WebbAlign™

WCEPS

Complexity and Difficulty in a Coherent Standards-Based Education System

Sara Christopherson, Director of WebbAlign
Dr. Norman Webb, WebbAlign Program Advisor

June, 2024

WebbAlign is a program of the Wisconsin Center for Education Products and Services (WCEPS). WCEPS is a non-profit organization affiliated with the University of Wisconsin – Madison.

www.webbalign.org

The premise of standards-based education is that setting high expectations for all students can effect positive change through systemic coherence. Academic standards include multiple types of expectations–for example, as relates to perseverance, disposition, collaboration, and creativity. However, a main purpose of academic standards is to define expectations for particular types of cognitive engagement with academic concepts and ideas. More specifically, a central goal of academic standards is to shift away from a focus on recall of knowledge and skill development and toward conceptual understanding and use of higher order thinking (Porter, 1989). This central goal is reiterated in today's standards across content areas. The term **cognitive complexity** is used to describe the type of thinking required by students' interactions with academic content. The terms *higher-order thinking* and *rigor* are often used to refer to high cognitive complexity.

Academic standards are the referent that anchors a coherent system (Fulmer, 2018; Webb, 1997). One of many aspects of systemic coherence is the need for sufficient consistency between the complexity of cognitive engagement specified by the standards and the extent to which that complexity plays out in all parts of the system – including curriculum, learning opportunities, and assessments. For example, to meet the expectations within the Every Student Succeeds Act (ESSA), states must adopt "rigorous" academic content standards and statewide assessments are expected to reflect the "depth" or "cognitive complexity" explicit in those standards (US ED, 2018).

Academic standards vary in their complexity. Although today's academic standards represent an overall shift toward higher complexity expectations, they include a range of expectations–from simple to complex–within each grade or grade band (Sato, et al., 2011; Christopherson, 2019). A key underlying assumption of the standards-based model is that a set of clearly and carefully defined expectations can be reliably interpreted by all stakeholders. Coherence relies on a common interpretation of the cognitive demands of the standards. Inconsistent interpretations hamper the equity-focused goals of a standards-based system.

Depth of Knowledge (DOK): A Tool to Support Coherence in a Standards-Based Education System

In a standards-based education system, the expected complexity of cognitive engagement is communicated through the language of academic standards and applies to all students. Therefore, a standards-based system requires that expectations for the complexity of cognitive engagement can be determined through a content analysis of academic content, including standards, prompts, questions, and tasks. In this way, the standards can inform instructional design and content development to elicit thinking that is as cognitively demanding as the expectations in the corresponding learning targets.

To operationalize the intended cognitive engagement as expressed in the standards, educators, content developers, and other stakeholders benefit from the use of practical frameworks and tools (Stein, et al., 2016). Many different tools have been developed to help guide interpretations of cognitive demands for different purposes and in different contexts. For example, some tools are focused specifically on task analysis, while others are focused on evaluation of student responses. Depth of Knowledge (DOK) is a tool that was developed to evaluate and promote content alignment. DOK facilitates interpretation, evaluation, and communication about the cognitive demands of the standards and other educational materials in ways that are compatible with the intent of the standards and the underlying ideas about the nature of learning. By categorizing the qualitatively different types of cognitive demands within the standards, it becomes possible to compare the demands of an expectation with the demands of a question, prompt, or task. Content developers, educators, evaluators, and others use DOK to work with greater efficiency and intentionality to promote content alignment in support of a coherent system.

The four levels of DOK were derived from the standards themselves by sorting academic expectations according to subject area and type. Four broad categories emerged:

- **DOK 1** included expectations for recall of, reproduction of, or fluency with taught knowledge or processes;
- **DOK 2** included expectations requiring application of underlying conceptual understanding and emphasizing relationships between and among ideas/concepts/processes;
- **DOK 3** expectations focused on non-routine and abstract problem-solving or inferencing, sometimes requiring authentic evaluative and argumentative processes; and
- **DOK 4** expectations were at least as complex as DOK 3 but required iterative processes as well as extended and metacognitive thinking over time to complete.

Four levels were found to be necessary and sufficient for purposes of categorizing the types of expectations in academic standards in a way that is useful for educators, content developers, and evaluators. The categories and definitions are intended as a tool; they are not intended to suggest that there are four natural or discrete categories of complexity. However, because academic standards emphasize the importance of particular types of cognitive engagement, specifying meaningful categories helps to identify the characteristics of these qualitatively different types of cognitive engagement – as described within the different sets of expectations. Rater agreement statistics across decades of use in evaluative studies show that the categories can be applied consistently with appropriate training.

What constitutes complexity varies by subject area. In other words, while it is possible to sort expectations into similar general categories, what each category of complexity 'looks like' is specific to the types of thinking and mental processing that occur in each content area. This perspective is consistent with today's academic standards: expectations within each subject area reflect domain-specific epistemic practices and ways of knowing. Thus, the subject-specific DOK definitions help to clarify each DOK level as it plays out in the standards for each subject area.

Key Factors that Influence the Complexity of an Expectation or Task

Variations of the word 'complexity' are commonly used in relation to many different units of analysis, including the processes of teaching, learning, and problem solving; student performance; and content area topics. *Cognitive* complexity is also differentiated from *text* complexity[1] as well as *linguistic* complexity[2]. The DOK language system is a tool used for content analysis, to differentiate types of cognitive complexity as expressed in expectations, questions, prompts, tasks, and other components of curriculum and assessment.

In general, the degree of complexity of an expectation or task depends on how students must interact with the disciplinary concepts and ideas to successfully demonstrate the intended outcome or complete the task. The evaluation is not focused on the learning processes leading up to or underlying the final outcome but instead on the demands of the outcome itself; the ''end point.' Some key factors that affect expectation or task complexity include the amount of mental processing required, the abstractness of the reasoning, the interdependence of parts, and the engagement with context (**Figure 1**). However, because different subject areas are grounded in different disciplinary practices, complexity 'looks' different in different subject areas.

Figure 1. Key factors that influence the complexity of an expectation or task, etc.

Low complexity	**High complexity**
Concrete ideas and concepts	Abstract/Hypothetical ideas and concepts
Discrete facts or ideas in isolation	Intricacies and/or dependencies between/among ideas
Low processing of concepts and skills	High processing of concepts and skills
Routine/typical problem type	Non-routine/novel problem type
Straightforward/literal thinking	Inferential/figurative thinking
No transfer	Far transfer
No metacognitive demands	Extensive metacognitive demands

[1] The terms *text difficulty* and *text complexity* are often used interchangeably. Within the CCSS, text complexity is defined as "how easy or difficult a particular text is to read" (Appendix A).
[2] Similarly, the terms *linguistic difficulty* and *linguistic complexity* are often used interchangeably.

DOK Levels Represent Levels of Complexity, but Not Levels of Priority, Sequence, or Progression

Expectations at each level of DOK can be of high priority. In general, the types of expectations that are prioritized are, by definition, those that are included in the academic standards for each subject area. The DOK levels also **do not** represent a leveled sequence of instruction or progression of learning or performance. Learning does not take place strictly from low complexity to high complexity. If learning progressed, as a rule, from low complexity to higher complexity, then we would expect low complexity expectations in the academic standards for lower grades and high complexity expectations in higher grades. Instead, today's standards include different types of expectations–including high complexity expectations–for students across all grade levels (Sato, et al., 2011; Christopherson, 2019). This reflects a contemporary understanding of learning as a process that follows non-linear and potentially interacting sequences (NASEM, 2007 p. 221; Songer and Gotwals, 2012).

In some cases, higher complexity expectations in lower grades purposefully precede lower complexity expectations in higher grades. For example, the Common Core Mathematics Standards emphasize conceptual strategies such as decomposing numbers in Kindergarten and Grade 1 – *before* students are expected to recall single-digit sums from memory in Grade 2. This sequencing is based on research that shows that traditional approaches of teaching basic skills first (e.g., memorizing addition facts) did not promote the intended conceptual understanding of mathematical operations, as was assumed. Instead, student outcomes were improved by *starting* with higher complexity engagement with math concepts, and only *later* expecting fluency with operations as well as recall from memory. As described in the Progressions, the Grade 2 expectation to recall sums from memory is intended to be achieved "not as a matter of instilling facts" but rather "as an outcome of a multi-year process that heavily involves the interplay of practice and reasoning" (Common Core Standards Writing Team, 2022). Overall, academic standards include a variety of types of expectations that are qualitatively different in terms of the complexity of engagement required. The use of DOK can help to focus attention on these important features of the standards, and ensure they are translated into learning and assessment opportunities for students.

Academic standards are meant to reflect research and best understanding of teaching and learning. A key aspect of this best understanding is that engaging students with high complexity tasks can improve outcomes, including as relates to basic skills (Darling-Hammond, et al. 2020; Agarwal, 2019; Smith, et al., 1996). Further, starting with low complexity tasks, such as memorization, can sometimes be "colossally inefficient" (Zimba, J., as quoted in Northern, 2016). The collection and sequence of expectations within and across the grade-level or grade-band standards reflects consideration of multiple factors related to teaching and learning, including efficiency and effectiveness. While the standards do not specify "how" to teach, they are intended to influence teaching practice.

Cognitive Complexity, Standards, and Underlying Learning Models

Many different models are used to help understand the nature of learning. Different models are necessary for different contexts. For example, learning to read is a qualitatively different endeavor than learning to add and subtract. Learning to read, add, and subtract are all different from learning how to analyze the credibility of a source of information when conducting research. Etcetera.

Although different learning models may be better fits for different circumstances, most of today's models live within the cognitivist and constructivist families. In general, learning tends to be considered a qualitative change but it is sometimes represented as including a change in amount, scope, or quantity. It is generally agreed that learning occurs in myriad ways, "does not customarily follow linear, sequential steps of development" (Gotwals and Songer, 2012), and although it is not necessarily directional at the finer grain, learning becomes more thorough or complete over time while reasoning and problem-solving strategies generally become more sophisticated over time (e.g., NASEM, 2007, Alonzo and Steedle, 2009, Steedle and Shavelson, 2009).

In contrast to a stepwise directional sequence, learning has also been envisioned as a non-linear "conceptual ecology," particularly within mathematics and science (Posner, et al., 1982; diSessa, A. 2002; Hammer and Kikorski, 2015), in which learning involves an interconnected network of ideas that get restructured and elaborated as needed. OECD Future of Education and Skills 2030 (2019) asserts that, from an international perspective, "approaches to curriculum design and learning progression [are] shifting from a "static, linear learning-progression model" to a "non-linear, dynamic model." Although background knowledge and prerequisite skills are important for higher complexity tasks, these skills can be envisioned as part of an interconnected network, rather than as the 'foundation' of higher complexity thinking (Darling-Hammond, et al., 2020).

Consistent with current thinking about learning, today's standards include qualitatively different types of expectations–including rigorous, high-complexity expectations–for all grades across the K-12 grade span. The standards also emphasize conceptual if not empirical learning progressions, represented as the development of increased sophistication of thinking across the grades. When considering sequencing decisions for assessment targets, Shepard, et al. (2013) point out that "postponing more complex reasoning about subject matter would be antithetical to the intentions of both the CCSS and learning progressions research." Attention to complexity as distinct from difficulty promotes coherence because the way(s) in which stakeholders conceptualize the unfolding of learning, as relates to difficulty and complexity, influences decisions about how learning and assessment opportunities are structured.

Differentiating Difficulty from Complexity

Within the literature, the terms 'difficulty' and 'complexity' are not necessarily defined and are often used interchangeably to refer to various units of analysis including expectations, tasks, topics, learning processes, and outcomes (e.g. NRC, 2001, Noroozi and Karami, 2022). Beckmann, et al. (2017) called out the "insufficient differentiation of complexity and difficulty" as a key factor that hampers theory building.

In general, academic standards do not organize or prioritize learning outcomes based on qualities of being 'difficult' or 'hard.' However, the word "difficulty" is used in multiple ways in the context of education including, for example, to describe processes of learning or problem solving. Decades of research have focused on "why some material is difficult to learn" and why some problems are harder than others – often using the lens of cognitive load theory (Sweller and Chandler, 1994; Kotovsky, et al., 1985; de Jong, 2010) but also through other lenses. Particularly within science education, conceptual change theory (introduced by Posner, et al. 1982) is grounded in the observation that some science ideas are "systematically extremely difficult" for students to learn (diSessa, 2014).

The term **perceived difficulty** is often used to represent judgmental determinations of how hard a task will be for a particular group of students. Because a wide range of factors pose obstacles that can interfere with students' opportunity to learn as well as to demonstrate what they know, efforts are made to limit the effect of these factors in content development processes. For example, item-writing guidelines emphasize how to minimize factors that could "unnecessarily" (TIMMS/PIRLS, 2019) or "artificially" (de Jong, 2010) increase the difficulty of an item or task. Content development guidelines also often invoke aspects of Universal Design for Learning (UDL), a framework designed to help instructional designers "reduce construct-irrelevant barriers" and promote access to the intended learning goals (CAST, 2018). Aside from obstacles to learning, task difficulty is often considered to relate very broadly to the effort required: the more effort required (irrespective of complexity), the more difficult the task.

The equity-focused goals of a standards-based education system require complexity to be a characteristic that is inherent to an expectation or task and communicated through the language of the expectation or task. Structuring conversations about cognitive complexity through the lens of DOK leads to rich conversations that enhance understanding of academic expectations and tasks. In contrast to complexity, the degree of difficulty is partly a characteristic of an expectation or task but also depends on the individual or the population. These individual or population factors include English learner status and opportunity to learn. Separating perceptions of difficulty from complexity is helpful, for example, to ensure access to the complex cognitive demands of the standards is maintained, including when linguistic supports are needed (Lee, 2018; Lee, 2019).

Understanding the factors that influence the complexity of an expectation or task, and differentiating complexity from difficulty, can help both content developers and teachers to work with greater intentionality toward shared goals of student learning within a coherent education system. Use of a common language adds clarity to the interpretation of standards, helps focus the development of learning and assessment tasks, and allows efficient and effective communication about shared goals. Teachers and content developers can work more purposefully toward goals of coherence by using DOK to guide development, evaluation, and revision of materials. Evaluating components of educational materials through the DOK lens can also help to ensure any unnecessary 'difficulty obstacles' are identified and minimized to better promote student learning.

When concepts of difficulty and complexity are not differentiated, tasks and items that are tricky or very difficult tend to be misclassified as complex, and false positive decisions may be made about content coherence or alignment between learning or assessment tasks and academic standards. Thus, disentangling perceptions of difficulty from complexity is commonly reported by teachers and content developers to be an "aha moment" (WCEPS, 2022). As an evaluative tool, DOK provides a common language that allows for actionable decisions about instructional design and content development that promote coherence.

Item Difficulty is Empirically Determined

Item difficulty is a psychometric term and is empirically determined based on student responses. Interpretation of scores on educational assessments typically relies on item difficulty, which is more widely studied and easier to operationalize empirically than are conceptualizations of rigor, complexity, or sophistication. Item difficulty ranges from very easy items to very hard items. Classical test theory defines item difficulty as the proportion of students who get an item correct. This is typically represented by *p*-value. A Rasch model defines item difficulty as the position on an ability scale (θ) at which a student has a 50% probability of answering an item correctly. This is known as the *b* parameter.

The complexity of a task does not always agree with item difficulty. This distinction is observed through the lens of multiple frameworks used to sort items by complexity (Valencia, et al., 2014, Schneider, et al, 2013; Christopherson, S., 2023). Such findings are consistent with the conceptual underpinnings of the DOK framework: while complex tasks are generally difficult, difficult tasks are not necessarily complex.

For example, items that require recall of the correct spelling of English words (e.g., tough, though, through) or correct use of punctuation may be difficult based on student responses even though the tasks are not complex. Similarly, it is possible for a complex item to have a relatively low item difficulty, for example, due to practice and experience as well as due to multiple inroads to solutions. The lack of full agreement between complexity and difficulty – from both conceptual and empirical angles – suggests that these two constructs are not the same. In fact, the two constructs can provide useful distinctions when thinking about assessment tasks and content standards. For example, all students must be afforded the opportunity to demonstrate proficiency as relates to the complexity of the assessment targets even if they do not attain a score that indicates proficient performance.

Because item difficulty depends on factors outside of the task itself, it may not be feasible to reliably perceive difficulty in ways fully consistent with empirical determinations. Even observable differences may be attributed to “subtle nuances” (Schneider, et al., 2013). Some studies have documented limited success in the estimation of empirical item difficulty based on perceived difficulty. However, additional study may be helpful to support item writers (van de Watering and van der Rijt, 2006; Subedi, et al., 2023). Lessons learned from empirical studies contribute to the ‘best practices’ and general understandings that have informed common components of item writing guidelines.

Summary: Toward Coherence

Content alignment refers to the relationship between a referent (e.g. academic standards) and a comparand (e.g. an assessment) (Fulmer, 2018; Webb, 1997). *Validity* refers to the appropriateness of inferences or plausibility of claims based on test scores. Evidence of content alignment critically supports a validity argument. Consistency of cognitive engagement between standards and assessments is a key piece of validity evidence based on test content, which is “at the heart of” content alignment (AERA, APA, & NCME, 2014). In turn, systemic coherence, or alignment, is the broader concept “at the heart of” the academic standards movement (Polikoff, 2020).

Differentiating difficulty and complexity helps promote coherence in a standards-based system. Academic standards are narrative documents. Interpreting and operationalizing the expectations within these academic standards requires a *content analysis* of the language of the standards. While academic standards are multifaceted, a core purpose is to specify expectations for particular types of cognitive engagement with academic concepts. Analyzing the language of a learning expectation, such as an academic standard, allows for an inference about the complexity of engagement required to successfully meet the expectation. Then, an analysis of the language of corresponding questions, prompts, and tasks can help determine the extent to which they provide opportunities to engage with the disciplinary ideas and concepts at the intended level(s) of complexity, and adjustments can be made as needed.

While all students are to be provided access to the full complexity of cognitive engagement as expressed in the standards, some aspects of these expectations may be more difficult for some individuals than for others. Expectations themselves vary in difficulty. Attention to complexity as distinct from difficulty can help ensure all students, across the full range of proficiency, are provided access to learning and assessment opportunities consistent with the complexity of engagement specified in the academic

standards for all students. These distinctions are valuable for purposes of assessment development and evaluation as well as to help focus classroom goals, support struggling as well as advanced learners, and, overall, to understand the different types of challenges students face as learning unfolds.
How we envision academic expectations, learning, and performance as relates to difficulty and complexity influences how we measure progress and achievement. Although academic standards emphasize complexity of cognitive engagement, practical implementation of assessment relies on metrics of item difficulty. Item difficulty and a student's achievement level are highly correlated in part because one is used to compute the other. While item difficulty surely captures, to some extent, some of what is intended for measurement, difficulty is not the target construct; it is used as a proxy. The question is whether item difficulty, and the use of corresponding metrics in the context of an overall test design, sufficiently represents the construct that is intended for measurement.

Collaborative input from both content and psychometric perspectives may help the field attain shared goals related to educational assessment. Attention to complexity along with and as distinct from difficulty can help test developers and test users strike a balance between providing clear, precise results and communicating the messier reality of learning and achievement as it occurs in the wild. Through the lens of DOK, difficulty and complexity are seen as related-but-distinct attributes of an expectation or task. Complex tasks are generally difficult because they tend to involve significant effort due to the characteristics of the tasks. However, difficult tasks are not necessarily complex. Use of the DOK lens for content analysis helps promote intentionality in practice—for example, to check that a task that is intended to be complex, consistent with the nature of the assessment target, is not accidentally just difficult or that a task that is very difficult is necessarily difficult, due to the nature of the assessment target.

How we envision academic expectations, learning, and performance as relates to difficulty and complexity also influences how teachers interpret and use the results of assessments. This distinction can help teachers to better understand the sources of student struggles, make instructional decisions appropriate for different contexts, as well as focus and prioritize lesson time. For example, it has been observed that instructional modifications for low-scoring students on interim assessments tend to be remedial. In contrast, evidence suggests that "A challenging curriculum is more effective in improving students' learning than a low-level remedial curriculum" and that learning is better achieved by "placing low achievers in advanced programmes rather than lowering the expectations" (OECD, 2012). Perspectives on difficulty and complexity influence decisions that can result in students being tracked in ways that limit their opportunity to learn. Through both design and messaging, curriculum and assessment systems may help disrupt these persistent patterns to help ensure that all students, including low-scoring students, are provided access to the full complexity of the grade-level expectations within academic standards.

Equitable opportunity to learn requires consistent interpretation of the language of the standards.
Use of a practical tool such as DOK can help both test developers and teachers differentiate difficulty and complexity to support content development as well as their interpretation and use of assessment results, connecting these attributes to a model of learning with greater system coherence.

Comments or questions? Get in touch: sara@wceps.org.

References

Agarwal, P. K. (2019) Retrieval Practice & Bloom's Taxonomy: Do Students Need Fact Knowledge Before Higher Order Learning? *Journal of Educational Psychology* 111:2 189-209

Alonzo, A. and Steedle, J. (2009). Developing and Assessing a Force and Motion Learning Progression. *Science Education*. 93. 389 - 421.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). *Standards for educational and psychological testing*.

Beckmann, J.F., Birney, D.P., and Goode, N. (2017) Beyond Psychometrics: The Difference between Difficult Problem Solving and Complex Problem Solving. *Frontiers in Psychology*. 8:1739

CAST (2018). Universal Design for Learning Guidelines version 2.2. Retrieved from http://udlguidelines.cast.org

Christopherson, S. (2019). Comparative alignment analysis of four interim assessment programs with Oklahoma State ELA, mathematics, and science academic standards. University of Wisconsin–Madison Wisconsin Center for Education Research. Retrieved from https://sde.ok.gov/sites/default/files/FINAL%20REPORT_OK%20Comparative_07102019.pdf

Christopherson, S. (2023). Differentiating Difficulty from Complexity to Support Intended Uses of Learning Progressions. Paper presented at the 2023 Annual meeting of the National Council on Measurement in Education

Common Core Standards Writing Team. (2022). *Progressions for the Common Core State Standards for Mathematics (February 28, 2023).* Tucson, AZ: Institute for Mathematics and Education Retrieved from https://mathematicalmusings.org/wp-content/uploads/2023/02/Progressions.pdf

Darling-Hammond, L., Flook, L., Cook-Harvey, C., Barron, B., & Osher, D. (2019). Implications for educational practice of the science of learning and development. *Applied Developmental Science*, *24*(2), 97–140.

de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science, 38*(2), 105–134.

diSessa, A. (2002) Why "Conceptual Ecology" is a good idea. Chapter within: M. Limón & L. Mason (Eds.), Reconsidering Conceptual Change. Issues in Theory and Practice, 29-60. Kluwer Academic Publishers. Printed in the Netherlands.

disessa, A. (2014). A History of Conceptual Change Research: Threads and Fault Lines. 10.1017/CBO9781139519526.007.

Fulmer, G.W., Tanas, J., and Weiss, K.A. (2018) The challenges of alignment for the Next Generation Science Standards. *Journal of Research in Science Teaching*, 55:7.

Gotwals, A.W., & Songer, N.B. (2013). Validity Evidence for Learning Progression-Based Assessment Items That Fuse Core Disciplinary Ideas and Science Practices. *Journal of Research in Science Teaching, 50*, 597-626.

Hammer, D. & Sikorski, T. (2015). Implications of Complexity for Research on Learning Progressions. Science Education. 99. 10.1002/sce.21165.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology, 17*(2), 248–294.

Lee, O. (2018) English Language Proficiency Standards Aligned With Content Standards. *Educational Researcher*, 47(5), 317-327.

Lee, O. (2019). Aligning English Language Proficiency Standards With Content Standards: Shared Opportunity and Responsibility Across English Learner Education and Content Areas. *Educational Researcher*, 48(8), 534-542.

Mullis, I.V.S., Martin, M.O., Cotter, K.E., and Centurino, V.A.S. (2019) TIMSS 2019 Item Writing Guidelines. TIMSS and PIRLS International Study Center. Lynch School of Education, Boston College.

National Research Council (2001) Knowing What Students Know: The Science and Design of Educational Assessment. Washington, DC: The National Academies Press. https://doi.org/10.17226/10019.

National Academies of Sciences, Engineering, and Medicine (2007) Taking Science to School: Learning and Teaching Science in Grades K-8. Washington, DC: The National Academies Press.

Noroozi S, Karami H. (2022) A scrutiny of the relationship between cognitive load and difficulty estimates of language test items. Lang Test Asia. 12(1):13

Northern, A.M. (2016) Does Common Core Math expect memorization? A candid conversation with Jason Zimba. Thomas B. Fordham Institute. Retrieved from https://fordhaminstitute.org/national/commentary/does-common-core-math-expect-memorization-candid-conversation-jason-zimba

OECD, 2012 - Equity and Quality in Education: Supporting Disadvantaged Students and Schools

Polikoff, M. (2020) The Present and Future of Alignment. Educational Measurement: Issues and Practice 39:2, 18-20

Porter, A. (1989). A Curriculum out of Balance: The Case of Elementary School Mathematics. *Educational Researcher*, *18*(5), 9-15.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of scientific conception: Toward a theory of conceptual change. Science Education, 66(2), 211–227.

Sato, E., Lagunoff, R., and Worth, P. (2011) SMARTER Balanced Assessment Consortium Common Core State Standards Analysis: Eligible Content for the Summative Assessment. WestEd. Retrieved from https://files.eric.ed.gov/fulltext/ED536959.pdf

Schneider, C.M., Huff, K.L., Egan, K.L, Gaines, M.L, and Ferrara, S. (2013). Relationships Among Item Cognitive Complexity, Contextual Demands, and Item Difficulty: Implications for Achievement-Level Descriptors. *Educational Assessment* 18:99-121.

Schnotz, W. and Kürschner, C. (2007) A Reconsideration of Cognitive Load Theory. *Educational Psychology Review* 19:469-508

Shepard, L., Daro, P., and Stancavage, F.B. (2013) The Relevance of Learning Progressions for NAEP. NAEP Validity Studies Panel. Retrieved from https://files.eric.ed.gov/fulltext/ED545240.pdf

Smith, M.S., and Stein, M.K. (1998) "Selecting and Creating Mathematical Tasks: From Research to Practice." *Mathematics Teaching in the Middle School* 3:344–50.

Steedle, J. and Shavelson, R. (2009). Supporting valid interpretations of learning progression level diagnoses. *Journal of Research in Science Teaching*. 46. 699 - 715. 10.1002/tea.20308.

Stein, M.K., Crowley, K., and Resnick, L (2016) Chapter 10: Education Policy and the Learning Sciences. The Case for a New Alliance. In Evans MA, Packer MJ, Sawyer RK, eds. *Reflections on the Learning* Sciences. Cambridge University Press; p. 219

Subedi, D., Wang, C., and Shin, D. (2023) Item Pool Development and Maintenance. Paper presented at the annual meeting of the National Council on Measurement in Education.

Sweller, J., and Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, 12, 185–233.

United States Department of Education (2018) A State's Guide to the U.S. Department of Education's Assessment Peer Review Process. Office of Elementary and Secondary Education. Washington, D.C.

Valencia, S., Wixson, K. K., and Pearson, P. D. (2014). Putting text complexity in context: Refocusing on comprehension of complex text. *Elementary School Journal, 115*(2), 270-289.

van de Watering, G. A., and van der Rijt, J. (2006). Teachers' and students' perceptions of assessments: a review and a study into the ability and accuracy of estimating the difficulty levels of assessment items. *Educational Research Review*, 2(1), 133-147.

Webb N.L. (1997) Determining Alignment of Expectations and Assessments in Mathematics and Science Education Wisconsin Center for Education Research UW-Madison

Webb N.L. (1997) Research Monograph No. 6 Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education. Wisconsin Center for Education Research UW-Madison

Webb N.L. (1997) Research Monograph No 8 Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education Council of Chief State School Officers Washington, DC

Webb N.L. (1999) Research Monograph No. 18 Alignment of Science and Mathematics Standards and Assessment in Four States, while working on the National Institute for Science Education University of Wisconsin-Madison Council of Chief State School Officers Washington, DC

Webb, N. L. (2003). Alignment study in language arts, mathematics, science, and social studies of state standards and assessments in four states. Washington, D. C.: Council of Chief State School officers.