Educational Policy

Prelude to the Common Core: Internationally Benchmarking a State's Math Standards
J. Christopher Woolard

The online version of this article can be found at:

Published by:

$SAGE

http://www.sagepublications.com

On behalf of:

Politics of Education Association

Additional services and information for *Educational Policy* can be found at:

Email Alerts: http://epx.sagepub.com/cgi/alerts

Subscriptions: http://epx.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations: http://epx.sagepub.com/content/27/4/615.refs.html

>> Version of Record - Jun 3, 2013

OnlineFirst Version of Record - Apr 24, 2012

What is This?

Prelude to the Common Core: Internationally Benchmarking a State's Math Standards

J. Christopher Woolard[1]

Abstract

As states struggle with the notion of international competitiveness, the quality and rigor of academic content standards has come into question. While Ohio's content standards are well regarded, the state initiated a process to revise the standards and eventually joined with the majority of states in adopting a voluntary set of national standards—the Common Core. This study uses the Surveys of Enacted Curriculum (SEC) methodology to examine Ohio's current math content standards in comparison to TIMSS, PISA, high performing international counterparts, and the recently released Common Core. Specifically, it examines whether the state's standards are "a mile wide and inch deep." Second, this study analyzes whether high performing countries' standards are more aligned to Trends in International Mathematics and Science Study (TIMSS) and Program for International Student Assessment (PISA) through the SEC lens of topic and cognitive expectations. Ohio's standards generally are less focused than the international comparisons, not very aligned to TIMSS and PISA, and have lower cognitive expectations. The CCSS have greatly increased that focus by reducing the number of topics in the analyzed grade levels while increasing the levels of cognitive expectations. These results provide a baseline for comparison to the full implementation of the Common Core. Once fully implemented, policy makers will have a reference point for evaluation of policy goals.

[1]Ohio Department of Education, Columbus, OH, USA

Corresponding Author:
J. Christopher Woolard, Ohio Department of Education, 25 S. Front St.,
Columbus, OH 43215, USA.
Email: christopher.woolard@ode.state.oh.us

Keywords

international benchmarking, content standards, standards-based reform, common core

Introduction

A standards-based educational system relies on coordination and alignment between three elements: standards, assessments, and instruction. Each of these three pillars is vital to insure that students receive a rigorous and high quality education. Teachers have specific guidance for expectations. Students receive instruction based on those expectations. Assessments measure against those expectations. States by and large have been developing their own standards and assessment systems. While these reforms have taken root, students in high performing international education systems have frequently outperformed their American counterparts.

The 2010 *EdWeek* Quality Counts report graded Ohio as an "A" for standards, assessments and accountability, which was third highest nationally (2010). Nonetheless, the state recently began development of revised standards. Concurrently, the Obama administration made education reform a key priority and focused extensive resources toward states—including the multibillion dollar Race to the Top competitive grant initiative. Rigorous, common, and internationally competitive expectations are a key component of Race to the Top and accordingly, the vast majority of states signed on to implement a voluntary set of national standards—the Common Core. Ohio was one of those states.

A report from the National Governor's Association (NGA), the Council of Chief State School Officers (CCSSO), and Achieve Inc. (2008) emphasized that governments are eager to compare themselves to high performers around the globe. Likewise, this study was designed to produce actionable research for policy makers by providing a baseline comparison of Ohio's system of standards to high performing international counterparts, especially in relation to the Trends in International Mathematics and Science Study (TIMSS) and Program for International Student Assessment (PISA). This comparison attempts to provide explanations for performance on international assessments by focusing on "potential opportunities to learn." This is accomplished by examining the alignment of standards to the international assessments and exploring the mile wide, inch deep phenomenon. State standards tend to include extensive amounts of topics making it difficult to cover any topic in great depth. This emphasis on quantity, not quality, often lacks focus and coherence from

grade to grade (Schmidt, Wang, & McKnight, 2005). In addition, this study will examine relevant grade levels of the Common Core to determine whether the new standards reflect an improvement in focus and depth of expectations.

The two international assessments are high profile efforts to compare students across countries. American students, including Ohioans, are frequently outperformed by their international counterparts. For example, the TIMSS 2007 results show that fourth-grade students in the United States performed higher than average but below several Asian and European countries. The results were similar in eighth grade—above average, but trailing high performing peers (Gonzales et al., 2008; Mullis, Martin, & Foy, 2008). The 2006 PISA mathematics results show that American students were outperformed by high achievers such as Finland, Korea, and Hong Kong (Organisation for Economic Co-operation and Development [OECD], 2007). Similar themes were evident in the 2009 PISA, that is, American students were above average but not at the level of the highest performers (OECD, 2010). These latter results were referred to by U.S. Education Secretary Arne Duncan as "an absolute wake-up call for America" (Armario, 2010, para. 4).

Clearly, U.S. education policy makers are paying attention to results on these international assessments. Numerous authors have examined the average performance of U.S. students (see numerous recent examples such as Baines, 2008; Min-Hsiung, 2009; Richland, Zur, & Holyoak, 2007) in regard to performance, instruction, and curriculum in relation to higher performing countries. To help understand the implications at a state level, this study examines Ohio's math standards in comparison to TIMSS, PISA, high performing international counterparts, and the Common Core. The Surveys of Enacted Curriculum (SEC) methodology is employed to provide data on depth, breath, and cognitive expectations. Specifically, this study examines two legs of the standards-based system, standards and assessments. Clearly, a standards-based system relies on alignment between all three legs. While the SEC methodology provides a mechanism for collecting data on instruction, it is beyond the scope of this study. This research uses standards as a proxy for instruction in framing an analysis of opportunity to learn. In other words, this study considers standards as a representation of the content that is delivered to students in Ohio and the comparison countries. Schmidt, Bates, and Leroi (2009) emphasize that standards represent *potential* opportunities to learn. Future research should build on this theme by collecting data on the enacted curriculum to examine the degree of alignment between instruction and the standards.

The results find evidence to support the mile wide, inch deep phenomenon, and while the standards are considered to be high quality, there is room to improve. The current standards will remain in place for at least 2 school years

until the Common Core and related assessments are phased in. Initial analysis of the Common Core in this study provides evidence that the process was successful in creating more focused and rigorous standards for the grades analyzed. This analysis can be expanded on with full implementation of the Common Core and future updates/revisions. The SEC methodology provides a mechanism for in-depth analysis of standards, assessment, and instruction across states and countries. This study models how the SEC methodology can be used for future policy research.

International Benchmarking

In 2007, Achieve, Inc. conducted a sweeping review of the education system in Ohio and presented a series of recommendations in its report *Creating a World-Class Education System in Ohio*. The very first recommendation suggested that Ohio benchmark its standards against the highest performing countries in the world:

> Research indicates that the best systems in the world create a high challenge for their child that includes high standards and rigorous, equitable assessments. This will require Ohio to go beyond the strong progress in this area over the last 10 years by aligning K-12 standards with knowledge and skills needed for success in postsecondary education and the global economy and by benchmarking its standards against those of high-performing states and especially nations that compete with the United States. (p. 4)

This study takes up that challenge.

Comparing the United States to High-Performing Countries

Scholars and policy makers have an extensive interest in international benchmarking (Phillips, 2009; Phillips & Dossey, 2008; Schmidt, McKnight, & Raizen, 1997) as states realize that they are not just in competition with each other but with advancing societies in the global economy. This is not a revelation, but recent economic conditions have made it more clear and urgent.

The NGA/CCSSO/Achieve report defines international benchmarking as not just setting performance targets but "to indentify and learn from top performers and rapid improvers—from nations and states that offer ideas for boosting their own performance" (2008, p. 5). As states are trying to raise standards and improve instruction and performance, benchmarking efforts

can help make that process more effective. A series of recommendations includes raising standards by adopting a then newly phrased common core of standards, and measuring "state-level education performance globally" (p. 6).

Scholars have made comparisons on performance, standards, and practice for much of the last decade. Student performance data from the TIMSS and PISA provide a window into these educational settings. One common theme of international benchmarking research is that the United States does not always perform at the highest levels and may be falling further behind. If policy makers do not address these issues, American students may be stuck in the middle of the pack. The phrase that best describes much of the recent research explaining poor U.S. results on international assessments is the "mile wide, inch deep" curriculum problem. That is, American schools tend to cover too many topics, at insufficient depth. For example, Schmidt et al. (1997) found that countries with more focused curricula performed better on TIMSS than the United States, which not only covers too many topics but repeats many of them in successive years. Similar findings were echoed in subsequent studies (Schmidt et al., 2001, 2005) that conclude state content standards are unfocused compared to high performing countries. Higher achieving countries focus more intensely on fewer topics as well as on reasoning skills and application (Darling-Hammond & McCloskey, 2008). Porter, Politkoff, and Smithson (2009) note that, in general, standards tend to ignore big ideas in favor of "laundry lists of small topics" (p. 240), while noting that analysis of this trend in individual states is lacking.

Schmidt et al. (2005) examined the coherence of content standards and found that not only do states often teach more topics but they also allow topics to linger without a progression of depth. Higher performing countries demonstrate more curriculum coherence by introducing a topic, and then progressing the depth of instruction from simple to more complicated expectations. They describe different conceptions of coherent standards that includes alignment, and more specifically "[articulates] over time as a sequence of topics and performance consistent with the logical and, if appropriate, hierarchical nature of the disciplinary content" (p. 528). In other words, the laundry list not only includes more topics but often lacks a logical and sequential progression. They ask the question of how U.S. standards compare to the rest of the TIMSS world and posit that national standards may be the solution for addressing coherence and rigor. Schmidt and Prawat (2006) reexamine that question later in more depth and argue that it may not be the centralization of standards that is important, but the credibility that comes with it.

The recent large scale emphasis on standards-based reform sets the context for these benchmarking activities. Education policy makers have gone

from decades of input-based decisions, to an outcome-based system emphasizing student learning and achievement. This is a major paradigm shift that has changed the roles of states, districts, and education systems as a whole. For such a system to work, standards-based reform hypothesizes that states must create systems with a focus on alignment between what is taught, what is required, and what is tested. Research efforts have been made to understand the relationship between curriculum and performance. For example, Hook, Bishop, and Hook (2007) found that a quality curriculum, as modeled after and aligned with high performers, could be transplanted in local schools as an effort to overcome achievement concerns.

Alignment is an important tenet of standards-based reform and rests on the principle that what students learn can be explained by what is taught (Porter et al., 2009). Alignment is a key concept when analyzing student performance on tests. A Council of Chief State School Officers (CCSSO) report examining alignment contained six assumptions of alignment; number six is particularly critical for this analysis: "Valid and meaningful data-based decision-making depends on the degree of alignment between standards and assessments" (La Marca, Redfield, Winter, Bailey, & Despriet, 2000, p. iii). Alignment analysis is a critical step in interpreting student test performance. According to La Marca (2001),

> a post hoc review of alignment should be conducted. This step is important where standards-based custom assessments are used and absolutely essential when states choose to use assessment products not specifically designed to measure their state standards. (p. 4)

The related argument is that state standards should be compared to international assessments such as TIMSS and PISA. While complete alignment would not be expected, it could be reasonably argued that higher performing countries would likely have content standards that are more aligned to the international assessments. In other words, students should be expected to perform better on assessments when they have been taught the respective subject matter (Gamoran, Porter, Smithson, & White, 1997). For this analysis, the question is how aligned are standards to the international assessments that states and countries are evaluated on. This does not mean to imply that alignment is purposeful, but based on this assumption it is reasonable to expect better results with better alignment and worse results with lower alignment. Results on international assessments should at least be considered in the light of how aligned the system is to those assessments.

Porter's model of alignment indicators offers numerous advantages in quantitatively analyzing alignment of standards, assessments, and instruction. His model permits analysis of any combination of these three pillars. According to Porter (2002),

> An index of alignment also can be used as a descriptive variable in assessing the coherence of a state's or district's curriculum policy system. At the heart of systemic reform is the concept of alignment. When a system is aligned, all the messages from the policy environment are consistent with each other, content standards drive the system, and assessments, materials, and professional development are all tightly aligned to the content standards. (p. 11)

In a standards-based system, assessments are expected to be aligned to the curriculum. In fact, NCLB requires that state assessments must be aligned with state academic content standards and uses a peer review process to evaluate such evidence (U.S. Department of Education [DOE], 2007). The theoretical framework is that an aligned system of standards and assessments will provide valid measures of student achievement on said standards. The standards-based system relies on the notion of alignment and it is a basic principle that achievement is directly related to the alignment of standards to the assessments. States strive to, and are held accountable for, the alignment of their tests to their content standards. Thus, a valid research concern for international comparisons is just how aligned are a particular state's (or country's) standards to the international assessments on which they are being evaluated.

Contributions to the Literature

Schmidt et al. (2005) emphasize that it is important to ask whether "US content standards [are] similar to those of the rest of the 'TIMSS world'?" (p. 531). This study contributes to the research by specifically testing principles from the literature at a state level. It provides a detailed micro-level analysis to contribute to the macro-level literature. This state-level analysis provides insights to Schmidt et al.'s question from one specific state by conducting an in-depth analysis of its current standards and preliminary analysis of the Common Core. The state standards are still relevant as they will be in place at least 2 years until the Common Core is fully implemented. Once the state fully implements the Common Core, this analysis could be expanded to all grade

levels and incorporate teacher data on the enacted curriculum. Finally, this study models how the SEC methodology can be used for international benchmarking work. The SEC's unique multidimensional metrics and ability to standardize comparisons of standards, assessments, and instruction across states and countries hold great promise for robust international comparisons.

Research questions. Based on the foundational principles and questions in the literature, this study will conduct an international benchmarking analysis for a specific state. Accordingly, this research will address critical themes including the breadth and depth of content standards, and SEC alignment, which considers both topic and cognitive expectations. The two primary research questions are as follows:

> *Research Question 1*: Are Ohio's standards "a mile wide and an inch deep," especially as compared to high performing countries?
> *Research Question 2*: Are high performing countries' standards more aligned to TIMSS and PISA than Ohio?

The comparison countries. The comparison countries are based on performance on international assessments as well as guidance from the literature. For example, Darling-Hammond and McClosky (2008) conclude we can learn lessons from the education systems in Finland, Sweden, United Kingdom, Hong Kong, and Australia. Phillips' (2009) statistical linking study ranked Ohio's performance in fourth and eighth grade as a "C+" ahead of the national average but behind several countries including Hong Kong, Singapore, and Japan. One frequently cited classification system identifies "A+" countries based on TIMSS performance (Schmidt et al., 2005; Valverde & Schmidt, 2000) including Singapore, Japan, and Hong Kong. Based on the literature, test performance, and data availability; England, Hong Kong, Japan, Singapore, and Finland are included as high-performing comparisons for this study.

Method and Data

Surveys of Enacted Curriculum (SEC) Method

The Surveys of Enacted Curriculum (SEC) is a research-based tool that provides data on the alignment between the enacted curriculum (what is taught), the intended curriculum (what is required), and the assessed curriculum (what is tested). Several studies have used the SEC methodology and contain detailed descriptions of the data coding process (Porter et al., 2009; Smithson &

Porter, 2001). The SEC methodology uses a neutral grid that considers two dimensions: topic and cognitive demand. This standardizing feature allows for the comparison of standards and assessments across districts, states, and countries. The SEC coding process examines individual items (indicators, benchmarks, assessment questions) on two dimensions: topic and cognitive demand (expectations for students).

The SEC mathematics matrix includes 16 coarse grain topics (number sense/properties/relationships, operations, measurement, consumer applications, basic algebra, advanced algebra, geometric concepts, advanced geometry, data displays, statistics, probability, analysis, trigonometry, special topics, functions, instructional technology) and 183 fine grain subtopics.

The second axis evaluates the level of cognitive expectation. The math matrix includes five levels: (a) memorize facts, definitions, formulas; (b) perform procedures; (c) demonstrate understanding; (d) conjecture, analyze, generalize, prove; and (e) solve nonroutine problems, make connections. Particular items may have more than one level of cognitive expectation.

Groups of content experts from state departments, research institutions, universities, and schools follow standard protocols to code documents on this neutral, standardizing grid. These experts work in teams to analyze each item of a specific document (a set of standards, an assessment, etc.). For the data in this study, groups contained at least four members, and employed a generalizabililty method to analyze documents. That is, their aggregate scores, not consensus scores, were used to determine the final data report for each document. Previous studies (Porter, 2002; Porter, Politkoff, Zeidner, & Smithson, 2008) have confirmed high levels of reliability among the generalizablity coefficients using these standard procedures.

Numerous variables, measures, and data are generated from these analyses. For this study, a few key measures are used. First, the data for each document can be aggregated to determine the total number of subtopics covered. That is, of the 183 possible subtopics, the analysis can identify how many a particular set of standards or assessment covers. For example, a set of fourth-grade standards that covers 100 subtopics has different implications than a set that covers 20 and may allow teachers to get more in-depth with their instruction. This variable allows researchers to examine the breadth of standards and assessments and, for this study's purposes, to test the mile wide hypothesis.

Second, the SEC methodology calculates an alignment index with values from 0 (*not aligned*) to 1.0 (*perfectly aligned*). This is accomplished by examining the proportion of each document's subtopic by cognitive demand grid in comparison to another document's subtopic by cognitive demand grid. The formula is as follows:

$$\text{Alignment Index} = 1 - \frac{\Sigma|x-y|}{2}$$

In this equation, X = cell proportions in one comparison matrix such a specific set of standards, an assessment, or teacher data; and Y = cell proportions in a corresponding comparison target. So for example, X could represent fifth-grade math standards and Y could represent the fifth-grade state assessment. The alignment index compares the proportion of content in each respective cell between two comparison documents. Thus a set of standards that has the exact same cell-by-cell proportion of subtopic by cognitive demand, to a comparison document would have a 1.0 alignment index. However, with the sheer number of cells involved, perfect alignment is highly unlikely

The SEC math grid has 915 total data points (183 subtopics × 5 levels of cognitive demand). For content standards, it measures the proportion of items (indicators, benchmarks, etc.) coded at each level. For assessments, it measures the proportion of test questions (or points if weighted) at each level. Items can be coded at multiple subtopics and cognitive expectations. For example, a test question that asks students to calculate proportions and present the data in a table might be coded as two different subtopics: "ratio and proportion" and "summarize data in a table or graph." In that case, the content would be assigned evenly across the two subtopics in the SEC grid. While much of the previous literature's emphasis has been on the SEC alignment that considers both dimensions (Porter, 2002), the methodology also produces alignment indices for both individual axes: a subtopic to subtopic alignment, and a cognitive demand to cognitive demand alignment. These marginal measures of alignment are generated by the same process as Porter's alignment index, that is, a marginal-by-marginal comparison of proportions. Accordingly, this study will consider two kinds of alignment:

1. *Topic Alignment*—the alignment of topics between the comparison documents, with each item (e.g., individual indicators) coded separately.
2. *SEC Alignment*—considers two dimensions, both topic and cognitive expectation.

Graphic content maps that display the relative emphasis of topics and subtopics on both axes are a unique feature of the SEC methodology. These maps are useful in visually displaying alignment and misalignment. The following analysis includes the course grain and fine grain SEC content maps

to provide more extensive details and an additional visual lens through which this robust data can be examined.

Data. This study uses data from several sources that has been analyzed with the SEC methodology including the Ohio Academic Content Standards (grade-band benchmarks and grade-level indicators). Ohio's academic content standards use both grade-level indicators and grade-band benchmarks. The state defines benchmarks as "the specific components of the knowledge or skill identified by [a standard]" and an indicator as "what students should know and be able to do. The indicators are the checkpoints that monitor progress towards the benchmarks" ("Academic Content Standards," 2010). In principle, the benchmarks are the general statements of knowledge and skills, while the indicators are the stepping stones to reaching the benchmarks. The test blueprints for state assessments use the benchmarks. The documents themselves were created separately. Thus there is some confusion among educators as some focus on indicators, some focus on benchmarks, and some focus on both. Accordingly, this research must include both in the analysis.

The second major source of data is SEC analysis of the 2007 TIMSS assessment (fourth and eighth grade) and 2006 PISA assessment for mathematics, which were SEC-coded in November 2008 by teams of state content experts with participation from the U.S. Department of Education (U.S. DOE), Institute of Educational Sciences (IES), and members of the CCSSO collaborative. A validity analysis of the TIMSS and PISA coding determined that the data represent a fair and valid description of the content assessed by those items (Blank & Smithson, 2009, p. 4). The third set of data represents SEC-analyzed academic content standards from high performing countries such as England, Hong Kong, Japan, Singapore, and Finland. Availability of full documents (standards and assessments) was a limiting factor. The fourth set of data represents SEC analysis of the Common Core State Standards (CCSS). These standards were coded shortly after public release in 2010. The coded documents represent the mathematics standards for grades four and eight, and an aggregated analysis of the high school standards. The Standards for Mathematical Practice are not included in these analyses.

Analysis

The 2010 *EdWeek* Quality Counts report graded Ohio as an "A" for standards, assessments, and accountability, which was third highest nationally. Although it should be noted that not all reviews have given high marks to Ohio's math standards. Finn, Julian, and Petrelli (2006) graded them a "D" and found issue with holes in some areas, ordering and overemphasis in others. The following

analyses examine these standards through the SEC lens based on the research questions discussed above.

Mile Wide, Inch Deep

These data can test the mile wide, inch deep hypothesis, that is, how many subtopics are in the standards at any particular grade level? Many policy makers and educators argue that too many topics make it difficult to thoroughly cover topics and spend time at higher cognitive levels. In addition, as La Marca et al. (2000) noted, "Standards that contain excessive detail also impede the development of assessments, making an acceptable degree of alignment difficult to achieve" (p. 5).

This first section of analysis tests the mile wide, inch deep hypothesis using SEC-coded data of Ohio's math standards compared to international counterparts. Thus it is hypothesized as follows:

> *Hypothesis 1*: Ohio's standards cover more topics than the standards of high performing countries.

Since Ohio uses both grade-level indicators and grade-band benchmarks, and educators focus on one or both when making instructional decisions, it is valuable to consider both in this analysis. This builds on Schmidt et al.'s (2005) use of the General Topic Trace Mapping (GTTM) model that examines the mean number of intended grade topics at each grade level. This analysis examines the quantity of the 183 possible subtopics addressed by each document.

Using the TIMSS and PISA respective test grading levels (fourth, eighth, and ninth grade), the results show that generally speaking, Ohio's math indicators tend to cover substantially more subtopics at fourth and eighth grades than high performing counties using grade-level standards such as Japan, Singapore, Hong Kong, and Finland. The Common Core standards reflect a major improvement in focus by reducing the number of topics covered in fourth and eighth grade.

Table 1 examines these data in detail and separates out the documents by grade level and grade band. Ohio's grade-level indicators cover more subtopics at fourth grade (67) and eighth grade (81) than the comparisons. Many of these differences are substantial, such as fourth grade in Hong Kong and eighth grade in Japan, each of which only address 26 of the SEC subtopics. The Ohio benchmarks fare better in comparison, especially to England,

Table 1. Number of Subtopics by Grade-Level and Grade-Band.

Grade-level standards	Subtopics	Grade-band standards	Subtopics
Fourth grade			
Ohio indicators	67	England (3-4)	61
Japan	55	*Ohio benchmarks (3-4)*	57
Common core	45		
Singapore	35		
Hong Kong	26		
Eighth grade			
Ohio Indicators	81	England (6-8)	116
Finland	50	Hong Kong (7-9)	100
Common core	49	*Ohio benchmarks (8-10)*	91
Singapore	27		
Japan	26		
Ninth grade (high school)			
Ohio indicators	70	England (9-10)	127
		Common core (HS)	105
		Hong Kong (7-9)	100
		Ohio benchmarks (8-10)	91
		Singapore (9-10)	73
		Japan (9-12)	32

which covers the most subtopics. However, again there are some substantial differences such as Japan, which covers 32 SEC subtopics at the high school level. Ohio's standards structure that includes both grade-level indicators and grade-band benchmarks further confounds the issue.

These results are consistent with the literature that argues that U.S. standards tend to cover too many topics compared to their counterparts. This trend is emphasized through the SEC lens. Ohio's grade-level indicators cover substantially more SEC subtopics than the international comparisons. However, analysis of the Common Core shows that in fourth and eighth grade, the newly adopted standards have addressed this lack of focus by reducing the number of subtopics that are required. In fourth grade, the number drops by almost 33% from 67 to 45. In eighth grade, the number of subtopics drops by almost 40% from 81 in the current indicators to 49 in the CCSS. In ninth grade, the CCSS focus is less clear since the new standards cover all high school grades making it difficult to compare to the respective Ohio grade-level indicator or benchmark in this analysis.

Alignment of Standards to International Tests

As previously discussed, a valid question for this type of analysis is an examination of how aligned standards are to the international assessments that states and countries are evaluated on. Results on international assessments should at least be considered in the light of how aligned the system is to those assessments. Thus the second hypothesis is as follows:

> *Hypothesis 2*: Higher performing countries' standards are more aligned to TIMSS and PISA than Ohio.

Topic alignment is a frequent consideration when examining "alignment." The SEC methodology strengthens the analysis as studies have shown strong, positive relationships between SEC alignment and achievement gains (Gamoran et al., 1997; Porter & Smithson, 2001). The SEC alignment index ranges from 0 (*no alignment*) to 1.0 (*complete alignment*). There is no magic number or threshold that signifies "good" alignment, but rather it is a relative concept. A higher alignment index score indicates better alignment. Blank and Smithson (2009) found that an analysis of all states with standards and corresponding assessment data had a mean overall SEC alignment of state assessment to standard/framework = .26, with the highest = .5 (n = 118). Analysis here will focus on relative alignment. That is, while there is no threshold for good alignment, when comparing across countries and documents, a higher alignment index indicates more overall alignment.

Fourth grade. In fourth grade, the countries with grade-level indicators tend to be more aligned to TIMSS than Ohio. As displayed in Table 2, Singapore and Japan's fourth-grade standards are more aligned to TIMSS than Ohio's fourth-grade indicators in both topic (Singapore, .48; Japan, .45) and SEC alignment (Singapore, .39; Japan, .30). Ohio's indicators and Hong Kong have fairly similar topic (.39 and .36, respectively) and SEC (both = .29) alignment. Some of these differences in the alignment index comparisons are small in scale. While Singapore has the highest SEC alignment index score, the others are not dramatically different. Nonetheless, the data show differences in relative alignment that are consistent with the initial hypothesis. Interestingly, Ohio's 3 to 4 benchmarks rank in the upper half of the standards on the topic comparison (.41), but drop to the bottom on the SEC comparison (.25), which also factors in cognitive expectations. The CCSS alignment to TIMSS is similar to the Ohio indicators and does not represent a substantial change in either topic (+.01) or SEC (−.01) alignment.

Table 2. Alignment of Fourth-Grade Standards to Fourth-Grade TIMSS.

Topic alignment		SEC alignment	
Singapore	.48	Singapore	.39
Japan	.45	Japan	.30
Ohio Benchmarks (3-4)	.41	*Ohio Indicators*	.29
Common core	.40	Hong Kong	.29
Ohio Indicators	.39	Common core	.28
Hong Kong	.36	England (3-4)	.25
England (3-4)	.36	*Ohio Benchmarks (3-4)*	.25

Figure 1 presents the coarse grain SEC analysis comparing Ohio's fourth-grade indicators to TIMSS and the CCSS. The SEC's unique maps provide the topic list on the Y axis and the cognitive demand on the X axis, with data points at each intersection. Figure 1 shows that Ohio's indicators focus heavily on performing procedures, especially in number sense/properties/relationship and measurement. While some areas appear similar to the TIMSS data, other areas of misalignment are readily apparent. Ohio generally covers more topic areas than TIMSS. TIMSS has areas of focus such as operations (23%) and geometric concepts (24%) that have less emphasis in the Ohio standards (10% and 14%, respectively). Conversely, the Ohio indicators have more probability (11%), which is not covered on TIMSS. The CCSS map appears much more aligned to TIMSS at a coarse grain level. However, the overall SEC alignment is determined at the fine grain level and, as mentioned above, the CCSS SEC alignment to TIMSS is very similar to the current fourth-grade indicators. The CCSS map does show the increased focus of the new standards and a shift away from Level 2 cognitive expectations.

Figure 2 drills down into the fine grain view of operations to further examine that area of misalignment. This map shows that the TIMSS items include topics that are not covered by the Ohio indicators. For example, 3.6% of the items involve ratio and proportion, which are not covered by the Ohio indicators at fourth grade. Almost 6% of items involve adding and subtracting whole numbers and integers, which are not emphasized by the standards (only .5%). The CCSS map shows more emphasis on operations including adding and subtracting whole numbers and integers (9.1%). However, the CCSS map is similar to the current Ohio indicators with no emphasis on ratio or proportion. From a potential opportunity to learn perspective, should fourth-grade students be expected to perform well on these assessment items if the state
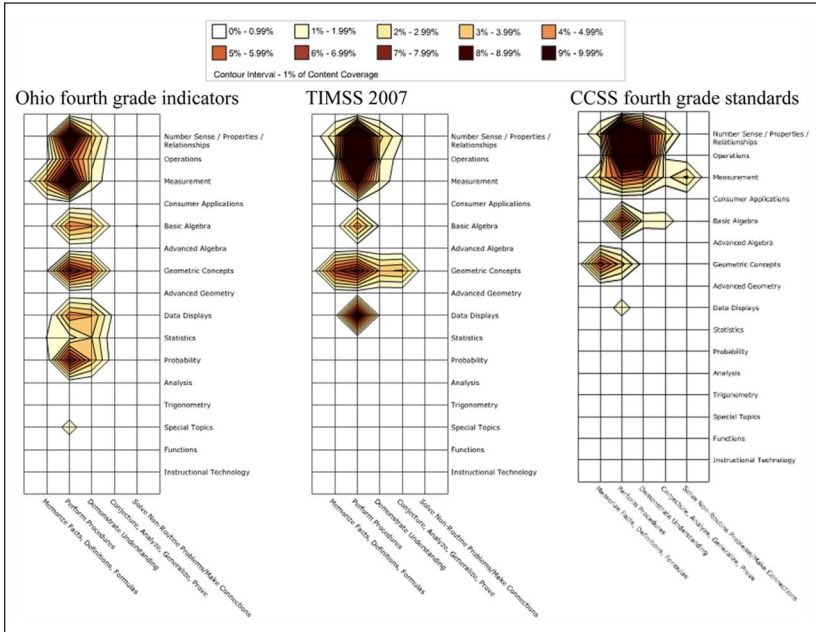
Figure 1. SEC map of Ohio indicators, TIMSS, and CCSS (fourth grade).

standards do not address them at the respective grade level? This may not be an issue if a subtopic was taught at a previous grade although the span of time between previous grade coverage and the assessment may itself be problematic. For example, if a test is based on third to fifth grade-band benchmarks, material that is taught at the beginning of third grade and tested at the end of the fifth grade may have different implications than material covered late in fifth grade then tested at the end of fifth grade.

Eighth grade. The eighth-grade analysis shows that Ohio's eighth-grade indicators are fairly well aligned to the TIMSS (topic = .41; SEC = .31) in comparison to the other countries. However, the 8-10 benchmarks are the lowest of the sample. Hong Kong (topic = .48; SEC = .38) is the most aligned to the eighth-grade TIMSS. The CCSS alignment to TIMSS is less than the Ohio indicators in both topic (.26 compared to .41) and SEC (.20 compared to .31) alignment; and slightly higher than the benchmarks. Table 3 displays these results.

Figure 3 displays the coarse grain SEC map of the Ohio eighth-grade indicators compared to the TIMSS and the CCSS. Similar to the fourth-grade
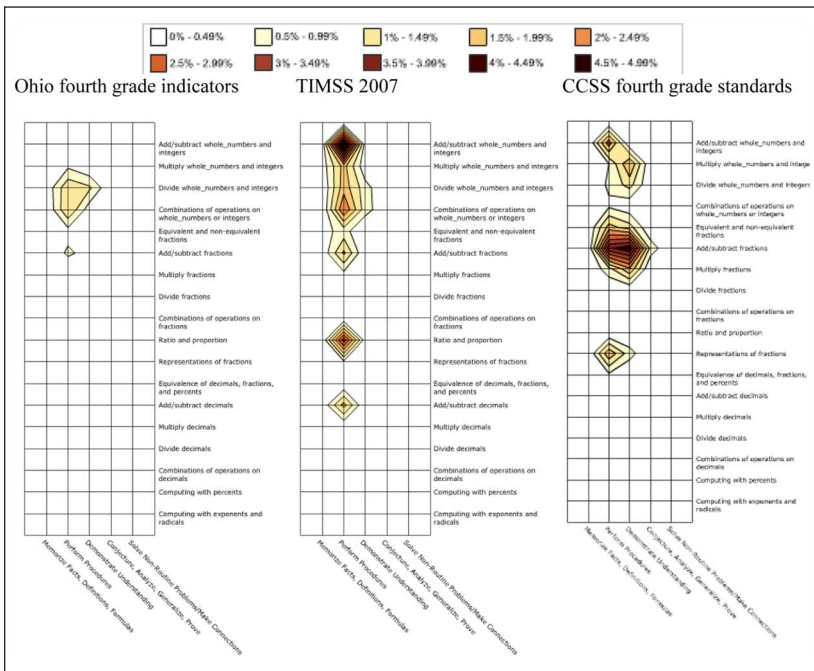
Figure 2. Fine grain SEC map of Ohio indicators, TIMSS, and CCSS (fourth-grade operations).

map in Figure 1, Ohio's indicators focus on performing procedures. The SEC maps show areas of incongruity when examining of alignment of topics. For example, Ohio's eighth-grade indicators have a heavy emphasis on measurement (16%) that is not as emphasized by TIMSS (8%) or the CCSS (1.7%). Vice versa, in a very substantial incongruity, TIMSS emphasizes operations (19.3%) that are not as heavily emphasized by Ohio (2.2%) or the CCSS (3.4%). The CCSS maps again display the increased focus and shift away from Level 2 cognitive expectations.

Figure 4 examines the eighth-grade maps in more detail specifically focusing on the fine grain view of operations. In this view, the incongruity is readily apparent as TIMSS tests several subtopics such as ratio and proportion (4.3%), multiplying fractions (1.9%), and computing with percents (2.6%) that are either not covered or barely emphasized (less than .5%) by the Ohio indicators or the CCSS. Again, this represents a misalignment in potential opportunities to learn and it may be reflected in student performance on TIMSS.

Table 3. Alignment of Eighth-Grade Standards to Eighth-Grade TIMSS.

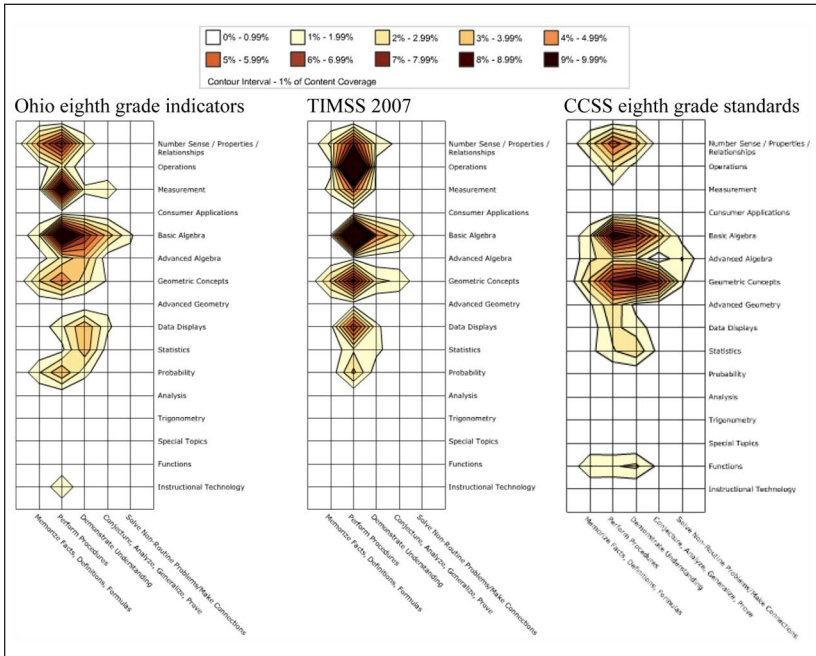|	Topic Alignment		SEC Alignment
Hong Kong (7-9)	.48	Hong Kong (7-9)	.38
Ohio indicators	.41	England	.36
England	.40	*Ohio indicators*	.31
Finland	.38	Singapore	.27
Singapore	.33	Finland	.25
Japan	.27	Japan	.22
Common core	.26	Common core	.20
Ohio benchmarks (8-10)	.24	*Ohio benchmarks (8-10)*	.17

Figure 3. SEC map of Ohio indicators, TIMSS, and CCSS (eighth grade).

Ninth grade. As shown in Table 4, Ohio's ninth-grade indicators have the lowest alignment to the PISA of all the standards in the sample (topic = .14; SEC = .12). The 8-10 benchmarks fared slightly better, but still had relatively
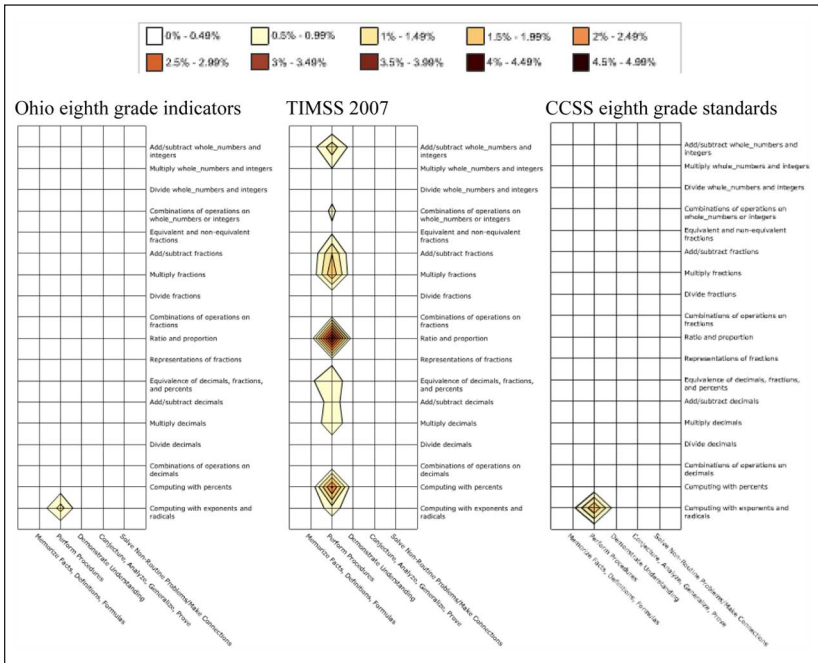
Figure 4. Fine grain SEC map of Ohio indicators, TIMSS, and CCSS (eighth-grade operations).

Table 4. Alignment of Ninth-Grade Standards to PISA.

	Topic alignment		SEC alignment
Hong Kong (7-9)	.33	Hong Kong (7-9)	.27
England (9-10)	.23	England (9-10)	.17
Ohio benchmarks (8-10)	.22	*Ohio benchmarks (8-10)*	.14
Japan	.18	Japan	.14
Singapore (9-10)	.16	Singapore (9-10)	.12
Ohio indicators	.14	*Ohio indicators*	.12
Common core	.14	Common core	.10

low alignment (topic = .22, SEC = .14). The high school Common Core was similar in both topic and SEC alignment to the Ohio indicators although it should be noted that the CCSS data reflect all the high school standards. Similar to the eighth-grade analysis, Hong Kong and England scored the highest of the comparison countries.
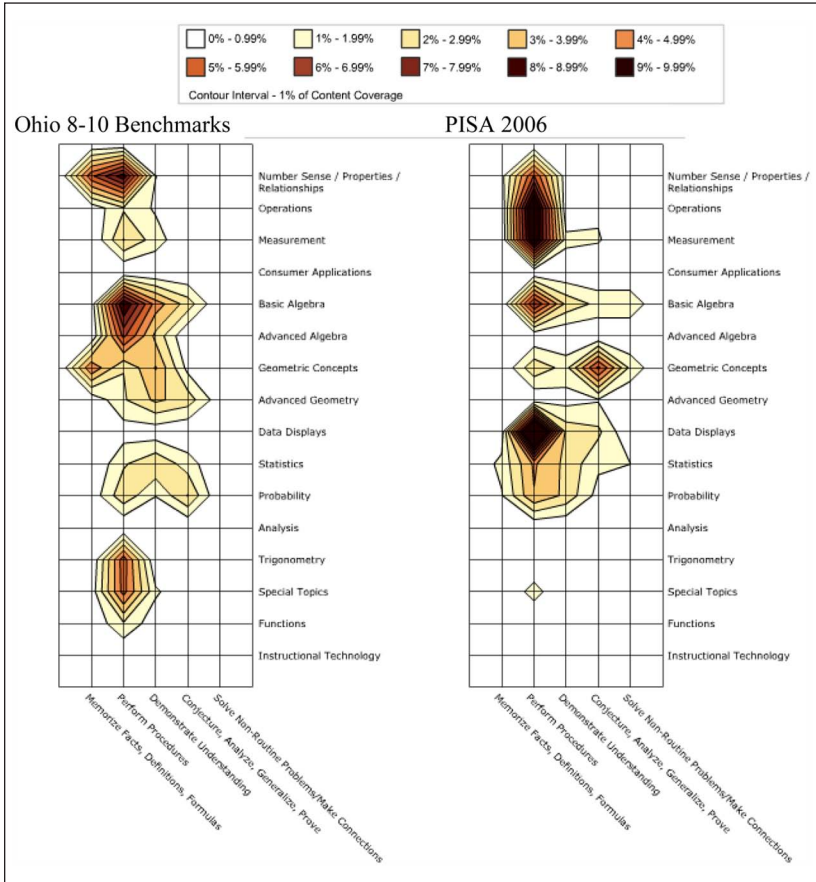
Figure 5. SEC map of Ohio 8-10 benchmarks and PISA.

Figure 5 displays the SEC analysis of the Ohio 8-10 benchmarks compared to PISA. The benchmarks were chosen for this comparison since the PISA may encompass multiple grade levels based on the age status of the students. Since the high school CCSS represents a drastically different age range than the Ohio comparison, it was not included. As noted in Table 4, the benchmarks are not very well aligned to PISA and some very specific areas of incongruity are evident in the SEC maps. For example, Ohio emphasizes trigonometry (4%) and other special topics (5%) that are only minimally addressed by PISA (0% and 1%, respectively). Conversely, PISA emphasizes topics such as
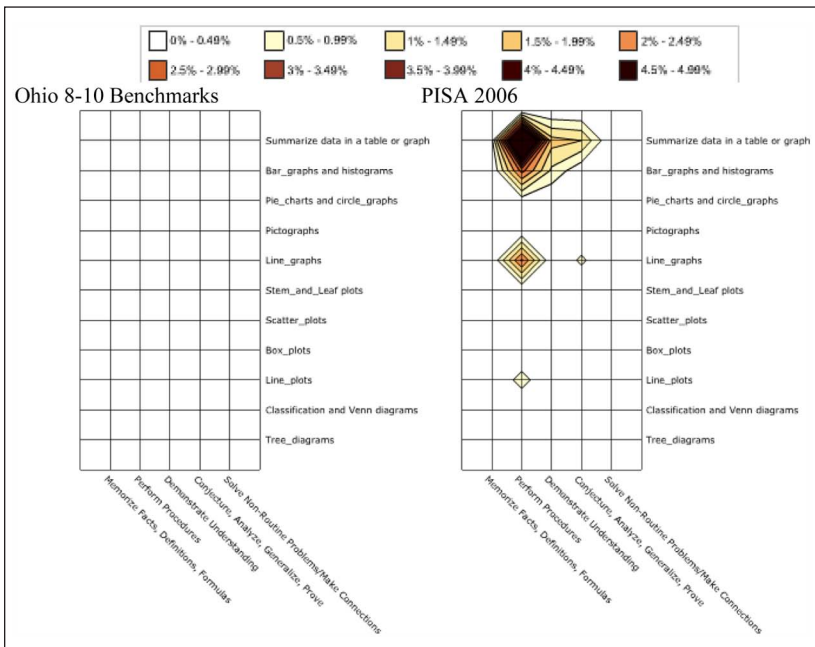
Figure 6. Fine grain SEC map of Ohio indicators and PISA (data displays).

operations (13%) and data displays (18.9%) that have less emphasis in the Ohio benchmarks, (4% and .5%, respectively).

Figure 6 examines the substantial incongruities in the fine grain comparison of data displays. PISA asks students to summarize data in tables and graphs (11.1%), use bar graphs and histograms (3.7%), and use line graphs (3.2%), whereas the Ohio benchmarks only minimally address any subtopic (.5%) within data displays. A lack of alignment clearly exists, which does not necessarily mean that the benchmarks are inadequate but may help explain why students do not perform at high levels. PISA asks students to address data display subtopics that their respective benchmarks do not require.

Dissecting depth of knowledge (cognitive expectations). The previous analysis points to differences in cognitive expectations—one axis of the SEC methodology. SEC alignment is lower than topic alignment for each comparison on each of the three assessments suggesting differences in cognitive expectations. One benefit of the SEC methodology is the ability to explore this factor. Accordingly, the depth of cognitive expectations across the countries' standards can be specifically examined to determine whether Ohio's standards

Table 5. Country Specific Cognitive Expectations by Grade.

	1	2	3	4	5
Fourth grade					
Ohio indicators	7.5	61.2	29.1	1.2	1.0
Ohio benchmarks	29.7	34.2	25.7	10.1	0.3
Common core	12.3	47.6	29.7	7.2	3.2
England	22.9	50.3	15.2	10.7	1.0
Japan	29.0	30.7	28.3	11.6	0.5
Hong Kong	5.3	50.0	37.1	7.5	0.0
Singapore	3.5	76.3	17.5	2.8	0.0
Eighth grade					
Ohio indicators	13.1	50.7	25.2	8.8	2.2
Ohio benchmarks	12.4	50.8	24.2	10.8	1.9
Common core	14.2	37.8	30.6	12.9	4.5
Japan	0.0	79.8	20.2	0.0	0.0
Finland	0.0	76.5	19.6	4.0	0.0
Singapore	9.3	81.4	0.0	9.3	0.0
England	7.3	73.4	13.3	6.0	0.0
Hong Kong	15.3	52.4	25.9	6.4	0.0
Ninth grade					
Ohio indicators	12.5	65.1	18.2	4.0	0.3
Ohio benchmarks	12.4	50.8	24.2	10.8	1.9
Common core	6.0	52.7	33.5	6.4	1.3
England	2.4	75.5	20.2	1.8	0.0
Singapore	3.5	92.6	2.8	1.1	0.0
Japan	0.0	88.4	11.6	0.0	0.0
Hong Kong	15.3	52.4	25.9	6.4	0.0

have the same level of depth. Table 5 displays the cognitive expectations of the various countries' standards in relation to Ohio's indicators and benchmarks.

In fourth grade, Ohio's indicators are primarily focused on Level 2: perform procedures (61.2%). Almost 30% of Ohio's 3-4 benchmarks are coded at Level 1: memorize facts, definitions, and formulas; which is similar to Japan (29%) but higher than Hong Kong (5.3%) and Singapore (3.4%). Hong Kong has the highest percentage (37.1) of standards coded at Level 3: demonstrate understanding. The Common Core standards represent an increase in the higher level cognitive levels for Ohio compared to the state indicators. With the fourth-grade math indicators, the CCSS represents a shift away

from Level 2 to more emphasis on Levels 4 and 5 (7.2% and 3.2%). <mark>The CCSS show much less emphasis on Level 1 expectations (12.3%) than the benchmarks.</mark>

In eighth grade, a majority of Ohio's indicators (50.7%) and benchmarks (50.8%) are coded at Level 2. While Ohio does have many Level 1 indicators (13.1%) at this grade level, it fares relatively well at the higher levels. That is, the number of Level 4 and 5 indicators and benchmarks are similar to and in some cases greater than the international comparisons. The SEC analysis of the eighth-grade CCSS emphasize a shift into higher level cognitive levels. For Ohio, the CCSS has less emphasis on Level 2 in comparison to the indicators and benchmarks, and more emphasis on each of the three higher levels 3, 4, and 5. The CCSS is less coded at Level 2 (37.8%) with higher percentages of items coded at Level 3, 4, and 5. In <mark>fact, the Common Core has the highest percentage of emphasis in these three highest levels when compared to all the high performing countries with 12.9% at Level 4: conjecture, analyze, generalize, and prove; and 4.5% at Level 5: solve nonroutine problems and make connections.</mark>

The ninth-grade results are similar as Ohio's indicators focus on Level 2: perform procedures (65.1%). Both the Ohio indicators (12.5%) and benchmarks (12.4%) have relatively high amounts of Level 1 items. Ohio's 8-10 benchmarks have the highest percentage of items coded at Level 4 (10.8%). The Common Core represents a shift to higher level expectations in comparison to Ohio's ninth-grade indicators. Specifically, there is a shift to more Level 3 standards (33.5%) and less Level 1 and 2. The Common Core has the highest percentage of emphasis of Level 3 standards of all the comparison countries.

The results of this analysis show that a majority of the current Ohio indicators, at the respective grade levels, are focused on Level 2: perform procedures, and Ohio tends to have more items coded at Level 1 than many of the comparison countries especially at Grades 8 and 9. <mark>The Common Core aims to have more rigorous set of standards. The SEC analysis of cognitive expectations provides some evidence to support this for Ohio.</mark>

Discussion and Policy Considerations

As Viadero (2006) noted, it is important to examine the data from international assessments and relate findings to policy interests. Min-Hsiung (2009) argued that researchers need to get beyond the symbolic rankings and understand the data. <mark>This study attempted to understand the potential opportunities to learn by benchmarking a state's math standards against high performing countries and TIMSS and PISA</mark>. The analysis reveals some interesting findings

that researchers and policy makers should consider when implementing the Common Core and evaluating recent reforms. The focus on current state standards is relevant for three reasons. First, the existing state standards were not immediately replaced and will remain state policy for at least 2 years until the Common Core is fully implemented. Second, states are permitted to add up to an additional 15% of content in addition to the Core. Third, standards are a political process and states can always revisit or replace their standards.

A Baseline Analysis

The first important policy consideration is a state-level baseline analysis for the Common Core. Studies such as Schmidt (2001) emphasize that international high performers are more focused and this may be a competitive advantage on international assessments. Darling-Hammond and McClosky (2008) describe how Finland's lean documents went from hundreds of pages of prescription to a few sets of core concepts. Standards-writing tends to be an incremental process where more and more items are added to avoid political conflict. Finn et al. (2006) find that standards fall short when vision is sacrificed for consensus and that committee processes lead to "kitchen-sinkism" (p. 12). Schmidt et al. (2005) describe "an exercise in democratic consensus-making" (p. 530) that requires political compromise and moves away from a structured process. Wirt and Kirst (2005) describe the process as disjointed incrementalism, with only marginal changes considered, piecemeal alteration, lack of a comprehensive strategy, and avoidance of conflict by "using vague language concerning standards and covering so many topics that no major interest group feels left out. Content priority is sacrificed to the political necessity in breadth of coverage" (p. 346). In other words, the political process often results in increasing the breadth of standards to achieve pluralist support. If a policy goal is to reduce the breadth of standards and focus content, then maintaining that focus may be challenging.

The NGA/CCSSO/Achieve report (2008) recommending creation of a Common Core specifically addressed the notion of focus and thus reducing the number of topics per grade level. This initiative aimed to provide more focus, rigor, and coherence to state standards. When considering the number of topics covered, Ohio's math standards tend to cover more topics than the comparison countries providing evidence of the mile wide, inch deep syndrome. Initial analysis of the Common Core in specific grades provides evidence that the process was successful in meeting those goals. In fourth and eighth grades, the CCSS represent a substantially more focused set of standards.

Policy makers should remember these policy goals when fully implementing the CCSS and for future revisions. While states voluntarily agree to adopt all the Common Core elements, they are also permitted under the 15% Rule to add up to an additional 15% of content. This allows states some local discretion, but should be used sparingly to maintain the focus of the standards. As pressures build around this, states will also consider future revisions. Policy makers need to stay true to this theory of action when considering marginal additions and future revisions.

Understanding Assessment Results

Policy makers are attuned to results on international assessments such as TIMSS and PISA as they are typically high profile in the media and often drive calls for reform. For example, a recent Associated Press column referred to PISA results as a "'Wake-up call': U.S. students trail global leaders" (2010). An Ohio newspaper reporting PISA results, argued "A country doing 'average' in education, is falling behind" (2010). Whether American students' performance on international assessments should be a policy goal is debatable, but regardless policy makers and the public do pay attention. La Marca's (2001) review of alignment claims, "It would be a disservice to students and schools to judge achievement of academic expectations based on a poorly aligned system of assessment" (p. 5). This study provides some potential explanation for the mediocre performance of American students. When considering performance on TIMSS and PISA, it should be noted that Ohio's current standards are not particularly well aligned to these tests—especially PISA. Other countries' standards, especially at fourth and ninth grade, are more aligned than Ohio.

Overall, Ohio's alignment to the international assessments in relation to the higher performing comparison countries ranges from *poor* (at worst) to *mixed* (at best). This is complicated by the inclusion of both Ohio indicators and benchmarks, which have varying degrees of alignment by grade level. Japan and Singapore are more aligned than Ohio and the other countries at the fourth grade. Hong Kong and England are best aligned at both eighth and ninth grades. The alignment of the new CCSS is similar at Grade 4 and lower than the current eighth grade indicators. These indicators were not necessarily designed to be aligned to the TIMSS and PISA, so this should be considered when analyzing student performance data. If policy makers want higher student performance on TIMSS and PISA, then the state standards and assessment system could consider a system that is more tightly aligned to those frameworks. If that is not a policy goal, then results on

future international assessments should be viewed through that moderated lens of expectations.

Future Research

This study has identified and generated several avenues of future research. First, the intent of this study was to conduct a state-level analysis of standards in relation to high performing countries and international assessments—using the SEC methodology. The Common Core promises a more focused and rigorous set of standards designed to make American students more competitive. The results of this analysis provide some preliminary evidence supporting that principle in terms of reducing the number of topics covered and increasing the levels of cognitive expectations. This baseline data should be compared to the fully implemented Common Core across all grade levels and expanded to examine other participating states. In addition, while this study specifically addressed questions of focus, future SEC research should specifically address coherence issues. As Schmidt et al. (2005) emphasize, it is not just about reducing the number of topics but also creating a coherent structure built on progressions. This is a point that the Common Core framers address in the introduction to the standards—focus is not enough, there must be coherent progressions (Common Core State Standards Initiative, 2010). For example, Ohio's grade-level indicators are generally most aligned to the grade levels immediately before and after (Woolard, 2007) In other words, across all grade levels, there is a correlation of subtopic/cognitive demand that is often repeated from grade to grade. Additional SEC research should examine whether that repetitiveness is reduced both in Ohio and in other states.

A second area where the SEC methodology can specifically add to the research base is by examining the enacted curriculum in the comparison countries. As much research has pointed out (Porter & Smithson, 2001; Schmidt et al., 2009), there is variation between the intended and implemented curriculum. This study used the SEC to compare potential opportunities to learn by focusing the alignment of the standards to assessments. Adding data on the enacted curriculum, while representing a resource intensive task, would provide additional insights on instructional practices and help understand differences across borders.

The SEC methodology can be useful in these types of studies. Future SEC development could increase the value of such analysis. This includes the coding of additional standards across all states, analysis of any 15% rule additions, and coding of other high performing international standards. The

methodology could benefit from additional research that examines the question of "what is the threshold of good alignment?"

Schmidt et al. (2005) note that "coherence and rigor might only be possible in the United States if curriculum standards are national in scope" (p. 556). With the implementation of the Common Core, that may come to fruition. This particular research has provided a baseline to examine against. Future research should build on this by further developing the use of the SEC methodology in addressing these questions, examining more states in detail, and revisiting these questions when the states fully implement their versions of the Common Core and the subsequent assessments.

Declaration of Conflicting Interests

Funding

References

Academic content standards frequently asked questions. (2010). Retrieved from http://www.ode.state.oh.us/GD/Templates/Pages/ODE/ODEDetail.aspx?page=3&TopicRelationID=1696&ContentID=1929&Content=88486

Achieve, Inc. (2007). *Creating a world-class education system in Ohio*. Retrieved from http://www.achieve.org/files/World_Class_Edu_Ohio_FINAL.pdf

A country doing 'average' in education, is falling behind. (2010, December 30). *Youngstown Vindicator*. Retrieved from http://www.vindy.com/news/2010/dec/30/a-country-doing-8216average8217-in-educa/?print

Armario, C. (2010, December 7). "Wake-up call": U.S. students trail global leaders. *MSNBC.com*. Retrieved from http://www.msnbc.msn.com/id/40544897/ns/us_news-life/

Baines, L. (2008). Learning from the world: Achieving more by doing less. *Education Digest: Essential Readings Condensed for Quick Review, 73*(5), 23-26.

Blank, R., & Smithson, J. (2009, February). *Alignment content analysis of TIMSS and PISA mathematics and science assessments using the Surveys of Enacted Curriculum methodology*. A CCSSO paper prepared for National Center for Education Statistics and American Institutes for Research, Washington, DC. Retrieved from http://www.ccsso.org/Documents/2009/Alignment_Content_Analysis_of_TIMSS_2009.pdf

Common Core State Standards Initiative. (2010). *Toward greater focus and coher-ence*. Retrieved from http://www.corestandards.org/the-standards/mathematics

Darling-Hammond, L., & McCloskey, L. (2008). Assessment for learning around the world: What would it mean to be internationally competitive? *Phi Delta Kappan 90*, 263-272.

Finn, C. E., Julian, L., & Petrelli, M. J. (2006). The state of the state standards 2006. *Thomas B. Fordham Foundation*. Retrieved from http://www.edexcellence media.net/publications/2006/200608_thestateofstatestandards2006/State%20 of%20State%20Standards2006FINAL.pdf

Frelindich, N. (1998). From Sputnik to TIMSS: Reforms in science education make headway despite Setback. *Harvard Education Letter*. Retrieved from http://www .project2061.org/publications/articles/articles/harvard.htm

Gamoran, A., Porter, A. C., Smithson, J., & White, P. A. (1997). Upgrading high school mathematics instruction: Improving learning opportunities for low-achieving, low-income youth. *Educational Evaluation and Policy Analysis, 19*, 325-338.

Gonzales, P., Williams, T., Jocelyn, L., Roey, S., Kastberg, D., & Brenwald, S. (2008). *Highlights from TIMSS 2007: Mathematics and science achievement of U.S. fourth- and eighth-grade students in an international context* (NCES 2009-001 Revised). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Hook, W., Bishop, W., & Hook, J. (2007). A quality math curriculum in support of effective teaching for elementary schools. *Educational Studies in Mathematics, 65*(2), 125-148.

La Marca, P. M. (2001). *Alignment of standards and assessments as an accountability criterion*. Retrieved from http://PAREonline.net/getvn.asp?v=7&n=21

La Marca, P. M., Redfield, D., Winter, P. C., Bailey, A., & Despriet, L. H. (2000). *State standards and state assessment systems: A guide to alignment*. Retrieved from http://www.eric.ed.gov/PDFS/ED466497.pdf

Min-Hsiung, H. (2009). Beyond horse race comparisons of national perfor-mance averages: Math performance variation within and between class-rooms in 38 countries. *Educational Research & Evaluation, 15*, 327-342. doi:10.1080/13803610903087003

Mullis, I. V. S., Martin, M. O., & Foy, P. (2008). TIMSS 2007 *International math-ematics report: Findings from IEA's trends in international mathematics and sci-ence study at the fourth and eighth grades*. Chestnut Hill, MA: Boston College.

National Governors Association, the Council of Chief State School Officers, & Achieve, Inc. (2008). *Benchmarking for success: Ensuring U.S. students receive a world-class education*. Retrieved from http://www.achieve.org/BenchmarkingforSuccess

Organisation for Economic Co-operation and Development. (2007). *PISA 2006: Sci-ence competencies for tomorrow's world*. Retrieved from http://www.oecd.org/ dataoecd/30/17/39703267.pdf

Organisation for Economic Co-operation and Development. (2010). *PISA 2009 Results: Executive summary*. Retrieved from http://www.pisa.oecd.org/dataoecd/34/60/46619703.pdf

Phillips, G. W. (2009). *The second derivative: International benchmarks in mathematics for U.S. states and school districts*. Washington, DC: American Institutes for Research.

Phillips, G. W., & Dossey, J. (2008). *Counting on the future: International benchmarks in mathematics for American school districts*. Washington, DC: American Institutes for Research.

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

Porter, A. C., Polikoff, M. S., & Smithson, J. (2009). Is there a defacto national intended curriculum? Evidence from state content standards. *Education Evaluation and Policy Analysis 31*, 238-268.

Porter, A. C., Politkoff, M. S., Zeidner, T., & Smithson, J. (2008). The quality of content analyses of state student achievement tests and state content standards. *Educational Measurement: Issues and Practice, 27*(4), 2-14.

Porter, A. C., & Smithson, J. (2001). *Defining, developing, and using curriculum indicators* (CPRE Research Report Series). Philadelphia: Consortium for Policy Research in Education: University of Pennsylvania.

Quality Counts 2010. (2010). *Ed Week*. Retrieved from http://www.edweek.org/ew/qc/2010/17src.h29.html

Richland, L. E., Zur, O., & Holyoak, K. J. (2007). Cognitive supports for analogies in the mathematics classroom: Variations in the effective use of analogies in math instruction across countries may contribute to performance differences in the TIMSS studies. *Science, 316*, 1128-1129.

Schmidt, W. H., Bates, P., & Leroi, G. (2009). Research Report—Opportunities to learn in PROM/SE classrooms: Teachers' reported coverage of mathematics content. *Promoting Rigorous Outcomes in Mathematics and Science Education (PROM/SE), 6*, 1-20.

Schmidt, W. H., McKnight, C., & Raizen, S. (1997). *A splintered vision: An investigation of U.S. science and mathematics education*. Dordrecht, Netherlands: Kluwer Academic.

Schmidt, W. H., & Prawat, R. S. (2006). Curriculum coherence and national control of education: Issue or non-issue? *Journal of Curriculum Studies, 38*, 641-658.

Schmidt, W. H., Wang, H., & McKnight, C. C. (2005). Curriculum coherence: An examination of US mathematics and science standards from an international perspective. *Journal of Curriculum Studies, 37*, 525-559.

Schmidt, W. H., McKnight, C. C., Houang, R. T., Wang, H., Wiley, D. E., Cogan, L. S., & Wolfe, R. G. (2001). *Why schools matter: A cross-national comparison of curriculum and learning*. San Francisco, CA: Jossey-Bass.

U.S. Department of Education. (2007). *Standards and assessments peer review guid-ance: Information and examples for meeting requirements of the No Child Left Behind Act of 2001*. Retrieved from http://www.ed.gov/policy/elsec/guid/saapr-guidance.pdf

Valverde, G. A., & Schmidt, W. H. (2000). Greater expectations: learning from other nations in the quest for 'world-class standards' in US school mathematics and science. *Journal of Curriculum Studies, 32*(5), 651-687.

Viadero, D. (2006, November 01). Potential of global tests seen as unrealized: Scholars urged to scour TIMSS, PISA for policy insights. *Education Week, 26*(13), 1-15.

Wirt, F., & Kirst, M. (2005). *The political dynamics of American education*. Richmond, CA: McCutcheon.

Woolard, J. C. (2007, April). *Measuring systemic alignment of a state's instruction, standards, and assessments: A baseline analysis*. Paper presented at the 2007 Annual Meeting of the American Educational Research Association, Chicago, IL. Retrieved from http://programs.ccsso.org/content/pdfs/woolard%2007%20 AERA.doc

Author Biography

J. Christopher Woolard is a research and evaluation manager for the Ohio Department of Education - Office of Policy and Research. His work focuses on data tools for educators, teacher evaluation, curriculum alignment, and state-level research. Chris received his PhD/MA (Political Science) from Miami University and BA (Education, Political Science) from Muskingum College. His current research interests include international benchmarking of state standards and assessments, and best practices in online education. Chris also teaches adjunct courses at Ohio University-Pickerington.