single school year. Thus, it is expected that a more representative sample of teacher and student data will yield a stronger relationship between alignment and achievement.

Alignment Analyses

Ohio's K-4 Benchmarks are used as the target for calculating the instructional alignment measures used in the analysis. Since these benchmarks were designed to guide construction of the reading assessments administered across the elementary grades in Ohio, it serves as an ideal target for the sample of elementary reading achievement data provided by the state.

The analysis results are presented in three stages. The first stage looks at simple correlations. While correlations provide fairly low-level information on associations, not speaking to causality or predictability, it is nonetheless a convenient way to look at potential associations across many variables. Correlation results also assist in narrowing the number of variables examined in the second and third stages of analysis. Since there are many potential variables of interest on both assessment and instructional measures available from the data set, correlations from the stage one analysis serve to narrow the field of variables examined more closely in the second and third stages, while giving some sense of the broader picture of the associations between alignment, achievement and other key variables.

In the second stage of analysis, a series single-level, multivariate linear regression equations are reported to provide an initial look at how well alignment appears to predict achievement and achievement gains, with and without controls for grade level, prior achievement and economic disadvantage (SES). For the sake of brevity, and because they represent the more general summary outcome measures, achievement results for Reading Number Correct (RNC) and Reading Scale Score (RSS) are the only outcome measures used in the models reported for stage 2 and stage 3 analyses. These also happen to be the measures with the strongest correlations, as reported in stage 1.

Since the alignment variable is collected (from teachers) at the classroom level, and achievement results are collected at the student level, stage 3 analyses employ hierarchical linear modeling to better distinguish class and student level effects. For stage 1 and stage 2 analyses, classroom averages of achievement measures were calculated to serve as a classroom level indicator of achievement.

Stage I Analysis: Correlations between achievement, instructional practice, and content.

We begin by looking at correlations between alignment and average classroom achievement for all 2005 assessment results that could be connected with SEC classrooms. Table 1 presents significant results ($p < 0.01$) across eleven achievement

measures, alignment to the K-4 Benchmarks for the state, and other key measures known to have a relationship to achievement.

Table 1: Correlations* between Achievement scores, SES, Grade Level, Alignment and Prior Achievement

	Disadvantaged	Grade Level	OH Gr. K-4 Benchmarks	Prior Achievement
rnc05		0.82	0.41	0.81
rvnc05	-0.18	0.16	0.22	0.40
rcnc05	-0.13	0.73	0.37	0.74
wanc05			0.20	0.44
rss05		0.66	0.42	0.71
rvss05	-0.14	0.72	0.36	0.77
rcss05	-0.14	0.71	0.39	0.79
wass05		0.63	0.38	0.74

* Correlation is significant at the 0.01 level (2 tailed).

rnc05 Reading Number Correct
 rvnc05 Reading Vocabulary Number Correct
 rcnc05 Reading Comprehension Number Correct
 wanc05 Word Analysis Number Correct
 rss05 Reading Scale Score
 rvss05 Reading Vocabulary Scale Score
 rcss05 Reading Comprehension Scale Score
 wass05 Word Analysis Scale Score

It is clear from Table 1, though not surprising, that there is strong positive relationship between achievement and prior achievement on each of the eight achievement measures reported. Grade level also shows a strong association with achievement, though this is largely an artifact of the measured characteristics. For example, correlations to the number correct (_NC) scores, could be the result of there being fewer items on the early primary grade assessments than for grades 3 and 4. The scale scores would be expected to be correlated with grade level since these scores are vertically scaled to show growth across school years. Thus grade level and prior achievement will be important control variables to consider when looking at regression models.

Whether or not a student is classified economically disadvantaged appears to have a weaker association with achievement results, though where it appears, it is negative as would be expected. This weak effect for economic disadvantage is likely a result of the sample being skewed in the direction of economic disadvantaged students. It should also be noted that since these results are based on classroom averages, thus the disadvantage measure used to calculate the correlation represents the proportion of students in the classroom that are classified as economically disadvantaged. (At the student level disadvantaged is a dummy variable where disadvantaged = 1.) With respect to alignment, while the level or correlation is not as high as prior achievement (or even grade level in this data set), correlation results show a strong persistent and positive relationship to student achievement across all eight achievement measures reported here.

It should be noted that these *classroom* level correlations tend to increase the level of the correlation as compared to correlations with the larger sample of student data. The tendency of classroom averages to show stronger correlations with achievement is almost certainly due to the elimination of within-class variance on achievement that results from

averaging by class. Nonetheless, the pattern of significance across achievement results remains about the same as when using *student* level correlations.

As previously stated, the stage 3 analyses will treat each of the variables at the appropriate level in the models in order to better examine the classroom effect of alignment on individual achievement scores (and more importantly, gains). Before turning to multi-level models however, simple regression equations reported as Stage 2 results provide a convenient look at several models with stepwise inclusion of each of the three control variables included in the full model. This will permit some preliminary examination of effect sizes and model fit to see the extent to which the various control measures contribute to the explanation of achievement variation across classrooms. This will in turn assist in limiting the number of multi-level models considered for stage 3 analysis.

Stage 2 Analysis: Single-level multivariate linear regression equations

While the correlation results are generally positive for Benchmark alignment, regression equations allow us to add other controlling variables, examine model fit, and look at effect sizes for alignment. In particular, controlling for prior achievement might help boost the effect of alignment and/or improve model fit. Presumably, alignment for a specific year would be better at predicting achievement *gains*, than achievement *scores*. This is premised however on the assumption that achievement *gains* are the result of new knowledge, and secondly that alignment measures indicate the opportunity to learn new knowledge. Both assumptions are problematic. Assessments, particularly large-scale assessments, are likely to draw upon content knowledge learned in previous years. Likewise the curriculum, especially English, language arts and reading has a tendency to review old knowledge, either for remediation, or to prepare the student for extending that knowledge further, or connecting with other, new knowledge. One need only look at the grade-by-grade progression of grade level expectations (GLE's) for any state that uses them to see this pattern of an overlapping and (hopefully) spiraling curriculum.

Table 2 presents results for two very simple regression equations, looking at the effect of alignment to K-4 Benchmarks on student achievement for Grades K-3 as measured by the Reading Number Correct (RNC) and Reading Scale Score (RSS) measures. While Benchmark alignment is statistically significant in predicting both outcome measures, the model fit is poor, with an adjusted R² of 0.176 for RNC and 0.194 for RSS. The effect size reported here is in terms of the percentage of the overall score accounted for by Benchmark alignment. One could also consider the standardized coefficient as a useful measure for considering effect size (.42 & .44 for RNC and RSS respectively).

Table 2: Alignment to K-4 Benchmarks as a predictor of Class Achievement

Class Average (n = 547)		Mean	Standard Deviation	Regression Coefficient	Standardized Beta	Significance Level	Effect Size**	Adjusted R ²
<i>predicting</i>	RNC05	25.63	7.66					
	Constant			10.02		0.000		
	K-4 Benchmarks	0.18	0.04	87.78	0.42	0.000	62%	0.176
	RSS05	597.52	29.45					
	Constant			534.63		0.000		
	K-4 Benchmarks	0.18	0.04	353.58	0.44	0.000	11%	0.194

* Constant

** Percent of Achievement Score

One measure likely to improve model fit, especially given the results from the first stage of analysis is grade level, which indicates that these scores tend to rise as grade level rises. In order to account for a grade level effect, the 2005 grade level is entered into the equation as a control variable. The results are reported in Table 6.

Table 3: Alignment to K-4 Benchmarks as a predictor of Class Achievement controlling for Grade Level

Class Average (n = 547)	Mean	Standard Deviation	Regression Coefficient	Standardized Beta	Significance Level	Effect Size**	Adjusted R ²
<i>predicting</i> RNC05	25.63	7.66			0.000		
Constant			9.17		0.000		
Grade05	1.90	0.88	7.15	0.82	0.000		
K-4 Benchmarks	0.18	0.04	16.07	0.08	0.002	11%	0.726
<i>predicting</i> RSS05	597.52	29345.00			0.000		
Constant			532.16		0.000		
Grade05	1.90	0.88	20.88	0.62	0.000		
K-4 Benchmarks	0.18	0.04	144.24	0.18	0.000	4%	0.510

* Constant

** Percent of Achievement Score

As expected, the model fit increases substantially when grade level is considered. This substantially reduces the apparent impact of alignment, as can be seen by the dramatic drop in the standardized coefficient. However, since alignment is conceived as a value-added measure, and our interest is in the contribution of alignment to achievement gains, it is more informative to look the percent of achievement gain for a typical student that alignment appears to account for. Viewing effect size from this perspective still yields only a modest effect (11% and 4% respectively for RNC and RSS), but at least in a metric somewhat more relevant to central question of predicting achievement *gains*.

Since alignment is primarily a value-added measure (representing opportunity to learn for a single school year), and achievement *gains* the relevant outcome measure, prior achievement becomes an essential control variable. One might also reasonably expect that including a prior achievement measure in the regression model will noticeably improve the mode fit (adjusted R²). For this set of data however, prior achievement data is problematic insofar as almost two-thirds of the sample (at the student level) is missing prior achievement data (note the sample size in Table 2 compared to Table 3). This problem is lessened somewhat by looking at classroom averages, as most of the 2nd and 3rd grade classrooms have prior achievement data for at least some students, though that raises another question – how well does the prior achievement for one portion of the class reflect prior achievement for the students without such data. The problem in either case is whether the sub-sample with prior achievement data is representative of the larger sample. Thus, regardless of whether the results are significant or not, it raises some question about the extent to which those results can be generalized.

Table 4 reports results when prior achievement is added to the regression model.

Table 4: Alignment to K-4 Benchmarks as a predictor of Class Achievement controlling for Prior Achievement and Grade Level

Class Average (n = 258)	Mean	Standard Deviation	Regression Coefficient	Standardized Beta	Significance Level	Effect Size**	Adjusted R ²
<i>predicting</i> RNC05	29.85	6.57					
Constant			-3.63		0.023		
RNC04	21.81	4.16	0.68	0.44	0.000		
Grade05	2.51	0.50	6.50	0.50	0.000		
K-4 Benchmarks	0.19	0.03	13.26	0.06	0.053	31%	0.728
<i>predicting</i> RSS05	612.40	22.82					
Constant			327.37				
RSS04	589.09	25.76	0.43	0.52	0.000		
Grade05	2.51	0.50	8.92	0.21	0.000		
K-4 Benchmarks	0.19	0.03	71.14	0.11	0.024	57%	0.448

** Percent of Achievement *Gain*

As can be seen from Table 4, the model fit only increases slightly for predicting RNC, and actually drops for RSS when prior achievement is controlled for. Alignment becomes only borderline significant for the RNC measure, accounting for 31% of the achievement gain of an average student. The results remain significant for the RSS measure however, and account for more than half (57%) of the achievement gain for an average student. As mentioned above however, these results need to be reviewed cautiously, due to the large amount of missing data on prior achievement.

In order to compensate somewhat for this missing data, and also test whether the sub-sample with prior achievement data is representative of the larger sample, two new variables were created for each of the two prior achievement measures used in the analysis. One variable (with the prefix, '*gmeans*') replaces missing data with a mean value. Since it has already been determined that achievement measures tend to increase as grade level increases, missing data was replaced with the grade level mean for the relevant measure, rather than the sample mean. The second variable created is a dummy variable to indicate the cases with missing data (1) or not (0) for each student included in the analysis (then averaged by class, yielding a measure representing the proportion of missing data for each class). This second variable assists in determining whether the missing data is affecting the results, which are reported in Table 5.

Table 5: Alignment to K-4 Benchmarks as a predictor of Class Achievement controlling for Prior Achievement***, & Grade Level

Class Average (n = 337)	Mean	Standard Deviation	Regression Coefficient	Standardized Beta	Significance Level	Effect Size**	Adjusted R ²
<i>predicting</i> RNC05	29.85	6.57					
Constant			-4.79		0.002		
<i>gmean</i> RNC04	21.73	3.83	0.65	0.38	0.000		
misRNC	0.45	0.36	-1.28	-0.07	0.021		
Grade05	2.51	0.50	6.58	0.50	0.000		
K-4 Benchmarks	0.19	0.03	24.61	0.12	0.000	50%	0.705
<i>predicting</i> RSS05	612.40	22.82					
Constant			333.69		0.000		
<i>gmean</i> RSS04	589.09	25.76	0.39	0.39	0.000		
misRSS	0.45	0.36	-9.63	-0.15	0.000		
Grade05	2.51	0.50	10.14	0.22	0.000		
K-4 Benchmarks	0.19	0.03	151.80	0.21	0.000	104%	0.414

** Percent of Achievement *Gain (minus missing data effect)*

*** Missing data on prior achievement replaced with grade level mean.

Replacing missing data with grade level means for prior achievement substantially increases the significance and effect size for alignment on both outcome measures, though the model fit is decreased somewhat. The dummy variable for missing data (prefix = 'mis') is significant for both RNC and RSS, suggesting that the missing data is influencing the results, and suggests caution in interpreting those results.

Table 6: Alignment to K-4 Benchmarks as a predictor of Class Achievement controlling for Prior Achievement***, Grade Level, and Economic Disadvantage

Class Average (n = 337)	Mean	Standard Deviation	Regression Coefficient	Standardized Beta	Significance Level	Effect Size**	Adjusted R2
<i>predicting RNC05</i>	29.85	6.57			0.041		
<i>Constant</i>			-3.60		0.033		
<i>gmeanRNC04</i>	21.73	3.83	0.63	0.37	0.000		
<i>misRNC</i>	0.45	0.36	-1.38	-0.08	0.013		
<i>Grade05</i>	2.51	0.50	6.67	0.51	0.000		
<i>ClsDisadvantage</i>	0.73	0.32	-1.05	-0.05	0.097		
<i>K-4 Benchmarks</i>	0.19	0.03	23.42	0.11	0.000	47%	0.707
<i>predicting RSS05</i>	612.40	22.82					
<i>Constant</i>			346.07		0.000		
<i>gmeanRSS04</i>	589.09	25.76	0.38	0.38	0.000		
<i>misRSS</i>	0.45	0.36	-10.16	-0.16	0.000		
<i>Grade05</i>	2.51	0.50	10.46	0.23	0.000		
<i>ClsDisadvantage</i>	0.73	0.32	-5.25	-0.07	0.090		
<i>K-4 Benchmarks</i>	0.19	0.03	145.86	0.20	0.000	98%	0.418

** Percent of Achievement Gain (minus missing data effect)

*** Missing data on prior achievement replaced with grade level mean.

Table 6 reports results when economic disadvantage is added to the model as a control variable. Using classroom averages for economic disadvantage is not significant, though this likely has more to do with the effects of averaging by classroom, and the nature of the sample described above, than an indication of little impact of economic disadvantage on achievement. To the extent that class disadvantage (proportion of students in the class that are classified as economically disadvantaged) does contribute to the explanation of variance in student achievement, it comes at the expense of the explanatory power of alignment, which would not be unexpected. A better indication of the effect of economic disadvantage should emerge as we consider HLM models where student level variables can be addressed more directly.

Stage 3 Analyses: Multi-level Models

In order to best reflect the combination of classroom and student level variables used for these analyses, multi-level models were examined for the two outcome measures and the full set of control variables reported in Table 7.

Table 7: Alignment to K-4 Benchmarks as a predictor of Student Achievement Gains, controlling for prior achievement*, grade level and economic disadvantage

5,515 Students nested in 348 classrooms		Mean	Standard Deviation	Regression Coefficient	Standard Error	Significance Level	Effect Size**
predicting RNC05		29.97	9.37				
Level 1							
	<i>Constant</i>			29.9	0.17	0.000	
	K-4 Benchmarks	0.19	0.03	17.17	5.61	0.003	34%
Level 2							
	Grade Level	2.50	0.50	5.6	0.36	0.000	
	Disadvantaged	0.71	0.46	-0.76	0.24	0.002	
	<i>gmean</i> RNC04	21.82	4.48	0.88	0.02	0.000	
	<i>mis</i> RNC04	0.44	0.50	-1.04	0.22	0.000	
predicting RSS05		613.20	38.92				
Level 1							
	<i>Constant</i>			612.83	0.82	0.000	
	K-4 Benchmarks	0.19	0.03	107.01	27.66	0.000	71%
Level 2							
	Grade Level	2.50	0.50	7.6	1.72	0.000	
	Disadvantaged	0.71	0.46	-3.96	1.15	0.001	
	RSS04	588.98	31.56	0.53	0.02	0.000	
	<i>mis</i> RSS04	0.44	0.50	-6.91	1.06	0.000	

* Missing data on prior achievement replaced with grade level mean.

** Percent of Average Annual Achievement Gain

Though the effect size for alignment is somewhat smaller in the multi-level equations, it is generally consistent with what was reported for single level regressions. Again, the missing data is a factor, and using only the sub-sample for whom prior achievement data is available does not yield significant results with multi-level modeling. Despite this caution, when all the various analyses and challenges to the data are taken as a whole, there is evidence to suggest that alignment to Benchmarks does seem to contribute to explaining variation in student achievement. We are confident that with a more robust and representative data set that we will be better able to more accurately report the size of this alignment effect.

As with the single level regression equations, missing data on prior achievement is replaced with the grade level mean. Results for multi-level models using only the sub-set of students for whom prior achievement was available did not yield significant results for alignment. As seen previously, the dummy variable included for missing data continues to have a significant role in the models. Thus caution must be taken in judging the results. Using a conservative estimate (subtracting the missing data effect from the alignment effect) of the effect size, the average achievement *gain* for the school year represented by the alignment variable in the model predicting RNC, accounts for about 34% of that gain. This is only slightly lower than the effect size reported with the single level regression model (39%). The effect size for RSS is slightly higher in the HLM

model (71% of the average achievement gain for the school year, compared to 68% in the single level multivariate linear regression model). Again, the pattern is roughly similar across single-level and multi-level regression models.

Conclusion

With the advent of No Child Left Behind legislation, achievement, accountability, and closing-the-gap have become the watchwords of educational improvement. Teachers are under persistent pressure to raise student scores, and are told over and over that the way to accomplish that is to bring instruction into better alignment with state and national standards. Standards, benchmarks, indicators, and grade-level-expectations are just some of the documents that have been developed with the intention of providing teachers a road-map for practice and student success. Indeed, the entire standards-based reform effort hinges on the notion that students provided an opportunity to learn standards-based academic content can succeed on state assessments. In such an environment an important tool for both teachers and researchers is one that can assist in pin-pointing standards-based content, and confirm the thesis that opportunity to learn standards-based content does indeed contribute to student performance.

Of course, opportunity to learn is not the only thing necessary for student success, though it is one important, probably essential, ingredient. Quality of instruction and teacher content knowledge and skill in adapting that knowledge to the educational environment are also undoubtedly important ingredients. We also know from years of research that students' intrinsic ability and their environmental setting also plays a significant role in student performance. Other, more subtle factors, such as test anxiety, learning style, and brain function also undoubtedly play a role in learning, and recent and on-going research is beginning to put greater focus on these types of influences.

At the same time, teachers' professional development opportunities are increasingly focused on teacher content knowledge and moving instructional practice toward more inquiry-based and application oriented strategies in order to raise the level of students' cognitive engagement with subject-matter content. Moreover, such initiatives are being required to provide research-based evidence and evaluation of programmatic impacts on teacher practice and student outcomes.

The contribution that a valid and reliable measure of instructional alignment offers both research and evaluation is a measure that permits one to control for opportunity to learn standards-based content. Just as researchers regularly control for prior achievement and students' economic context, we would argue that alignment plays an equally important role in explaining achievement gains, and should be as regularly employed in trying to model student learning as the more common measures of prior achievement and SES. By adding alignment to their statistical models, researchers and program evaluators are much better situated to study the effects of other, no less important elements of student learning and instructional practice.

Does our language and methodology provide that valid and reliable measure? At present, a cautious, but no less optimistic stance seems justified. While our results are not as strong as we might like, given the challenges to the data set out above the fact we were able to show any positive and significant relationship between alignment and achievement gains at all feels like a small victory. More work and more research are obviously required. We will continue to refine our methods and our instruments, and await more data in order to further investigate the role of instructional alignment. That said, we are confident that the tools and methods we offer teachers, administrators, evaluators and researchers have great promise in serving their efforts toward achieving success for all students.

References:

- Blank, R.K., Porter, A.C., & Smithson, J.L. (2001). *New Tools for Analyzing Teaching, Curriculum and Standards in Mathematics & Science*. Washington, DC: Council of Chief State School Officers.
- Gamoran, A., Porter, A.C., Smithson, J., & White, P.A. (1997, Winter). *Upgrading high school mathematics instruction: Improving learning opportunities for low-achieving, low-income youth*. *Educational Evaluation and Policy Analysis*, 19(4).
- Osthoff, E. and Karatzas, C (2005). *Ohio Project on Surveys of Enacted Curriculum: Spring 2005 Evaluation Report*. Washington DC: Council of Chief State School Officers.
- Osthoff, E. (2006). *Ohio Project on Surveys of Enacted Curriculum: Spring 2006 Evaluation Report*. Washington DC: Council of Chief State School Officers.
- Porter, A.C. & Smithson, J.L. (2001). *Defining, developing, and using curriculum indicators*. CPRE Research Report Series RR-048. Philadelphia, PA: University of Pennsylvania, Consortium for Policy Research in Education.
- Porter, A.C. & Zeidner, T. (2007). *Investigating the Quality of SEC Content Analyses of State Student Achievement Tests and Content Standards*. Paper presented at the 2007 Annual Meeting of the American Educational Research Association, Chicago.