The purpose of this paper is to report findings related to curriculum implementation of two types of mathematics textbooks that have different organizations of mathematical content and to discuss implications of these findings. Research has established that students learn what they are given the opportunity to learn (for mathematics, see Floden, 2002; Hiebert, 1999; Hiebert & Grouws, 2007). Ultimately teachers are the decision-makers with regard to specific content taught, but they rely heavily on curricular materials, especially textbooks, to inform such decisions. Thus it is important to determine how teachers use textbooks in the classroom. Findings reported in this paper are based on data collected as part of an NSF-funded research project: Comparing Options in Secondary School Mathematics: Investigating Curriculum (COSMIC). The COSMIC project involves a three-year longitudinal comparative study of integrated mathematics curricula and subject-specific mathematics curricula on student learning at the high school level. One group of students (n ≈ 1300) studied from an integrated content approach (e.g., Core-Plus Mathematics) and the other students (n ≈ 1300) from a subject-specific content approach (students followed an Algebra I, Geometry, Algebra II sequence). Curriculum implementation was characterized using multiple measures including classroom observations and teacher surveys which are the measures discussed in this paper. See the project website (cosmic.missouri.edu) for additional study details.

**Perspective**

When examining what students are learning when interacting with specific mathematics curricula, we cannot assume these curricula are being implemented in any specific way (Scott, 1994). In particular, teachers modify curricula based on their personal beliefs, content knowledge, and classroom experience. “Two classrooms in which the same curriculum is supposedly being ‘implemented’ may look very different; the activities of teacher and students in each room may be quite dissimilar, with different learning opportunities available, different mathematical ideas under consideration, and different outcomes achieved” (Kilpatrick, 2003). Teachers may choose to move through the textbook sequentially or not; they may choose to cover most of the chapters of the textbook or not; they may supplement the textbook with materials from other resources or not, and so forth. All of these decisions and others affect the measure of fidelity to which the curriculum is implemented and the mathematics that students are
given the opportunity to learn. Each of these factors was carefully assessed via triangulated measures in this study.

Although we know that textbooks are the centerpiece of mathematics instruction in U.S. schools (Grouws & Smith, 2000), we know very little about the relationship between implementation levels, specific textbook types, and content specific learning. Textbook implementation was examined in this study through two conceptual lenses: content fidelity and presentation fidelity. The content lens examined what mathematics content in the intended curriculum was taught as part of classroom instruction. This construct was measured by multiple methods including ratings made by trained observers during classroom visits.

The presentation lens examined how the curriculum materials were implemented. This construct considered the way students were expected to engage with the textbook material during the mathematics class period, as determined by textbook analysis and interviews with textbook authors. This aspect of implementation is particularly important when studying what are often referred to as “reform-oriented” textbooks because the program philosophy of these textbooks call for changes in the way mathematics is typically taught, such as students more regularly working in small group settings, engaging more frequently in discussion of ideas, often developing their own informal algorithms and procedures, and so on. Each of these facets of presentation implementation was assessed in this study, but their measurement required multiple instruments, including the ratings made by classroom observers that are reported in this paper.

It is important to note that measures of implementation are not direct measures of the quality of teaching. That is, a low measure of implementation does not necessarily imply poor teaching. The measure of content fidelity, for instance, represents the extent to which the content of a particular textbook is taught and there may be occasions where it is prudent for a teacher to deviate from the content included in a textbook. For example, it could be that a teacher is aware that students have a deep knowledge of some particular mathematical topic, so it is appropriate not to spend additional instructional time on it. In summary, we remind the reader that documenting the extent of implementation is necessary in order to make valid interpretations of student achievement in relation to a specific textbook curriculum (National Research Council, 2004) and that there are few curriculum studies that take appropriate account of curriculum implementation (O’Donnell, 2008).

The Study

Research Questions

In this paper, we report findings centered on three clusters of research questions:

- To what extent do teachers exhibit fidelity of implementation to the textbook with respect to content and presentation? What is the relationship between textbook type and the Content Fidelity index? Between textbook type and the Presentation Fidelity index? Are there differences in the relationships based on textbook type?

- What is the relationship between fidelity of implementation and a teacher’s years of classroom teaching experience? Are the preceding findings valid across textbook types or are they dependent on the content organization and philosophy of the textbook? Do teachers with more years of experience tend to supplement the textbook more frequently or less frequently
than teachers with less teaching experience?

- What are the relationships between fidelity of implementation and professional development? Are the relationships different across textbook types?

**Participants**

Participants in this study were teachers in schools using textbooks that embodied both content organization approaches, but with different groups of students. Students were given their choice between enrolling in classes utilizing the subject-specific approach or those utilizing the integrated approach. The sample for this report is comprised of 112 classes (58 integrated and 54 subject-specific) in 5 U.S. states with 34 teachers and approximately 2600 students. Because the study occurs within schools that offer two types of programs, it was the case that 10 teachers taught a class(es) using one curriculum type during some class periods of the school day and the other curriculum type during other class periods. Student learning was assessed through the use of standardized measures of achievement as well as internally developed instruments to assess students’ depth of knowledge, skills acquisition, and conceptual development.

**Data Sources**

Curriculum implementation was measured using instruments that separately reflected the perspective of the teacher and the perspective of the researcher. To gain the teacher perspective we had teachers report information by completing a written survey about such things as teaching experience, professional development, views of the curriculum, and so forth. From a researcher perspective, curriculum implementation was measured utilizing classroom observations. The classroom visitation protocols were developed after interviewing textbook authors to determine what the developers of each curriculum type expected of teachers in implementing their curricular materials. In interviewing authors, we asked questions regarding the key lesson components (e.g., Launch, Explore, Share and Summarize, and Apply) of each textbook with particular attention to what constitutes a “faithful” implementation of each lesson component according to textbook authors. Their responses to these interview questions were used to develop classroom visit protocols specific to each textbook used in the participating school districts.

During each classroom observation, the observer took notes to record anecdotal evidence of particular presentation features being implemented. Upon completion of the class visit, the observer completed a comprehensive summary form that included judgments about the implementation of the textbook components of the specific curricula being used (extent of use or attention received). Finally, the observer assigned overall level of content and presentation fidelity scores (5 point scales) using field notes and a well-defined scoring rubric (See Figure 1). High Content Fidelity consists of the mathematical content being enacted as written in the textbook and the textbook used as the primary source of content for the lesson. High Presentation Fidelity indicates the enacted curriculum being consistent with the expectations of the authors as expressed in the textbook author interviews we conducted and the authors’ pedagogical philosophy as reflected in the notes and suggestions to teachers in the Teacher’s Edition of the textbook. Each rating was independent of the other allowing for the possibility of a high fidelity rating on one aspect of implementation and a lower fidelity rating on the other for any given lesson. Data from 129 classroom observations were collected (3 observations for most teacher participants but only 2 observations for 3 teachers because of scheduling difficulties).
In an effort to maintain a high degree of consistency in the coding of classroom visits, extensive training of observers was conducted and continued until a sufficiently high degree of reliability was achieved on Content Fidelity and Presentation Fidelity ratings. Furthermore, during the data collection, a confirmatory reliability test was conducted by double coding 15 lessons. These lessons were neither randomly selected nor purposefully selected but instead chosen based on feasible observation schedules. We double coded at least one lesson in most schools. Individual researchers took field notes and immediately following the lesson each researcher worked in isolation to complete the protocol including the content and presentation judgments. When all coding was completed, researchers compared codes, negotiating disagreements until they were resolved. Consensus codes were used in subsequent analysis of implementation data but the set of original assigned codes were used to gauge the ongoing reliability of the classroom visit coding.

Overall, the reliability of the coding of Content Fidelity and Presentation Fidelity was high. With regard to Content Fidelity, 14 of the 15 rating pairs from the two observers were identical. In the one instance when the two observers disagreed, their individual ratings were within one unit of each other. The results for Presentation Fidelity revealed 10 of the 15 rating pairs were identical. Similar to Content Fidelity, the remaining 5 pairs were all within one unit of each other.

Results

In the following paragraphs we provide data regarding the three clusters of research questions.
We begin by providing a descriptive characterization of the Content Fidelity and Presentation Fidelity indices. We follow this with information regarding the relationships among the fidelity indices, teaching experience, and professional development.

**Content Fidelity and Presentation Fidelity Indices**

During classroom observations, observers recorded judgments regarding the degree to which the textbook influenced the content taught and the manner of presentation of the mathematics lessons. For each teacher a Content Fidelity and Presentation Fidelity mean rating was computed by taking the mean across the three classroom observations. Analyses of Content Fidelity ratings showed that the content of lessons was primarily attributable to the textbook (see Table 1). Across teachers, the mean Content Fidelity rating was 4.16 (on a 5 point scale). Moreover, these data indicate that Content Fidelity ratings were relatively high regardless of whether the teacher was teaching from an integrated or a subject-specific textbook and the means were not significantly different statistically across textbook types.

<table>
<thead>
<tr>
<th>Teacher Content Fidelity and Presentation Fidelity Indices</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="#" alt="Table" /></td>
</tr>
</tbody>
</table>

* n > 34 because 10 teachers were assigned to teach both subject-specific courses and integrated courses within a school day.

With regard to Presentation Fidelity, the overall ratings were lower than the Content Fidelity ratings. The mean overall Presentation Fidelity rating was 3.11 (SD = 0.89) across the 44 teachers suggesting that the manner in which the lesson was taught was less consistent with the authors’ expectations than was the content of lessons taught. Furthermore, the presentation ratings for the teachers using an integrated textbook were slightly lower than that of teachers using the subject-specific textbooks, but not statistically significantly different. As in the previous analyses, substantial variation was present within these ratings. For example, one teacher of integrated curriculum was rated consistently with a presentation rating of 5 across three lessons, whereas another teacher of integrated curriculum received ratings of 1, 2, and 2 across three lessons.

**Relationships Between Content Fidelity and Presentation Fidelity Indices**

The Kendall’s Tau correlation between the teacher content ratings and the teacher presentation ratings was $\tau = 0.50$ ($p < 0.01$). Thus, there is a moderate, statistically significant relationship between these two lesson implementation dimensions. In other words, those teachers who tended to have higher content ratings also tended to have higher presentation ratings. However, the size of this correlation suggests that Content and Presentation Fidelity should continue to be examined as separate dimensions of implementation in research studies where the textbook is a variable of interest.
As one component of the demographic information collected on the surveys, teachers were asked to provide the number of years they had taught at any grade level or in any subject area. The range of experience reported spanned a broad range from first-year teachers to a teacher with 28 years of experience. The mean years of overall teaching experience of the teachers in this study was 9.65 (SD = 7.5) years. When the mean was calculated with respect to teachers of particular curriculum types, teachers teaching the integrated curriculum had a mean of 10.21 (SD = 7.8) years while teachers teaching subject-specific courses had a mean of 9.13 (SD = 7.2) years of experience. As a follow-up to this question, teachers were asked how many years they had taught middle or high school mathematics. These means were 8.79 (SD = 7.2) and 7.70 (SD = 6.5) for teachers using an integrated textbook and teachers using a subject-specific textbook, respectively. Thus, among these teachers, most had taught mathematics during the majority of their teaching experience and there was only a minimal difference in the years of experience between the two types of curricula.

When examining the relationship between years of teaching experience and fidelity of implementation, across teachers regardless of the curriculum type, there was a small to moderate negative trend in the correlations between teaching experience and the content fidelity measure (see Table 2). Specifically, more experienced teachers tended to exercise more professional judgment with regard to content than did less experienced teachers. In other words, more experienced teachers were more likely to adjust the content of the lesson than were less experienced teachers, and this was particularly true when teachers were teaching classes where the textbook embodied an integrated approach to the organization of the mathematics content. Based on previous work, we found considerable teacher variation in the Content Fidelity dimension (McNaught, Grouws, & Tarr, 2008). Accordingly, these results considered in relation to our previous work suggest that more experienced teachers exercise this judgment about content with some frequency. As the correlations in Table 2 indicate, there was little relationship between years of teaching experience and Presentation Fidelity.

Table 2

<table>
<thead>
<tr>
<th></th>
<th>All teachers</th>
<th>IN Teachers</th>
<th>SS teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experience with Content Fidelity</td>
<td>-0.24*</td>
<td>-0.44**</td>
<td>-0.05</td>
</tr>
<tr>
<td>Math Exp with Content Fidelity</td>
<td>-0.25*</td>
<td>-0.36*</td>
<td>-0.12</td>
</tr>
<tr>
<td>Experience with Presentation Fidelity</td>
<td>-0.15</td>
<td>-0.12</td>
<td>-0.22</td>
</tr>
<tr>
<td>Math Exp with Presentation Fidelity</td>
<td>-0.07</td>
<td>-0.05</td>
<td>-0.12</td>
</tr>
</tbody>
</table>

*aKendall’s Tau Correlations
* - p < .05
** - p < .01

Another component of the demographic information collected on the teacher surveys asked
teachers to indicate the total amount of time they had spent in professional development in mathematics or in the teaching of mathematics for both short-term (last 12 months) and for a longer term (last 3 years). Teachers responded by choosing from the following categorical options: (1) None; (2) Less than 6 hours; (3) 6-15 hours; (4) 16-35 hours; or (5) More than 35 hours.

Our examination of the relationships between professional development and the Content and Presentation Fidelity indices basically indicated no relationship. This was the case for both short-term and longer-term professional development. Correlation coefficients by teachers and fidelity indices are shown in Table 3. Discussion of this lack of relationship is addressed in the next section.

**Table 3**

*Correlations* of Hours of Professional Development with Textbook Implementation Indices

<table>
<thead>
<tr>
<th></th>
<th>All teachers</th>
<th>IN Teachers</th>
<th>SS teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>PD 1 with Content Fidelity</td>
<td>-0.01</td>
<td>-0.02</td>
<td>-0.05</td>
</tr>
<tr>
<td>PD 2 with Content Fidelity</td>
<td>-0.15</td>
<td>-0.14</td>
<td>-0.15</td>
</tr>
<tr>
<td>PD 1 with Presentation Fidelity</td>
<td>-0.05</td>
<td>-0.13</td>
<td>-0.03</td>
</tr>
<tr>
<td>PD 2 with Presentation Fidelity</td>
<td>-0.10</td>
<td>-0.13</td>
<td>-0.04</td>
</tr>
</tbody>
</table>

*a* Kendall’s Tau Correlations  
*b* Professional development in the last 12 months  
*c* Professional development in the last 3 years

**Discussion**

In this study we conceptualized fidelity of implementation along two dimensions, *content* and *presentation* fidelity. We found that teachers tend to be more faithful to the textbook in terms of Content Fidelity than in terms of Presentation Fidelity. However, across teachers there is considerable variation on both dimensions. For example, we had 10 of 129 lessons for which the lesson taught was rated a 1 (scale of 1 to 5, with 1 = low and 5 = high), but on the other hand, we had 74 lessons for which the lesson taught was rated a 5. These results indicate that many teachers deviate from the textbook content and our field notes indicate they do this in a variety of ways. This result will be viewed negatively by textbook authors as they take care in sequencing topics, choosing the number and kind of exercises to include, and so on. The reasons for these deviations are uncertain at this time, but state-mandated testing and the associated course level expectations may play a role in these teacher decisions. The results clearly indicate, however, that simply knowing which textbook is used in mathematics classrooms is not sufficient to characterize the content of the mathematics instruction students experience.

The Presentation Fidelity ratings, which were lower than the Content Fidelity ratings, provide evidence that adopting a particular curriculum with a recommended instructional approach does not automatically yield changes in teachers’ instructional practices. These results support the contention of Ball and Cohen (1996) that teachers transform the textbook as they prepare and teach lessons. Many new curricular programs include presentation strategies that are not well developed as part of many teacher education programs (e.g., an investigative approach to instruction, use of small cooperative groups) and this may help to explain the wide range in Presentation Fidelity ratings we observed across teachers. For example, we had 13 of 129 lessons for which the teacher was rated a 1 on this measure (scale of 1 to 5, with 1 = low and 5 = high).
On the other hand, we had only 15 lessons for which the teacher’s rating was a 5. Noteworthy, there were five times as many lessons with a content rating of 5 than there were lessons with presentation rating of 5. This disparity is worthy of further research if we are to understand why it is more difficult to achieve presentation fidelity than content fidelity.

The results from this study indicate that more experienced teachers are more likely to deviate from the textbook content than are less experienced teachers. This finding may suggest that teachers with additional teaching experience more often sense that, in relation to their students, it is appropriate to amend the textbook content in a variety of ways including adding, deleting, or supplementing aspects of the mathematical ideas being taught. This was especially true when teachers were teaching from an integrated content approach. Whether teachers were considering the composition of students in their current class when deciding to modify the content of the textbook was not established in this study but it warrants further research, as does the contrary finding of little relationship between teaching experience and Presentation Fidelity.

The findings of little or no relationship between implementation indices and professional development might be considered a bit surprising. Reasons for the lack of relationship may be attributed to the quantity and quality of the professional development provided. For example, the professional development may not have been specific to the textbook used, may have been of insufficient length to change teaching behavior or perhaps it was not sustained with follow-up activities. Alternatively, it is possible that, for some teachers, teaching behavior is a deeply rooted cultural activity that is not easily changed regardless of the amount or quality of professional development provided. These explanations, while plausible, nevertheless require ongoing research into teachers’ use of curricular materials to ascertain the extent of their validity.

References


