High-Stakes Testing and Curricular Control: A Qualitative Metasynthesis
Author(s): Wayne Au
Source: *Educational Researcher*, Vol. 36, No. 5 (Jun. - Jul., 2007), pp. 258-267
Published by: American Educational Research Association
Stable URL: http://www.jstor.org/stable/30137912
Accessed: 11/11/2014 20:34

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

American Educational Research Association is collaborating with JSTOR to digitize, preserve and extend access to *Educational Researcher*.

http://www.jstor.org

This content downloaded from 128.192.114.19 on Tue, 11 Nov 2014 20:34:03 PM
All use subject to JSTOR Terms and Conditions

High-Stakes Testing and Curricular Control: A Qualitative Metasynthesis

by Wayne Au

Using the method of qualitative metasynthesis, this study analyzes 49 qualitative studies to interrogate how high-stakes testing affects curriculum, defined here as embodying content, knowledge form, and pedagogy. The findings from this study complicate the understanding of the relationship between high-stakes testing and classroom practice by identifying contradictory trends. The primary effect of high-stakes testing is that curricular content is narrowed to tested subjects, subject area knowledge is fragmented into test-related pieces, and teachers increase the use of teacher-centered pedagogies. However, this study also finds that, in a significant minority of cases, certain types of high-stakes tests have led to curricular content expansion, the integration of knowledge, and more student-centered, cooperative pedagogies. Thus the findings of the study suggest that the nature of high-stakes-test-induced curricular control is highly dependent on the structures of the tests themselves.

Keywords: curriculum theory; high-stakes testing; qualitative metasynthesis; template analysis.

With the advent of federally mandated high-stakes testing since the No Child Left Behind Act of 2001, many important questions have been raised regarding the implementation of this policy tool at the classroom level. In this article, I focus on one such question: What, if any, is the effect of high-stakes testing on curriculum? To answer this question, I begin by exploring the meanings of two key terms, "curriculum" and "high-stakes testing," and by offering a brief review of some of the literature regarding the relationship between the two. Then, using the method of qualitative metasynthesis, I undertake a comparative study of 49 qualitative studies of high-stakes testing to better understand testing's impact on curriculum.

Curriculum

There exists a wide range of definitions of the term "curriculum" (Beauchamp, 1982; Jackson, 1996; Kliebard, 1989). Historically, the word has its roots in the Latin word *currere*, which means a course to be run (Eisner, 1994), and was first used at the University of Glasgow in the 17th century to describe "a formal course of study that the students completed" (Harden, 2001, p. 335). This definition is perhaps the simplest and easiest for most to recognize because it is evident in the way schools are generally organized around a course of predetermined, required subject matter classes that students must pass to graduate. Thus most scholars and educators would at least recognize that curriculum encompasses a body of content knowledge to be learned in some way, shape, or form.

However, to stop at the level of content obscures other crucial aspects of curriculum because subject matter content within schools implies not only *selection* but also *transmission* of knowledge. As McEwan and Bull (1991) state,

> Subject matter is always an expression of a desire to communicate ideas to others. . . . Differences within the form and content of various expressions of subject matter reflect an understanding of differences in the backgrounds of potential audiences and the circumstances of the subject matter's formulation. (p. 331)

Indeed, all content is pedagogical. It implies the communication of ideas to an audience and does so through the structuring of knowledge (Segall, 2004a, 2004b). The concept of curriculum, therefore, also implicates the structure of knowledge embedded in curricular *form*—the form of how knowledge is organized and presented within a curriculum (Apple, 1995), as well as pedagogy—the intended form of communication of selected content. Thus the trilogy of (a) subject matter content knowledge, (b) structure or form of curricular knowledge, and (c) pedagogy are three defining aspects of "curriculum." This basic conception of curriculum is what I use for the present analysis.

High-Stakes Testing

A test is high-stakes when its results are used to make important decisions that affect students, teachers, administrators, communities, schools, and districts (Madaus, 1988). In very specific terms, high-stakes tests are a part of a *policy design* (Schneider & Ingram, 1997) that "links the score on one set of standardized tests to grade promotion, high school graduation and, in some cases, teacher and principal salaries and tenure decisions" (Orfield & Wald, 2000, p. 38). As part of the accountability movement, stakes are also deemed high because the results of tests, as well as the ranking and categorization of schools, teachers, and children that extend from those results, are reported to the public (McNeil, 2000).

The Research Debate

The question of whether high-stakes testing affects curriculum has been highly contested in the field of educational research. For instance, at a time when high-stakes testing policies were inconsistently implemented across individual states, Airasian (1987)

Educational Researcher, Vol. 36, No. 5, pp. 258–267
DOI: 10.3102/0013189X07306523
© 2007 AERA. http://er.aera.net

This content downloaded from 128.192.114.19 on Tue, 11 Nov 2014 20:34:03 PM
All use subject to JSTOR Terms and Conditions

and Madaus (1988) offered some of the earliest assertions that the tests would control classroom practice. M. L. Smith (1991) followed with one of the few early empirical studies, finding that high-stakes tests promote "multiple choice teaching." More recent research on high-stakes testing is more conflicted. Some research finds that high-stakes tests merely represent one limited factor, among others, influencing classroom practice (see, e.g., Cimbricz, 2002; Firestone, Mayrowetz, & Fairman, 1998; Grant, 2003), have little to no influence on what teachers do in the classroom (see, e.g., Gradwell, 2006; van Hover, 2006), or lead to improved learning experiences and positive educational outcomes (see, e.g., Braun, 2004; Williamson, Bondy, Langley, & Mayne, 2005). Other research challenges these claims, however, finding that high-stakes testing undermines education because it narrows curriculum, limits the ability of teachers to meet the sociocultural needs of their students, and corrupts systems of educational measurement (see, e.g., Amrein & Berliner, 2002a, 2002b; Lipman, 2004; McNeil, 2000; McNeil & Valenzuela, 2001; Nichols & Berliner, 2005, 2007; Watanabe, 2007). Given the wide range of research evidence, and given the ubiquity of high-stakes testing in education in the United States, the purpose of this study is to develop a broader, more complex understanding of the ways that these tests influence curriculum at the classroom level.

Method

For the purposes of this study I have chosen to analyze examples of qualitative research because of their focus on human interaction and attention to the day-to-day functioning of schools and classrooms (Valenzuela, Prieto, & Hamilton, 2007). To review the body of evidence reported in qualitative studies, I draw on the methodology of qualitative metasynthesis (DeWitt-Brinks & Rhodes, 1992; Noblit & Hare, 1988; Sandelowski, Docherty, & Emden, 1997; Thorne, Jensen, Kearney, Noblit, & Sandelowski, 2004), also referred to as qualitative meta-analysis (McCormick, Rodney, & Varcoe, 2003). Qualitative metasynthesis is part of a tradition of metaresearch that involves synthesizing the results of qualitative studies to gain a better understanding of the general nature of a given phenomenon (DeWitt-Brinks & Rhodes, 1992; Thorne et al., 2004).

In this study I make use of a specific form of qualitative metasynthesis known as template analysis (Crabtree & Miller, 1999; King, 1998, 2006). In this form of thematic meta-analysis, textual data are coded using a template of codes designed by the researcher. These codes are often hierarchical in nature, starting with broad themes and moving toward more narrow or specific ones. In this case the textual data used are from the collection of qualitative studies gathered by the researcher. In template analysis the coding template is developed in two stages based on themes that arise from the body of textual data. In the first stage the researcher begins by developing an initial template based on a combination of a priori codes and an initial reading and coding of a subset of the textual data. In the second stage, the initial template is then applied to the whole data set, and codes are added to the template as new themes arise. This leads to the creation of the final template. The final template is then used to interpret the textual data set as a whole, and the findings are presented in some form (King, 1998, 2006).

Data Collection

The data set consists of 49 qualitative studies. These studies were gathered from a search completed in June of 2006 using the Educational Resources Information Center (ERIC), Academic Search, and Education Full Text databases, as well as the library book database at the University of Wisconsin, Madison. Initially, the search terms "high-stakes testing" and "state-mandated testing" were used to identify potential studies for use in my qualitative metasynthesis. This rather large initial pool was then narrowed to studies (a) based on original, scholarly research, (b) using qualitative methods, (c) taking place in the United States, and (d) specifically addressing the relationship between high-stakes tests and either curriculum or instruction, or both. Because this study focuses on the relationship between high-stakes testing and curriculum at the K–12 classroom level, the sample excludes studies that examine the relationship between high-stakes testing and retention, studies that focus on the role of high-stakes testing and access to teacher education programs (e.g., Praxis II), studies that focus on the tests themselves (e.g., discourse analyses of the actual test content), and policy studies that use qualitative methods to compare pressures between states. In addition, because of their ambiguous and complicated positions in school hierarchies, studies that focus on student teachers are also excluded.

Based on the self-identification of the researchers, the data gathered and analyzed from the 49 studies used in the qualitative metasynthesis performed here include at least 740 "teachers" identified as participants; 845 "educators" or "teachers and administrators" (not broken out into "teachers" alone) identified as participants; 96 schools identified as the focus of study; 38 districts identified as the level of focus of study; and covers at least 19 states (Arizona, Colorado, Florida, Illinois, Kansas, Kentucky, Maine, Maryland, Massachusetts, Michigan, Minnesota, New York, North Carolina, Ohio, Oregon, Texas, Vermont, Virginia, and Washington). In addition, of the 49 qualitative studies used in this metasynthesis, 15 focus on elementary education, 23 focus on secondary education, and 11 are K–12 analyses. Alternatively, while several of the included studies (23) are more general in focus, 14 are history/social studies–specific (3 elementary and 11 secondary), 9 are English/language arts–specific (1 elementary and 8 secondary), and 3 are math/science–specific. (See Table 1 for a complete listing of the studies analyzed here.)

Data Analysis

For this study I tracked the citation information, research sites, scope, and methods of inquiry of the 49 qualitative studies, including the dominant themes in each study's findings. I then coded dominant themes using the above definition of curriculum as the framework for my initial template of analysis. Thus my thematic coding began with three broad categories: Subject Matter Content, Pedagogy, and Structure of Knowledge. Consistent with the template analysis methodological framework, the full elaboration of my coding template evolved during the course of the research. For instance, it has been widely asserted over the past 20-plus years that high-stakes tests cause a narrowing or contraction of nontested subject areas. I was aware of research substantiating this assertion prior to beginning the template analysis and thus assumed that I would need to code the studies that reported the theme of contraction of subject matter content. Based on my previous understandings and

This content downloaded from 128.192.114.19 on Tue, 11 Nov 2014 20:34:03 PM
All use subject to JSTOR Terms and Conditions

Table 1
Qualitative Metasynthesis Studies and Codes

Article	Codes[a]
Agee, 2004	SAC, PCT, KCF
Anagnostopolous, 2003a	SAC, SAE, PCT, KCF
Anagnostopolous, 2003b	SAC, PCT, KCF
Barton, 2005	SAE, KCI
Bol, 2004	PCT, KCF
Bolgatz, 2006	KCI
Booher-Jennings, 2005	PCT
Brimijoin, 2005	SAC, PCT
Clarke et al., 2003	SAC, SAE, PCT, PCS, KCI
Costigan, 2002	PCT
Debray, Parson, & Avila, 2003	SAC
Fickel, 2006	SAE, PCT, KCI
Firestone, Mayrowetz, & Fairman, 1998	SAC, PCT, KCF
Gerwin & Visone, 2006	SAC, PCT, KCF
Gradwell, 2006	No changes[b]
Grant, 2003	No changes[b]
Grant et al., 2002	SAC, PCT, KCF
Groves, 2002	SAC, PCT
Hillocks, 2002	SAC, KCF
Landman, 2000	SAC, PCT, KCF
Libresco, 2005	SAE, PCS, KCI
Lipman, 2002	SAC, PCT, KCF
Lomax et al., 1995	SAC, PCT, KCF
Luna & Turner, 2001	SAC, SAE, PCT, KCF
McNeil, 2000	SAC, PCT, KCF
McNeil & Valenzuela, 2001	SAC, PCT, KCF
Murillo & Flores, 2002	SAC
Passman, 2001	SAC, PCT, KCF
Perreault, 2000	SAC, PCT
Renter et al., 2006	SAC
Rex, 2003	SAC, PCT
Rex & Nelson, 2004	SAE, PCS, KCI
Salinas, 2006	SAE, SAC
Segall, 2003	SAE, SAC, PCT, KCF, KCI
Siskin, 2003	SAC
Sloan, 2005	SAC, PCT, KCF
Smagorinsky, Lakly, & Johnson, 2002	SAC, PCT, KCF
Smith, A. M., 2006	SAE, SAC, PCT, KCF
Taylor et al., 2001	SAC, PCT, KCF
Valenzuela, 2000	SAC, PCT, KCF
van Hover, 2006	SAC, KCF
van Hover & Heinecke, 2005	PCT, KCF
Vogler, 2003	SAC, SAE
Williamson et al., 2005	PCT
Wolf & Wolf, 2002	SAE, PCS, KCI
Wollman-Bonilla, 2004	SAE, PCS, KCI
Wright & Choi, 2005	SAC
Yeh, 2005	SAE, PCS, KCI
Zancanella, 1992	SAC, PCT, KCF

[a]See Table 2 and the text discussion of it for explanations of the codes.
[b]These two studies reported no curricular changes in response to high-stakes testing.

on an initial analysis of qualitative studies, I produced an initial template of codes. However, after I undertook the template analysis, for instance, in addition to finding the theme of narrowing/contraction

Table 2
Qualitative Metasynthesis Code Template

Curriculum	Code
Content	SAC—Subject matter content alignment, contraction
	SAE—Subject matter content alignment, expansion
Knowledge form	KCF—Form of knowledge changed, fractured
	KCI—Form of knowledge changed, integrated
Pedagogy	PCT—Pedagogic change to teacher-centered
	PCS—Pedagogic change to student-centered

of curriculum to align with high-stakes tests, I also encountered the theme of subject matter content expansion. This finding required the addition of a new thematic code. As I read and reread the 49 qualitative studies, I added thematic codes as the patterns emerged and used them to develop the final template of codes for metasynthesis.

The thematic codes in Table 2 can be explained as follows. The first set of thematic codes seeks to track whether teachers, as individual actors at the classroom level, aligned their classroom content to the high-stakes tests. If they did, the thematic codes then mark the nature of this alignment—either subject matter content expansion or subject matter content contraction. In looking for subject matter contraction, I studied the research findings for occurrences of teachers and schools reducing the amount of instructional time and course offerings in either tested or nontested subject areas. An example of findings being coded for content matter expansion can be found in the research of Renter and colleagues (Renter et al., 2006), who found that schools were reducing the amount of instruction in science and social studies because those subjects were not a focus of the high-stakes tests. Conversely, in looking for subject matter expansion, I analyzed the data for reports of teachers and schools increasing the teaching of either tested or nontested subjects in response to high-stakes tests. Vogler (2003) is an example of a study that was coded for test-related content expansion because he found that social studies teachers in his study added language arts/literacy instruction to their social studies curriculum in response to high-stakes tests, which tested for writing but not for social studies content knowledge.

The second set of thematic codes tracked whether the high-stakes tests affected curricular knowledge forms. This theme was perhaps the most elusive of the three because it required that I follow how teachers organized the knowledge in their classrooms in relation to high-stakes testing. If a study reported that there was a shift in how teachers structured the knowledge they taught, I then coded for whether classroom knowledge forms became more fragmented and isolated into discrete, test-driven bits or became more expansive, inclusive in integrated wholes. As an instance of a study being coded for knowledge fragmentation, one study in this metasynthesis found that math and science were increasingly being taught as a collection of procedures and facts, as opposed to being taught as conceptual, thematic, and higher-order mathematic and scientific thinking (Lomax, West, Harmon, Viator, & Madaus, 1995). Such test-influenced instruction thus essentially fragmented

This content downloaded from 128.192.114.19 on Tue, 11 Nov 2014 20:34:03 PM
All use subject to JSTOR Terms and Conditions

the content knowledge into individuated and isolated procedures and facts for use on the high-stakes test. Other examples can be found where researchers reported that subjects such as social studies were broken up into collections of historical data (see, e.g., Grant et al., 2002) or subjects such as writing were reduced to the production of formulaic and procedural five-paragraph essays (see, e.g., Hillocks, 2002). Conversely, more integrated knowledge forms were coded in studies that found, for instance, some teachers focusing on more conceptual, higher-order thinking that sought to develop more holistic understanding of mathematics (see, e.g., Firestone et al., 1998) or studies that found language arts teachers focusing more conceptually on the process of writing as opposed to step-by-step procedural essay writing (see, e.g., Hillocks, 2002).

Third, I looked at the theme of teachers' pedagogy in response to high-stakes tests. If a study reported that teachers changed their instructional practice because of the testing, then I coded for the theme of teacher-centered instructional strategies or the theme of student-centered instructional strategies. In tracking these themes, I analyzed the studies' findings for evidence of teachers' increasing their use of direct instruction or increasing their use of more interactive pedagogies in response to the tests. For instance, in their research into high-stakes-testing-related social studies instruction, Gerwin and Visone (2006) found that teachers in their study showed dramatic increases in the amount of teacher-centered, fact-driven instruction in subjects included in state-mandated tests. Studies such as this were coded as demonstrating increased teacher-centered pedagogy. Studies reporting teachers' increasing the amount of student-centered, constructivist instruction in response to high-stakes tests, for example, some studies of language arts classrooms where teachers increased their use of interactional and student-led activities (see Wollman-Bonilla, 2004), were coded accordingly.

Once coding was completed, I analyzed the codes for patterns and anomalies on three levels. First, looking at the data as a whole collection, I tracked the predominant themes in terms of individual codes, essentially asking, What do these studies tell us about the overall effects of high-stakes testing on curriculum in terms of content, form, and pedagogy? Within this first level of analysis, I then sought to find relationships between the trends at the level of the single codes and other contextual variables found within the research, looking for overlaps between grade levels and subject areas and the trends found among individual themes.

At the second level I analyzed theme pairings. This involved tracking the number of times that particular codes appeared in corresponding pairs to determine if any relationships existed between changes in content, knowledge structures, and pedagogy. At this level of analysis, I also tracked whether the pairings corresponded to particular grade levels or subject areas.

Finally, at the third level, I analyzed theme triplets, seeking any potential connections between all three areas of content, pedagogy, and knowledge form in relation to the effects of high-stakes testing on classroom practice.

In addition to these three levels of analysis, I looked at the anomalies or weaker thematic relationships. Some studies simply came up with singular findings that did not match or support the trends and patterns of the larger metasynthesis; some groups of studies (such as are found within the social studies) were more conflicted in their findings.

Study Reliability

Reliability is a known issue within template analysis (King, 1998, 2006; Pawson, Greenhalgh, Harvey, & Walshe, 2005), and I have used two strategies to ensure the reliability of the findings of this study. First, to empirically determine the interrater reliability of my own coding, two colleagues independently coded findings of a sample subset of 10 studies. The findings of these coders were then checked against my own, resulting in the following interrater reliability percentages: subject matter content contraction, 86.7%; subject matter content expansion, 83.3%; knowledge fragmentation, 93.3%; knowledge integration, 96.7%; teacher-centered pedagogy, 90%; student-centered pedagogy, 86.7%. The overall interrater reliability for this study was 89.4%.

Second, reliability in template analysis is also improved when researchers are explicitly reflexive about both the process of their research and their positioning in relation to their study (King, 1998, 2006; Pawson et al., 2005). Thus it is important to explain my research orientation. I approach this study from within the critical realist tradition, which holds that a real world exists objectively outside human perception, that this world is to varying extents knowable through human cognition, and that this world is in fact changeable relative to our knowledge of it. Furthermore, critical realism recognizes human subjectivity in the understanding of the externally existing world, and as such views knowledge as a social process and as fallible. In these ways, critical realism simultaneously rejects both positivist objectivist *and* relativist subjectivist theories of knowledge in favor of an epistemology that in essence synthesizes aspects of both—an objectively existing world and a socially mediated understanding of that world (Benton & Craib, 2001; Bhaskar, 1989). Consequently, my use of template analysis combined with critical realism makes this study a form of realist review (Pawson et al., 2005).

My critical realist positioning also influences this study in that the use of the word "critical" points to a particular set of political commitments on the part of the researcher. Critical realists seek to understand the world to change it for the better, seek to reflexively understand social mechanisms to promote social equality (Benton & Craib, 2001; Bhaskar, 1989). A similar political commitment underlies the impetus for this study, because I, as a social justice educator, scholar, and activist, have sought to understand the relationship between education and power (see, e.g., Au, 2005, 2006; Au & Apple, 2004). As such, I am interested in the relationship between high-stakes testing and inequalities associated with race and socioeconomic status (see, e.g., Hunter & Bartee, 2003; Kim & Sunderman, 2005; Sirin, 2005). However, although ultimately inseparable from my overall research agenda, for the purposes of this study I have attempted to put my political commitments aside in favor of a focused empirical analysis of how high-stakes testing affects curriculum. Thus, although these effects may have implications for educational equality and social justice, I have made a conscious choice here to bracket those implications as beyond the scope of this specific study and analysis.

Study Limits

Before presenting the findings, it is important to recognize that this study has a specific focus and is therefore limited in at least two particular ways. First, in this metasynthesis I inquire into the frequency

This content downloaded from 128.192.114.19 on Tue, 11 Nov 2014 20:34:03 PM
All use subject to JSTOR Terms and Conditions

Table 3
Summary Findings: Effects of High-Stakes Testing on Curriculum

Curricular Change	Number of Studies, $N = 49$	Percentage of Total	Exemplar of Dominant Theme
Subject matter	41	83.7%	A Colorado teacher: "Our district has told us to focus on reading, writing, and mathematics. Therefore, science and social studies . . . don't get taught." Taylor et al., 2001, p. 30
Contraction	34	69.4%	
Expansion	14	28.6%	
Knowledge form	34	69.4%	A Massachusetts teacher: "You know, we're not really teaching them how to write. We're teaching them how to follow a format. . . . It's like . . . they're doing paint-by-numbers." Luna & Turner, 2001, p. 83
Fractured	24	49%	
Integrated	10	20.4%	
Pedagogy	38	77.6%	A Kansas teacher: " . . . I don't get to do as many fun activities, like cooperative learning activities or projects. . . . [T]his year I've done a lot more direct teaching than being able to do student-led learning. . . ." Clarke et al., 2003, p. 50
Teacher-centered	32	65.3%	
Student-centered	6	12.3%	

Note. Individual code totals do not necessarily equal the total for any one category because some studies exhibit multiple, even contradictory, codes; for example, subject alignment contraction and subject alignment expansion may appear in the same study.

and types of curricular change induced by high-stakes testing. Consequently, my inquiry excludes instances where high-stakes testing does not affect the curriculum. As this study's findings will show, the body of research analyzed here focuses predominantly on test-related events, as opposed to test-related nonevents. In this regard, even though a handful of studies included here specifically focus on a lack of test-related instructional changes (see, e.g., Bolgatz, 2006; Gradwell, 2006; Grant, 2003), the findings of this qualitative metasynthesis are inherently skewed toward what the researchers in the majority of these studies chose to focus on in their research: classroom-level changes due to high-stakes tests.

A second way in which the findings of this qualitative metasynthesis are limited relates to the time periods reported on. The studies analyzed here report inconsistently on how curriculum changes in response to high-stakes testing relative to time. Thus some studies focus on periods of curricular change in the months, weeks, or days leading up to high-stakes tests, and others focus on test-related curricular change more generally. Consequently, it was difficult to ascertain whether high-stakes testing was affecting the curriculum all year or simply in time periods immediately preceding the tests. I would argue, however, that these two limits do not take away from the power of the findings presented here. Rather, the limits simply refine the focus of this qualitative metasynthesis, which provides a snapshot and general depiction of the types and frequency of changes made to curricula in high-stakes testing environments.

Findings

As Table 3 indicates, the findings of this study suggest that there is a significant relationship between the implementation of high-stakes testing and changes in the content of a curriculum, the structure of knowledge contained within the content, and the types of pedagogy associated with communication of that content. These changes represent three types of control that high-stakes tests exert on curriculum: content control, formal control, and pedagogic control.

Content Control

The dominant theme found in the qualitative research regarding high-stakes testing and curriculum is that of content alignment. More than 80% of the studies contained the theme of curricular content change, whether by contraction or expansion. Furthermore, as Table 3 shows, in an overwhelming number of the qualitative studies, participants reported instances of the narrowing of curriculum, or curricular contraction to tested subjects. This phenomenon was the most prominent way in which "teaching to the test" manifested in curricula, as nontested subjects were increasingly excluded from curricular content. A more detailed analysis finds that the narrowing of curricular content was strongest among participants in the studies that focused on secondary education, with the most narrowing found in studies of social studies and language arts classrooms. In addition, in another expression of curricular alignment, a significant minority of studies reported some form of content expansion as a result of high-stakes testing, with most of these coming from studies focusing on secondary education and social studies classrooms. As the above evidence suggests, whether in the form of content contraction or content expansion, high-stakes testing leverages a significant amount *content control* over curriculum.

Formal Control

Table 3 also indicates that, in a significant number of the qualitative studies, participants reported changes to the form that curricular knowledge took in response to high-stakes testing. The dominant theme in this category suggests that there is a relationship between high-stakes testing and teachers' increasing the fragmentation of knowledge. Such fragmentation manifested in the teaching of content in small, individuated, and isolated test-size pieces, as well as teaching in direct relation to the tests rather than in relation to other subject matter knowledge. However, it is important to note that, as shown in Table 3, a minority of studies found that high-stakes testing had led to the increased integration of knowledge in the classroom. Thus, within the body of qualitative research, a dominant theme is that, whether leading to fragmentation or integration of knowledge, high-stakes testing affects curricular form, that is, it leverages *formal control* over the curriculum.

Pedagogic Control

A third dominant theme that appears in the qualitative research is pedagogic change. As shown in Table 3, a significant number of participants in qualitative studies reported that their pedagogy

This content downloaded from 128.192.114.19 on Tue, 11 Nov 2014 20:34:03 PM
All use subject to JSTOR Terms and Conditions

changed in response to high-stakes tests and that a significant majority of the changes included an increase in teacher-centered instruction associated with lecturing and the direct transmission of test-related facts. In addition, as Table 3 indicates, a small but important number of studies exhibited the theme of increased student-centered instruction as an effect of high-stakes testing. Further analysis shows that, in this metasynthesis, a cluster of test-related, teacher-centered pedagogy exists surrounding instruction in both language arts and social studies classrooms. Whether in the form of increased teacher-centered instruction or increased student-centered instruction, the evidence suggests that high-stakes testing exerts significant *pedagogic control* over curriculum.

Theme Pairings

An analysis of theme pairings generally mirrors the above findings but also provides a more nuanced outline of potentially significant relationships between dominant themes.

As Table 4 indicates, the most prominent theme pairing suggests that there is a relationship between the narrowing of curriculum and an increase in teacher-centered instruction as teachers respond to pressures created by high-stakes testing environments. The next highest occurrence of theme pairing suggests that increased teacher-centered pedagogy and increased fragmentation of knowledge forms are likely to coincide in response to high-stakes testing. The third most frequent theme pairing suggests a relationship between curricular content narrowing and the fragmentation of knowledge forms, which are likely to occur together in response to high-stakes testing.

The findings further suggest that there are weaker but significant relationships between the expansion of subject matter and an increase of a more integrated structure of knowledge in response to high-stakes testing, as well as a contraction or narrowing of curricular content and a simultaneous content expansion. Three other significant theme pairings appear in the study, two of which are seemingly contradictory to the dominant trends outlined above. As Table 4 shows, theme pairing of curricular expansion and an increase in teacher-centered pedagogy in response to high-stakes testing was also found. Other findings showed increases in student-centered pedagogy paired with an increase in the integration of knowledge in response to high-stakes testing.

Theme Triplets

A total of 28 studies in this qualitative metasynthesis produced codes within each area of curriculum identified here. I now turn to the final level of analysis, examining these theme triplets to determine if there are any potential relationships between all three thematic areas. Overwhelmingly, the prevalent theme triplet in the qualitative research was the combination of contracting curricular content, fragmentation of the structure of knowledge, and increasing teacher-centered pedagogy in response to high-stakes testing. This theme triplet appears 21 times (75%) among the 28 studies that produced themes in all three areas, suggesting a relationship between the themes in response to high-stakes testing. The second most frequently occurring theme triplet, that of curricular content expansion, increasing integration of knowledge, and increasing student-centered instruction, appears 6 times (21.4%) in the study. This triplet is indeed the exact opposite of the dominant triplet.

Table 4
Summary of Selected Theme Pairings

Theme Pairing	Occurrence
Content contraction/teacher-centered pedagogy	26/37 (70.3%)
Teacher-centered pedagogy/knowledge fragmentation	23/35 (65.7%)
Content contraction/knowledge fragmentation	22/34 (64.7%)
Content expansion/knowledge integration	9/34 (26.5%)
Student-centered pedagogy/knowledge integration	6/35 (17.1%)
Content contraction/content expansion	7/43 (16.3%)
Content expansion/teacher-centered pedagogy	6/37 (16.2%)
Content expansion/student-centered pedagogy	6/37 (16.2%)

Discussion

Despite some researchers' claims to the contrary, the findings of this study suggest that high-stakes tests encourage curricular alignment to the tests themselves. This alignment tends to take the form of a curricular content narrowing to tested subjects, to the detriment or exclusion of nontested subjects. The findings of this study further suggest that the structure of the knowledge itself is also changed to meet the test-based norms: Content is increasingly taught in isolated pieces and often learned only within the context of the tests themselves. Finally, in tandem with both content contraction and the fragmentation of knowledge, pedagogy is also implicated, as teachers increasingly turn to teacher-centered instruction to cover the breadth of test-required information and procedures. Thus I have identified three different, interrelated types of curricular control associated with high-stakes testing: *content, formal,* and *pedagogic.* The control over knowledge content and the form the knowledge takes are related to and associated with control of pedagogy as well.

As I noted in Tables 3 and 4, however, several less frequently occurring themes seemed to contradict the predominant findings of this study. The data suggest that in a small number of cases, high-stakes testing was associated with an increase in student-centered instruction, content integration, and subject matter expansion. For instance, there are seven simultaneous occurrences of the themes of content contraction and content expansion related to high-stakes tests, most of which come from secondary social studies and language arts (see, e.g., Anagnostopolous, 2003b; Luna & Turner, 2001; Segall, 2003; A. M. Smith, 2006; Vogler, 2003). In these cases, teachers are both adding some content to meet the demands of the tests and contracting content in other areas. In addition, because the stakes of state-mandated social studies testing vary greatly from state to state (Grant & Horn, 2006), the findings indicate that high-stakes-test-induced curricular expansion has taken place in social studies classrooms as teachers integrate reading-test-related literacy skills into their own social studies curricula (see, e.g., Vogler, 2003). Indeed, this phenomenon of expanding curricular

This content downloaded from 128.192.114.19 on Tue, 11 Nov 2014 20:34:03 PM
All use subject to JSTOR Terms and Conditions

content due to the integration of test-required literacy skills or test-specific content accounts for the majority of the instances of curricular expansion (see, e.g., Barton, 2005; Clarke et al., 2003; Libresco, 2005; Rex & Nelson, 2004; Wolf & Wolf, 2002; Wollman-Bonilla, 2004; Yeh, 2005).

There appears to be a similar relationship regarding the small numbers of increases in student-centered pedagogies relative to high-stakes testing. Almost all occurrences of the theme of increases in student-centered pedagogy occur with instances of subject matter expansion. These cases revolve around teachers whose test-based instruction involves the development of critical literacy skills (see, e.g., Clarke et al., 2003; Libresco, 2005; Rex & Nelson, 2004; Wolf & Wolf, 2002; Wollman-Bonilla, 2004; Yeh, 2005). For instance, New York State's history exam involves a mix of multiple-choice questions and a document-based essay question (DBQ; Grant, 2003). Social studies teachers, in preparing students for DBQs, have the charge of teaching a specific critical literacy skill set instead of being forced to focus solely on a rigidly imposed collection of historical facts (see, e.g., Bolgatz, 2006; Clarke et al., 2003; Grant, 2003; Libresco, 2005). It is likely that teachers in these studies thus find the potential for increased flexibility in the content and pedagogy they use to teach social studies in their respective high-stakes environments. Furthermore, because social studies instruction figures prominently in the above contradictory findings, and because the only two studies to argue that testing does not influence any aspect of curriculum also focus on this subject area (Gradwell, 2006; Grant, 2003), it is also possible that social studies represents a special case in relation to high-stakes testing and curricular control (Au, in press).

The above discussion indicates a likely relationship between the construction of the high-stakes tests themselves and the curricular changes induced by the tests. Research supports the existence of such a relationship. As Yeh (2005) finds, teachers in Minnesota report that their pedagogy is not negatively affected by high-stakes tests because they feel the tests there are well designed and do not promote drill and rote memorization. Another example comes from Hillocks (2002), who analyzes the teaching of writing in relation to the writing examinations delivered in Texas, Illinois, New York, Oregon, and Kentucky. One of Hillocks's main findings is that states with poorly designed systems of writing assessment promote a technical, mechanical, five-paragraph essay form, and that teachers' pedagogy adapts to that form in those states. The findings of these studies suggest that test construction matters in terms of teachers' curricular responses to high-stakes tests (see also Clarke et al., 2003).

Conclusion

In this study, using a form of qualitative metasynthesis called template analysis, I have reviewed the findings of 49 qualitative studies addressing the impact of high-stakes testing on curriculum. As Tables 3 and 4 indicate, the evidence presented here strongly suggests that as teachers negotiate high-stakes testing educational environments, the tests have the predominant effect of narrowing curricular content to those subjects included in the tests, resulting in the increased fragmentation of knowledge forms into bits and pieces learned for the sake of the tests themselves, and compelling teachers to use more lecture-based, teacher-centered pedagogies. Another significant finding of this study is that, in a minority of cases, high-stakes tests have led to increases in student-centered pedagogy and increases in content knowledge integration. Combined, these findings indicate that high-stakes testing exerts significant amounts of control over the content, knowledge forms, and pedagogies at the classroom level.

The curricular control found in this study further suggests that high-stakes testing represents the tightening of the loose coupling between policymakers' intentions and the institutional environments created by their policies (Burch, 2007). This conclusion should not be surprising to educational researchers and practitioners because systems of educational accountability built on high-stakes, standardized tests are in fact intended to increase external control over what happens in schools and classrooms. As Moe (2003) explains, the rationale behind systems of high-stakes accountability is quite clear:

> The movement for school accountability is essentially a movement for more effective top-down control of the schools. The idea is that, if public authorities want to promote student achievement, they need to adopt organizational control mechanisms—tests, school report cards, rewards and sanctions, and the like—designed to get district officials, principals, teachers, and students to change their behavior. . . . Virtually all organizations need to engage in top-down control, because the people at the top have goals they want the people at the bottom to pursue, and something has to be done to bring about the desired behaviors.
>
> The public school system is just like other organizations in this respect. (p. 81)

The intentions of promoters of high-stakes test-based educational reforms are thus apparent in the policy designs, which are purposefully constructed to negate "asymmetries" between classroom practice and the policy goals of those with political and bureaucratic power (Wößmann, 2003).

Given the central findings of this study, however, a crucial question is raised: Are test-driven curriculum and teacher-centered instruction good or bad for teachers, students, schools, communities, and education in general? Considering the body of research connecting high-stakes testing with increased drop-out rates and lower achievement for working-class students and students of color (see, e.g., Amrein & Berliner, 2002b; Groves, 2002; Madaus & Clarke, 2001; Marchant & Paulson, 2005; Nichols, Glass, & Berliner, 2005), the findings of this study point to the need for further analysis of how curricular control may or may not contribute to educational inequality.

NOTE

I would like to thank Diana Hess, Simone Schweber, Keita Takayama, Ross Collin, Eduardo Cavieres, Quentin Wheeler-Bell, the three anonymous *ER* reviewers, and *ER* editor Gregory Camilli for their invaluable feedback on this article.

REFERENCES

References marked with an asterisk indicate studies included in the metasynthesis.

*Agee, J. (2004). Negotiating a teaching identity: An African American teacher's struggle to teach in test-driven contexts. *Teachers College Record, 106*(4), 747–774.

This content downloaded from 128.192.114.19 on Tue, 11 Nov 2014 20:34:03 PM
All use subject to JSTOR Terms and Conditions

Airasian, P. W. (1987). State mandated testing and educational reform: Context and consequences. *American Journal of Education, 95*(3), 393–412.

Amrein, A. L., & Berliner, D. C. (2002a, December). *An analysis of some unintended and negative consequences of high-stakes testing.* Tempe, AZ: Educational Policy Studies Laboratory, Arizona State University. Retrieved February 12, 2006, from http://www.asu.edu/educ/epsl/EPRU/epru_2002_Research_Writing.htm

Amrein, A. L., & Berliner, D. C. (2002b). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives, 10*(18). Retrieved September 27, 2005, from http://epaa.asu.edu/epaa/v10n18

*Anagnostopolous, D. (2003a). The new accountability, student failure, and teachers' work in urban high schools. *Educational Policy, 17*(3), 291–316.

*Anagnostopolous, D. (2003b). Testing and student engagement with literature in urban classrooms: A multi-layered perspective. *Research in the Teaching of English, 38*(2), 177–212.

Apple, M. W. (1995). *Education and power* (2nd ed.). New York: Routledge.

Au, W. (2005). Power, identity, and the third rail. In P. C. Miller (Ed.), *Narratives from the classroom: An introduction to teaching* (pp. 65–85). Thousand Oaks, CA: Sage.

Au, W. (2006, November). Against economic determinism: Revisiting the roots of neo-Marxism in critical educational theory. *Journal for Critical Education Policy Studies, 4*(2). Retrieved December 12, 2006, from http://www.jceps.com/?pageID=article&articleID=66

Au, W. (in press). Social studies, social justice: W(h)ither the social studies in high-stakes testing? *Teacher Education Quarterly.*

Au, W., & Apple, M. W. (2004). Interrupting globalization as an educational practice. *Educational Policy, 18*(5), 784–793.

*Barton, K. C. (2005). "I'm not saying these are going to be easy": Wise practice in an urban elementary school. In E. A. Yeager & O. L. Davis Jr. (Eds.), *Wise social studies teaching in an age of high-stakes testing* (pp. 11–31). Greenwich, CT: Information Age Publishing.

Beauchamp, G. A. (1982). Curriculum theory: Meaning, development, and use. *Theory Into Practice, 21*(1), 23–27.

Benton, T., & Craib, I. (2001). *Philosophy of social science: The philosophical foundations of social thought.* New York: Palgrave.

Bhaskar, R. (1989). *Reclaiming reality: A critical introduction to contemporary philosophy* (2nd ed.). New York: Verso.

*Bol, L. (2004). Teachers' assessment practices in a high-stakes testing environment. *Teacher Education and Practice, 17*(2), 162–181.

*Bolgatz, J. (2006). Using primary documents with fourth-grade students: Talking about racism while preparing for state-level tests. In S. G. Grant (Ed.), *Measuring history: Cases of state-level testing across the United States* (pp. 133–156). Greenwich, CT: Information Age Publishing.

*Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas accountability system. *American Educational Research Journal, 42*(2), 231–268.

Braun, H. (2004). Reconsidering the impact of high-stakes testing. *Education Policy Analysis Archives, 12*(1). Retrieved March 29, 2006, from http://epaa.asu.edu/epaa/v12n1

*Brimijoin, K. (2005). Differentiation and high-stakes testing: An oxymoron? *Theory Into Practice, 44*(3), 254–261.

Burch, P. E. (2007). Educational policy and practice from the perspective of institutional theory: Crafting a wider lens. *Educational Researcher, 36*(2), 84–95.

Cimbricz, S. (2002). State testing and teachers' thinking and practice. *Education Policy Analysis Archives, 10*(2). Retrieved September 4, 2006, from http://epaa.asu.edu/epaa/v10n2.html

*Clarke, M., Shore, A., Rhoades, K., Abrams, L. M., Miao, J., & Li, J. (2003, January). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from interviews with educators in low-, medium-, and high-stakes states.* Boston: National Board on Educational Testing and Public Policy, Lynch School of Education, Boston College. Retrieved March 20, 2006, from http://www.bc.edu/research/nbetpp/reports.html

*Costigan, A. T., III. (2002). Teaching the culture of high stakes testing: Listening to new teachers. *Action in Teacher Education, 23*(4), 28–34.

Crabtree, B. F., & Miller, W. L. (1999). Using codes and code manuals: A template organizing style of interpretation. In B. F. Crabtree & W. L. Miller (Eds.), *Doing qualitative research (2nd ed., pp. 163–178).* Thousand Oaks, CA: Sage.

*Debray, E., Parson, G., & Avila, S. (2003). Internal alignment and external pressure. In M. Carnoy, R. Elmore, & L. S. Siskin (Eds.), *The new accountability: High schools and high-stakes testing* (pp. 55–85). New York: RoutledgeFalmer.

DeWitt-Brinks, D., & Rhodes, S. C. (1992, May 20–25). *Listening instruction: A qualitative meta-analysis of twenty-four selected studies.* Paper presented at the annual meeting of the International Communication Association, Miami, FL.

Eisner, E. W. (1994). *The educational imagination: On the design and evaluation of school programs* (3rd ed.). New York: Macmillan.

*Fickel, L. H. (2006). Paradox of practice: Expanding and contracting curriculum in a high-stakes climate. In S. G. Grant (Ed.), *Measuring history: Cases of state-level testing across the United States* (pp. 75–103). Greenwich, CT: Information Age Publishing.

*Firestone, W. A., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis, 20*(2), 94–113.

*Gerwin, D., & Visone, F. (2006). The freedom to teach: Contrasting history teaching in elective and state-tested course. *Social Education, 34*(2), 259–282.

*Gradwell, J. M. (2006). Teaching in spite of, rather than because of, the test: A case of ambitious history teaching in New York State. In S. G. Grant (Ed.), *Measuring history: Cases of state-level testing across the United States* (pp. 157–176). Greenwich, CT: Information Age Publishing.

*Grant, S. G. (2003). *History lessons: Teaching, learning, and testing in U.S. high school classrooms.* Mahwah, NJ: Lawrence Erlbaum.

*Grant, S. G., Gradwell, J. M., Lauricella, A. M., Derme-Insinna, A., Pullano, L., & Tzetzo, K. (2002). When increasing stakes need not mean increasing standards: The case of the New York state global history and geography exam. *Theory and Research in Social Education, 30*(4), 488–515.

Grant, S. G., & Horn, C. L. (2006). The state of state-level history tests. In S. G. Grant (Ed.), *Measuring history: Cases of state-level testing across the United States* (pp. 9–27). Greenwich, CT: Information Age Publishing.

*Groves, P. (2002). "Doesn't it feel morbid here?" High-stakes testing and the widening of the equity gap. *Educational Foundations, 16*(2), 15–31.

Harden, R. M. (2001). The learning environment and the curriculum. *Medical Teacher, 23*(4), 335–336.

*Hillocks, G., Jr. (2002). *The testing trap: How state writing assessments control learning.* New York: Teachers College Press.

Hunter, R. C., & Bartee, R. (2003). The achievement gap: Issues of competition, class, and race. *Education and Urban Society, 35*(2), 151–160.

Jackson, P. W. (1996). Conceptions of curriculum and curriculum specialists. In P. W. Jackson (Ed.), *Handbook of research on curriculum: A project of the American Educational Research Association* (pp. 3–40). New York: Simon & Schuster Macmillan.

Kim, J. S., & Sunderman, G. L. (2005). Measuring academic proficiency under the No Child Left Behind Act: Implications for educational equity. *Educational Researcher, 34*(8), 3–13.

This content downloaded from 128.192.114.19 on Tue, 11 Nov 2014 20:34:03 PM
All use subject to JSTOR Terms and Conditions

King, N. (1998) Template analysis. In G. Symon & C. Cassell (Eds.), *Qualitative methods and analysis in organizational research: A practical guide* (pp. 118–134). London: Sage.

King, N. (2006, October 23). *What is template analysis?* University of Huddersfield School of Human and Health Sciences. Retrieved April 27, 2007, from http://www.hud.ac.uk/hhs/research/template_analysis/whatis.htm

Kliebard, H. M. (1989). Problems of definition in curriculum. *Journal of Curriculum and Supervision, 5*(1), 1–5.

*Landman, J. (2000, January 7). *A state-mandated curriculum, a high-stakes test: One Massachusetts high school history department's response to a very new policy context.* Doctoral qualifying paper, Harvard Graduate School of Education, Cambridge, MA. (ERIC Document Reproduction Service No. ED440915)

*Libresco, A. S. (2005). How she stopped worrying and learned to love the test . . . sort of. In E. A. Yeager & O. L. Davis Jr. (Eds.), *Wise social studies teaching in an age of high-stakes testing* (pp. 33–49). Greenwich, CT: Information Age Publishing.

*Lipman, P. (2002). Making the global city, making inequality: The political economy and cultural politics of Chicago school policy. *American Educational Research Journal, 39*(2), 379–419.

Lipman, P. (2004). *High stakes education: Inequality, globalization, and urban school reform.* New York: RoutledgeFalmer.

*Lomax, R. G., West, M. M., Harmon, M. C., Viator, K. A., & Madaus, G. F. (1995). The impact of mandated standardized testing on minority students. *Journal of Negro Education, 64*(2), 171–185.

*Luna, C., & Turner, C. L. (2001). The impact of the MCAS: Teachers talk about high-stakes testing. *English Journal, 91*(1), 79–87.

Madaus, G. F. (1988). The influence of testing on the curriculum. In L. N. Tanner (Ed.), *Critical issues in curriculum: Eighty-seventh yearbook of the national society for the study of education* (pp. 83–121). Chicago: University of Chicago Press.

Madaus, G. F., & Clarke, M. (2001). The adverse impact of high-stakes testing on minority students: Evidence from one hundred years of test data. In G. Orfield & M. L. Kornhaber (Eds.), *Raising standards or raising barriers? Inequality and high-stakes testing in public education* (pp. 85–106). New York: Century Foundation Press.

Marchant, G. J., & Paulson, S. E. (2005, January 21). The relationship of high school graduation exams to graduation rates and SAT scores. *Education Policy Analysis Archives, 13*(6). Retrieved February 8, 2006, from http://epaa.asu.edu/epaa/v13n6/

McCormick, J., Rodney, P., & Varcoe, C. (2003). Reinterpretations across studies: An approach to meta-analysis. *Qualitative Health Research, 13*(7), 933–944.

McEwan, H., & Bull, B. (1991). The pedagogic nature of subject matter knowledge. *American Educational Research Journal, 28*(2),316–334.

*McNeil, L. M. (2000). *Contradictions of school reform: Educational costs of standardized testing.* New York: Routledge.

*McNeil, L. M., & Valenzuela, A. (2001). The harmful impact of the TAAS system of testing in Texas: Beneath the accountability rhetoric. In G. Orfield & M. L. Kornhaber (Eds.), *Raising standards or raising barriers? Inequality and high-stakes testing in public education* (pp. 127–150). New York: Century Foundation Press.

Moe, T. M. (2003). Politics, control, and the future of school accountability. In P. E. Peterson & M. R. West (Eds.), *No child left behind? The politics and practice of school accountability* (pp. 80–106). Washington, DC: Brookings Institution Press.

*Murillo E. G., Jr., & Flores, S. Y. (2002). Reform by shame: Managing the stigma of labels in high-stakes testing. *Educational Foundations, 16*(2), 93–108.

Nichols, S. L., & Berliner, D. C. (2005, March). *The inevitable corruption of indicators and educators through high-stakes testing* (No. EPSL-0503-101-EPRU). Tempe, AZ: Education Policy Studies Laboratory, Arizona State University. Retrieved September 27, 2005, from http://www.asu.edu/educ/epsl/EPRU/documents/EPSL-0503-101-EPRU-exec.pdf

Nichols, S. L., & Berliner, D. C. (2007). *Collateral damage: How high-stakes testing corrupts America's schools.* Cambridge, MA: Harvard Education Press.

Nichols, S. L., Glass, G. V., & Berliner, D. C. (2005, September). *High-stakes testing and student achievement: Problems for the No Child Left Behind Act* (No. EPSL-0509-105-EPRU). Tempe, AZ: Education Policy Studies Laboratory, Arizona State University. Retrieved September 27, 2005, from http://www.asu.edu/educ/epsl/EPRU/documents/EPSL-0509-105-EPRU.pdf

Noblit, G. W., & Hare, R. D. (1988). *Meta-ethnography: Synthesizing qualitative studies* (Vol. 11). Newbury Park, CA: Sage.

Orfield, G., & Wald, J. (2000). Testing, testing: The high-stakes testing mania hurts poor and minority students the most. *Nation, 270*(22), 38–40.

*Passman, R. (2001). Experiences with student-centered teaching and learning in high-stakes assessment environments. *Education, 122*(1), 189–199.

Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2005). Realist review: A new method of systematic review designed for complex policy interventions. *Journal of Health Services Research & Policy, 10*(1), 21–84.

*Perreault, G. (2000). The classroom impact of high-stress testing. *Education, 120*(4), 705–710.

*Renter, D. S., Scott, C., Kober, N., Chudowsky, N., Joftus, S., & Zabala, D. (2006, March 28). *From the capital to the classroom: Year 4 of the No Child Left Behind Act.* Washington, DC: Center on Education Policy. Retrieved March 28, 2006, from http://www.cep-dc.org

*Rex, L. A. (2003). Loss of the creature: The obscuring of inclusivity in classroom discourse. *Communication Education, 52*(1), 30–46.

*Rex, L. A., & Nelson, M. C. (2004). How teachers' professional identities position high-stakes preparation in their classrooms. *Teachers College Record, 106*(6), 1288–1331.

*Salinas, C. (2006). Teaching in a high-stakes testing setting: What becomes of teacher knowledge. In S. G. Grant (Ed.), *Measuring history: Cases of state-level testing across the United States* (pp. 177–193). Greenwich, CT: Information Age Publishing.

Sandelowski, M., Docherty, S., & Emden, C. (1997). Qualitative metasynthesis: Issues and techniques. *Research in Nursing & Health, 20*(4), 365–371.

Schneider, A. L., & Ingram, H. (1997). *Policy design for democracy.* Lawrence: University of Kansas.

*Segall, A. (2003). Teachers' perceptions of the impact of state-mandated standardized testing: The Michigan Educational Assessment Program (MEAP) as a case study of consequences. *Theory and Research in Social Education, 31*(3), 287–325.

Segall, A. (2004a). Blurring the lines between content and pedagogy. *Social Education, 68*(7), 479–482.

Segall, A. (2004b). Revisiting pedagogical content knowledge: The pedagogy of content/the content of pedagogy. *Teaching and Teacher Education, 20,* 489–504.

Sirin, S. R. (2005). Socioeconomic status and student achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417–453.

*Siskin, L. S. (2003). Outside the core: Accountability in tested and untested subjects. In M. Carnoy, R. Elmore, & L. S. Siskin (Eds.), *The new accountability: High schools and high-stakes testing* (pp. 87–98). New York: RoutledgeFalmer.

*Sloan, K. (2005). Playing to the logic of the Texas accountability system: How focusing on "ratings"—not children—undermines quality and equity. In A. Valenzuela (Ed.), *Leaving children behind: How*

This content downloaded from 128.192.114.19 on Tue, 11 Nov 2014 20:34:03 PM
All use subject to JSTOR Terms and Conditions

"Texas-style" accountability fails Latino youth (pp. 153–178). Albany: State University of New York.

*Smagorinsky, P., Lakly, A., & Johnson, T. S. (2002). Acquiescence, accommodation, and resistance in learning to teach within a prescribed curriculum. *English Education, 34*(3), 187–213.

*Smith, A. M. (2006). Negotiating control and protecting the private: History teachers and the Virginia standards of learning. In S. G. Grant (Ed.), *Measuring history: Cases of state-level testing across the United States* (pp. 221–247). Greenwich, CT: Information Age Publishing.

Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher, 20*(5), 8–11.

*Taylor, G., Shepard, L., Kinner, F., & Rosenthal, J. (2001, September). *A survey of teachers' perspectives on high-stakes testing in Colorado: What gets taught, what gets lost.* Boulder, CO: CRESST/CREDE/University of Colorado at Boulder. Retrieved February 24, 2006, from http://education.colorado.edu/epic/COStdRprts/COstdrpreptitle.asp

Thorne, S. E., Jensen, L., Kearney, M. H., Noblit, G. W., & Sandelowski, M. (2004). Qualitative metasynthesis: Reflections on methodological orientation and ideological agenda. *Qualitative Health Research, 14*(10), 1342–1365.

*Valenzuela, A. (2000). The significance of the TAAS test for Mexican immigrant and Mexican American adolescents: A case study. *Hispanic Journal of Behavioral Sciences, 22*(4), 524–539.

Valenzuela, A., Prieto, L., & Hamilton, M. P. (2007). No Child Left Behind and minority youth: What the qualitative evidence suggests. *Anthropology & Education Quarterly, 38*(1), 1–8.

*van Hover, S. D. (2006). Teaching history in the old dominion: The impact of Virginia's accountability reform on seven secondary beginning history teachers. In S. G. Grant (Ed.), *Measuring history: Cases of state-level testing across the United States* (pp. 195–219). Greenwich, CT: Information Age Publishing.

*van Hover, S. D., & Heinecke, W. (2005). The impact of accountability reform on the "'wise practice" of secondary history teachers: The Virginia experience. In E. A. Yeager & O. L. Davis Jr. (Eds.), *Wise social studies teaching in an age of high-stakes testing* (pp. 89–105). Greenwich, CT: Information Age Publishing.

*Vogler, K. E. (2003). An integrated curriculum using state standards in a high-stakes testing environment. *Middle School Journal, 34*(4), 10.

Watanabe, M. (2007). Displaced teacher and state priorities in a high-stakes accountability context. *Educational Policy, 21*(2), 311–368.

*Williamson, P., Bondy, E., Langley, L., & Mayne, D. (2005). Meeting the challenge of high-stakes testing while remaining child-centered: The representations of two urban teachers. *Childhood Education, 81*(4), 190–195.

Wößmann, L. (2003). Central exit exams and student achievement: International evidence. In P. E. Peterson & M. R. West (Eds.), *No child left behind? The politics and practice of school accountability* (pp. 292–324). Washington, DC: Brookings Institution Press.

*Wolf, S. A., & Wolf, K. P. (2002). Teaching *true* and *to* the test in writing. *Language Arts, 79*(3), 240.

*Wollman-Bonilla, J. E. (2004). Principled teaching to(wards) the test? Persuasive writing in two classrooms. *Language Arts, 81*(6), 502–511.

*Wright, W. E., & Choi, D. (2005, December). *Voices from the classroom: A statewide survey of experienced third-grade English Language Learner teachers on the impact of language and high-stakes testing policies in Arizona.* Tempe, AZ: Educational Policy Studies Laboratory, Arizona State University. Retrieved February 12, 2006, from http://www.asu.edu/educ/epsl/EPRU/epru_2005_Research_Writing.htm

*Yeh, S. S. (2005, October 28). Limiting the unintended consequences of high-stakes testing. *Education Policy Analysis Archives, 13*(43). Retrieved February 8, 2006, from http://epaa.asu.edu/epaa/v13n43/

*Zancanella, D. (1992). The influence of state-mandated testing on teachers of literature. *Educational Evaluation and Policy Analysis, 14*(3), 283–295.

AUTHOR

WAYNE AU is an assistant professor in the Department of Secondary Education at California State University, P.O. Box 6868, Fullerton, CA 92834-6868; *wau@fullerton.edu*. He is also an editor for the journal *Rethinking Schools*, and his research interests include social justice education and critical educational theory.

Manuscript received November 11, 2006
Revisions received May 11, 2007, June 6, 2007, and June 16, 2007
Accepted June 20, 2007

This content downloaded from 128.192.114.19 on Tue, 11 Nov 2014 20:34:03 PM
All use subject to JSTOR Terms and Conditions