NOTE: This paper is a preprint, it is not the copy of record and may not exactly replicate the final, authoritative, peer-reviewed version of the article. The final article will be available, upon publication:

Nichols, S. & Thacker, I., (2026). Chapter 15: Assessment and Learning. In T. Marshik, M. I. Axelrod, M. B. Leibham, & S. Hupp (Eds), *Investigating Educational Psychology: Pseudoscience, Fringe Science, and Controversies.* Routledge.

Chapter 15: Assessment and Learning.

Sharon Nichols and Ian Thacker

Assessment is the systematic process of defining, measuring, and interpreting student learning. It involves collecting information about what students know, understand, and can do in relation to learning goals and helps educators, parents, and students make informed decisions about instruction, curriculum, and student progress. Assessments can be developed by teachers or by individuals outside of schools, such as those who create standardized tests. Assessments serve several purposes: they can guide real-time instructional adjustments (formative), evaluate whether learning objectives have been met (summative), uncover students' prior knowledge and gaps (diagnostic), and can even be used judge the effectiveness of teachers or schools. Given these multiple purposes, assessments are vital tools for teaching and learning; and they play an important role in how students, teachers, parents, and other stakeholders respond to student progress. The goal of this chapter is to focus on two core claims related to assessment, testing, teaching, and student learning. The claims we consider are: (a) teaching to the test benefits student learning, and (b) grades signify and promote learning.

Examining the Claims

Claim#1: Teaching to the Test Benefits Student Learning

One of the purposes of tests and testing is to assess what students know and don't know. To that end, it seems reasonable that many would think teaching "to" the test is beneficial. After all, isn't that the entire goal of schooling in the first place? To teach learners what will be on the test so that they will be able to demonstrate what they know? The degree to which "teaching to the test" matters or is problematic depends largely on the goals of instruction, the type of knowledge being taught, the purpose of the assessment, and the extent to which students are expected to *transfer* their learning out of the classroom.

Primer: A Lesson on Transfer

Before deconstructing the claim, we begin with the assertion (and assumption) that one of the most important goals of education is to help students take the knowledge, skills, and attitudes they learn inside the classroom and apply them to what they do outside the classroom—

otherwise referred to as *transfer*. We believe that education, and therefore, teaching, is about helping students understand *when and how to apply* knowledge, skills, attitudes, and dispositions learned through the process of schooling to what they choose to do outside of school (McKeough et al., 2013). Reciprocally, it is equally important that students be able to apply what they bring from outside of school to what they do inside of school, such as offering students opportunities to convert currencies in mathematics, which is an established skill among immigrant families (Moll et al., 2006).

Unfortunately, myths in education are abundant, and there are some pertaining to the topic of transfer that we should address here. For example, there is a myth that some general, content-free cognitive skills can be taught and subsequently transferred broadly across all domains, such as the problem-solving skills learned through chess. More specifically, some people believe that the skills learned through chess will improve creativity, intellect, and memory in unrelated contexts (i.e., if I learn the strategy of chess I will automatically know when and how to apply those kinds of strategies to other contexts, e.g., De Bruyckere et al., 2020), yet empirical evidence does not support this view. Instead, research indicates that the ability to transfer learning depends on the extent to which the original and target contexts share underlying knowledge structures (Christodoulou, 2014; De Bruyckere et al., 2020; Willingham, 2019)—for example, people who are proficient in a language tend to learn languages faster when the target language is similar in structure, grammar, and lexicon (Puig-Mayenco et al., 2020). Thus, transfer is firmly rooted in content knowledge.

When teaching for transfer, teachers must ask themselves two essential questions. First, "what knowledge is being taught?" And second, "in what contexts, situations, and/or culturally based practices do we want students to demonstrate this knowledge?" If we only ask the first question ("what do we want students to know?"), and answered the question with the topic of, for example, subtraction, then, it stands to reason that all we would focus on is that particular outcome—subtraction. But if we *also* ask, "in what contexts and in what ways do we want students to be able to demonstrate what they know," then we are also concerned with real-world applications of subtraction knowledge. We want students to know math, but we also want them to know when and how to apply it—to recognize the situations that are calling for it. So, ideally, we want both. For example, if we want students to demonstrate their knowledge of subtraction of double digit numbers, then we might teach them how to do double digit subtraction through repetition and drill exercises and then create a test that has a series of double digit subtraction problems that *look like* the kinds of problems students were given while they were learning the skill (e.g., "solve: 99 – 50 = ?"). However, if we are also interested in seeing how well students not only understand how to do two-digit subtraction but also when and where to use that knowledge, then we might teach (and test) that knowledge differently. For example, in addition to double digit subtraction drills, we might *also* expose students to word problems ("how much change would you give if a candy bar costs 50 cents and someone gave you 99 pennies?"), helping them to see various situations that might call for two-digit subtraction skills.

The topic of transfer is important background information for the following sections in which we explore different types of assessments and the ways in which we think about what it means to "teach to the test." Next, we consider this issue across common test types: teacher-made tests, standardized, and high-stakes standardized tests.

Teaching to the Test and Classroom-Based Assessments

Tests serve a variety of roles and purposes in the classroom (Black & Wiliam, 1998; Brookhart, 1997; Crooks, 1988). They can be teacher-made or developed by outsiders as in standardized tests, and they can be high- or low-stakes to the students. Teacher-made tests have varied formats (multiple choice, essay, short answer, performance-assessments), roles (graded, ungraded, quiz, unit test) and purposes (to identify student needs, clarify achievement expectations, motivate, and evaluate instructional effectiveness) and are used frequently by teachers to guide instructional decision-making (Crooks, 1988; Stiggins & Bridgeford, 1985; Stiggins & Conklin, 1992). For most students, teacher-made tests are a low-stakes situation since passing/failing a single classroom test does not lead to life-altering consequences (i.e., students have multiple opportunities throughout any given year to demonstrate proficiencies). When it comes to most forms of teacher-made, classroom-based tests, teaching to the test is less of a concern. Afterall, a significant goal of these tests is to gauge how well students have learned the material the teacher just covered. When it comes to classroom-based, teacher-designed tests, the goal teachers have in mind is typically something referred to as ***near transfer*** meaning they are using tests to gauge how well students learned the specific information and form of knowledge that was just taught (Marini & Genereux, 2013). Here, it may be more acceptable for teachers to "teach to the test" (in general) because the main goal of the test is to see how well students have learned the recent content. For example, in social studies, teachers might want students to memorize key dates during a lesson. In this situation, learning involves repetition of information presented in a variety of ways (flash cards, fill in the blank). Here the test would include similar items to those they practiced and as such reflects a "teaching to the test" approach in which instruction and testing items are relatively well aligned. This is true across all other content areas taught in school—when the goal is to see if students have learned the lesson, topic, or process (how to play A on a flute, or how to make a candle holder in metal shop), then it would make sense the teacher spend time directly teaching and then testing for those skills, knowledge or outcomes. Teaching to the test is largely acceptable in these instances.

By contrast, for information that is more complex, or that requires deeper cognitive processing skills such as evaluation, analysis, and application, the test might not exactly match the instruction. Here, teachers want students to develop deeper understandings that enable them to recognize what is being asked even if it is presented to them in a new way (**far transfer**). In this situation, the perils of teaching to the test are more consequential. If a teacher instructs a student to memorize one fact—the only reason to go to war is to protect geographic boundaries—then that student (a) will do exceptionally if asked to describe why countries go to war when their geographic boundaries are at risk, but (b) will fail if the question is posed in any other way (e.g., "Discuss several reasons countries go to war," or "What are pros and cons of going to war?"). This problem is further illustrated with the current emphasis on standardized tests.

Teaching to the Test and Standardized Tests

Standardized tests are developed outside of the classroom, are administered infrequently, are often given to large bodies of students, and have *historically* had a minimal direct influence on teacher practice (Stiggins & Bridgeford, 1985). Standardized tests fall into one of two main categories that define how they are ultimately used, namely criterion based and normative based. Criterion based tests are designed to measure student learning against some predefined target body of knowledge. In the US, students take the National Assessment for Education Progress (or

NAEP) test, performance on which informs policymakers how "well" students are learning in specific content areas and grade levels (e.g., 4th grade math, 8th grade reading). These standardized tests are about how well students are learning certain topic areas, results of which are used for a variety of purposes including to make decisions about the quality of schools or effectiveness of teachers. Currently, every state in the USA has adopted some type of criterion test, and results help that state determine how much and how well students are learning various topic areas.

Standardized tests may also be "normative based" in which they are designed to measure students' achievement against each other. We have many examples of these kinds of tests such as the SAT or GRE or any test students take to apply to school. These tests are designed in such a way to gauge how students perform next to one another and are often used to make selection decisions, such as determining college admissions, eligibility for gifted programs, or placement into advanced courses.

The goals and purposes of most standardized tests (both criterion and normative referenced) are to gauge students' **general** knowledge (or **far transfer**) in each topic area and therefore, they are meant to be relatively "immune" to teaching to the test pressures. By way of an example, consider that every year in the United States, third graders take statewide standardized tests designed to probe something called "third grade reading knowledge" and "third grade math knowledge." Compared to classroom-based tests, the aim of these assessments is much grander in scope and design. Also, given they are often designed by entities outside of the school systems, and therefore teachers and students are not aware of the specific items, it is more difficult to "teach to the test." As long as the test is designed well (includes items that cover the possible universe of items and content areas that were taught) and the conditions under which teaching and learning in preparation for the test are "normal" (all students exposed to same topics with similar opportunities to practice their knowledge and learning), then we can trust the result of the test to meaningful portray what students learned. However, if any of these conditions is violated (test designed poorly, conditions of learning are altered), then we begin to worry about the negative consequences "teaching to the test" might yield.

High Stakes Testing-Changing The Conditions of Teaching and Learning

In the U.S., high-stakes tests are standardized tests (mostly criterion referenced) that are used to hold teachers and students "accountable" for what they learn. The rationale for this use of tests is that by attaching significant rewards or serious threats to student test scores, teachers and their students will be motivated to work harder, better, and learn more (as measured by tests) (Ryan, 2004). Although the practice of high stakes testing is not new (Giordano, 2005), -tests have been used to distribute rewards and sanctions to teachers in urban schools since the mid-1800s (Tyack, 1974) and for many schools throughout the United States since at least the 1970s (Herman & Haertel, 2005). The passage of the No Child Left Behind Act (2002, NCLB) [and continuing under the Every Student Succeeds Act to the time of this writing (2015, ESSA)] essentially codified the practice for all teachers and students in public schools across the nation. NCLB mandated that all states must test students annually in core subjects areas and use the results for accountability purposes. The pressures associated with ensuring students score well on these tests (performance on which may determine teacher pay raises, public school ratings) means the **conditions under which students learn and prepare for these tests have been altered**. Whereas prior to NCLB standardized tests and testing did not include these pressures,

since 2002, the number, role, and purposes of standardized tests students take has radically shifted how teachers teach and students learn. And although the next iteration of NCLB (Every Student Succeeds Act) somewhat lessened the role of tests in accountability, most states continue to use tests as part of accountability mechanisms.

Under these laws, we have learned that the pressures associated with ensuring students score well have led to fundamental changes in ways teachers teach such that they engage more directly in preparing students to take the test (Au, 2007, 2022; Jones et al., 1999; Nichols & Berliner, 2007). For example, data reveal that subject areas not on the test (music, art) are increasingly cut from the curriculum, while subject areas that are tested receive more time throughout a school day and year (Heilig et al.,, 2010). Surveys conducted by the Center on Education Policy (CEP) with a nationally representative sample of school districts found that from 2000-2001 to 2006-2007 (and the federally mandated institution of high-stakes testing), 62% of districts had increased the amount of time dedicated to English Language Arts (ELA) and Math (most likely to be tested) in elementary schools. And, on average, districts reported an increase of 141 and 89 weekly minutes added to instructional time for ELA and math (respectively) (McMurrer, 2008; Nichols & Berliner, 2007), a trend that has persisted over time (Hoyer & Sparks, 2017).

These changing conditions under which teachers teach and students learn raise questions about the meaningfulness of test scores. If I learn about math under "normal" conditions that are similar across all schools and districts, then my math score is more representative of what I can do in math (as it is for other students). However, if I learn about math being taught for more hours a day, with lessons being repeated regularly and with me being given lots of practice tests, then what is the "meaningfulness" of my resultant test score as it compares to other students who might not experience this similar type of learning environment? The fact that learning conditions have changed dramatically raise questions about resultant test scores and the meaningfulness of student learning (Nichols & Berliner, 2007). In some cases, it could be argued these are the effects of greater "teaching to the test" efforts due to the pressures to ensure students score well.

Claim #2: Grades signify and promote learning

The second claim we consider is that "*grades signify and promote learning*." For grades to *signify* learning, learning and grades would have the same meaning (e.g., a good grade implies that substantial learning has occurred, and conversely, learning implies a good grade). For grades to *promote* learning means that the presence of a grading system enhances learning (e.g., students learn more when an assignment is graded compared to when is not). While it might be easy to assume that learning and grades are one and the same or that they boost learning given the large emphasis of grades in schools, the evidence is largely not consistent with Claim 2.

The Multiple Purposes of Grades

The claim that grades both *signify* and *promote* learning has many problems. First, we discuss issues with the idea that grades signify learning and then issues with the claim that they promote learning.

Grades may potentially signify some aspects of learning, but they also signify a lot more. First, grades have multiple purposes; there is no consensus on a singular purpose of grading (Brookhart, 2011). Grades can be used to track student progress and mastery of content. They

can also be used to communicate student achievement to stakeholders, such as parents, school admin, college admissions committees, and potential employers (Bailey & McTighe, 1996). As such, grades are often used to capture a wide range of cognitive and "non-cognitive" dimensions such as motivation, self-regulation, and social skills (Brookhart et al., 2016). For example, it is common practice in the USA for teachers to design their own grading criteria at both the primary, secondary, and tertiary level (e.g., by creating their own rubrics, or by grading using implicit criteria; Guskey & Link, 2019; Lipnevich et al., 2020). Sometimes, teachers design their grading criteria to reward students for good behavior, attendance, effort, study habits, products (e.g., grades for assignment *completion*), test performance, or learning processes (e.g., grades for collaborative *processes*), and decide on how these criteria are weighted (Guskey & Link, 2019; Lipnevich et al., 2020; Wentzel, 1993). For this reason, grades often capture far more than cognitive dimensions and mastery of skills but also capture non-cognitive dimensions as well (e.g., motivation, self-regulation, or social skills). To further complicate matters, there is variability in what teachers include in their grading systems. While some may include non-cognitive factors, and allow students multiple attempts to master learning goals, others may not (e.g., Lipnevich et al., 2020).

Furthermore, grades on their own may not always be effective in helping students learn or acquire new skills (Klapp, 2015). For example, Lipnevich and Smith (2009) found that college students who received detailed written feedback on an essay geared towards improving writing performed better than students who received only a grade. Moreover, students who received both a grade and written feedback did not improve as much as students who received feedback with no grade, suggesting that grades can sometimes undermine the effect of more specific feedback.

Additional studies support the same idea: that grades are not necessarily linked to learning or motivation. Chamberlin, et al., (2023) compared students who received narrative evaluations and pass/fail grades to traditional letter grades, finding that grades did not enhance motivation and instead undermined students' reported intrinsic motivation and feelings of autonomy, as measured by the Academic Motivation Scale (Vallerand et al., 1992). This suggests that letter grades may not be effective in helping students improve the quality of their work, which is also consistent with literature on extrinsic motivators.

We would also like to point out that *learning is not the same thing as performance*. There are many definitions of learning across multiple disciplines (Barron et al., 2015), but to select one that is compatible with many traditions in education research, Ormrod (2020) defines learning as "a *long-term* change in mental representations or associations as a result of experience" (p. 4). Performance, on the other hand, can be a *short-term* change in behavior. For example, an "A student" who memorizes a collection of Spanish verbs before the test may not retain information as well as the "C student" who is inspired a Spanish lecture and begins reading the untranslated version of Don Quixote; this student may be less prepared for the test, but they actively apply Spanish skills over time. Indeed, higher performance on an exam does not necessarily indicate that a student learned something—a student with low levels of prior knowledge who works very hard and improves their content knowledge dramatically may perform lower than their peers who entered the classroom with higher levels of prior knowledge and didn't learn much from instruction. As such, while grades may contribute to extrinsic motivation and quantity-based performance, there is no guarantee that the performance represents a long-term change.

Incentives and Performance

While the evidence largely does not support the claim that grades signify and promote learning, depending on how the term "learning" is defined, and depending on how grades are implemented and understood by students, there are some aspects of this claim that have supporting evidence. For example, grades are sometimes framed as an incentive, the dangling carrot that drives students to complete work. Indeed, grades can be viewed as a reward (or sometimes as a punishment) given from a teacher to a student in exchange for achievement—also called an *extrinsic motivator* (Lepper & Henderlong, 2000).

Extrinsic incentives, in general, are associated with performance—depending on how performance is defined. For example, Cerasoli et al., (2014) demonstrated in a meta-analysis across 125 school-based (and 54 work- or fitness-based) studies that extrinsic motivators, such as the presence of grades or other incentives (e.g., money, promotions, awards, health benefits, praise), predict performance on tests or other readily measurable tasks (e.g., task completion, essay quality). However, an interesting finding of this meta-analysis was that extrinsic incentives were a better predictor of *quantity-based* performance*,* such as the number of arithmetic problems answered, number of books read, or test score achieved (see, e.g., Locke & Latham, 2002). In contrast, *intrinsic motivation*—the motivation derived from enjoyment and personal investment in a task—was a better predictor of *quality-based* performance, such as the quality of an essay, creativity demonstrated in a project, or depth of understanding (see e.g., Vansteenkiste et al., 2006). In other words, grades might incentivize quantity of work in the short-term, but extrinsic motivators can condition learners to be more passive and less invested than the intensity of effort demonstrated by intrinsically motivated learners.

But what if a learner is graded for a task that they already enjoy? While early evidence on the suggested that incentivizing intrinsically motivated tasks can undermine student motivation (Lepper et al., 1973); mounting evidence seems to show that the effects depend on the details of the situation (Lepper & Henderlong, 2000), where oftentimes, extrinsic motivation tends to *complement* intrinsic motivation. For example, Lin et al. (2003) found that intrinsic and extrinsic motivation in combination predicted the higher levels of performance. College students with moderate (but not extreme) levels of extrinsic motivation to get good grades who also had high levels of intrinsic motivation achieved better grades than students with higher or lower levels of extrinsic motivation. Once again though, we caution readers to remember that higher *grades* do not necessarily equate to higher *learning*.

Summary and Conclusion

Assessments, while in some ways essential for understanding what students can and cannot do, are never perfect. Assessments come in many forms, are used for a variety of purposes, and can reveal a lot or very little about teaching and learning. In American society, we ask grades and assessments to do a lot. When it comes to standardized tests specifically, policymakers have endorsed a system where these tests are expected to act as both "detectors" of change (are students learning over time? Are teachers effective at teaching) and "effectors" of change (as carrots and sticks to incentivize more effective and efficient teaching and learning in schools). The role of grades also has multiple uses and meanings (grades signify how much a

student has learned, but also, they can promote effort but not learning). Therefore, it makes sense there is confusion about their efficacy and use in education. We end with the takeaway that readers should be critically conscientious about assessment results, taking care not to conflate the indicator of learning with inferences about learning itself.

References

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher, 36*(5), 258-267.

Au, W. (2022). *Unequal by design: High-stakes testing and the standardization of inequality.* Routledge.

Bailey, J. M., & McTighe, J. (1996). Reporting achievement at the secondary level: What and how. In T. R. Guskey (Ed.), Communicating student learning: 1996 Yearbook of the Association for Supervision and Curriculum Development (pp. 119–140). Alexandria, VA: ASCD.

Barron, A.B., Hebets, E.A., Cleland, T.A., Fitzpatrick, C.L., Hauber, M.E., & Stevens, J.R. (2015). Embracing multiple definitions of learning. *Trends in Neuroscience, 38* (7), 405-407.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7-74. doi:10.1080/0969595980050102

Brookhart. S. M. (1997). A theoretical framework for the role of classroom assessment in motivating student effort and achievement. *Applied Measurement in Education, 10*, 161-180.

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., ... & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research, 86*(4), 803-848.

Brookhart, S. M. (2011). *Grading and learning: Practices that support student achievement.* Bloomington, IN: Solution Tree Press.

Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. Psychological Bulletin, 140(4), 980–1008.

Chamberlin, K., Yasué, M., & Chiang, I. C. A. (2023). The impact of grades on student motivation. Active Learning in Higher Education, 24(2), 109-124.

Christodoulou, D. (2014). Seven myths about education. Routledge.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*, 438-481.

De Bruyckere, P., Kirschner, P. A., & Hulshof, C. (2020). If You Learn A, Will You Be Better Able to Learn B? Understanding Transfer of Learning. American Educator, 44(1), 30.

Giordano, G. (2005). *How testing came to dominate American schools: The history of educational assessmen*t. New York, NY: Peter Lang.

Guskey, T. R., & Link, L. J. (2019). *Exploring the factors teachers consider in determining students' grades.* Assessment in Education: Principles, Policy & Practice, 26(3), 303-320.

Heilig, J. V., Cole, H., & Aguilar, A. (2010). From Dewey to No Child Left Behind: The Evolution and Devolution of Public Arts Education. *Arts Education Policy Review*, *111*(4), 136–145. https://doi-org.libweb.lib.utsa.edu/10.1080/10632913.2010.490776

Herman, J. L., & Haertel, E. H. (Eds.). (2005). *Uses and misuses of data for educational accountability and improvement: The 104th yearbook of the National Society for the Study of Education, Part II*. Malden, MA: Blackwell.

Hoyer, K. M. & Sparks, D. (2017). *Instructional time for third- and eighth-graders in public and private schools: School Year 2011-2012*. Stats in Brief. NCES 2017-076. http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2017076

Jones, M. G., Jones, B. D., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999). The impact of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan, 81*(3), 199-203.

Klapp, A. (2015). Does grading affect educational attainment? A longitudinal study. Assessment in Education: *Principles, Policy & Practice, 22*(3), 302-323.

Lepper, M. R., & Henderlong, J. (2000). Turning "play" into "work" and "work" into "play": 25 years of research on intrinsic versus extrinsic motivation. In C. Sansone & J. M. Harackiewicz (Eds.), Intrinsic and extrinsic motivation: The search for optimal motivation and performance (pp. 257-307). San Diego, CA: Academic Press.

Lin, Y. G., McKeachie, W. J., & Kim, Y. C. (2003). College student intrinsic and/or extrinsic motivation and learning. Learning and individual differences, 13(3), 251-258.

Lipnevich, A. A., & Smith, J. K. (2009). Effects of differential feedback on students' examination performance. *Journal of Experimental Psychology: Applied*, 15(4), 319.

Lipnevich, A. A., Guskey, T. R., Murano, D. M., & Smith, J. K. (2020). What do grades mean? Variation in grading criteria in American college and university courses. Assessment in education: Principles, policy & practice, 27(5), 480-500.

Locke, E. A., & Latham, G. P. (1990). Core findings. In E. A. Locke & G. P. Latham (Eds.), *A theory of goal setting and task performance* (pp. 27-62). Englewood Cliffs, NJ: Prentice Hall.

Marini, A., & Genereux, R., (2013). The challenge of teaching for transfer. In A. McKeough, J. L. Lupart & A. Marini (Eds.), *Teaching for transfer: Fostering generalization in learning* (1-19). Routledge.

McKeough, A., Lupart, J. L., & Marini, A. (2013) (Eds.), *Teaching for transfer: Fostering generalization in learning*. Routledge.

McMurrer (2008). *Instructional time in elementary schools: A closer look at changes for specific subjects*. Washington, DC: Center on Education Policy.

Moll, L., Amanti, C., Neff, D., & Gonzalez, N. (2006). Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. In *Funds of knowledge* (pp. 71-87). Routledge.

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

Ormrod, J. E. (2020). *Human learning* (8th edition). Pearson.

Puig-Mayenco, E., González Alonso, J., & Rothman, J. (2020). A systematic review of transfer studies in third language acquisition. *Second Language Research, 36*(1), 31-64.

Ryan, J. (2004). The perverse incentives of the No Child Left Behind Act. *New York University Law Review, 79*, 932-989.

Stiggins, R. J., & Conklin, N. F. (1992). *In teachers' hands: investigating the practices of classroom assessment*. Albany, NY: State University of New York Press.

Stiggins, R. J., & Bridgeford, N. J. (1985). The ecology of classroom assessment. *Journal of Educational Measurement, 22*(4), 271-286.

Tyack, D. (1974). *The One Best System: A History of American Urban Education.* Cambridge: Harvard University Press.

Vallerand, R. J., Pelletier, L. G., Blais, M. R., Briere, N. M., Senecal, C., & Vallieres, E. F. (1992). The Academic Motivation Scale: A measure of intrinsic, extrinsic, and amotivation in education. *Educational and Psychological Measurement, 52*(4), 1003-1017.

Vansteenkiste, M., Lens, W., & Deci, E. L. (2006). Intrinsic versus extrinsic goal contents in self-determination theory: Another look at the quality of academic motivation. *Educational Psychologist, 41*, 19 –31. doi:10.1207/s15326985ep4101_4

Wentzel, K. R. (1993). Does being good make the grade? Social behavior and academic competence in middle school. *Journal of Educational Psychology, 85*(2), 357–364.

Willingham, D. T. (2019). How to Teach Critical Thinking. NSW Department of Education. http://www.danielwillingham.com/uploads/5/0/0/7/5007325/willingham_2019_nsw_critical_thinking2.pdf