Running head: ESTIMATING ALIGNMENT INDEX CRITICAL VALUES

Estimating critical values for strength of alignment among curriculum, assessments, and instruction

Gavin W. Fulmer

Division of Research on Learning in Formal and Informal Settings

National Science Foundation

Arlington, VA 22230

May 1, 2010

Paper presented at the 2010 annual meeting of the American Educational Research Association

This material is based on work supported by the National Science Foundation while under employment of the Foundation through an Independent Research and Development plan. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Abstract

School accountability decisions based on standardized tests hinge on the degree of alignment of the test with a state's standards. Yet no established criteria were available for judging strength of alignment. Previous studies of alignment among tests, standards, and teachers' instruction have yielded mixed results that are difficult to interpret and compare across studies. This simulation study determined critical values for Porter's (2002) alignment index, suitable for hypothesis testing at alpha levels of .05 and .10. The critical values will be useful to researchers or policymakers who seek to judge strength of alignment among standards, assessments, and instruction. The presentation will also describe directions for future research in establishing objective criteria for the interpretation of alignment indices.

Estimating critical values for strength of alignment among curriculum, assessments, and instruction

The 2001 passage of the No Child Left Behind Act (NCLB; Public Law 107-110) solidified a heightened focus on accountability in school reform that began in the 1990s. NCLB required states to develop rigorous content standards and standardized tests to measure students' progress and, thus, to increase accountability for state and local education entities. The US Department of Education under President Obama appears to support a continued focus on student achievement and teacher quality (Tomsho, 2009).

The greater importance of standardized tests in student outcomes and school accountability underscores the value of alignment among standards and assessments (Linn, Baker, & Betebenner, 2002). Accurate claims about a student's mastery of the standards rely upon the degree to which the test measures those standards adequately (Bhola, Impara, & Buckendahl, 2003). In addition, teachers' classroom instruction should be reflective of the state's standards and the assessment (Blank, 2002).

However, while researchers argue that greater alignment is superior (e.g., Porter, 2002), there do not exist any objective criteria for assessing alignment strength. Additionally, no studies in the current literature have identified critical values that are appropriate for hypothesis testing on alignment indices. The present study fills this void by presenting critical values for hypothesis tests on indices of alignment, based on data simulation techniques. It then uses the critical values to reanalyze results from prior studies on strength of alignment among standards, assessments, and instruction.

Background

Just as the passage of NCLB increased the focus on student achievement and school accountability, it also heightened concerns with states' content standards and assessments. The law required states to establish content standards and achievement tests, if not already in existence, and to

demonstrate adequate yearly progress (AYP) for students in a variety of categories. Prior to the passage of NCLB, states and territories varied in the grade level, content areas, cognitive demands, and stringency of passing requirements (Linn, Baker, & Betebenner, 2002). This made comparison of AYP across states difficult (Joint Study Group on Adequate Yearly Progress, 2002) even if one assumes that patterns of non-test-taking are random and not strategically selected (Lemke, Hoerandner, & McMahon, 2006). Since the passage of NCLB, states have been required to mathematics and English language arts at grades 3 through 8, and again in high school. However, the cognitive demands and stringency of passing requirements have remained idiosyncratic (e.g., Linn, 2008).

Furthermore, NCLB requires that each state's assessment be aligned with the state's standards, in order to promote accurate AYP determinations. This has prompted much research on alignment among content standards, standardized tests, and instruction (Bhola, Impara, & Buckendahl, 2003; Porter, 2002; Porter, Smithson, Blank, & Zeidner, 2007; Webb, 2007). Ongoing analyses of alignment are necessary to gauge changes in alignment over time (Author, Year), as states modify their annual assessments through item revision and creation, and update their curricula and standards. As some states move to develop common core standards and assessment instruments (Glod, 2009), the importance of accurate and effective measurement of alignment between standards and tests will only increase.

Measures of alignment have had mixed results. The National Research Council found that alignment between state standards and standardized tests was poor (Rothman, 2003). Research showed great variation in the degree of alignment among tests, standards, and instruction. Using an alignment scoring process that allows indices ranging from 0 to 1 (Porter, 2002), some prior study has found alignments as low as 0.15 (Porter, Smithson, Blank, and Zeidner, 2007) and as high as 0.80 (Liu & Fulmer, 2008). Most would agree that greater alignment is appropriate, particularly if the purpose is to ensure that an assessment aligns well with the standards it purports to measure. On the

other hand, items on a test represent only a sample of the domains of desired content and performance specified in standards. Therefore, alignment could not be expected to be perfect. Furthermore, discrepancies between the emphases of an assessment and state standards may be acceptable if the test encourages teachers and students to prepare for higher-order thinking than the standards describe (e.g., Liu & Fulmer, 2008).

While the measurement of alignment has increased in frequency, the lack of a reliable metric for alignment indices has eluded researchers. This has greatly limited the ability of researchers to make valid conclusions about the strength of alignment. For a study that calculates an alignment of 0.15, could the researcher decide whether this value could have occurred by chance? If another state has an alignment of 0.50 on this scoring rubric, would that be more or less likely? One solution is to explore where a particular value falls in the distribution of alignments and determine a corresponding p-value. This uses the probabilistic language of hypothesis testing (e.g., Shavelson, 1996, p. 243), which is familiar to many scholars. Since a closed-form expression for the distribution of alignments is not known, this study uses numerical methods to estimate critical values corresponding to alpha levels .05 and .10 so that researchers can determine whether their alignment measures are likely to have occurred by chance.

The following section explains the process for selecting a rubric for scoring alignment and demonstrates its use. Next, the methodology section describes the estimation algorithm for determining the critical values of alignment indices. Then, the results section presents the critical values and a reexamination of alignment results in previous studies. The final section discusses implications of these critical values for the field and directions for future research.

Selecting and Using an Alignment Rubric

Blank (2002) lists four models for determining alignment: Webb (cf. Webb, 2007), Surveys of Enacted Curriculum (SEC, as described by Porter, 2002), Achieve (cf. Rothman, Slattery, Vranek, & Resnick, 2002), and Council for Basic Education (CBE). Of these four, Webb and Porter were

deemed more likely candidates for the present study. Ultimately, the Porter alignment index was selected. More information on Webb and Porter alignment methods are described below, followed by a justification for selecting Porter's index.

Webb's model uses four alignment criteria (2007): Categorical Consistency, Depth-of-Knowledge Consistency, Range of Knowledge Correspondence, and Balance of Representation. Webb recommends using multiple raters, cooperating over several days, to produce an alignment value for each of the four criteria. In addition, the Webb alignment method is used for the purpose of comparing alignment of an assessment to a particular content standard; it takes the content standard as a given.

The Porter (2002) alignment index uses two variables for coding. Prior research has typically used content and cognitive complexity, such as a revised Bloom's taxonomy of educational objectives (Anderson & Krathwohl, 2001), as the two variables. Each element from the comparison documents (e.g., items from the test) are rated on the two alignment variables, and alignment is calculated among the tables. The Porter alignment is a much simpler process than Webb's in terms of the amount of coding required. As a result, the coding process is faster and inter-rater reliability is easier to calculate. In addition, Porter's index is independent of standard and assessment; each document is coded using the same rubric, not a rubric established based on a content standard. This lets researchers' and policymakers' decisions about degree of alignment be informed by alignment among tests, standards, or instruction across multiple jurisdictions.

Porter's alignment index can also be used to compare documents on any two categorical variables, not just on content and cognitive complexity as has been common. The only restrictions are that the two variables for coding must be categorical, and both variables must be applicable to the standards or test items to be coded. For example, one could choose dimensions for coding test items and standards to be (1) language complexity and (2) gender neutrality. This could be applied so long

as every element for coding (whether items from a test or statements from standards) could be adequately coded into these variables.

In general, the size of the coding tables is smaller than the tables for standards documents or test-development plans. For comparison, Porter (2002) might summarize all items and statements on the mathematics content of "number properties and operations" in one row, whereas the mathematics framework for the 2009 National Assessment of Educational Progress (NAEP; National Assessment Governing Board, 2009) includes 31 content statements under the same content (pp. 9-13). Therefore, Porter's alignment index analysis process reduces the dimensionality of such comparisons. Because of its relative simplicity in calculation and broad applicability, it is preferable for the present purpose. All subsequent analyses use Porter's index.

The Porter alignment index analyzes the extent of alignment between two tables of frequencies (i.e., a table for coding of a standards document, and a table for coding of the assessment). It produces a single alignment index, ranging from 0 to 1, to indicate how closely the distribution of points in the first table (standards) aligns with the second table (assessment). The Porter alignment index, *P*, is computed in four steps. (1) Create tables of frequencies for the two documents being compared. These are labeled *A* and *B*. (2) For each cell in tables *A* and *B*, compute the ratio of points in the cell with the total number of points in the respective table. Label the tables of ratios as *a* and *b*. (3) For every row *j* and column *k* in tables *a* and *b* (the tables of ratios), calculate the absolute value of the discrepancy between the ratios in cells a_{jk} and b_{jk}. (4) Compute the alignment index using the following equation:

$$P = 1 - \frac{\sum_{k=1}^{K} \sum_{j=1}^{J} \left| a_{jk} - b_{jk} \right|}{2}$$

In the equation, *J* is the number of rows and *K* is the number of columns in each table, and a_{jk} and b_{jk} are the ratios of points in the cells at row *j* and column *k* for the respective ratio tables, *a* and

b. Let the total number of cells in the table be called N ($=J\times K$). Figure 1 demonstrates the procedure for calculating alignment between a pair of tables, for sample data where $J=2$ and $K=2$ (and, thus, $N = 4$).

INSERT FIGURE 1 ABOUT HERE.

It is important to note that the total number of cells in the A and B tables can have an effect on the alignment index. For any fixed number of test points or standards statements, a greater number of cells in the table will yield a range of likely values that is lower than for tables with fewer cells. For example, if both tables consist of just one cell, then the alignment will always be 1 because the ratios will correspond perfectly. As the number of cells increases, there is much more room for discrepancy between the ratios, and the values for the index are likely to be lower.

This table-size dependence of the alignment index would be a significant stumbling block in comparing alignment indices across studies. In Liu *et al*. (2008), all alignment analyses used the same 5 rows and 6 columns to enable such comparisons. However, if studies use different coding dimensions when assessing alignment, it is difficult to know whether a higher or lower alignment is meaningful, or is a consequence of the table size. This heightens the need for established criteria for assessing the strength of alignment indices.

In addition to a dependence on size, the alignment index also depends on the number of curriculum/standards statements or test items being coded. As an example, consider a standards document that only has one statement. The table of ratios of points for the standards (e.g., table *a*) will be all zeros, except for one cell. As more items or statements are included, the range of likely alignments may increase.

The dependence on quantity and coding dimensions of the standards and tests is a reason that one cannot use critical values for correlation coefficients when comparing alignment indices. The underlying structure upon which the calculation of the alignment index rests is categorical, rather than continuous as in the case of biserial correlation. In addition, correlations range from -1 to 1, with

zero being the center of the distribution and, therefore, the most likely value to occur if the data points were scattered at random without any relationship between the variables. By contrast, the alignment index ranges from 0 to 1, and the center of the distribution of indices depends on the size of the tables being compared, so that the most likely value to occur by chance is not zero. Therefore, comparing the alignment index with a null hypothesis of zero would be inappropriate. It is more essential to assess how far an observed alignment index is from the center of the distribution of indices could occur by chance.

Methodology

The present study comprised three phases. In the first phase, a computational algorithm was used to simulate alignment indices under different conditions by varying: a) number of cells in tables; and b) number of "points" in a standards document. A fixed number of test points, 100, was used to simplify the algorithm. The second phase involved examination of the resultant distributions of the simulated indices and determinations of quantiles. In the third phase, the critical values were used in a reexamination of previous research studies' findings on observed alignment among curriculum, standards, and instruction.

The simulation procedures were conducted using the R software package (version 2.9.1), a free, open-source implementation of the S language (Ihaka & Gentleman, 1996). R was selected because of its accessibility and the wealth of documentation available on its use (R Development Core Team, 2005). Furthermore, the random number generator (RNG) in R produces numbers from a uniform distribution using the Mersenne-Twister algorithm, and is considered superior to the RNG in other software packages readily available to most researchers, such as Microsoft Excel or Visual Basic for Applications (McCullough, 2008).

For the first phase, the decision to compare two ratio tables that were generated randomly was made intentionally. Such a comparison does not assume that there is any single, reference standard; so, the process and results are more general, and can be informative for researchers and

policymakers from any jurisdiction. That is, the present study focused on identifying likely ranges of alignment indices, making no assumptions about the distribution of points in the "true" standards or test. Fitting the exploratory nature of this study, the uniform distribution was selected because it treats all cells as equally likely. Using the uniform distribution also means that table size could be interpreted as the total number of cells, regardless of dimension: that is, tables with 20 cells have equivalent results whether the dimensions are 4×5, 2×10, or 1×20. A pilot study to demonstrate this equivalence, but is not reported here to conserve space.

For tables with rows J and columns K, let the number of cells be $N=J\times K$. The algorithm placed points, at random, into N cells for a table, and then repeated the process for another table of equal dimension. For the test tables, the number of points was fixed at 100. For the standards tables (table *a*), the number of points varied by condition: 30, 60, 90, and 120 points. The values for the standards tables were chosen to reflect differences among jurisdictions. For example, Florida's previous standards had 36 statements for mathematics content at grade 8 and the current "Next Generation" Sunshine State Standards have 19 statements (Florida Department of Education, 2005). On the other hand, New York had 85 standards statements for its high school physics Regents exam (University of the State of New York, 1996). While Florida appears to have far fewer standards statements, extracting elements for coding would be a process of interpreting specific content and cognitive processes from them, just as is necessary when developing an assessment framework.

For each unique N and number of standards points, the alignment index calculation was reiterated 5000 times. A pilot study compared results using a number of iterations (e.g., 500, 5000, and 50000), finding that additional iterations did not noticeably alter the distributions of simulated alignment indices by cell. The range of table sizes used was from 2×5 through 10×10, which included 33 unique sizes (N'*s*). The decision to calculate through a table size of 10×10 was arbitrary. Additionally, the algorithm was run on a table of size 19×6 to allow reanalysis of one study with unusually large table size (Liang & Yuan, 2008). In this way, the algorithm allowed comparisons of

results according to number of cells in the table, and according to the number of standards statements.

In the second phase, the simulated data was analyzed using graphical examination and development of quantiles. Histograms of the data are compared under the varying conditions of number of cells and standards points. The set of simulated was also used to determine quantiles that reflect critical values at the alpha levels of .10 and .05 for one- or two-tailed tests (i.e., 0.025, 0.050, 0.100, 0.900, 0.950, and 0.975). The quantiles for the mean and first standard deviation were also computed. Complete R scripts are available upon request.

In the third phase, the means and critical values from the simulated data were used to reexamine results from previous alignment studies. In some instances the dimensions of the tables for determining alignment were not reported in the original work. In such cases, the author used the coding dimensions from the cited authors' previous studies, or assumed that the coding dimensions were the same across different content areas (e.g., math and reading) in the same report.

Results

The results yielded a suitable set of means and critical values for alignment indices. The results also demonstrated the expected distribution pattern of alignment indices (Table 1): the mean alignment index is lower for tables of greater size; and the mean index is lower in cases with fewer points in the standards.

INSERT TABLE 1 ABOUT HERE.

Figure 2 presents histograms with overlaid curves that demonstrate the effects that table size and number of standards points have on the distributions. As can be seen, the distributions have the characteristic normal shape, but with narrower peaks. A normal distribution would be expected under the Central Limit Theorem, which demonstrates that the distribution of a sufficient number of random variables will be normal, even if the underlying distributions from which they are drawn are non-normal (Ross, 2004, p. 210). The narrow peaks are likely an artifact of the restricted range of the

indices [0,1] compared to the normal number line $(-\infty,+\infty)$. Moreover, notice that the mean is different for each distribution, and is never zero. This is meaningful for the interpretation of hypothesis tests on alignment indices. A hypothesis test based on these criteria is not a test of whether alignment is non-zero. The mean values in Table 1 should be used as null hypothesis values. Table 2 presents the critical values for alignment indices calculated from the simulation. The critical values in Table 2 are used to determine if the observed alignment index is statistically greater or less than the corresponding value in Table 1.

INSERT FIGURE 2 ABOUT HERE.

INSERT TABLE 2 ABOUT HERE.

Repetitions of the alignment simulation yielded results that were identical to the third decimal (thousandths) in all cases, and identical to the fourth decimal (ten-thousandths) in most cases. The standard deviations among repeated simulations tended to be slightly higher for the larger table sizes—that is, where the size of the table allows more variability in alignment—but were still small (less than 0.002). This low variability indicates that the critical values determined through this algorithm are fairly stable across repetitions, suggesting high precision in the alignment simulation algorithm.

Use of the Reference Criterion Values

To use the reference criterion values in Table 2, first identify the appropriate alignment index for comparison, based on the two-dimensional coding scheme the researchers used. Previous examples have used content and cognitive level (e.g., Liu & Fulmer, 2008). Then, identify the appropriate quantile. For two-tailed tests at an alpha level of α, the quantiles are $(\alpha/2)$ and $(1-\alpha/2)$. Typically, researchers use a two-tailed test with alpha level .05. In that case, the quantiles are 0.025 and 0.975. If one wishes to use a one-tailed test, then the quantile to be used would be either (α), or

(1-α), depending on the hypothesized direction (below or above the mean, respectively). Finally, compare the alignment index to the criterion value. An example will help demonstrate the process.

Example. Porter (2002) examined alignment between a state's standardized mathematics test and its curriculum standards using a classification process from the Surveys of Enacted Curriculum (SEC) with 6 content areas and 5 cognitive levels. The observed alignment index was 0.37 (Porter, 2002, p. 6). Suppose that there were 30 standards points involved. The mean simulated alignment index for a 5×6 comparison, with 30 cells, is 0.805 (Table 1). If a two-tailed test was used, at the .05 alpha level, the researcher would look to the 0.025 and 0.975 quantiles. For a 5×6 table, with 30 cells and 30 standards points, these criterion values are 0.737 and 0.867, respectively (Table 3). The observed alignment value is well below 0.737 (the 0.025 quantile). Therefore, the alignment is significantly lower than would be expected by chance at the .05 level.

Alignment in Previous Studies

Several previous studies have used Porter's alignment index to examine alignment among curriculum, standardized tests, and teacher instruction in mathematics and science. A reexamination of the results from those studies was conducted to further demonstrate the effectiveness of critical values for the alignment index in determining strength of alignment.

INSERT TABLE 7 ABOUT HERE

First, consider the results presented by Porter (2002). Porter applied the SEC mathematics coding, which used six content areas and five cognitive levels. Table 7 presents an adaptation of Porter's results on alignment between state assessments and mathematics standards documents for each state and the National Council of Teachers of Mathematics (NCTM). The diagonal represents the degree of alignment between each state and its own standards. As can be seen, all of the alignment indices presented are significantly lower than might be expected by chance (ranging from

0.30 to 0.47, with a critical value of 0.737 for 30 standards points[1]). Therefore, one would conclude that alignment among assessment and standards is very low for these states.

Liu *et al.* (2008) used a coding structure with 5 content categories and 6 cognitive levels to compare alignment of physics curriculum and assessments for China, Singapore, and New York State. China and Singapore had alignments of 0.67, which are significantly lower than the mean at the .05 level (below the critical value of 0.737), whereas New York's alignment index of 0.80 was equivalent to the mean (between the lower and upper critical values). Moreover, Liang and Yuan (2008) used a coding structure based on the Chinese teachers' guide, with 19 content distinctions and 6 cognitive levels. The observed alignment for China was 0.41, which was still significantly lower than would be expected by chance (lower than the critical value of 0.446; not shown in Table 2).

Note that alignment can vary over time. Liu and Fulmer (2008) examined the alignment between the New York State standards and state tests in high school level physics and chemistry over a two year period. While the physics tests retained higher validity (around 0.80), the alignment of the chemistry test varied more over time: 0.597 in January 2004; 0.713 in January 2006; and 0.692 in June 2006. Additionally, the alignment indices of the physics tests were equivalent to the mean, but the chemistry tests had alignments significantly lower than the mean. The variations over time support the need for ongoing analyses of alignment with revisions to high-stakes tests, as well as for comparisons across content areas.

The alignment of instruction has also been a focus of study. Porter, Smithson, Blank, and Zeidner (2007) presented longitudinal results from an intervention to increase alignment between teachers' instruction and their state's standards. Table 8 presents indices of alignment between instruction and the tests and standards for the treatment and control groups of teachers. All of the

[1] Note that higher numbers of standards points increase the critical value, so results below the critical value at 30 standards points would also be significant if there were more standards points.

reported alignment indices are significantly lower than would be expected by chance. Therefore, teachers' instruction was poorly aligned with their respective state standards and tests.

INSERT TABLE 8 ABOUT HERE

While the instruction-assessment and instruction-standards alignment indices were low, observed alignment among teachers' instructional practices has shown relatively higher results. Porter (2002) showed that teachers' instructional practices across jurisdictions had alignment indices ranging from 0.56 to 0.84 (Table 9). In general, these alignment indices are higher than found in Tables 7 and 8, suggesting that teachers' instruction is better aligned across jurisdictions, compared to the alignment of instruction with tests or standards, or even between the states' tests and the standards they purport to measure. Furthermore, some indices in Table 9 were statistically equivalent to the mean (though none were above the mean). This comparison assumes that the reported alignment indices follow the same table dimensions (5×6, or 30 cells); some table sizes for coding were not reported.

INSERT TABLE 9 ABOUT HERE

Discussion

Previous research has explored alignment among standards documents, assessments, and instruction using Porter's (2002) index. However, there existed no objective criteria for evaluating the strength of alignment. This was due in part to the categorical nature underlying the measure of alignment, as well as the dependence of alignment indices on the dimensions of the coding rubric used. The purpose of the present article was to fill this void in the literature by identifying mean and critical values for the alignment index, and to reexamine observed alignment values from previous research using these criteria. The results provide researchers and policymakers the first opportunity to compare if observed alignment indices differ significantly from what could occur by chance.

In so doing, the present report demonstrates how researchers can determine whether observed alignments are "high" or "low." As was shown with the comparison of alignments among Chinese

and New York State physics (Liu *et al*., 2008), some observed alignments were clearly lower than the mean at the .05 level, whereas others were significantly above the mean. Additionally, Liu and Fulmer (2008) showed that alignment between a test and the relevant curriculum may change with different testing instances, with some test forms being better aligned than others. By determining alignment for a particular test, the critical values allow researchers and policymakers to move beyond the qualitative recognition that alignments differ across jurisdictions or over time. Rather, policymakers and school administrators can consider whether those differences are meaningful in the context of meeting goals for satisfactory degree of alignment with learning goals.

In addition, the reexamination of prior studies reinforces earlier findings that alignment of instruction across jurisdictions was much stronger than alignment of instruction with standards or tests (Porter, 2002; Porter, Smithson, Blank, & Zeidner, 2007). While previous study had identified that the alignments differed, this study's critical values indicate that the instruction-instruction alignment was statistically significantly higher. This lends further weight to studies on why teachers' instruction aligns better across states, compared to the alignment of the teachers' instruction with the standards or tests of their own state. This may suggest that state or local educational policies and teacher education and professional development programs have relatively lower influence on instructional practices than culturally ubiquitous images of classroom instruction held by teachers (e.g., Calderhead & Robson, 1991). Further research on this is warranted.

It is important to note that the present article has focused on determining critical values for assessing strength of alignment between two documents. In previous work, the tables of interest have often been content standards and standardized tests. But Porter's alignment index measures the agreement between two documents without regard to the adequacy of either source. While it is tempting to argue that higher alignment is always superior, that decision must also be informed by critical evaluation of the qualities of both documents. If researchers and policymakers determine that a set of standards does not adequately represent desired student learning outcomes or teacher

instructional practices—whether in content, cognitive demand, or other areas of interest to stakeholders—then it is possible that low alignment would be acceptable or even preferable. Demands for greater alignment are only of value if at least one of the items being compared is considered a valid point for reference.

Another issue that this study uncovers is a need for conventions for reporting alignment indices and their calculation. As the present study has shown, the average alignment index that may occur by chance is dependent on the size of the frequency tables being compared and the number of test items or standards statements involved in the comparison. Therefore, any effort to gauge the strength of alignment is affected by the scoring rubric that is used to code the test items or other document. Some of the prior research reviewed for this study did not report the alignment coding systems used, or described only the coding scheme used for one content area (such as reading) but left out the coding system used for other subject areas. In such instances, it would not be possible for readers to judge the strength of alignment reported. Future studies in alignment should provide sufficient information about the alignment coding process they adopt. At a minimum, the dimensions of the tables compared should be reported. It would be even more beneficial to the field for researchers to present both the categories used and frequency tables of the coding.

Despite its contributions to the literature on alignment, this study is not without limitations. Readers will note that the critical values have normal distributions with relatively small ranges. The normality of the distribution is an expected outcome under the Central Limit Theorem, but the narrow range is not. The narrow range may be an artifact of the restricted range of the Porter alignment index, which has boundary values of 0 and 1. Recall that the normal distribution with mean of 0 and standard deviation of 1 has no minimum or maximum value; its tails are asymptotic to zero height, but stretch to positive and negative infinity along the number line. Because the distribution of the alignment indices is limited to such a narrow section of the number line, the shape

of the curve may also be narrower around the mean value than might be expected if the boundaries of the range were further apart.

Another explanation for the limited range is that the simulated values were not sufficiently random. This could lead to tight clustering of values. While the random number generator (RNG) in R has been deemed superior to other common software packages (McCullough, 2008), further effort in this area is worthwhile to replicate these findings. Future research could employ a RNG in another software package or the application of a non-machine RNG, such as that produced from atmospheric noise. Another valuable addition to the literature would be to explore closed-form solutions for the distribution of Porter alignment indices, which would eliminate the need for tables of critical values.

Most importantly, the present study is but a first step in understanding and comparing indices of alignments. Future research should expand on the present work to explore critical values in differences between observed alignments. That is, the present study identified the distributions of alignment indices under a variety of conditions, but did not examine whether differences between observed alignments were statistically significant. This is an open area for study that will help researchers and policymakers understand whether marginal changes in alignment across jurisdictions or years are significant. The present study lays a foundation for that future work.

References

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Bhola, D. S., Impara, J. C., & Buckendahl, C. W. (2003). Aligning tests with states' content standards: Methods and issues. *Educational Measurement: Issues and Practice* (fall), 21-29.Collins, A. (1998).

Blank, R. K. (2002). *Models for alignment analysis and assistance to states*. Council of Chief State School Officers Summary Document. Washington, DC: Council of Chief State School Officers.

Blank, R. K., Smithson, J., Porter, A., Nunnaley, D., & Osthoff, E. (2006). Improving instruction through schoolwide professional development: Effects of the data-on-enacted-curriculum model. *ERS Spectrum*, 6 (1), 9-22.

Calderhead, J. & Robson, M. (1991). Images of teaching: Student teachers' early conceptions of classroom practice. *Teaching and Teacher Education*, 7 (1), 1-8.

Florida Department of Education. (2005). *Next Generation Sunshine State Standards*. Retrieved from www.floridastandards.org.

Glod, M. (2009). 46 States, D.C. plan to draft common education standards. *Washington Post*. [Accessed online June 1, 2009 from http://www.washingtonpost.com/wp-dyn/content/article/2009/05/31/AR2009053102339.html]

Ihaka, R. & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, *5* (3), 299-314.

Joint Study Group on Academic Yearly Progress. (2002). Making Valid and Reliable Decisions in Determining Adequate Yearly Progress. Washington, DC: Council of Chief State School Officers.

Lemke, R. J., Hoerandner, C. M., & McMahon, R. E. (2006). Student assessments, non-test-takers, and school accountability. *Education Economics, 6*(14), 235-250.

Liang, L. L., & Yuan, H. (2008). Examining the alignment of Chinese national physics curriculum guidelines and 12th-grade exit examinations: A case study. *International Journal of Science Education*, 30 (13), 1823-1825.

Linn, R. L. (2008). Educational accountability systems. In K. E. Ryan & L. A. Shepard (Eds.), *The future of test-based educational accountability* (pp. 3-24), New York, NY: Routledge.

Linn, R. L., Baker, E., L., and Betebenner, D. W. (2002). Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher*, 31 (6), 3–16.

Liu, X. & Fulmer, G. W. (2008). Alignment between science curriculum and assessments in selected New York State Regents exams. *Journal of Science Education and Technology*, *17*(4), 373-383.

Liu, X., Zhang, B. H., Liang, L. L., Fulmer, G. W., Kim, B., and Yuan, H.Q. (2008). Alignment between the physics content standards and standardized tests: A comparison among US-NY, Singapore, and China-Jiangsu. *Science Education*.

McCullough, B. D. (2008). Microsoft Excel's 'not the Wichmann–Hill' random number generators. *Computational Statistics and Data Analysis*, 52(10), pp. 4587–4593.

National Assessment Governing Board. (2009). *Mathematics Framework for the 2009 National Assessment of Educational Progress*. Washington, DC: US Department of Education.

No Child Left Behind Act of 2001. Public Law Number 107-110 (2001).

Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, 31(7), 3-14.

Porter, A. C., Smithson, J. Blank, R. K., & Zeidner, T. (2007). Alignment as a teacher variable. *Applied Measurement in Education*, 20(1), 27-51.

R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0 [URL http://www.R-project.org].

Ross, S. M. (2004). *Introduction to probability and statistics for engineers and scientists* (3rd ed). Burlington, MA: Elsevier Academic Press.

Rothman, R. (2003). *Imperfect matches: The alignment of standards and tests*. Washington, DC: National Research Council.

Rothman, R., Slattery, J. B., Vranek, J. L., and Resnick, L. B. (2002). *Benchmarking and alignment of standards and testing*. Los Angeles, CA: Center for the Study of Evaluation.

Shavelson, R. J. (1996). *Statistical Reasoning for the Behavioral Sciences*. 3rd ed. Needham Heights, MA: Allyn & Bacon.

Tomsho, R. (2009). U.S. ties new funds to schools data. *Wall Street Journal*. [Accessed April 2, 2009 from http://online.wsj.com/article/SB123860983393679075.html]

University of the State of New York (1996). *Learning standards for mathematics, science, and technology*. Albany, NY: The State Education Department.

Webb, N. L. (2007). Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education*, 20(1), 7-25.

Woolard, J. C. (2007). *Measuring systemic alignment of a state's instruction, standards, and assessments: A baseline analysis*. Paper presented at the annual meeting of the American Educational Research Association; Chicago, IL.

Figures

Figure 1

Frequency tables

Table A

5	5
3	7

Table B

6	4
5	5

Ratio tables

Table *a*

0.25	0.25
0.15	0.35

Table *b*

0.3	0.2
0.25	0.25

Absolute Discrepancies

$|a_{jk} - b_{jk}|$

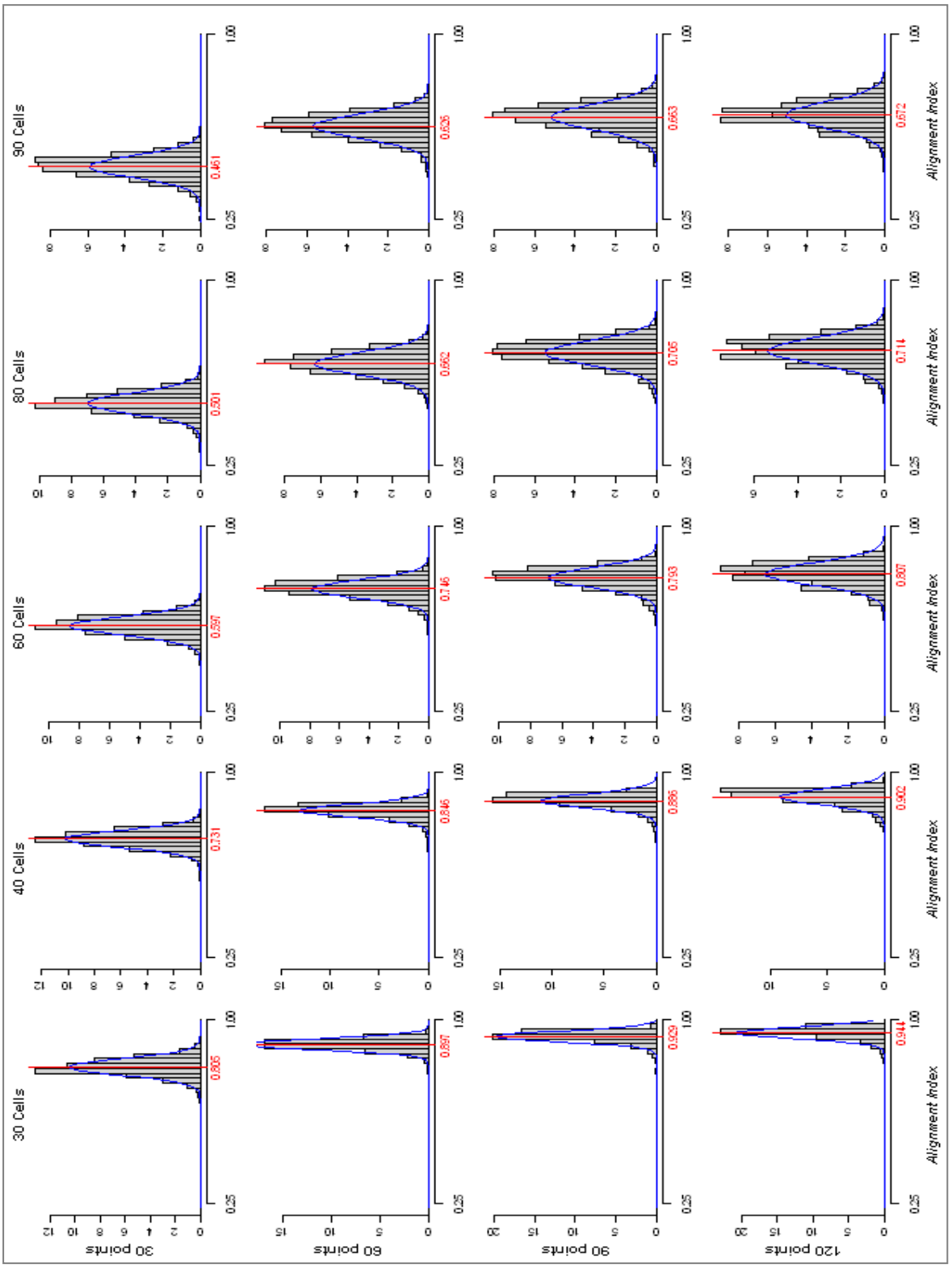
0.05	0.05
0.1	0.1

Alignment

$$1 - \frac{\sum_{k=1}^{K}\sum_{j=1}^{J}\left|a_{jk} - b_{jk}\right|}{2}$$

0.85

Figure 2


30 Cells
40 Cells
60 Cells
80 Cells
90 Cells
30 points
60 points
90 points
120 points
Alignment Index
0.805
0.897
0.929
0.944
0.731
0.846
0.886
0.902
0.597
0.746
0.793
0.807
0.501
0.662
0.705
0.714
0.461
0.626
0.663
0.672

Captions

Figure 1. Example calculation of the Porter alignment index for a pair of 2x2 tables.

Figure 2. Histograms with overlay curves demonstrating distributions of alignment indices from five selected table sizes.

Tables

Table 1. Mean alignment indices by number of cells and number of standards points

Cells	Standards Points			
	30	60	90	120
10	0.9464	0.9782	0.9875	0.9916
12	0.9297	0.9720	0.9833	0.9890
14	0.9172	0.9643	0.9792	0.9859
15	0.9117	0.9615	0.9771	0.9842
16	0.9024	0.9572	0.9743	0.9821
18	0.8842	0.9494	0.9695	0.9778
20	0.8674	0.9428	0.9635	0.9737
21	0.8596	0.9370	0.9606	0.9709
24	0.8394	0.9231	0.9504	0.9633
25	0.8335	0.9188	0.9469	0.9602
27	0.8217	0.9098	0.9400	0.9534
28	0.8168	0.9060	0.9367	0.9507
30	0.8054	0.8974	0.9291	0.9438
32	0.7935	0.8867	0.9212	0.9366
35	0.7709	0.8712	0.9077	0.9245
36	0.7627	0.8669	0.9033	0.9195
40	0.7314	0.8462	0.8860	0.9021
42	0.7161	0.8350	0.8761	0.8933
45	0.6941	0.8203	0.8621	0.8790
48	0.6721	0.8051	0.8474	0.8654
49	0.6648	0.7999	0.8431	0.8605
50	0.6587	0.7958	0.8384	0.8553
54	0.6321	0.7752	0.8194	0.8366
56	0.6204	0.7653	0.8106	0.8256
60	0.5968	0.7465	0.7925	0.8072
63	0.5802	0.7326	0.7797	0.7922
64	0.5747	0.7290	0.7730	0.7869
70	0.5451	0.7039	0.7478	0.7600
72	0.5351	0.6946	0.7369	0.7497
80	0.5006	0.6615	0.7053	0.7137
81	0.4951	0.6600	0.6987	0.7107
90	0.4614	0.6258	0.6633	0.6716
100	0.4309	0.5908	0.6276	0.6337
114	0.3897	0.5506	0.5820	0.5845

Table 2. Standard deviations of alignment indices by number of cells and number of standards points

Cells	Standards Points			
	30	60	90	120
10	0.0041	0.0024	0.0016	0.0010
12	0.0072	0.0028	0.0020	0.0011
14	0.0068	0.0036	0.0019	0.0013
15	0.0111	0.0050	0.0016	0.0014
16	0.0102	0.0035	0.0022	0.0015
18	0.0104	0.0044	0.0028	0.0023
20	0.0100	0.0062	0.0031	0.0037
21	0.0088	0.0056	0.0033	0.0038
24	0.0112	0.0060	0.0047	0.0048
25	0.0108	0.0059	0.0054	0.0046
27	0.0145	0.0076	0.0061	0.0071
28	0.0150	0.0083	0.0069	0.0074
30	0.0149	0.0100	0.0086	0.0085
32	0.0150	0.0103	0.0088	0.0091
35	0.0159	0.0102	0.0104	0.0110
36	0.0144	0.0107	0.0109	0.0119
40	0.0168	0.0111	0.0120	0.0140
42	0.0155	0.0113	0.0124	0.0151
45	0.0148	0.0122	0.0136	0.0162
48	0.0161	0.0127	0.0152	0.0178
49	0.0152	0.0135	0.0149	0.0183
50	0.0166	0.0132	0.0155	0.0186
54	0.0166	0.0143	0.0166	0.0198
56	0.0168	0.0153	0.0176	0.0213
60	0.0179	0.0165	0.0179	0.0214
63	0.0183	0.0169	0.0191	0.0220
64	0.0189	0.0178	0.0190	0.0239
70	0.0192	0.0187	0.0211	0.0233
72	0.0186	0.0190	0.0225	0.0229
80	0.0195	0.0215	0.0223	0.0250
81	0.0211	0.0215	0.0238	0.0250
90	0.0204	0.0231	0.0241	0.0250
100	0.0216	0.0241	0.0251	0.0250
114	0.0221	0.0263	0.0272	0.0250

Table 3. Reference values for indices of alignment, by number of cells – 30 standards points

Cells	Quantile					
	0.025	0.050	0.100	0.900	0.950	0.975
10	0.9167	0.9250	0.9250	0.9667	0.9667	0.9750
12	0.9000	0.9048	0.9095	0.9476	0.9524	0.9571
14	0.8803	0.8894	0.8924	0.9409	0.9439	0.9515
15	0.8667	0.8778	0.8889	0.9333	0.9444	0.9444
16	0.8609	0.8710	0.8754	0.9275	0.9362	0.9377
18	0.8417	0.8500	0.8583	0.9083	0.9125	0.9208
20	0.8267	0.8333	0.8400	0.8933	0.9000	0.9067
21	0.8143	0.8227	0.8333	0.8863	0.8903	0.8961
24	0.7867	0.7963	0.8074	0.8704	0.8778	0.8849
25	0.7788	0.7879	0.8000	0.8636	0.8727	0.8788
27	0.7616	0.7741	0.7852	0.8561	0.8667	0.8719
28	0.7540	0.7667	0.7788	0.8544	0.8644	0.8726
30	0.7372	0.7500	0.7643	0.8500	0.8565	0.8667
32	0.7258	0.7357	0.7500	0.8361	0.8500	0.8548
35	0.7000	0.7143	0.7258	0.8125	0.8254	0.8361
36	0.6909	0.7077	0.7213	0.8033	0.8154	0.8254
40	0.6632	0.6754	0.6885	0.7727	0.7846	0.7941
42	0.6462	0.6575	0.6719	0.7576	0.7692	0.7794
45	0.6269	0.6393	0.6515	0.7353	0.7463	0.7571
48	0.6000	0.6143	0.6286	0.7143	0.7260	0.7361
49	0.5909	0.6056	0.6197	0.7083	0.7206	0.7286
50	0.5857	0.6000	0.6143	0.7027	0.7123	0.7231
54	0.5588	0.5714	0.5857	0.6761	0.6892	0.6986
56	0.5441	0.5571	0.5735	0.6667	0.6770	0.6892
60	0.5211	0.5342	0.5479	0.6438	0.6571	0.6667
63	0.5000	0.5143	0.5303	0.6267	0.6389	0.6486
64	0.4932	0.5070	0.5231	0.6234	0.6351	0.6456
70	0.4627	0.4783	0.4932	0.5949	0.6071	0.6173
72	0.4507	0.4658	0.4805	0.5844	0.6000	0.6118
80	0.4143	0.4304	0.4459	0.5529	0.5663	0.5783
81	0.4054	0.4211	0.4390	0.5476	0.5625	0.5732
90	0.3684	0.3836	0.4026	0.5176	0.5316	0.5429
100	0.3333	0.3544	0.3718	0.4881	0.5000	0.5125
114	0.2927	0.3077	0.3289	0.4494	0.4639	0.4773

Table 4. Reference values for indices of alignment, by number of cells – 60 standards points

Cells	Quantile					
	0.025	0.050	0.100	0.900	0.950	0.975
10	0.9667	0.9690	0.9714	0.9857	0.9881	0.9881
12	0.9583	0.9611	0.9639	0.9806	0.9833	0.9833
14	0.9486	0.9518	0.9550	0.9739	0.9757	0.9770
15	0.9433	0.9467	0.9500	0.9733	0.9733	0.9767
16	0.9377	0.9430	0.9474	0.9675	0.9702	0.9719
18	0.9230	0.9321	0.9372	0.9615	0.9641	0.9667
20	0.9100	0.9200	0.9288	0.9583	0.9610	0.9625
21	0.9000	0.9114	0.9205	0.9531	0.9568	0.9580
24	0.8842	0.8952	0.9043	0.9405	0.9429	0.9460
25	0.8783	0.8892	0.9000	0.9363	0.9402	0.9431
27	0.8667	0.8785	0.8876	0.9295	0.9339	0.9379
28	0.8610	0.8708	0.8821	0.9276	0.9326	0.9371
30	0.8481	0.8584	0.8696	0.9222	0.9278	0.9333
32	0.8339	0.8451	0.8571	0.9116	0.9174	0.9225
35	0.8126	0.8257	0.8382	0.8978	0.9037	0.9088
36	0.8062	0.8214	0.8345	0.8942	0.9000	0.9048
40	0.7833	0.7959	0.8115	0.8754	0.8812	0.8875
42	0.7683	0.7822	0.7988	0.8649	0.8724	0.8779
45	0.7516	0.7674	0.7833	0.8521	0.8595	0.8644
48	0.7326	0.7500	0.7667	0.8384	0.8467	0.8533
49	0.7256	0.7429	0.7611	0.8361	0.8447	0.8500
50	0.7225	0.7412	0.7577	0.8313	0.8394	0.8455
54	0.6977	0.7161	0.7337	0.8133	0.8218	0.8311
56	0.6857	0.7053	0.7222	0.8059	0.8141	0.8222
60	0.6626	0.6812	0.7000	0.7902	0.8006	0.8090
63	0.6454	0.6666	0.6848	0.7778	0.7880	0.7972
64	0.6389	0.6611	0.6800	0.7757	0.7870	0.7963
70	0.6137	0.6333	0.6521	0.7548	0.7657	0.7758
72	0.6034	0.6232	0.6412	0.7451	0.7564	0.7680
80	0.5637	0.5828	0.6016	0.7190	0.7318	0.7439
81	0.5641	0.5817	0.6007	0.7167	0.7302	0.7405
90	0.5258	0.5430	0.5623	0.6859	0.7009	0.7137
100	0.4833	0.5021	0.5230	0.6549	0.6722	0.6859
114	0.4459	0.4621	0.4812	0.6167	0.6333	0.6483

Table 5. Reference values for indices of alignment, by number of cells – 90 standards points

Cells	Quantile					
	0.025	0.050	0.100	0.900	0.950	0.975
10	0.9811	0.9822	0.9833	0.9911	0.9922	0.9933
12	0.9758	0.9771	0.9784	0.9882	0.9889	0.9902
14	0.9703	0.9722	0.9739	0.9844	0.9859	0.9872
15	0.9667	0.9698	0.9714	0.9841	0.9841	0.9857
16	0.9610	0.9654	0.9681	0.9809	0.9826	0.9839
18	0.9372	0.9537	0.9630	0.9778	0.9796	0.9815
20	0.9302	0.9404	0.9535	0.9727	0.9747	0.9758
21	0.9226	0.9349	0.9486	0.9712	0.9733	0.9748
24	0.9079	0.9190	0.9315	0.9643	0.9667	0.9684
25	0.9048	0.9167	0.9273	0.9614	0.9633	0.9657
27	0.8908	0.9051	0.9179	0.9564	0.9590	0.9615
28	0.8876	0.8988	0.9132	0.9544	0.9576	0.9598
30	0.8748	0.8876	0.9013	0.9500	0.9528	0.9570
32	0.8648	0.8788	0.8929	0.9441	0.9475	0.9505
35	0.8444	0.8602	0.8758	0.9333	0.9369	0.9400
36	0.8390	0.8535	0.8698	0.9302	0.9347	0.9375
40	0.8111	0.8328	0.8496	0.9159	0.9214	0.9250
42	0.8000	0.8192	0.8371	0.9078	0.9134	0.9180
45	0.7852	0.8018	0.8194	0.8979	0.9044	0.9103
48	0.7556	0.7778	0.7990	0.8857	0.8930	0.8998
49	0.7556	0.7768	0.7959	0.8824	0.8906	0.8979
50	0.7537	0.7696	0.7894	0.8786	0.8869	0.8931
54	0.7222	0.7444	0.7667	0.8630	0.8726	0.8804
56	0.7111	0.7333	0.7556	0.8556	0.8648	0.8725
60	0.7000	0.7202	0.7393	0.8393	0.8507	0.8586
63	0.6778	0.7000	0.7222	0.8299	0.8407	0.8498
64	0.6778	0.6889	0.7159	0.8237	0.8359	0.8459
70	0.6444	0.6657	0.6879	0.8026	0.8161	0.8249
72	0.6333	0.6554	0.6735	0.7940	0.8066	0.8158
80	0.6000	0.6218	0.6410	0.7651	0.7783	0.7889
81	0.5889	0.6111	0.6333	0.7589	0.7707	0.7811
90	0.5556	0.5773	0.5889	0.7273	0.7421	0.7556
100	0.5111	0.5333	0.5556	0.6930	0.7108	0.7224
114	0.4667	0.4889	0.5111	0.6521	0.6692	0.6847

Table 6. Reference values for indices of alignment, by number of cells – 120 standards points

Cells	Quantile					
	0.025	0.050	0.100	0.900	0.950	0.975
10	0.9872	0.9878	0.9891	0.9942	0.9949	0.9955
12	0.9841	0.9848	0.9856	0.9924	0.9932	0.9939
14	0.9800	0.9813	0.9826	0.9897	0.9905	0.9913
15	0.9778	0.9787	0.9806	0.9889	0.9898	0.9907
16	0.9664	0.9765	0.9789	0.9873	0.9882	0.9892
18	0.9404	0.9583	0.9743	0.9844	0.9859	0.9866
20	0.9333	0.9417	0.9620	0.9821	0.9833	0.9845
21	0.9311	0.9410	0.9556	0.9803	0.9816	0.9828
24	0.9167	0.9306	0.9407	0.9764	0.9778	0.9792
25	0.9146	0.9250	0.9377	0.9741	0.9759	0.9770
27	0.8994	0.9139	0.9287	0.9699	0.9721	0.9736
28	0.8917	0.9083	0.9245	0.9682	0.9703	0.9718
30	0.8833	0.8990	0.9141	0.9650	0.9683	0.9700
32	0.8667	0.8892	0.9051	0.9596	0.9623	0.9640
35	0.8500	0.8720	0.8891	0.9513	0.9548	0.9575
36	0.8417	0.8583	0.8818	0.9484	0.9519	0.9549
40	0.8167	0.8417	0.8569	0.9360	0.9409	0.9456
42	0.8083	0.8250	0.8481	0.9296	0.9356	0.9400
45	0.7917	0.8083	0.8250	0.9185	0.9256	0.9308
48	0.7750	0.7917	0.8147	0.9081	0.9167	0.9214
49	0.7667	0.7833	0.8080	0.9057	0.9140	0.9193
50	0.7583	0.7750	0.8000	0.9019	0.9091	0.9148
54	0.7333	0.7583	0.7750	0.8861	0.8951	0.9021
56	0.7250	0.7417	0.7667	0.8776	0.8881	0.8958
60	0.7083	0.7250	0.7495	0.8616	0.8720	0.8801
63	0.6917	0.7083	0.7333	0.8471	0.8594	0.8672
64	0.6833	0.7000	0.7250	0.8444	0.8562	0.8651
70	0.6581	0.6750	0.6917	0.8201	0.8329	0.8441
72	0.6417	0.6583	0.6833	0.8116	0.8242	0.8342
80	0.6000	0.6250	0.6417	0.7788	0.7931	0.8050
81	0.6000	0.6167	0.6417	0.7750	0.7891	0.8042
90	0.5667	0.5833	0.6000	0.7391	0.7551	0.7683
100	0.5250	0.5500	0.5667	0.7000	0.7166	0.7312
114	0.4833	0.5000	0.5167	0.6500	0.6728	0.6877

Table 7. Alignment of Assessments with Standards

Standard	Assessment			
State	B	D	E	F
B	0.37*	0.39*	0.37*	0.45*
D	0.35*	0.37*	0.36*	0.40*
E	0.36*	0.33*	0.43*	0.31*
F	0.32*	0.35*	0.30*	0.41*
NCTM	0.34*	0.40*	0.33*	0.47*

Note. Adapted from Porter (p. 6), Seventh-Grade Math – Goals 2000 Study.

* Alignment is significantly different from respective mean at the .05 level using a two-tailed test, assuming a 5×6 coding rubric (i.e., 30 cells) and 30 standards points.

Table 8. Mean alignment of instruction with tests and standards over a longitudinal study

	Math Test		Math Standard		Science Test		Science Standard	
	TRT	CTL	TRT	CTL	TRT	CTL	TRT	CTL
Year1	0.204*	0.207*	0.302*	0.304*	0.156*	0.146*	0.180*	0.168*
Year3	0.189*	0.191*	0.328*	0.309*	0.170*	0.165*	0.197*	0.185*

Note. TRT= Treatment group; CTL=Control group. Table adapted from Porter, Smithson, Blank, and Zeidner (2007, p. 42).

* Alignment is significantly different from respective mean at the .05 level using a two-tailed test. Tests and standards are assumed to follow a 5×6 coding rubric in mathematics (i.e., 30 cells) and a 5×8 coding rubric in science (i.e., 40 cells), and 30 standards points.

Table 9. Alignments of Instruction with Instruction

State	*H*	*J*	*K*	*L*	*E*	*O*	*G*	*I*	*M*
H									
J	0.73*								
K	0.59*	0.66*							
L	0.56*	0.64*	0.67*						
E	0.65*	0.71*	0.78	0.70*					
O	0.71*	0.80	0.63*	0.65*	0.70*				
G	0.71*	0.81	0.66*	0.67*	0.71*	0.84			
I	0.73*	0.82	0.63*	0.66*	0.68*	0.79	0.80		
M	0.68*	0.77	0.61*	0.62*	0.66*	0.73*	0.76	0.79	
N	0.62*	0.69*	0.58*	0.61*	0.62*	0.71*	0.70*	0.67*	0.65*

Note. Average alignment = 0.69*. Table adapted from Porter (2002, p. 7), Eighth-Grade Mathematics from the State Collaborative on Assessment and Student Standards (SCASS) Study.

* Alignment is significantly different from respective mean at the .05 level using a two-tailed test, assuming a 5×6 coding rubric (i.e., 30 cells) and 30 standards points.