Four Textbooks on Assessment: A Qualitative Comparison

Virginia Navarro

This paper presents a qualitative analysis of the assessment chapters from four popular educational psychology texts. Teacher knowledge about assessment tools and how they support learning is critical to improving student outcomes. The goals of the analysis are to (a) differentiate author 'voices' by looking at style and content choices, (b) track embedded links between theory and practice, (c) note the level of multicultural consciousness in addressing testing issues and (d) compare cited research. Comparisons will be developed from textual evidence gleaned from a close reading of the two assessment chapters in each text that cover classroom and standardized assessment. Interpretive content analysis inevitably reflects the researcher's ideas about what preservice teachers need to know about assessment; however, this analysis provides a fine-grained comparative critique that encourages educational psychology instructors to re-look at the importance of theoretical frameworks, links between theory and practice, multicultural consciousness and cited research in choosing an educational psychology textbook for their classes.

Teacher education programs rely on educational psychology textbooks to equip future teachers with a knowledge base about teaching/learning processes and human development that draws from both research and practitioner wisdom. The marketplace offers multiple educational psychology textbooks, and instructors must decide which book will best support learning goals in their contexts. When making course textbook decisions, instructors may confer with colleagues, preview publisher copies, check out supplements, and review student feedback. Once a specific textbook's features and tools have become familiar to professors, adjuncts, and teacher assistants, departments often stay with an author through several editions because instructors are comfortable with the content. In fact, there is even resistance to updating content from those who have used a text for several years (Ormrod, 2006).

This article offers a qualitative comparison of the assessment chapters in four well-known educational psychology textbooks in order to highlight perceived strengths and weaknesses. My goals include the following (a) to differentiate author 'voices' by looking at style and content choices, (b) to track embedded links between theory and practice, (c) to note the level of multicultural consciousness in addressing testing issues and (d) to compare the research cited by each textbook's author(s). Selected quotes from each text are included to support my analysis, but I recognize that my perspective is shaped by my beliefs about

teacher preparation needs and my experience teaching educational psychology over ten years.

In order to situate myself in this research project and make any potential biases as transparent as possible, I will share a bit about my experience as a teacher educator: In the past several years, I have taught undergraduate courses on child and adolescent development for preservice teachers, an online graduate course in the psychology of early childhood, and adolescent development courses for post-degree certification students. From 1997-2001, I also taught the basic educational psychology course at both graduate and undergraduate levels using earlier versions of the Ormrod and Woolfolk texts. Currently I teach qualitative research methods and sociocultural theory courses for doctoral students. As I gravitate towards a more post-modern consciousness, my critique of positivistic epistemological stances has increased, and I believe that teachers need to be guided toward a realistic view of the pros, cons, and inherent limitations of any assessment tool.

My focus on teaching in urban contexts is reflected in my work with the Urban Network to Improve Teacher Education (UNITE) and my work as co-director of the Career Transition Certification Program (CTCP), a partnership with St. Louis Public Schools. Participation in the Teaching Educational Psychology Special Interest Group (TEPSIG) of the American Educational Research Association (AERA) has supported my examination of "best practices" in teaching educational

psychology. For example, in my courses, although I do mini-lectures on course content, I try to create a dialogic environment though group projects, choices for text readings, case study analysis, electronic Discussion Boards, and assignments tied to activities in schools.

As an English major in the 1960s, I was trained to analyze texts for internal patterns without referencing outside biographical and historical factors. Deconstructing language meanings continues to fascinate me, but now I apply the more contextual theories of Vygotsky (1986), Bakhtin (1991), and others to think about the multi-voicedness of all texts. The choice of citations used by these authors ventriloquate other voices across time (Wertsch, 2001) and textbook content echoes authors of the past with re-accentuation. Because of my orientation towards sociocultural theory, I bring a critical sense about the importance of context to thinking about the teaching/learning process. Yet I am also committed to discerning how various theoretical perspectives might shape how each author chooses to present topical information. Qualitative content analysis allows me to compare actual sentences and phrases, as well as space allocation, related to a limited number of topics and texts. Analyzing what authors write and do not write, what research they include and exclude, what support materials are offered, etc., can clarify if there are real or only perceived differences among textbooks.

It seems that in each new edition of most educational psychology textbooks, the lists of sub-topics covered grows longer, which makes me ponder if authors and instructors are really practicing the "less is more" maxim that we so often preach to preservice teachers. I have chosen, in this analysis, to focus on these textbooks' coverage of assessment because of my feeling that the role of assessment in teaching has taken a radical new direction since the 2001 *No Child Left Behind* legislation. Many scholars have debated the validity and reliability of state and national test scores (Madaus, 1991; Baker, O'Neil, & Linn, 1993; Shepard, 2000; Amrein & Berliner, 2002; Popham, 2005a, 2005b). Ideological differences run deep between those who argue for professional teachers equipped to document student learning in contextual classrooms, and those who believe that incentives for high stakes test scores tied to market driven school choice will improve educational systems (Cochran-Smith & Fries, 2001).

The rhetoric of accountability, in my opinion, has involved systematic discrediting of teacher knowledge and a facile attack on teacher education that masks opposition to the perceived liberal bent of higher education. NCLB's mandate for Adequate Yearly Progress (AYP) does not reflect the reality that learning progress is often uneven, rather than incremental, even in the most stable school districts (Ding & Navarro, 2004). If teachers cannot learn to document student learning in skilled ways, we become overly reliant on high-stakes, extrinsic measures of performance to guide decision-making. Teacher educators need textbooks that foster complex thinking about assessment in their students. Because documenting learning in multiple and complex ways is a teacher's responsibility, teachers today require an increasingly sophisticated understanding of both tools and context when assessing learning.

PREVIOUS TEXTBOOK COMPARISONS

The present study builds on earlier inquiries about educational psychology textbooks that compared issues such as social/emotional and moral development, classroom management, individual differences, exceptionality, diversity, and assessment. Feldhusen (1977), for example, identified common topics across 26 texts in his study and then had instructors rate the importance of each topic. Snowman (1997) also did a content analysis of ten educational psychology texts published after 1990, establishing what percentage of the total pages was devoted to individual topics. Snowman (1997) summarizes his findings in this way:

> In general, six topics account for more than half of the pages in the typical text. The six topics, with average numbers of pages and average percent of total pages in parentheses, are instructional methods and practices (45.5, 11.2%), motivation (40.9, 10.1%), information processing theory (38.6, 9.5%), classroom measurement and evaluation (33.3, 8.2%), classroom management (29, 7.1%), and intellectual differences (26.9, 6.6%) . . . Most texts contain anywhere from 12 to 18 chapters that reflect about a dozen major topic areas, and unlike textbooks prior to the 1970s, there is a strong emphasis on explaining and illustrating how classroom teachers can use psychological principles to help students achieve educational goals and objectives. (pp. 157-159)

Snowman's suggestion that there should be a two-semester course to allow time for more in-depth coverage of selected concepts (p. 162) seems further than ever from reality as states slash pedagogical requirements for certification. Ten years later, in 2007, Snowman's identified topics are still prominent in newer editions. Tellingly, however, the average percentage of pages devoted to assessment has increased over 50%, from the 8.2% cited above to 12.7%, in the four texts chosen for this study.

Wininger & Norman (2005) offer a more recent content analysis of 20 educational psychology texts, looking at preservice teachers' exposure to the concept of formative assessment. They conclude that teacher candidates are exposed to a minimal level of information on this topic even though the NCLB mandate has greatly increased their need to understand basic concepts such as why feedback is critical to learning success. Three of the texts discussed here use the term "formative evaluation" (Eggen & Kauchak, Ormrod, and Snowman & Biehler) while Woolfolk uses the term "formative assessment," saying, "The purposes of formative assessment are to guide the teacher in planning and to help students identify areas that need work" (Woolfolk, 2005, p. 534). According to Snowman & Biehler (2006), "formative evaluation includes such data points as quizzes, homework assignments, in-class worksheets, oral reading, responding to teacher questions, and behavioral observations" (pp. 457-458).

State departments of education in the current conservative political environment are exerting tremendous pressure to streamline education coursework and privilege subject content knowledge. As a result, there are fewer required courses in foundations and educational psychology for preservice teachers. Practitioners often challenge the usefulness of theoretical courses and report that they need more classroom management skills to succeed. Colleagues in Departments of Curriculum & Instruction and Educational Leadership are also voicing concerns about the utility of traditional educational psychology for teachers (Kiewra & Gubbels 1997; Snowman 1997; Chase 1998; Zechmeister & Zeichmeister, 2000; Hanich & Deemer, 2005). Snowman (1997) acknowledges that textbook authors' are working to bridge theory and practice more intentionally by inserting new features:

> The authors of these ten textbooks went to great lengths to blunt the traditional criticism that educational psychology provides the prospective teacher with nothing of *practical* value. First, each text contained one or more chapters on planning and carrying out classroom instruction. Second, each text had one or more features that described and illustrated how psychological knowledge might be used to deal with a variety of educational concerns. (p. 6)

These features have names such as "Suggestions for Teaching in Your Classroom," "Implications for Teachers," "Theory into Practice," "Teachers on Teaching," and "Teacher's Casebook." While textbook authors are to be applauded for responding to critics', and students' calls for relevance, students need help to avoid seeing such "application' sections as simple prescriptions; they need to understand clearly that principles of educational psychology must be interpreted and applied locally in culturally competent ways.

METHOD AND PROCESS

I chose a purposive sample of four educational psychology textbooks for this in-depth analysis:

➢ Eggen, P. & Kauchak, D. *(2004). Educational psychology: Windows on classrooms.* Columbus, OH: Pearson, Merrill-Prentice Hall.

➢ Ormrod, J. E. (2003). *Educational psychology: Developing learners.* Columbus, OH: Pearson, Merrill-Prentice Hall.

➢ Snowman, J. & Biehler, R. (2006). *Psychology applied to teaching.* New York: Houghton-Mifflin.

➢ Woolfolk, A. (2005). *Educational psychology (9th ed.).* Boston: Allyn & Bacon/Pearson.

The choice of these four particular texts out of multiple possibilities was dictated in part by Kelvin Seifert's study of these same four texts for his article in this special issue. However, Eggen & Kauchak, Ormrod, Snowman & Biehler, and Woolfolk, are arguably the best-selling authors in the field as well. As we write, two of these four books are ready to introduce new editions (Woolfolk and Eggen & Kauchak), and all have published multiple editions over the years. An overview of these four textbooks, including chapter titles,

editions, chapter length, and percent of total pages   appears in Table 1.

TABLE 1

EDUCATIONAL PSYCHOLOGY TEXTBOOK PROFILES: ASSESSMENT CHAPTERS

AUTHOR	BOOK TITLE & DATE	ED.	PUBLISHER	CHAPTER TITLE	TOTAL PGS	% of BOOK
Paul Eggen, University of North Florida Don Kauchak, University of Utah	*Educational Psychology: Windows on classrooms* (2004, 2001, 1999, 1997, 1994, 1992)	6th ed.	Merrill Prentice Hall/Pearson	CH 14: Assessing Classroom Learning	492-539 (47 pages)	8.3%
				CH 15: Assessment through Standardized Testing	540-567 (27 pages)	4.7%
					74 pages	74/568= **13%**
Jeanne Ellis Ormrod, University of New Hampshire	*Educational Psychology: Developing learners* (2003,2000, 1998,1995)	4th ed.	Merrill Prentice Hall/Pearson	CH 15: Basic Concepts and Issues in Assessment	510-551 (41 pages)	6.9%
				CH 16: Classroom Assessment Strategies	552-591 (39 pages	6.6%
					80 pages	80/591= **13.5%**
Jack Snowman, Southern Illinois University Robert Biehler	*Psychology Applied to Teaching* (2006) Other eds. not listed	11th ed.	Houghton Mifflin Company	CH 14: Assessment of Classroom Learning	455-490 (35 pages)	6.5%
				CH 15: Understanding and Using Standardized Tests	492-521 (29 pages)	5.4%
					64 pages	64/536= **11.9%**
Anita Woolfolk, Ohio State University	*Educational Psychology: Active learning edition* (2005, 2004)	9th ed.	Allyn & Bacon /Pearson	Cluster 14: Evaluation, Measurement, and Assessment *Module 39* Def. & Concepts *Module 40* Standardized Tests	502-533 (31 pages)	5.5%
				Module 41 Trad. Approaches *Module 42* Innovations in Classroom Ass	534-565 (31 pages)	5.5%
					62 pages	62/565= **11%**

As seen in this table, the chapters on assessment range from 9 percent to 13.5 per cent of the total text volume in these texts. Since all four texts have one chapter devoted to evaluation terminology, concepts and standardized testing; and a second chapter that features informal teacher assessment, test construction, and grading issues, I decided to divide the analysis into two sections. I will discuss the four chapters on the first of these topics, addressing the books in alphabetical order, and then tackle a comparison across the second block of content using the same order. In this section, in order to give the reader an overview of the topics involved, I developed a table that provides a more in-depth outline of the organizational structure for the teacher assessment chapters in two of the four texts.

My guiding questions in this analysis were: What theoretical orientation is revealed though each textbook's choice of emphasis? How well is theory translated into applied practice? What interpretive stance is taken toward high-stakes tests and formative assessment? What research is cited to support arguments? These topics represent particular areas of interest to me as a teacher educator. While I acknowledge that comparisons on other topics might yield different results, careful qualitative comparison of each text's presentation of these key conceptual domains yielded some provocative differences. I chose to discuss each textbook's chapters in a parallel set, rather than organizing by theme or questions, so the reader can get a clearer sense of each authorial voice. Comparison of authors' style and content as a way of discerning their theoretical orientation, however, runs throughout the analyses, rather than being pulled out as a separate section.

FOUR VOICES ON THE DISCOURSE OF ACCOUNTABILITY AND STANDARDIZED TESTING

Eggen & Kauchak

Chapter 15 "Assessment through Standardized Testing," in Eggen's & Kauchak's (2004) *Educational psychology: Windows on classrooms* identifies three broad areas of coverage in the outline: 'Standardized Tests', 'Understanding and Interpreting Scores', and 'Issues in Standardized Testing." The chapter begins with the case of a fourth grade teacher helping a parent interpret Stanford achievement test scores, which are provided in a Figure. The chapter offers assignment questions in the margin for each section.

Unfortunately, these student questions seem to prompt recall rather than analysis and application; a typical question is "Identify at least one similarity and two differences between standardized tests and teacher-made tests" (p. 543). In other words, these questions simply require students to parrot back adjoining text passages.

For me, Eggen & Kauchak appeared the least concerned, and even somewhat naïve, about the misuse or potential bias issues in testing. For example, in the section on using tests for selecting and placing students in high school, they write, "Scores from the math section of a standardized test can help the math faculty place students in classes that will best match their background and capabilities" (p. 544). No cautionary note is inserted about the potential harmful effects of tracking, or on the effects on test scores of prior education at different feeder schools or the sophistication of students' test-taking skills, etc. Likewise, these authors seem to assume that ACT and SAT scores are an accurate and acceptable means of helping admissions officers make college acceptance decisions. No reference is included from voices that problematize such ideas, no mention is made about the correlation between SAT or ACT scores and SES, and nothing is said about their lack of predictive validity for school success beyond the freshman year in college. Although they do later cite Shepard's (1993) study, stating that the correlation between SAT and college grades is only .42, while high school grades have a .46 correlation with college grades (p. 549), the authors frame ACT and SAT scores this way:

Because the tests are objective and reliable, they *eliminate* teacher bias and the inconsistency of grading practices from different teachers and schools. In this regard, they add valuable information in predicting future success (p. 548, italics added).

These assertions reveal a level of true belief in tests that strikes me as problematic in today's post-modern world. Issues such as middle-class cultural capital and the influence of internalized deficit labels on performance are never raised in this text. In my classes, I generally try to address this idea by bringing in copies of *Highlights,* a subscription magazine for children, for my students to analyze. Magazine features align with the tasks and format of intelligence tests and might function as the original test prep kit for preschoolers. Acknowledging the uneven playing field for students

entering formal school is important in today's society.

Eggen & Kauchak also mention a Berliner (1984) study that documents that some elementary math tests have only a 47% to 71% correlation with the actual taught curricula. In this section, the student question in the margins reads, "Describe what a teacher might do if a lack of fit is uncovered between a standardized achievement test and the teacher's curriculum?" (p. 455). I guess the choices are to teach to the test or to choose a different test. Although it is important to align standards, tests, and curriculum, I miss any questions from these authors about whose knowledge counts most, a question at the heart of how meaning is assigned to the results of high stakes tests.

The statistical description sections in this text are detailed and clear. Like most of the texts in this study, this one includes a Figure that has a normal curve on top and then underneath the range of standard deviations percentile, stanines, z-scores, t-scores, and SAT scores so students can visualize how the pieces and terms relate to one another. There is also a full explanation of the concepts of confidence interval and standard error of measurement.

The three bulleted items in the "Issues" section of Eggen's & Kauchak's chapter focus on (a) the accountability movement, (b) testing teachers, and (c) cultural minorities and standardized testing. Despite their overall disposition toward 'tests as truth,' I found the Eggen & Kauchak examples that illustrate bias in content, bias in testing procedures, and bias in test use to be realistic and concrete. The syntax of some sentences in this section, however, still implies a somewhat skeptical attitude toward the conclusions of those challenging the meaning of tests, e.g. "Critics contend" or "Evidence *suggests* that test results are *sometimes* used in ways that discriminate." According to the authors, advocates of minimum competency tests argue "that the tests are the *fairest and most effective* means of achieving the aim of democratic schooling: a quality education for all students" (p. 559, italics added). This translates for me as a sincere belief in the myth of meritocracy because it glosses the contextual constraints of poverty and racism– including affordable housing, literate parents, and resourced schools - issues that impinge on test outcomes in complex and understudied ways.

A final item of interest in the "Analyze Test Content" section involves a moral dilemma faced by a teacher, who while previewing a test, realizes that a test item will "clearly disadvantage a particular group of students" (p. 563). A distinction is then made between curriculum teaching, that is ethical, and item teaching, which is "indefensible." I had not heard these terms before and I applaud the authors for presenting tools to help teachers think about such real-world dilemmas of practice. Finally, the authors send a clear message that "No one test should be used as the sole basis for educational decisions about individual students" (p. 564).

Ormrod

The topics listed in Ormrod's overview of Chapter 15 in her 2003 edition of *Educational psychology: Developing learners* include items such as "Using Assessment for Different Purposes," "Important Qualities of Good Assessments," "Standardized Tests," "Types of Test Scores," "High-Stakes Testing and Accountability," "Confidentiality and Communication of Results," and "Taking Student Diversity into Account." Notice that the first two topics set up a conceptual framework before actual tests are discussed. I find this typical of Ormrod's style. She organizes material carefully, with her audience of preservice teachers clearly in mind. Too often authors and teachers assume a knowledge base in these students that is illusionary. I like the list of opening questions in this chapter because they frame the issue of assessment in personal terms, such as "What do we mean by *assessment,* and what forms can it take in classroom settings?" (p. 511). In this way, Ormrod models how to probe prior knowledge before instruction, which can assist the teacher in designing instructional strategies appropriate to a specific class and context. Second, Ormrod clearly raises the issue of who should know test results and how best to communicate them. Ethical decision-making is part of every teacher's day, so I think it is important to include items highlighting such issues in these chapters.

Ormrod's section on standardized tests (pp. 526-541) describes two ways they are useful: They provide a comparison of our students against students elsewhere, and they provide a means of tracking general progress over time. Comments on the SAT echo Eggen & Kauchak in that scores are linked with predicting college success and measuring general capacity to learn. Ormrod, however, does problematize the concept of "intelligence" and brings up the idea that education

should develop aptitudes, not just measure presumably innate qualities.

Her discussion of norm-referenced and criterion referenced tests is clear, with ample exemplars of the points raised. Overall, I feel Ormrod has a balanced sense of pros and cons of testing, but does not fully develop the conversation around the centrality of situated contexts in designing performance systems. The strength of the chapter is in not talking down to preservice teachers while also elaborating on the concepts, such as SD, z-scores, etc., so they are understandable. For instance, the section that describes the dangers of "grading on the curve" (p. 535) provides an excellent example of unpacking potential negative consequences of norm-referenced assessments.

In the section on "High-Stakes Testing," Ormrod clearly positions the chapter content with headers such as "Problems with." and "Potential Solutions for". The issue of Social-Economic-Status (SES) emerges in the discussion, and the failure of "proficiency" to mean the same thing across states, or even within states, is aired. The solutions are practical in nature and include educating the public more fully on the limits of high stakes test scores' meanings while assuring that multiple measures are always part of any decision-making.

The discussion of confidentiality and communication about assessment results section is a value added feature of the text in my opinion. We need to help teachers be more sensitive to their moral, legal, and ethical responsibilities to families. The Family Education Rights and Privacy Act (FERPA) poses serious implications for ethical and transparent practice; teachers need to know the legal rights of families.

For younger children (K-3) so-called "readiness" tests are poor predictors of success in school, according to Ormrod, yet our culture is so fixated on testing that pressure to perform against peers starts early. Her compare/contrast Table 15.4 (p. 544) provides a to-do and not-to-do list for avoiding student test anxiety. A side box notes that test bias can be around gender, SES, culture and ethnicity, which is certainly true. But when Ormrod writes about a question that culturally offends a student, her concern seems that it distracts them from doing their best work on the test, rather than about the deeper issue of institutional racism that reifies stereotypes in ways that affect learners' self-images. Special-needs students can get tests translated into their primary language, unlike

regular classroom students, but Ormrod cautions against interpreting translated test results because they have little reliability or validity data to support them.

The case study called "The Math Test," in which a frustrated middle school teacher blames kids for poor scores without offering them any help, demonstrates how *not* to use test scores. This teacher demands that parents must sign corrected work without probing into what went wrong. Anger will not help teachers to problem-solve, but a close look at the test items might offer clues on how to remediate. Ormrod goes on to argue that tests are not bad in themselves, but rather that meanings attributed to test scores by persons can misrepresent results. I personally am not quite so ready to absolve the tests because they too are cultural products that reflect the social order, but I get the point about not being too quick to shoot the messenger.

The series of Boxes called "Into the Classroom" offers concrete ideas for teachers to put the theories presented in the chapter into "best practice". Six such strategies are recommended in order to use assessment to best promote students' learning and achievement: (1) pretest to determine where to begin instruction, (2) choose assessments that match knowledge and skills you want students to achieve, (3) align assessments to how you want students to process, (4) use assessments as learning experiences (Ex: science teacher who has students collect and analyze local water samples), (5) use tests to give specific feedback, and (6) provide criteria for student self-evaluation.

Practical insights such as the danger of presenting incorrect information on a test that may be remembered instead of the correct response are peppered throughout each sub-topic along with examples from practice. Of particular value are the scanned informal instruments from actual teachers that are analyzed for flaws in construction. The marginal asides often connect the current topic with sections in previous chapters – a careful weaving together of text pieces.

Modeling the use of learning strategies, Ormrod creates an acronym or mnemonic to help students remember the qualities of good assessment RSVP: reliability, standardization, validity, and practicality. Ormrod's underlying position seems to be that tests are objective and worth trusting when used appropriately. Speaking of reliability, Ormrod writes that teachers

must be confident that our conclusions will be essentially the same regardless of whether we give the assessment on Monday or Wednesday, whether the weather is sunny or rainy, and whether we evaluate students' responses while in a good mood or a foul frame of mind. (p. 519)

However, in the classrooms I hang around, context has more of an impact on performance outcomes than is suggested here. In other words, context does matter in how students perform on tests, and learning in one's zone of proximal development is a dynamic rather than a static process based on getting the same test/retest results. Even internalized notions of academic identity can influence the degree to which what one knows can be translated into test performance in a situated assessment setting. When, where, and under what conditions will inevitably skew test results to some degree.

Ormrod develops the concepts of content, predictive, and construct validity extensively and with examples. A few of her assertions on construct validity are worth examining. The first implied that, by observing students' on-task or off-task behavior, one could infer motivation to learn "academic subject matter" (p.525), yet I know that bored and traumatized students can be off-task and still have strong motivation to learn. The second assertion focused on the need to compare the performance of two groups who are known to be different. This example involved a comparison of non-disabled 12-year-olds and 12-year-olds "identified as having mental retardation" (p. 525). Even though using person-centered language, the author's justification for comparing intelligence scores between these two groups seemed weak at best to me.

For me the strength of Ormrod's books have always been in the pedagogy of the text itself, along with multiple up-to-date realistic examples from practice and careful organizational weaving together of the content across chapters. The learning aides are more than blobs to break up print; they emphasize practical ideas to implement theory and organize ideas.

Snowman & Biehler

The chapter by in Snowman & Biehler's (2006) *Psychology applied to teaching* on understanding and using standardized tests is organized under three sub-headings: "Standardized Tests (ST)," "Using Standardized Tests for Accountability Purposes: High-Stakes Testing," and "Standardized Testing and Technology." This text does a good job overall of pointing the reader to web sites for further information. The authors begin, however, with the nature, prevalence, and uses of standardized testing without setting the stage as Ormrod did to probe what the preservice teachers are already thinking about these topics. Unlike Eggen & Kauchak (2004) and Ormrod (2003), Snowman & Biehler do not rely on case studies to frame chapter content in applied ways. A piece of what I label value-added information notes, however, that

147 million standardized tests are administered annually to public school students. Of this total, about 51 million are administered as part of state-mandated assessment programs, 85 million are given as part of district-level programs, and 11 million are given to special populations. (Clarke, Madeus, Horn, & Ramos, 2001, as cited in Snowman & Biehler, p. 492)

What struck me in this statement is that local entities administer the most standard assessments by far. I also wondered about where NAEP data fit into this schema. These figures leave little doubt that U.S. children are being fully assessed!

A feature in this text that I particularly liked is the "pause and reflect" marginal questions. They personalize the material by probing students' thinking about tests and to what degree they reflect what one really learns or is capable of learning. The chapter is fairly detailed about explaining concepts such as reliability, validity, norm and criterion referenced tests, etc. For example, under the "Aptitude Tests" section the authors cite Richard Snow's (1992) article that argues for broadening this concept of aptitude to embrace characteristics such as "extroversion, conformity, independence, production of mental images, attention span, beliefs, and fear of failure" (p.497). The authors also seem to support the view that "we should abandon the view that one's ability is the cause of one's achievement" (p. 497). They cite Sternberg's idea that we might better think of aptitudes as "developing expertise." Such ideas resonate with my own notions about aptitude versus ability garnered from Perkins (1994) book called *Outsmarting IQ.*

Like Ormrod (2003), Snowman & Biehler critique norm-referenced tests as promoting competitive comparisons rather than fostering feedback for individual learning. When presenting

the usual sections on stanine scores, z- and t-tests, etc., the margin notes mention a web tool called glossary Flashcards that helps students learn and review vocabulary – a useful idea.

Like Eggen & Kauchak (2004), these authors begin the section on using standardized tests by referencing the report, *A Nation at Risk* (1983). Although they describe the doom and gloom mantra of American education generated by such reports, they also include counter research (Bracey, 2002, 2003; Berliner & Biddle, 1997). Acknowledging that multiple perspectives and interpretations exist on a topic is especially important to develop preservice teachers' sense of critical thinking around text content.

Features of NCLB are also described in detail, including the concept of Adequate Yearly Progress (AYP) and recent developments on modifications to AYP, such as the 95% participation rate being averaged over two or three years. I think Snowman & Biehler provide the most extensive information about the politics and reality of NCLB and in a fairly even-handed way. They do not gloss over the many problems with implementation, but they also echo other writers in listing the pro side of the argument for high-stakes testing: i.e. goal clarity, improved quality control, and beneficial effects for teaching and student learning. On the down side of NCLB, they cite structural limitations, misinterpretation and misuse of test results, a one-size-fits-all approach to motivation, and undesirable side effects. I felt these sections of Snowman & Biehler were outstanding in that the latest research is reported, summaries of information located in other parts of the text are included, and ideas are presented in clear and thoughtful ways. As they note, "High-stakes test systems tend to ignore the contributions of input and process variables, or what might be called opportunity to learn" (p. 509). A value added feature is discussion of WYTIWYG (what you test is what you get). If you prepare students too specifically for a particular test, there is often little carryover or transfer to other measures of the same content. What this means to me is that concept mastery is not easily measured through traditional testing.

In the section on the effects of high-stakes testing, a study done in Chicago by Roderick & Engel (2001) is cited to demonstrate that "no single policy or approach will work for all students because classroom learning is an extremely complex phenomenon that is only partly under the teacher's control" (Snowman & Biehler, 2006, p.

512). I applaud their recommendations, including holding policymakers accountable for the effects of their high-stakes accountability systems. Again the "pause and reflect" questions challenge readers to take action by writing letters, informing school board members, set up workshops, etc. It is refreshing to see a text actually encourage a level of activism and advocacy for teachers.

Snowman & Biehler are also the only authors to offer a section on standardized testing and technology with recommended web links (pp. 515-517) such as www.smartthinking.com, which provides tutoring and digital whiteboards, and TestU (www.testu.com/), which offers diagnostic pretests. Evidently some districts and states are making these services available to students without cost, but I am cautious about the wisdom of promoting commercial outlets in textbooks. Also of interest is computer adaptive testing, a system that adjusts type and difficulty of questions based on earlier responses. Video games have honed expectations for immediate feedback and progressive levels of difficulty as the player can handle them, and adaptive testing systems mirror this capacity of games. In the final section of Chapter 15, the authors offer six directives on how to use standardized tests. As in other texts discussed, they emphasize that test scores are only one piece of information about students and that as a teacher, you are committed to using test scores not to classify students, but to help them learn.

Woolfolk

In Woolfolk's (2005) Active Learning Edition of *Educational Psychology*, the edition I used for this analysis, the terms "Cluster" and "Module" are used as organizers rather than "Chapter," so I will consider her Modules 41 and 42 as a unit to compare with others' chapters on classroom assessment; and her Modules 39 and 40 as another unit to compare with chapters on standardized testing.

Module 39, Definitions and Concepts, begins this way: "All teaching involves evaluation. At the heart of evaluation is judgment, making decisions based on values" (p. 504). I think Woolfolk gets the overall conceptual framework right here, as readers are guided to think about how norm-referenced or criterion-referenced tests create value to educators' work. I love the way she gives us a set of raw data (p. 506) and then lets the learner construct meanings around finding the mean, median, and

mode, demonstrating how these measures can mean different things, given different scenarios.

Concepts such as standard deviation (SD) are discussed in depth, including providing the formula for establishing SD on a set of scores (no other book provided detailed mathematical models), and explaining why SD is important in understanding scores. Unlike Ormrod (2003), who provides visuals of actual standardized test printouts, Woolfolk relies on more traditional normal curve graphs to show a set of scores and how they fall on the curve. I felt that Woolfolk's explanation of confidence intervals made good sense for preservice teachers, who often think of scores as discrete points on a continuum.

Challenges to construct-related evidence for validity are discussed, and it is emphasized that data is gathered over years to establish construct validity. In the feature called "Guidelines" Woolfolk lists five principles for increasing reliability and validity with multiple prompts under each heading; this reminded me of the information in the Snowman & Biehler (2006) chapter and is evidence of efforts to bridge theory and practice. Woolfolk more than the others continually brings the student's gaze outward into the larger community by using "Family and Community Partnership" boxes.

Module 40, called Standardized Tests, covers achievement, diagnostic, and aptitude tests as well as accountability issues and testing of teachers. Value added information in this short module includes Table 40.1 that compares features on five frequently used tests: Iowa Basic, MAT-7, SAT-9, TerraNova CAT, and TerraNova CTBS; websites where more information is available are listed. Table 40.2 describes the testing programs for nine states (again with web site support links). Beginning teachers are too often clueless about what tests their district uses and why. A final value added feature is the "Guidelines: Accommodations in Testing" (p. 531) that lays out a whole page of detailed suggestions under specific needs such as setting, timing, scheduling, presentation and response accommodations. Both modules read smoothly and strike me as particularly thoughtful and reflective.

issue of testing teachers is handled as a Point/Counterpoint box with both sides, pro and con, making their best argument. A list of ten ideas for testing programs (p. 525) argues for such things as matching the district standards, being part of a larger plan, testing complex thinking, and providing alternative assessment strategies and opportunities for retesting when the stakes are high. The list goes on to say that assessment systems should include all students, provide appropriate remediation, assure that students have adequate opportunity to learn content, and take into account the student's language. Good advice is "Use test results for children, not against them" (Haladyna, 2002).

TEACHER ASSESSMENT OF CLASSROOM LEARNING: TOOLS AND PRACTICE

This next section on classroom assessment of learning will also go through the four texts in alphabetical order, similar to the previous section on high-stakes testing. Table 2 looks at the number of references used by each author and Table 3 lists the Table of Contents for two texts so the reader can get a quick overview of how they generally align. As I reviewed the Tables of Contents for all four texts, I found more variation in choice of material for the classroom assessment chapters than for the standardized testing chapters, yet all four texts do cover the basic topics of formative and summative assessment, traditional paper/pencil and performance evaluations, portfolios, and norm-referenced versus criterion-referenced testing. As indicated in Table 2 below, these chapters run from 35 to 50 pages of text and have 80 to 105 references to support and explain their content.

Although Woolfolk's (2005) Modules 39 and 40 include fewer research references (39) than Eggen & Kauchak (2004) (64), Ormrod (2003) (69), or Snowman & Biehler (2006) (67), Woolfolk does include the full citation for each article in the margin rather than at the end of the book. I like this because it encourages students to pay more attention to where the cited material originates. Often there are a group of articles in a special journal edition highlighted this way.

Eggen & Kauchak

Eggen & Kauchak's chapter on classroom assessment starts with a case study of a fourth grade teacher who, on the basis of a pretest on fractions (which is included so we can see student work), decides to begin teaching on equivalent fractions using consumable chocolate wrapped candies as her manipulatives. Her post-test then shows her students' success in learning. With 105 references in a 47-page chapter, these authors draw on literature from the past two decades to look closely at classroom assessment and best

practices. I agree with their assertion that many teachers do not have a strong knowledge base and lack confidence in their assessment practices. New teachers rank assessment fourth, behind classroom management, motivation, and dealing with individual differences, as a problem area (Veenman, 1984). That Veenman's findings are still relevant was confirmed for me recently when a colleague shared a conversation she had had with an elementary teacher about how perhaps homework should not count for such a high percentage of class points. The teacher replied, "If I do not count homework, what will I grade on?". . . a discouraging moment for a teacher educator.

TABLE 2:

NUMBER OF CITED REFERENCES INCLUDED BY CHAPTER AND AUTHOR
(Shaded areas represent classroom assessment chapters)

Author	Chapter Title	No. of References.	No. of Pages
Paul Eggen, Don Kauchak	CH 14: Assessing Classroom Learning	105	47
	CH 15: Assessment through Standardized Testing	64	27
(totals)		**169**	**74**
Jeanne Ellis Ormrod	CH 15: Basic Concepts and Issues in Assessment	69	41
	CH 16: Classroom Assessment Strategies	89	39
(totals)		**158**	**80**
Jack Snowman, Robert Biehler	CH 14: Assessment of Classroom Learning	80	35
	CH 15: Understanding and Using Standardized Tests	67	29
(totals)		**147**	**64**
Anita Woolfolk	Cluster 14: Evaluation, Measurement, and Assessment		
	Modules 39 and 40 Definitions/Concepts & Stan. Tests	31	31
	Modules 41 and 42 Traditional Approaches & Innovations in Classroom Assessment	60	31
(totals)		**91**	**62**

The purpose of assessment is framed by Eggen & Kauchak in terms of (a) increasing learning, and (b) increasing motivation: "Frequent assessment, linked to well-planned goals, encourages students to pace themselves and keep up with their studies" (Eggen-Kauchak, p. 495).

Measurement is defined as "the process of gathering information about learning" and evaluation as "the process of making decisions on the basis of measurements" (p. 495). Figure 14.1 (p. 496) lists both traditional formats (true-false, multiple choice, matching, fill in blank, short answer, paragraph response to specific or open-ended questions, and essay) and alternative formats (performance task, timed trial, exhibition, reflective journal, oral presentation, collaborative project, audiovisual presentation, debate, and simulation) (cited from Cheek, 1993, pp. 6-10). The authors also caution teachers against bias in grading, particularly in the early grades where personality, appearance, neatness, etc. might sway assumptions about quality of work.

TABLE 3

OUTLINES OF TWO CLASSROOM ASSESSMENT CHAPTERS

EGGEN & KAUCHAK (2004)	SNOWMAN & BIEHLER (2006)
CH 14: ASSESSING CLASSROOM LEARNING	*CH 14: ASSESSMENT OF CLASSROOM LEARNING*

1) Classroom assessment (CA)
 a) Functions of CA
 b) Measurement and evaluation
  i) Formal and informal measurement
  ii) The need for systematic assessment
 c) Validity: making appropriate evaluation decisions
 d) Reliability: consistency in measurement
2) Traditional assessment strategies
 a) Teachers' assessment patterns
 b) Valid test items
 c) Constructing valid test items: instructional strategies
  i) Multiple-choice items
  ii) Matching items
  iii) True-false items
  iv) Completion items
  v) Essay tests: measuring complex outcomes
  vi) Using rubrics
 d) Commercially prepared test items
3) Alternative assessment
 a) Performance assessment (PA)
 b) Designing pa: instructional strategies
  i) Specifying the type of performance
  ii) Selecting the focus of assessment
  iii) Structuring the evaluation setting
  iv) Designing evaluation procedures
  v) Performance evaluation strategies
 c) Portfolios
 d) Putting traditional & alternative assessments into perspective
4) Effective assessment practices: instructional strategies
 a) Planning for assessment
  i) Tables of specification: increasing validity through planning
 b) Preparing students for assessments
  i) Teaching test-taking strategies
  ii) Reducing test anxiety
  iii) Specific test-preparation procedures
 c) Administering assessments
 d) Analyzing results
 e) Accommodating diversity in classrooms: reducing bias in assessment
  i) Carefully wording items
  ii) Making provisions for non-native speakers
  iii) Accommodating diversity in scoring
5) Grading and reporting: the total assessment system
 a) Designing a grading system
  i) Formative and summative evaluation
  ii) Norm-referenced and criterion-referenced evaluations
  iii) Tests and quizzes
  iv) Alternative assessments
  v) Homework
 b) Assigning grades: increasing learning and motivation
  i) Raw points or percentages?
 c) Technology and learning: using technology to improve assessment
  i) Planning and constructing tests
  ii) Analyzing test data
  iii) Maintaining student records
6) Technology and portfolios
	1) The role of assessment in teaching
	 a) What is assessment?
	  i) Measurement
	  ii) Evaluation
	 b) Why should we assess students' learning?
	  i) Summative evaluation
	  ii) Formative evaluation
	  iii) Diagnosis
	  iv) Effects on learning
	2) Ways to Measure Student Learning
	 a) Written tests
	  (for each category Characteristics, Advantages,& Disadvantages)
	  i) Selected response tests
	  ii) Short-answer tests
	  iii) Essay tests
	  iv) Constructing a useful test
	 b) Performance Tests (PT)
	  i) What are PTs?
	  ii) Types of PTs
	   (1) Direct writing assessments
	   (2) Portfolios
	   (3) Exhibitions
	   (4) Demonstrations
	   (5) Characteristics of PT
	  iii) Emphasis on active responding
	  iv) Degree of realism
	  v) Emphasis on complex problems
	  vi) Close relationship b/t teaching and testing
	  vii) Use of scoring rubrics
	  viii) Use of formative evaluation
	  ix) Responsiveness to cultural diversity
	   (1) Some concerns about performance assessment
	3) Ways to Evaluate Student Learning
	 a) Norm-referenced grading (nature, strengths, weaknesses)
	 b) Criterion-referenced grading (nature, strengths, weaknesses, mastery approach)
	4) Improving your Grading Methods: Assessment Practices to Avoid
	5) Technology for Classroom Assessment
	 a) Electronic grade books and grading programs
	 b) Technology-based Performance Assessment
	 c) Digital Portfolios
	 d) Performance and Portfolio Assessment Problems
	6) Effective Assessment Techniques

On the other hand, I felt Eggen & Kauchak's sections on validity and reliability exhibited a patronizing tone toward teachers, as if they routinely graded on "appearance rather than substance" (p. 497). The feel of the prose is that the teacher is "other," but if teachers will only "look for ways to improve" and "conscientiously revise items" they may be able to develop the recommended "reflective attitude" the authors say is "needed for success." Figure 14.3, "Characteristics of teacher-made tests," lists six common flaws of teacher-made items. The authors attribute these common failings to the fact that teachers respond to complexity by simplifying their work, use completion and matching items which are the easiest to grade, resist revising items, and avoid essays that take more time to grade (p. 499-500). I find this attitude too global and sweeping and not particulary motivating in terms of helping new teachers improve. Eggen & Kauchak state that

> All teachers, elementary and secondary, lack confidence in their ability to write good test items and use assessment to improve learning. . . Because of inadequate training, teachers frequently have difficulty writing clear and precise items at a level above knowledge and recall. (p. 500)

Although several studies are cited to support these assertions, the tone still feels a bit hostile. That being said, the authors do go on to give specific concrete suggestions for how to go about composing good test items that are helpful, though other authors still do a better job, I think, of giving readers applied examples for this content.

This chapter in Eggen & Kauchak also includes links to additional technology, such as videos on teachers use of rubrics. A value-added portion of this section involves sharing the Oregon Six Traits Writing Rubric: ideas, organization, voice, word choice, sentence fluency, and conventions. Eggen & Kauchak cite Airasian (2000) when listing cautions about using textbook pre-made tests (p. 509), agreeing with him that alignment with learning goals and classroom events must be the priority in assessment. To get a comparative overview of the table of contents for two of the texts (Eggen & Kauchak (2004) and Snowman & Biehler (2006)), I have included a list that show detailed topics for the teacher assessment chapters.

In Eggen & Kauchak, seven pages are devoted to alternative assessment methods, slightly fewer than the 11 pages on traditional assessment, but substantial nonetheless, with several Figures showing sample evaluation tools. Performance assessment connected with constructivist beliefs about learning as holistic and the concern that traditional testing does not encourage higher order thinking. The marginal questions in this chapter, however, remain unhelpful on the critical and creative thinking barometer; for example, one reads, "Are essay items performance assessments? Defend your answer, using the information from this section" (p. 511).

The information on designing evaluation procedures, including systematic observation, checklists, and rating scales, is somewhat helpful but a bit abstract for preservice teachers to get excited about. The portfolio section is not extensive but does emphasize the involvement of students in evaluating their own work to encourage reflective and metacognitive development, a position I totally support. After listing the pros and cons of alternative assessments using multiple citations for each point raised, the authors conclude that, "assessment isn't as simple as it appears on the surface; effective measurement and evaluation require sensitive and intelligent teachers" (p. 518). "Classroom Connections", a feature to connect theory and practice, lists three important ways to insure valid and reliable instruments with examples provided for each level of schooling. The suggestions include these points: (1) Increase validity through careful planning prior to assessment; (2) Use alternate assessments to increase validity; and (3) Use portfolios and performance assessments to develop learner self-regulation (p. 518).

Most of the texts in this study refer to the concept of building matrices to insure that tests align with the learning goals and covered subject matter, but Eggen & Kauchak go further to develop explicit ideas on how to build a table of specifications, including providing a sample table. They reconnect us with the fourth grade class we met at the beginning of the chapter to learn more about how the teacher has prepared the kids to take the test on fractions and how she administers the test. Demystifying testing for students means explaining to them why this test format was chosen for this particular material, how to best study for the test, and how the test will be scored. It amazes me that, even in excellent schools, teachers fail to scaffold these minimal criteria for success. Test anxiety in our culture is understandably high, given the paranoia over high stakes test results and what

they mean for a person's identity and future educational opportunities. Eggen & Kauchak continue to draw on this case as they discuss analyzing test results. I feel this whole section, with its connection to an actual case study, is strong and raises many issues that will be faced by teachers.

Although the modifications to testing procedures suggested in relation to diversity issues seem reasonable, Eggen & Kauchak gloss over the larger contextual reasons why students of color often experience weaker teachers, less home support, and less affirmation of themselves as learners. Carefully wording items, making provisions for non-native English speakers, and accommodating diversity in scoring (breaking apart content knowledge, problem-solving and the ability to use language) do not suffice to address the social justice issues that create an uneven playing field for many children in our wealthy nation.

The final section of the chapter is on grading and reporting as part of a total assessment system. This introduces sections on formative and summative, norm-referenced and criterion-referenced evaluation tools. Eggen & Kauchak state that norm referencing is "virtually non-existent in elementary schools and rarely seen in middle and secondary schools" (p. 529); I feel the other textbooks in this study offer a more realistic critique of common classroom norm-referenced testing practices, including the ubiquitous practice of "grading on the curve". Woolfolk (2005), for example, begins her chapter with an in-class assignment for groups to design a grading system for a grade or subject area and defend it to peers. Transfer is not likely unless we build in applied use of concepts.

The "Assigning grades: Increasing leaning and motivation" (p. 530) section in Eggen & Kauchuk includes another case study, comparing a math teacher's and science teacher's approach to weighting tests, quizzes, homework, performance, and projects. The danger of converting assignments of unequal value to a percent grade and then averaging those percents is clearly laid out. The chapter closes with a "Windows on Classrooms" case study followed by two Praxis preparation exercises: (1) constructed response questions on this final case and (2) a document-based analysis of the first case. These exercises are well-designed to provoke higher level thinking and authentic dialogue when students share responses, while the Praxis connection seems to appear across all the Pearson published texts in some form.

Ormrod

The opening case study in Ormrod's (2003) text shows two students, Ellen and Rosyln, studying for a geography test on the same chapter, but for two different teachers: one who emphasizes facts, the other who is a constructivist teacher looking for knowledge application on the test. Assessment in classrooms is approached through the processes of cognition and metacognition. The RSVP idea mentioned earlier (Reliability, Standardization, Validity and Practicality) appears again as a mnemonic to organize Ormrod's sections on both informal and formal assessments. Samples of student work are sprinkled throughout the chapter, and the various analyses of teacher responses continually showcase issues of how teachers' perceptions and assumptions might distort their evaluation of student work – a problem dealt with in more judgmental language by Eggen & Kauchak.

In describing the "halo effect" Ormrod cites a Darley & Gross (1983) study that had undergraduates evaluate a video of a fourth grade girl's oral achievement test based on several criteria. In one condition the undergraduates had been made to believe the girl was from a low SES family; in another that she had high SES. Those who believed she had high status rated her above average; those who believed her to have low status rated the same performance as below average. We need to continually bring such studies to students' attention because the implications for all of us are enormous.

Ormrod's section on how to construct paper-pencil assessments is outstanding, detailed, thoughtful, and full of essential information for teachers to know and practice. In a recent article, Ormrod (2005) discussed her reasons for using so many student and teacher artifacts in her new edition:

> Artifacts not only make such concepts [learning and cognition, effects of prior knowledge, metacognition and self-regulation, motivation, affect, and instructional strategies] more concrete for future teachers but also situate the concepts in actual classroom tasks and activities. (p. 217). (bracketed material added)

The many ways that test items can misrepresent the real question we want to ask is documented with lots of examples for each type of

test item. A brief fill-in-the-blank item on Chapter 15 is given to help students experience the frustration of how sentences out of context can have more than one word that is appropriate (p. 562). It is important to show both positive and negative exemplars of test questions and to give reasons for what is right or wrong about them. This is new terrain for many students despite sixteen years of schooling.

Analyzing the many hand-written stories and teacher-made test examples included by Ormrod engages preservice teachers in imagining themselves in the role of teacher, I believe. Seeing actual student work makes the day-to-day job of working with young people more accessible and increases motivation to learn the concepts involved in the analysis. I have mixed feelings, however, about Ormrod's advice, to assess writing skills "separately from the content of students' responses to the extent possible" (p. 568). I think opportunity for revision might be appropriate at times, but it is important for students to know that form does count as well as content. Detailed feedback, of course, can turn an evaluation into a teachable moment.

Ormrod includes a well-developed section on performance assessment, including design, choosing activities, and evaluating outcomes. In discussing individual versus group performance, she offers a field-based cooperative group project that will help high school students "think as an urban planner would think" (Newmann, 1997 in Ormrod, p. 572). She links the concept of dynamic assessment to Vygotsky's notion of the zone of proximal development, in which a child is provided instruction within the potential developmental level rather than at the actual independent level. According to Vygotsky, learning pulls development along, and individual children can stretch their minds with help from a more knowledgeable other to wrestle with various degrees of conceptual knowing. Ideas unfurl over time as children grow into the use of cultural tools of knowing. There are definite limits to what testing can accomplish, but preservice teachers need to have an up-to-date toolbox to be effective. I feel Ormrod's section on how to score student responses offers solutions to some of Eggen's & Kauchak's (2004) concerns about the reliability and validity of alternative performance evaluations. Citing Stiggins (2001) and Thorndike (1997), Ormrod reiterates that using more than one rater to evaluate a student's performance is a good idea (if sometimes unrealistic because of overloaded teacher-days). Probably the closest approximation is to have

students evaluate peers with a rubric to confirm what the teacher saw or to raise other points. Throughout the chapter, Ormrod encourages involving students in assessing their own work, others' work, and teacher pedagogy – always a good idea.

The only author to discuss item analysis at any length, Ormrod provides the formula for item discrimination that will tell a teacher if the item helps distinguish high scorers from low scorers (internal consistency reliability). Learning how to assess and grade well has moral implications as students, teachers, and districts continue to be judged by testing outcomes. Ormrod advocates for criterion, not norm-referenced testing, in high school and cautions teachers about assigning credit for improvement or effort. I agree with her that extra credit has to be offered equitably to all students and should be used very sparingly as an option.

In the final section on diversity, Ormrod reviews information from earlier chapters on gender, race and ethnicity and includes a large table with suggestions for assessing students with special educational needs. I came away from Ormrod's chapter with a sense of organized and practical information that reflects respect for the P-12 students as well as a realistic understanding of the knowledge base of traditional preservice teachers.

Snowman & Biehler

From the outset, the goals for assessment in Snowman's & Biehler's chapter are clearly indicated as a two-fold process: "using teacher-made measures to assess mastery of the teacher's specific objectives and using professionally prepared standardized tests to measure the extent of a student's general knowledge base and aptitudes" (p. 455). I confess that the way these goals are expressed made me wonder if the authors are implying that teacher tests are not considered "professional" because they are not designed by statisticians.

Under "Ways to Measure Student Learning," the first point says, "Written tests measure degree of knowledge about a subject" (p. 455). Although I believe strongly in the importance of assessment for learning, I feel Snowman & Biehler reflect our society's tendency toward a concept of learning that focuses almost exclusively on the individual learner, rather than the complex cultural social system that assigns by subtle cues the role of non-learner to some groups of children. We have to be very

skeptical and organize multiple strategies for assessment so we do not misdiagnose any student's will and talent to learn.

Snowman & Biehler lay out four reasons to assess: summative (clear feedback on performance level), formative (monitor progress, remediate), diagnosis (find out what is interfering with learning), and effects of learning (increases positive outcomes) (pp. 457-58). The many formats for testing are then enumerated by listing their characteristics, advantages and disadvantages. Using Popham's (2003) frame, Snowman & Biehler discuss attributes of a useful classroom test: significance, teachability, describability, reportability, and nonintrusiveness. Like Ormrod (2003), they offer the reader fairly detailed information on performance evaluations (eight pages versus three on traditional formats). Examples and models are provided to clarify concepts. Statements such as, "Students with learning disabilities are likely to experience the greatest benefit from being given a scoring rubric and being shown how to use it" (Jackson & Larkin, 2002), reveal a general sense of advocacy and fairness toward students.

I am always glad to read texts that acknowledge that "meaningful learning occurs within a cultural context with which one is familiar and comfortable" (p. 466). While Snowman & Biehler argue for a closer relationship between instruction and assessment, they open their chapter outline with a claim that traditional written tests are valid measures of learning and that we need standardized tests by professionals to measure general base knowledge. For me these perspectives seem somewhat contradictory in warrant. Two empirical studies by Supovitz & Brennan, (1997) and Supovitz (1998) are cited as evidence for increased achievement by minority children on performance tests such as portfolios over standard paper-and-pencil tests. While listing concerns about performance assessments from some quarters, Snowman & Biehler comment that with this sort of testing "a teacher becomes more of a collaborator and facilitator than a gatekeeper" (p. 467). This made me think of Ormrod's (2003) ideas about Vygotsky's ZPD and dynamic assessment systems. If learning is the ultimate goal, which teacher role will best bring individual children forward developmentally? I guess I would argue it depends on context. At any rate, the section ends by cajoling readers to sign up for an elective measurement course in their programs.

In their section about ways to evaluate student learning, Snowman & Biehler lay out the usual norm versus criterion-referenced arguments. Their examples arouse readers' sympathy for the built-in pros and cons of all testing formats. The overall message continues to be that complex assessment strategies can support better learning for all students. I like better the way Ormrod (2003) lays out the many differing purposes of assessment. Assessments can be motivators, mechanisms for review, learning experiences, or feedback to promote learning. Assessment practices can also guide instructional decision-making, diagnose learning and performance problems, promote self-regulation, and determine what students have learned (pp.514-517). With such a range of applications, it becomes clear why a deep knowledge base about assessment is critical for teachers.

Snowman & Biehler also comment on the irony of some common responses to a mastery approach to learning, such as accusing teachers of 'dumbing down' curriculum and lowering standards, when they allow students to continue to work towards defined levels of success (p. 473). The detailed 14-point guide on how to set up a mastery learning classroom is practical and useful.

Snowman & Biehler's Chapter 14, on "avoiding pitfalls" in grading, is a value-added feature. These authors pull no punches as they talk about "worshipping averages" and "using zeroes indiscriminately." Their fervor in denouncing such unprofessional practices almost seem to be a product of personal encounters with poor assessment practices. Taking Snowman & Biehler's litany of mistakes to heart could make a big difference for future teachers.

Another value-added feature of this text is the section on ways to integrate technology into the assessment process: e.g. electronic grade books, technology-based performance assessments, simulations, hypermedia, web quests, digital portfolios, and/or digital rubrics (Niguidula, 2005). The web sites and shared free teacher tools for grading, making rubrics, etc., seem useful, but may be hard to keep up-to-date, although publishers are getting better at having hotlinks as part of the online support packages.

The applied "Suggestions for Teaching in Your Classroom" feature provides a seven-page re-cap that lists seven effective assessment techniques with exemplars and bulleted indicators.

For students to effectively plan how they will master your objectives, they need to know as early as possible how many tests they will have to take, when the tests will occur, what types of items each test will contain, and what content they will be tested on. (p. 483)

As mentioned earlier, Snowman & Biehler also thoroughly cover the table of specifications concept as a way to plan tests to align with goals and class content.

Woolfolk

The title of Woolfolk's Module 41, "Getting the most from traditional assessment approaches: Testing and grading" seems to recognize the persistence of paper-and-pencil tests in schools. Why do teachers hold fast to the model of fill in the blanks and matching? I have known cases where a test came out of a co-operating teacher's files to be duplicated exactly by a student teacher because there was no electronic copy to modify. Such experiences seem to convey a belief that, regardless of how the curricular inputs might change, the instrument to test outcomes can remain fixed and static. Before we chastise teachers, however, we need to examine the demands of their work and the lack of collaborative time for them to develop and refine lessons. As mentioned before, Eggen & Kauchak (2004) implied that the path of least resistance is normative when it comes to assessment design, but I think that if school culture better supported professional development and collaboration in the assessment area, many teachers would gladly improve their assessment practices.

Woolfolk's textbook has value added features such as "Connect & Extend" marginal boxes that continually link the content to the Teaching Portfolio and Praxis test. The chapter begins with an excellent "Teachers' Casebook" selection that asks students to identify what assignments and projects they would develop for a grade and asks them in small groups to develop a class handbook describing the grading policy. Such critical thinking activities go beyond probing prior knowledge into constructing new knowledge by problem-solving.

After discussing the characteristics of formative and summative assessments, Woolfolk lays out examples of good and bad test questions to illustrate the directives given for design. In an even-handed way, she charts the pros and cons of each type of test item. The value-added part of her chapter for me was her portrayal of the effects of failure on students: "Students receiving low grades are more likely to withdraw, blame others, decide that the work is "dumb" or feel responsible for the low grade but helpless to make improvements" (p. 542), and the very different effects of constructive feedback. Four questions are identified to guide teacher comments: "What is the key error? What is the probable reason the student made this error? How can I guide the student to avoid the error in the future? What did the student do well that could be noted?" (p. 541). Although these questions seem obvious, they too often get obscured in educational jargon.

Woolfolk, like others, comments on the negative use of percentage averaging and how remarkably resistant to change this practice has become, despite the availability of better methods of scoring. I think Woolfolk is the only one to explain the contract system of grading. The "Guidelines" box on Using Any Grading System also has a nice summary of issues to consider including "giving the student the benefit of the doubt" (p. 547), a hard lesson for new teachers to learn. Like Ormrod (2003), Woolfolk always includes material on working with special needs populations and talks about the FERPA and Educational Amendment Acts of 1974 in terms of open communication with families. Families and communities must be cultivated as allies, not transformed into adversaries because they question or challenge educational practices.

Woolfolk begins her final Module on assessment by posing a question to the reader about a new teacher interviewing at an innovative school and being asked what they know about using portfolios, performances, projects, and rubrics to assess learning. I worry that many of our students, judging from their electronic portfolios, would not be able to articulate the benefits and execution of such systems. Woolfolk includes a Point/Counterpoint box from Wiggins (1991) to argue both sides of the traditional versus authentic assessment debate. She also shares a personal story about her advocacy as a graduate student for alternative doctoral comprehensive exams. Many sample rubrics and guideline lists are peppered throughout this chapter. In regards to diversity, there is not much space devoted to the topic but related issues are discussed and linked to equity of opportunity and support.

Overall Woolfolk seems knowledgeable about the latest developments in assessment but she does not use citations as extensively as some of the other authors when developing her points on this topic. There are many design features within these modules that have solid information, but Module 42 seems a bit fragmented with so many interruptions to the eye. The choice of content and style suggest that Woolfolk herself might be more in favor of newer authentic assessments than some of the other authors considered here.

SAME IDEAS BUT FROM DIFFERENT SOURCES

After listing all the references for each chapter, I sorted them alphabetically, expecting to see a fair amount of overlap. I had noticed the reliance on a core of authors who were referenced across texts. In the chapters on standardized testing, however, the only identical reference I found in all four texts was Linn & Gronlund (2000). Various articles by Popham (1990; 1993; 1995; 1998; 2000; 2001; 2002; 2003) and Airasian (1996; 2000; 2001) are frequently cited. At least two authors used Heubert & Hauser (1999), Anastasi & Urbina (1997), Olson (2000; 2001; 2002; 2003), and Messick (1989).

For the chapters on Classroom Assessments the alignment of referenced texts was no stronger. Bangert-Drowns, Kulik, & Kulik (1991) are cited in all four texts as is Airasian but for different articles (1994; 2000; 2001). Gronlund too shows up in the four studied texts (1988; 1993; 1998; 2003). Popham's work is regularly cited from a variety of references (1990; 1995; 2001; 2002; 2003) as is Stiggins (1992; 2001; 2002). Haladyna is mentioned in three texts with differing co-authors (2001; 2002), as is Guskey (2001; 2002; 2003). Pintrich & Schunk (2002) show up in two classroom assessment chapters. As Table 2 shows there is a fair range in the number of citations across chapters. Yet with a total of some 573 authors cited, some many times in a single chapter, over these eight chapters on assessment, only a handful of overlapping citations among these texts could be documented.

This phenomenon may be worth more study, since the field of education is often accused of lacking a cohesive set of findings shared by the discipline. Given that the topics in the chapters do align fairly closely, it is worth noting that the opinions and facts presented draw on substantially divergent readings. To confirm this lack of overlap, I tried to combine the authors used by each author across the two chapters, to see if the broader frame

of assessment might make the common texts more visible, but the patterns were similar.

What might account for the uniqueness of the majority of the references cited by individual authors? Are some more aligned with APA literature and others more in tune with AERA publications? Are there just so many refereed journals in education that we cannot possibly remain current in all venues relating to a topic? The idiosyncratic nature of the knowledge base that supports our corpus of knowledge about assessment in educational psychology may be a matter or concern or celebration, but I think it is an area that needs further examination.

SUMMING UP

This qualitative analysis of four popular educational psychology texts posed the following questions: How well do the authors translate theory into applied practices? How do they represent ideas about evaluation and assessment in an era of high stakes accountability? What theoretical orientation is revealed in the choice of material to include or exclude? What research do they cite to support assertions made? Are there substantial differences among the four texts in content or message?

My interpretive analysis suggests that these popular texts have a similar scope of topics, but with varying degrees of elaboration. Each text strikes a somewhat different tone that indicates some differing values and perspectives on assessment. Eggen & Kauchak (2004) seem the most positivistic in their beliefs about the meaning of statistically sound test scores, while Woolfolk (2005) probably represents the most social constructivist perspective. Although Ormrod (2003) remains fairly individualistic in her psychological theory of learning, she is probably the best for scaffolding student learning through structure and detailed supportive learning tools; I would give her chapters high marks for 'walking the talk' so to speak. Snowman & Biehler (2006) begin their chapters by asserting a strong belief in traditional assessment, but they also demonstrate awareness of recent critiques of standardized assessments and present a balanced picture on controversial issues.

Overall, the two assessment chapters in the four texts represent 11-13.5% of the total text content, a substantial increase from Snowman's 1997 mean of only 8.2%. This is not surprising

since NCLB legislation in 2001 has brought a laser-like focus on assessment and accountability issues in schools.

Theory-to-practice efforts are clear in all the books; however, I came away thinking that Woolfolk (2005) and Ormrod (2003) make such links more explicit and understandable, suggesting activities to help students understand how to apply educational psychology insights in a real world. Everyone but Snowman & Biehler (2006) opens their chapters with case studies; Eggen & Kauchak (2004) even draw on an initial case throughout the chapter as new ideas are introduced.

At the end of the day, I feel all four texts have unique value-added content as well as some strengths and weaknesses in their choice of material, tone of delivery, organization of supports, and theory to practice bridges. I conclude that coverage of assessment topics is not radically different across these texts, but the spin and degree of elaboration on a topic definitely reveals some critical variation in intent, beliefs, practitioner orientation, and theoretical framework. Assessment practice should be integral to a teacher's thinking about learning because close observation and reflection on the results of both informal and formal assessments can be powerful tools for improving teaching practice and student learning. Additionally, the goals and instructional design of learning units must be carefully aligned with assessment procedures. Too often classroom assessments show stronger correlations with literacy levels and test-taking skills than actual content mastery of the curriculum. Another problem to avoid is grading disproportionately on form rather than intellectual rigor. Without sophisticated assessment, targeting differentiated instruction to meet individual learning goals cannot be achieved.

I recognize that the interpretive nature of my analysis, based on topics that interested me, is partial and subjective, yet I have tried to provide textual evidence for my claims and to attend to both strengths and weaknesses in each chapter analyzed. I hope that reading this study provokes others to do similar fine-grained analyses of other topics in educational psychology textbooks in order to help instructors identify books that will align with specific goals and priorities.

If we could pick and choose sections from multiple texts to create our "personal preference" educational psychology text (technology is already making this an option), we might benefit even more from parallel studies of different concepts to help us re-look at the texts we use, the reasons we use them, the alignment of text content to teaching goals, and the situated learning needs of our students. Stimulating such conversations is the purpose of this work as we continue to dialogue about "best practices" in teaching educational psychology. All the authors of these four textbooks have my admiration and appreciation for their ability to utilize a broad knowledge base to write a comprehensive textbook for the field.

REFERENCES

Airasian, P. (2000). Classroom assessment (3rd ed.). New York: McGraw Hill.

Amrein, A.L. & Berliner, D.C. (2002, March 28). High-stakes testing, uncertainty, and student learning. Educational Policy Analysis Archives, 10(18). Retrieved May 25, 2022 from http://epaa.asu.edu/epaa/v10n18/.

Anderson, L. M. & Blumenfield, P. (1995). Educational psychology for teachers: Reforming our courses, rethinking our roles. Educational Psychologist 30, 143-157.

Artiles, A. J. & Trent, S. C. (2000). From individual acquisition to cultural-historical practices in multicultural teacher education. Remedial & Special Education 21(2), 79-91.

Baker, E. L., O'Neil, H. F., & Linn, R. L. (1993). Policy and validity prospects for performance-based assessment. American Psychologist, 48, 1212-1218.

Berliner, D. (1994). Making our schools more effective: Proceedings from three state conferences. San Francisco: Far West Laboratory.

Berliner, D. & Biddle,B.J. (1997). The manufactured crisis: Myths, frauds, and the attack on America's public schools. White Plains, NY: Longman.

Berliner, D. & Calfee, R. (1996). Handbook of Educational Psychology. New York: Macmillan Library Reference USA.

Bracey, G.W. (2002). The 12th Bracey report on the condition of public education. Phi Delta Kappan, 84(2) 135-150.

Bracey, G.W. (2003). The 13th Bracey report on the condition of public education. Phi Delta Kappan, 54(2) 148-164.

Brownlee, J., N. Purdie, et al. (2001). Changing Epistemological Beliefs in Pre-service Teacher Education Students. Teaching in Higher Education 6(2): 247-268.

Chase, C. I. (1998). A Chat Room -Educational Psychologists but No Curriculum Methodologists? Educational Psychology Review 10(2): 239-248.

Cheek, D. (1993). Plain talk about alternative assessment. Middle School Journal, 25(2), 6-10).

Cochran-Smith, M. & Fries, M.K. (2001). Sticks, stones, and ideology: The discourse of reform in teacher education. Educational Researcher, 30(8), 3-15.

Darley, J.M. & Gross, P.H. (1983). A hypothesis-confirming bias in labeling effects. Journal of Personality and Social Psychology, 44, 20-33.

Ding, C. & Navarro, V. (2004). An examination of student mathematics learning as assessed by SAT 9: A longitudinal look. Studies in Educational Evaluation 30, 237-253.

Educational Psychologist; Special Issue (1996; 2001).

Eggen, P. & Kauchak, D. (2004). Educational Psychology: Windows on classrooms. Columbus, OH: Pearson, Merrill-Prentice Hall.

Feldhusen, J.F. (1977). Issues in teaching undergraduate educational psychology courses. In D. F. Treffinger, J.K. Davis, & R.E. Ripple (eds.), Handbook of Teaching Educational Psychology. New York: Academic Press.

Foster, H & Iannaccone, C. (1994). Multicultural content in special education introductory textbooks. Journal of Special Education 28(1), 77. Retrieved Saturday March 11, 2006 from the Academic Search Premier database.

Haladyna, T.H. (2002). Essentials of standardized achievement testing: validity and accountability. Boston: Allyn & Bacon.

Hanick, L.B. & Deemer, S. (2005). The relevance of educational psychology in teacher education programs. Clearing House, 78 (5), pp.189-191.

Hoffman, D.H. & Elliott, S. (October 2, 1998). Web conferencing case dialogues: A supplement to traditional journal writing assignments in educational psychology. Paper presented at the Midwest Association of Teachers of Educational Psychology, Oxford, Ohio.

Jackson, C.W. & Larkin, M.J. (2002). Rubric: Teaching students to use grading rubrics. Teaching Exceptional Children, 35(1), 40-45.

Kiewra, K. A. & Gubbels, P.S. (1997). Are educational psychology courses educationally and psychologically sound? What textbooks and teachers say. Educational Psychology Review 9(2): 121-148.

Knapp, N.F. ((2005). "They're not all like me!" The role of educational psychology in preparing teachers for diversity. The Clearing House 78(5) 202-206.

Madaus, G, F. (1991). The effects of important tests on students: Implications for a National Examination System. Phi Delta Kappan, 73, 226-231.

Medina, M.A., Morrone, A. & Anderson, J. (2005). Promoting social justice in an urban secondary teacher education program. The Clearing House 78(5), 207-212.

Navarro, V. & Marquez-Zenkov, K. (Eds.). (2005). Staying connected: Re-framing the contours of partnership work in education. An edited volume of case study narratives on forming, sustaining, and celebrating school-university collaboration. Milwaukee, WI: Urban Network to Improve Teacher Education and The Holmes Partnership.

Niguidula, D. (2005). Documenting learning with digital portfolios. Educational Leadership 63(3), 44-47.

Ormrod, J. E. (2003). Educational Psychology: Developing learners. Columbus, OH: Pearson, Merrill-Prentice Hall.

Ormrod, J. (2005). Using student and teacher artifacts as case studies in educational psychology. The Clearing House78(5), 213-217.

Ormrod, J. (2006). Paper presented at the American Educational Research Association Meeting, San Francisco, CA.

Perkins, D. (1994). Outsmarting IQ: The emerging science of learnable intelligence. New York: Free Press.

Peterman, F. (Ed.). (2005). Designing Performance Assessment Systems for Urban Teacher Preparation. New York: Erlbaum.

Popham, W. J. (2003). The seductive allure of data. Educational Leadership, 60(5), 48-51.

Popham, W. J. (2005a). The age of compliance. Educational Leadership, 63(2), 84-85. Retrieved April 21, 2006, from ERIC database.

Popham, W. J. (2005b). NAEP: Gold standard or fool's gold? Educational Leadership, 62(8), 79-81. Retrieved April 21, 2006, from ERIC database.

Ransom, J.C. (1979). The new criticism. Westport, CT: Greenwood Press.

Seifert, K. (2005). Learning about peers: A missed opportunity for educational psychology. The Clearing House 78 (5) 239-243.

Shepard, L. (1993). Evaluating test validity. In L. Darling-Hammond (Ed.). Review of Research in Education, 19, 405-450. Washington, DC: AERA.

Shepard, L (2000). The role of assessment in a learning culture. Educational Researcher, 29 (7), 4-14

Shepard, L. (2005). Linking formative assessment to scaffolding. Educational Leadership 63(3), 66-71.

Snowman, J. & Biehler, R. (2006). Psychology applied to teaching. New York: Houghton-Mifflin.

Snowman, J. (1997). Educational psychology: What do we teach? What should we teach? Educational Psychology Review 9(2). 151-169.

Supovitz, J. & Brennan, R.T. (Fall, 1997). Mirror, mirror on the wall, Which is the fairest test of all? An examination of the equitability of portfolio assessment relative to standardized tests. Harvard Educational Review 67(3), 472-506.

Supovitz, J. (1998). Gender and racial/ethnic differences on alternative science assessments. Journal of Women and Minorities in Science and Engineering, 4(2), 129-140.

Veenman, (1984). Perceived problems of beginning teachers. Review of Educational Research, 54, 143-178.

Verdi, M. & Johnson, J. (2005). Teaching educational psychology in an online environment. The Clearing House 78 (5), 235-238.

Wertsch, J. (2001). Voices of the mind: A sociocultural approach to mediated action. Cambridge, MA: Harvard Press.

Wiggins, G. (1991). Standards, not standardization: Evoking quality student work. Educational Leadership, 48(5), 18-25.

Wininger, S.R. & Norman, A. D. (2005). Teacher candidates' exposure to formative assessment in educational psychology textbooks: A content analysis. Educational Assessment 10(1). 19-37.

Woolfolk, A. (2005). Educational Psychology, 9th ed. Boston: Allyn & Bacon/Pearson.

Worthen, B. (1993). Critical issue that will determine the future of alternative assessment. Phi Delta Kappan, 74, 12-27.

Yanowitz, K. & Weathers, K. (2004). Do boys and girls act differently in the classroom? A content analysis of student characters in educational psychology textbooks. Sex Roles 51(1/2), 101-108. Retrieved Saturday March 11, 2006 from the Academic Search Premier database.

Zechmeister, J. & Zechmeister, E. (2000). Introductory psychology textbooks: An objective analysis. Teaching of Psychology, 27(1), 6-11. Retrieved Saturday March 11, 2006 from the Academic Search Premier database.

Virginia Navarro, an Associate Professor in the Division of Teaching & Learning at the University of Missouri-St. Louis, currently chairs the Teaching Educational Psychology SIG for AERA. She writes about the social construction of identity, higher education teaching and learning, and school/university partnership work. As Co-Director of a Career Transition program, Virginia prepares urban educators as advocates for youth within a social justice framework.