

01 **Chapter 1**

02 **Prelude: Assessment for the 21st Century**

03
04
05 **Valerie J. Shute and Betsy J. Becker**

06
07
08
09
10
11
12
13 **Abstract** We recently hosted an assessment symposium at Florida State University
14 which served as the basis for this book and focused on how to integrate assessment
15 and instruction to improve student learning and education. The chapters in this book
16 address the general issue of integrating assessment and instruction, and additionally
17 provide innovative solutions to hard questions such as: What would an assessment,
18 suitable for the needs of the twenty-first century, look like? How could it be stan-
19 dardized? Should it be standardized? How could it satisfy the current obsession with
20 “metrics”? What is the role of the professional teacher in making twenty-first cen-
21 tury assessments possible? What constraints would be faced by those who would
22 implement such innovations in assessment practice?

23
24 **Keywords** Assessment · Education · Learning · Measurement

25
26
27
28 **1.1 Introduction**

29
30
31 Measurements are not to provide numbers but insight. Ingrid Bucher

32 The quote above inspired the title of the assessment symposium at Florida State
33 University which served as the basis for this book. The symposium—*Assessment*
34 *for the Twenty-First Century: Insight*—focused on how to integrate appropriate
35 assessment and instruction to improve student learning and education, especially
36 to suit the needs of the twenty-first century. We were motivated by the belief
37 that the goal of assessment should shift from obtaining numbers and rankings
38 to providing insight—on learners and learning, as well as on instructors and
39 instruction.

40
41
42 V.J. Shute (✉)
43 Educational Psychology and Learning Systems Department, Florida State University,
44 Tallahassee, FL, USA
45 e-mail: vshute@fsu.edu

46 The symposium had two intertwining tracks: (a) innovative assessment ideas
47 and technologies to support twenty-first century educational needs, and (b) spe-
48 cific educational and assessment issues and needs (e.g., teacher and student testing
49 and accountability). Florida State University's College of Education hosted the
50 2-day symposium as a part of the college's annual Dean's Colloquium Series.
51 Keynote speakers included James Gee of Arizona State University, Russell Almond
52 of Educational Testing Services (Princeton, NJ) and Mari Pearlman of Pearlman
53 Education Group, LLC. Our keynote speakers were innovative researchers who
54 have written on assessment for twenty-first century skills, and who understand broad
55 policy implications and the challenges of educational reform. Additional speakers,
56 broadly comprising educational researchers, policy makers, and practitioners, pre-
57 sented assessment-related research, and others responded to the invited speakers'
58 presentations, to connect their ideas to the context and concerns of researchers and
59 of state education departments. All speakers' and respondents' papers are included
60 in this book.

61 This book covers a range of topics. It does not just focus, for instance, on
62 the policy issues of large-scale assessment. Instead the book expounds upon the
63 provocative forward-thinking proposals of the keynote speakers and serves as a
64 launching pad for careful analyses of practical problems of implementation of
65 assessment, technical psychometric issues, and policy issues.

66 Our goal for the symposium was to bring together groups who don't normally
67 convene. We wanted policymakers and educational professionals to have the oppor-
68 tunity to learn about the latest research by scholars from across the nation; and
69 also wanted educational researchers to learn about critical educational issues impor-
70 tant to the states and to practitioners. The symposium successfully opened up lively
71 discussions among speakers, respondents, and attendees.

72 We now examine why we believe that this type of conversation—begun at the
73 symposium and continued with this book—is important. After identifying some
74 problems that may benefit from new thinking about assessment, we spend a little
75 time defining key aspects (and terms) of assessment, as well as describing differ-
76 ent types of assessment, to provide a basis for understanding subsequent chapters.
77 Finally, we close with an overview of the chapters in this book, as well as our own
78 thoughts on moving forward with assessment.

81 1.2 The Big Problems

83 Knowledge is no longer an immobile solid; it has been liquefied. It is actively moving in all
84 the currents of society itself. John Dewey

85 This presentient quote by Dewey (1916, p. 40), nearly 100 years ago, is particu-
86 larly relevant now. The world is evolving and effectively shrinking, due mainly to the
87 interconnections made possible via the Internet and other communication technolo-
88 gies. Our twenty-first century existence confronts us with problems of enormous
89 complexity (e.g., meltdowns on Wall Street, nuclear proliferation, pharmaceuticals
90

91 in the water supply, and poverty). Those who confront these issues in the twenty-first
92 century (e.g., policy makers) need to think critically, to identify and examine relevant
93 research, and to understand how systems, in general work, because solutions
94 will be highly complex and interconnected. When confronted by problems, especially
95 new issues for which solutions must be created out of whole cloth, the ability
96 to think creatively, critically, and collaboratively, and then communicate effectively
97 is essential.

98 Learning and succeeding in a complex and dynamic world is not easily measured
99 by the well-worn, multiple-choice response formats on simple knowledge
100 tests. We need to re-think assessment, identify new skills and standards relevant for
101 the twenty-first century, and then determine how to best assess students' acquisition
102 of the new competencies—which may in fact involve *others* doing this assessment
103 (e.g., the community of peers suggested in Chapter 2 by Gee, this book). Moreover,
104 the envisioned new competencies should include not only cognitive variables (e.g.,
105 critical thinking, reasoning skills) but also noncognitive variables (e.g., team-
106 work, tolerance, tenacity) as the basis for new assessments to support learning.
107 Each of these may be embedded and supported within valued domains—such as
108 mathematics and science.

109 Learning is an important part of everyday life, and is a lifelong endeavor. This is
110 especially true for knowledge workers in both developed and developing countries.
111 Avoiding lifelong learning and training is not an option for most people who work
112 in areas where rapid change is a norm in terms of how people work, what they are
113 expected to do, and the tools and information with which they work. For instance,
114 people working in information technology (IT) areas need to constantly acquire
115 new knowledge and skills about new products and ideas, and to perform new tasks.
116 People may change their careers multiple times before retirement, requiring new
117 learning for new work contexts, as well as new social networks and contacts (e.g.,
118 Higgins, 2001). Indeed, all of us face complex problems both at work and in our
119 daily lives. The complexity of these problems and the huge quantities of available
120 information require substantial learning and continuing education as well as the
121 development of learning management capabilities over our lifetimes (Georghiades,
122 2004; Sungur, 2007).

123 With all of these changes taking place in the world, it might be comforting that
124 education has changed little in the past several decades. We don't see it that way.
125 "Old school" (pun intended) philosophies and approaches are often inappropriate in
126 today's rapidly changing and information-rich world. Students need to develop new
127 competencies that are quite different from those needed by earlier generations in
128 order to deal successfully with the deluge of data and information in the twenty-first
129 century (e.g., information communication and technology skills). Many education-
130 ally valuable skills that are potentially suitable for success in the twenty-first century
131 are not currently being acknowledged, let alone assessed. Toward this end, we must
132 agree on what skills we value and promote these skills for a society requiring knowl-
133 edge workers, not simply service workers. Then we must determine how best to
134 measure those skills.

In addition to needing to identify a new set of skills for “twenty-first century competencies,” we believe that the very nature of assessment should be changing. Over a dozen years ago, the National Research Council (NRC, 1996) made a similar plea, which has yet to be adequately addressed. Table 1.1 presents a modified version of the NRC call for changes in the focus on assessment needed to support educational reform for the twenty-first century.

Table 1.1 Changing assessment foci

Less focus on assessing	More focus on assessing
Learning outcomes	Learning processes
What is easily measured	What is most highly valued
Discrete, declarative knowledge	Rich, authentic knowledge and skills
Content knowledge	Understanding and reasoning, within and across content areas
What learners do <i>not</i> know	What learners understand and can do
By teachers alone	By learners engaged in ongoing assessment of their work and that of others

1.3 Defining Assessment Ideas and Terms

Assessment should not merely be done *to* students; rather, it should also be done *for* students, to guide and enhance their learning. NCTM (2000)

In this section we briefly define and disambiguate important assessment terms that often get confounded. For instance, what exactly is the difference between “measurement” and “assessment”? Let’s start with the basic idea of measurement. Whenever you need to measure something accurately, you probably grab an appropriate tool to determine how heavy, light, tall, short, fast, slow, hot, cold, bright, dark, straight, or curved something is. We measure to obtain information (data), which may or may not be useful, depending on the accuracy of the tools we use, as well as our skill at using them. Measuring things like a person’s height, a room’s temperature, or a car’s speed is technically not an assessment and is instead simply collecting information relative to an established standard. How does this relate to education?

1.3.1 Educational Measurement

Educational measurement, in the context of this chapter, refers to the application of a measuring tool (or standard scale) to determine the degree to which educationally-valuable knowledge, skills, and other attributes have been, or are being acquired.

181 It thus entails the collection and analysis of data from learners. According to the
182 National Council on Measurement in Education Web site (<http://www.ncme.org/>),
183 this includes “theory, techniques, and instrumentation available for measurement
184 of educationally-relevant human, institutional, and social characteristics.” A *test* is
185 education’s equivalent of a ruler, thermometer, or radar gun. But note that a test
186 does not improve learning any more than a thermometer cures a fever; both are sim-
187 ply tools. Tests alone can’t enhance educational outcomes. Rather, tests can guide
188 improvement (presuming they are valid and reliable) if they motivate adjustments
189 to the educational system. Examples of educational adjustments include providing
190 the basis for bolstering curricula, ensuring support for struggling learners, guiding
191 professional development opportunities, and distributing limited resources fairly.

192 Again, we measure things to get information, which may be quantitative or qual-
193 itative.¹ How we choose to *use* the data is a different story. For instance, back
194 in the early 1900s, students’ abilities and intelligence were extensively measured.
195 However, this wasn’t done to help them learn better or otherwise to progress.
196 Instead, the main purpose of testing was to track students into appropriate paths,
197 based on the belief that their aptitudes were inherently fixed. That is, a dominant
198 belief during that period was that intelligence was part of a person’s genetic makeup,
199 thus testing was aimed at efficiently assigning students into high, middle, or low
200 educational tracks according to their supposedly innate mental abilities (Terman,
201 1916). In general, a fundamental shift to practical education occurred in the coun-
202 try during the early 1900s, countering “wasted time” in schools and abandoning
203 the classics as useless and inefficient for the masses (Shute, 2007). Early educa-
204 tional researchers and administrators inserted into the national educational discourse
205 the metaphor of the school as a “factory” (Kliebard, 1987). This metaphor is no
206 longer apt.

207 208 209 **1.3.2 Assessment** 210

211 Assessment involves much more than just measurement. That is, in addition to
212 systematically collecting and analyzing information (i.e., measurement), it also
213 involves interpreting and acting on information about learners’ understanding and/or
214 performance in relation to educational goals.² Measurement, then, can be viewed as
215 a precursor to or special case of assessment. Assessment information may be used
216 by a variety of stakeholders (e.g., teachers, administrators, students, parents) and for
217 a variety of purposes, such as to help improve learning outcomes, programs, and ser-
218 vices, and also to establish accountability. Furthermore, an assortment of procedures
219

221 ¹ For a fuller, more balanced perspective on educational measurement, see Messick (1989) and
222 Oosterhof (2009) which extend educational measurement beyond statistical conceptualizations and
223 numbers to include qualitative information as well.

224 ² Others, such as Guion (1998), see assessment as including less formal means of evaluating
225 individuals.

226 is associated with the different purposes. For example, if your goal was to enhance
227 an individual's learning, and you wanted to determine her progress toward an edu-
228 cational goal, you could: (a) administer a quiz; (b) view a portfolio of her work;
229 (c) ask the student (or peers) to evaluate her progress; (d) watch the person solve
230 a complex task; (e) review her lab reports or journal entries, and so on. You'd then
231 need to use the information gathered via these assessments to help guide her further
232 learning activities.

233 Finally we consider who is doing the assessing. Very often, it is the teacher.
234 However, *self-assessment* may be a viable option, as well as an important skill, espe-
235 cially if a valued educational goal is to produce self-directed and productive lifelong
236 learners. Promoting learners' self assessment in relation to setting reasonable learn-
237 ing goals involves supporting (assessing) knowledge of specific goals and learners'
238 progress toward them. It also involves supporting learners' metacognitive skills of
239 reflection and revision. Alternatively, *peer assessment* involves individuals collabo-
240 rating with one another to solve, explain, or understand a problem or task. A variety
241 of benefits (e.g., cognitive, social, motivational) accrue from encouraging learners
242 to work collaboratively. An effective teacher should emphasize a high and equal
243 level of interaction among group members, giving all an opportunity to negotiate
244 meaning, acquire new strategies and skills, and develop higher-order thinking skills.
245 However, as collaboration becomes an increasingly important aspect of twenty-first
246 century learning, it introduces not only opportunities, but also serious challenges
247 for assessment which will need to be resolved with innovative research (e.g., Jeong,
248 2005; Macdonald, 2003; Shute, Jeong, Spector, Seel, & Johnson, 2009).

251 252 **1.3.3 Determining Assessment Quality**

254 Because assessment is a process by which information is obtained relative to a
255 known objective, and since inferences are made about what a person knows (unob-
256 servable) on the basis of responses to assessment tasks (observable), there's always
257 some uncertainty in inferences made on the basis of assessments. So, an impor-
258 tant goal in educational measurement is to collect really good information about
259 the learner(s) and to minimize uncertainty or error. Consequently, key aspects of
260 assessment quality are consistency and validity.

261 The broad term *consistency* is used here rather than the more familiar term
262 *reliability* because it includes not only the quantitative aspects of reliability (e.g.,
263 correlations between parallel forms of tests), but also qualitative aspects of assess-
264 ment (e.g., consistency in a teacher's description of a learner's performance on two
265 comparable tasks). To illustrate, consider the produce scale at your local grocery
266 store. If you weigh two pounds of carrots in the morning, and the scale is consis-
267 tent, the same scale should register the same weight for the carrots an hour later.
268 Similarly, classroom tests and standardized exams should be stable, and it shouldn't
269 make much difference whether a learner takes the assessment at 10:00 AM or 11:00
270 AM. Another measure of consistency (i.e., internal consistency) relates to the items

271 within a test. For instance, if you create an Algebra 1 test, you'd assume that if a
272 learner correctly solves a difficult linear equations problem, then he should solve
273 other linear equation problems correctly. Similarly, the notion of generalizability
274 is often used with performance assessments and portfolios, and addresses the ade-
275 quacy with which you can generalize from a sample of observations to the universe
276 of observations from which it was randomly sampled.

277 As with consistency, there are a number of different types of validity; but in
278 general, *validity* refers to the extent to which the assessment accurately measures
279 what it is supposed to measure and the accuracy of the inferences made from test
280 results. For instance, if you wanted to assess learners' math problem solving skills,
281 but you gave them a personality questionnaire to complete, that would not be a
282 valid assessment of their math skills. Regarding the relationship between validity
283 and consistency, even if an assessment is judged to be consistent and stable (see
284 above), it may not, in fact, be a valid measure. Let's use a scale analogy again, only
285 now it's your bathroom scale. Suppose that you step on your scale 10 times in a row
286 and your scale, without fail, indicates that you weigh 150 pounds. The *consistency*
287 of your scale may be very good, but it may not be accurate (valid) if you actually
288 weigh 165 pounds. Because teachers, parents, school districts, and so on currently
289 make decisions about learners based on assessment results (e.g., grades, retention,
290 graduation), the validity inferred from the assessments is essential, and it's even
291 more crucial than the consistency. So, consistency is a prerequisite for validity. That
292 is, inconsistency in observations always threatens their validity. On the other hand,
293 simply having consistency in what is observed does not ensure the validity of those
294 observations.

297 1.4 Kinds of Assessment

299 When the cook tastes the soup, that's formative; when the guests taste the soup, that's
300 summative. Robert Stake

302 Different types of assessment are often presented in contrast to one another.
303 The two most familiar types of assessment are summative and formative, and the
304 choice and use of a particular type of assessment depends on the educational pur-
305 pose. Schools generally make heavy use of summative assessment (also known
306 as assessment *of* learning). These can be quite useful for accountability purposes
307 (e.g., assessments for grading and promotion purposes) but only marginally—if at
308 all—useful for supporting individual learning. In contrast, learner-centered mea-
309 surement models rely mostly on formative assessment, also known as assessment
310 *for* learning. Formative assessments can be very useful in guiding instruction and
311 supporting individual learning, but not for overall high-stakes decisions. Also, the
312 assessment-for-learning model is often implemented in a non-standardized and
313 hence less rigorous manner than summative assessment, and thus may have more
314 limited validity and consistency (Shute & Zapata-Rivera, [in press](#)). This is not to say
315 such assessments don't have value. Rather, the less standardized, informal nature of

316 formative assessment can be seen as a call for researchers to develop new techniques
317 to capitalize on these assessments' value and utility (e.g., a synthetic approach using
318 many formative assessments might provide an aggregate picture that cannot be seen
319 in single individual assessments). Strong formative assessment research is urgently
320 needed given changes in the types of learning and learning outcomes we are valu-
321 ing as twenty-first century competencies as well as the new, broader, complex and
322 integrated set of contexts in which learning is taking place and applied.

323 *Summative assessment* reflects a more traditional approach to assessing edu-
324 cational outcomes. This involves using assessment information for high-stakes,
325 cumulative purposes, such as for grades, promotion, certification, and so on. A sum-
326 mative assessment is usually administered after some major event, like the end of
327 the school year or marking period; or before a big event, like college entry. Benefits
328 of this approach include the following: (a) it allows for comparing learner perfor-
329 mances across diverse populations on clearly defined educational objectives and
330 standards; (b) it provides reliable data (e.g., scores) that can be used for accountabil-
331 ity purposes at various levels (e.g., classroom, school, district, state, and national)
332 and for various stakeholders (e.g., learners, teachers, and administrators); and (c) it
333 can inform educational policy (e.g., curriculum or funding decisions).

334 *Formative assessment* involves using assessments to support teaching and learn-
335 ing. Formative assessment is incorporated directly into the classroom curriculum
336 and uses results from learners' activities as the basis on which to adjust instruction
337 to promote learning in a timely manner. A simple example would be a teacher giving
338 a "pop quiz" to his students on some topic or lesson, immediately analyzing their
339 scores, and then re-focusing his lesson to straighten out a misconception shared by
340 a substantial number of students in the class. This type of assessment is adminis-
341 tered more frequently than summative assessment and has shown great potential for
342 harnessing the power of assessments to support learning in different content areas
343 and for diverse audiences (e.g., Black & Wiliam, 1998; Hindo, Rose, & Gomez,
344 2004; Schwartz, Bransford, & Sears, 2005). In addition to providing teachers with
345 evidence about how their class is learning so that they can revise instruction appro-
346 priately, formative assessment directly involves learners in the process, such as by
347 providing feedback that will help them gain insight about how to improve.

348
349
350

351 1.5 Discussion

352

353 Our symposium was organized to explore the many possible faces of future
354 assessment—and asked hard questions, such as: What would an assessment, suitable
355 for the needs of the twenty-first century, look like? How could it be standard-
356 ized? Should it be standardized? How could it satisfy the current obsession with
357 "metrics"? What is the role of the professional teacher in making twenty-first cen-
358 tury assessments possible? What constraints would be faced by those who would
359 implement such innovations in assessment practice?

360

361 This book strategically brings together views on innovation in assessment
362 along with perspectives concerning the opportunities and barriers presented by the
363 innovative ideas from those involved in research and large scale assessment. The two
364 main themes are represented, respectively, by our invited speakers, and by respon-
365 dents from state departments of education and university researchers involved in
366 research on assessment issues. Our keynote chapters focus on innovation in three
367 different realms—authentic assessment in contextualized environments (Jim Gee),
368 evidence centered assessment design (Russell Almond), and the role of teachers and
369 the connection between teaching and learning in assessment (Mari Pearlman). These
370 three realms represent three fundamental areas of promise for future assessment
371 systems. Each keynote chapter is followed by a response paper.

372 Additional chapters focus on more specific issues, as well as barriers and poten-
373 tials for implementation of innovative assessment options. For instance, the chapter
374 by Mark Shermis addresses important research concerning automated essay scoring
375 and its potential for widespread use in statewide testing systems. Joseph Martineau
376 and Vincent Dean describe their ideas for making assessment relevant to stu-
377 dents, teachers and schools by explicitly considering transparency, standards, and
378 measures/scales. Alysia Roehrig and Eric Christesen summarize their research on
379 designing and developing a reliable and valid tool for assessing the quality of
380 teaching in grades K-12. Their tool can be used in the professional development
381 of teachers to foster students’ literacy achievement and motivation. Allan Jeong
382 describes an innovative tool he developed called jMAP that can be used to exter-
383 nalize and assess learners’ mental models. In his chapter, he presents findings from
384 two studies that illustrate how jMAP was used to support the assessment of causal
385 understanding, and to identify areas for future research and development. And
386 finally, Dan Hickey and colleagues discuss important issues relating to the assess-
387 ment of new media and technology proficiencies. In their chapter, they introduce
388 a design-based “participatory assessment framework” comprised of multiple lev-
389 els of increasingly formal outcomes, and urge researchers to focus first on defining
390 the contexts underlying social participation before attempting to assess individual
391 proficiencies.

392 Each kind of assessment has a role to play in improving teaching and learning,
393 and needs to be part of a total, balanced and blended assessment system. Using
394 different kinds of assessment will allow us discern learners’ knowledge, skills, and
395 other attributes from multiple perspectives, providing a clearer and more complete
396 picture of each learner (Fletcher, 2007). And the more we know about learners, the
397 better we can provide them with optimal support at the time they really need it.
398 Moreover, it’s crucial to involve learners in the assessment process through peer-
399 and self-assessment. These alternative assessment approaches stimulate the use of
400 higher-order thinking skills and help learners to understand more deeply (Shute,
401 2008).

402 We conclude with a set of principles of good assessment based on merged recom-
403 mendations from Kellough and Kellough (1999), Mislevy, Steinberg, and Almond
404 (2003), and Shute (2008):

- 406 • Understand and specify in advance of teaching the achievement targets (i.e.,
407 competencies) that learners are supposed to attain.
- 408 • Inform the learners, simply and clearly, about the competencies (as well as the
409 associated rubrics), from the very beginning of the teaching and learning process.
- 410 • Use classroom assessments to bolster learners' confidence and help them assume
411 responsibility for their own learning, toward the goal of engendering lifelong
412 learners.
- 413 • Translate assessment results into frequent, descriptive feedback (not judgmental,
414 subjective, or norm-referenced feedback), providing learners with specific
415 insights on how to improve.
- 416 • Continuously adjust instruction (whether classroom- or computer-based) relative
417 to the results of the formative assessments.
- 418 • Engage learners in regular self-assessment with standards held constant so that
419 they can watch themselves grow over time and feel empowered.

420 We posit that the most important and powerful feature of assessment is the use
421 of results to make improvements and decisions. This is true whether the assess-
422 ment is used to support personal learning or for accountability purposes. Another
423 important feature of assessment is to make learning—processes and products—
424 visible to all stakeholders. That is, a person's knowledge (and other mental states
425 and traits) is invisible to others, and sometimes to oneself (e.g., tacit knowledge).
426 Using an evidence-based assessment can contribute toward improved teaching and
427 learning (see Chapter 4 by Pearlman, this book), as well as help explicate eviden-
428 tiary arguments supporting claims about that knowledge (see Chapter 6 by Almond,
429 this book).

430 Knowing when to use a particular type of assessment and how to interpret
431 the results is not easy. Similarly, designing assessments using an evidence-based
432 approach is non-trivial. But consider the potential end result: i.e., assessments that
433 exert substantial influence on the quality of information provided to teachers and
434 learners to support instructional decision-making and meaningful learning. This
435 chapter has briefly touched on different assessment topics and approaches, call-
436 ing for a rational understanding of what we value in terms of competencies to
437 be instructed and assessed. Knowing what a learner knows comes from obtaining
438 quality evidence, which in turn is obtained from carefully designed assessments.
439 The ideas herein, but more importantly, throughout this book, are intended to sup-
440 port teachers, learners, and policy makers, and perhaps even inspire educational
441 researchers toward new, exciting projects.

444 References

- 446 Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Educational Assessment:*
447 *Principles, Policy and Practice*, 5(1), 7–74.
- 448 Dewey, J. (1916). *Democracy and education*. New York: Simon & Shuster.
- 449 Fletcher, G. (2007, October 19). *Assessing learning from a holistic approach: Creating a balanced*
450 *system of learning assessment*. Paper presented the Congreso Internacional Evaluacion Factor
de Calidad Educativa, Queretaro.

- 451 Georghiades, P. (2004). From the general to the situated: Three decades of metacognition.
452 *International Journal of Science Education*, 26(3), 365–383.
- 453 Guion, R. M. (1998). *Assessment, measurement and prediction for personnel decisions*. Mahwah,
454 NJ: Erlbaum.
- 455 Higgins, M. C. (2001). Changing careers: The effects of social context. *Journal of Organizational*
456 *Behavior*, 22(6), 595–618.
- 457 Hindo, C., Rose, K., & Gomez, L. M. (2004). Searching for Steven Spielberg: Introducing iMovie
458 to the high school English classroom: A closer look at what open-ended technology project
459 designs can do to promote engaged learning. In *Proceedings of the 6th international conference*
460 *on learning sciences* (pp. 606–609). Mahwah, NJ: Erlbaum.
- 461 Jeong, A. (2005). A guide to analyzing message-response sequences and group interaction patterns
462 in computer-mediated communication. *Distance Education*, 26(3), 367–383.
- 463 Kellough, R. D., & Kellough, N. G. (1999). *Secondary school teaching: A guide to methods and*
464 *resources; planning for competence*. Upper Saddle River, NJ: Prentice Hall.
- 465 Kliebard, H. (1987). *The struggle for the American curriculum, 1893–1958*. New York: Routledge
466 and Kegan Paul.
- 467 Macdonald, J. (2003). Assessing online collaborative learning: Process and product. *Computers*
468 *and Education*, 40(4), 377–391.
- 469 Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103).
470 New York: Macmillan.
- 471 Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational
472 assessment. *Measurement: Interdisciplinary Research and Perspective*, 1(1), 3–62.
- 473 National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for*
474 *school mathematics*. Reston, VA: NCTM.
- 475 National Research Council (NRC). (1996). *National science education standards*. Washington,
476 DC: National Academy Press.
- 477 Oosterhof, A. (2009). *Developing and using classroom assessments* (4th ed.). Upper Saddle River,
478 NJ: Pearson.
- 479 Schwartz, D. L., Bransford, J. D., & Sears, D. L. (2005). Efficiency and innovation in transfer. In
480 J. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 1–51).
481 Greenwich, CT: Information Age Publishing.
- 482 Shute, V. J. (2007). Tensions, trends, tools, and technologies: Time for an educational sea change.
483 In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 139–187).
484 New York: Lawrence Erlbaum Associates, Taylor & Francis Group.
- 485 Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189.
- 486 Shute, V. J., Jeong, A. C., Spector, J. M., Seel, N. M., & Johnson, T. E. (2009). Model-
487 based methods for assessment, learning, and instruction: Innovative educational technology at
488 Florida State University. In M. Orey (Ed.), *2009 educational media and technology yearbook*
489 (pp. 61–80). Westport, CT: Greenwood Publishing Group.
- 490 Shute, V. J., & Zapata-Rivera, D. (2010). Educational measurement and intelligent systems. In
491 E. Baker, P. Peterson, & B. McGaw (Eds.), *Third edition of the international encyclopedia of*
492 *education* (pp. 75–80). Oxford, UK: Elsevier Publishers.
- 493 Sungur, S. (2007). Modeling the relationships among students' motivational beliefs, metacogni-
494 tive strategy use, and effort regulation. *Scandinavian Journal of Educational Research*, 51(3),
495 315–326.
- 496 Terman, L. M. (1916). *The measurement of intelligence*. Cambridge, MA: Riverside Press.