Differentiating Difficulty from Complexity to Promote Intended Uses of Learning Progressions

Sara C. Christopherson
Wisconsin Center For Education Products and Services (WCEPS)

Author note:
The information presented in this paper reflects the Depth of Knowledge (DOK) Framework and resulting tool as it was conceptualized by Dr. Norman Webb during his time at the Wisconsin Center for Education Research with input from content and education expert committees and funded in part by the National Institute for Science Education (NISE). Some ideas about DOK have been clarified and refined over time. The author of this paper is Director of WebbAlign, a program of the non-profit Wisconsin Center for Education Products and Services (WCEPS), affiliated with the University of Wisconsin-Madison. WCEPS' WebbAlign program serves as home and steward to Webb's DOK.

Paper presented at the 2023 Annual Meeting of the National Council on Measurement in Education

Address correspondence to:
Sara C. Christopherson
sara@wceps.org

Abstract

Academic standards include rigorous expectations for all grades across the K-12 grade span, represented as complex cognitive engagement with disciplinary ideas. The standards also emphasize concomitant K-12 learning progressions, represented as the development of increased sophistication of thinking across the grades. 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. Sorting out how these entangled ideas (of rigor, complexity, sophistication, difficulty, etc.) converge and diverge may help the field attain shared goals related to assessment development as well as interpretation and use of assessment scores to provide instructionally useful guidance as relates to learning progressions. How we envision learning progressions influences how we measure them and how teachers interpret and use the results of assessments. For example, to ensure results of assessments based on learning progressions do not result in "gate-keeping" (i.e., holding students back from learning opportunities due to a perception that they are "not ready"), actionable suggestions can reinforce the idea that all students be provided access to appropriately complex learning opportunities, with support as needed to address difficult aspects of expectations.

Keywords: Assessment, Coherence, Complexity, Difficulty, DOK, Instruction, Learning model, Learning progression, Sophistication of thinking, Validity

A strong bridge between the knowledge, ideas, and "conventional wisdoms" of cognitive science and of measurement science is critical from the perspectives of validity as well as utility. Bringing together these separate-but-inseparable disciplines was a core focus of the seminal publication Knowing What Students Know (NRC, 2001), which noted "exciting possibilities" and called for assessments that could be used by stakeholders to determine the progress of student learning. Assessments that provide insights into learning progressions, along with actionable feedback, could help move all students toward attaining the shared goals detailed within learning expectations (NRC, 2001, Shepard, 2013). These assessments could help to address the "implicit tension" and "dilemma" teachers face as they help students work toward grade-level learning expectations at the same time as they attend to the specific learning needs of individuals (OECD, 2019). This paper explores key considerations for systemic coherence, with a focus on difficulty and complexity as relates to the types of cognitive engagement within the underlying learning models, and identifies areas that warrant further attention.

How we conceptualize learning progressions influences how we measure them and how test results are interpreted and used in the classroom (NRC, 2001). To promote the instructional utility of assessments based on learning progressions, the underlying model of learning for the assessment needs to have some congruence with the underlying model of learning for classroom instruction, anchored in academic standards which themselves were purposefully developed to attend to vertical articulation of expectations, representing conceptual if not empirical learning progressions (CCSS, 2010, NGSS, 2013, NCSS, 2013; Shepard, et al., 2013). Divergent ideas about learning progressions have developed from different vantage points and pose a challenge for coherent practical implementation (Shepard, et al., 2013). In general, however, learning progressions are commonly described as involving the development of more "sophisticated" thinking (e.g., Smith, et al., 2006, Duschl, et al., 2007, NRC, 2007, Deane, et al., 2012). When translated into curriculum and assessment, learning is often represented as a progression along a low-to-high difficulty and/or low-to-high complexity pathway, sometimes without differentiating between the two concepts. The implications of these interpretations on students' opportunity to learn may be significant. Consider, for example, how decisions are made about who has access to advanced coursework or who is ready to "move on." Evidence suggests positive outcomes for students provided access to "higher level" opportunities, for example, to an algebra course before high school (e.g., Smith, 1996). Similarly, if students are determined to be "low-performing", providing "lower level" coursework has been observed to negatively influence academic success (e.g., Kifer, 1993). How decisionmakers conceptualize ideas of rigor and how it relates to learning progressions affect students' opportunity to learn.

A key challenge to interpreting different representations of learning progressions is that the terms 'difficulty' and 'complexity' are not necessarily defined and are often used interchangeably, (e.g. NRC, 2001, Noroozi and Karami, 2022). Given that difficulty and complexity, as defined in this paper, are both conceptually and empirically different, their common interuse suggests that further consideration of their differences could be helpful for moving forward work on learning progressions in general as well as helpful for teachers in their efforts to use test results to inform instructional decisions. For purposes of this paper, the definitions provided within are used even when interpreting descriptions that may use divergent conceptualizations.

While current academic standards do not provide guidance about within-grade learning progressions, successful implementation of the standards requires consideration of within-grade as well as across-grade learning. Beyond the K-12 sequence of the standards, more detailed progressions are needed for classroom contexts (Shepard, et al., 2013). For assessments based on more detailed learning progressions to be useful to inform classroom instruction, there needs to be some coherence between the learning models that underlie each component of the system. Hence, the question proposed for consideration here is not necessarily whether we are measuring learning progression "correctly" but how the way we are measuring learning progression relates to the way learning progression is conceptualized outside of assessment, including within the K-12 standards—and what we can do to promote coherence, maximizing the utility of learning progression based assessments as well as upholding the ethical oath of "do no harm."

Because of the centrality of "rigor" in the learning models that undergird current standards, system coherence requires examining the way(s) that our academic standards represent different types of domain-specific cognitive engagement. This qualitative attribute is variously referred to as cognitive demand, depth, cognitive complexity, or just 'complexity' (Polikoff, Porter, & Smithson, 2011, CCSSO, 2014;). In addition to learning expectations related to cognitive engagement with academic content, many other categories of valued goals guide education, for example, related to social and emotional learning, attitude, and cultural competency. Current sets of academic standards tend to separate out goals related to cognitive engagement with academic content from goals related to affect, etc., although there are many compelling arguments to integrate them. The discussion here is limited to the realm of K-12 learning expectations primarily related to cognitive engagement with academic content.

The development of rigorous academic standards

To make sense of the ideas of rigor, complexity, and learning progressions as relates to today's academic standards, it is helpful to retrace how we got here: A core purpose of standards-based education reform related to ensuring 'high' expectations were set for all students within a state. By the end of the 1990s, nearly all states had shifted to use of K-12 academic standards for core content areas as part of systemic reform strategies (NCES, 2003). As the national conversation continued about standards-based reform, there were concerns that academic expectations varied by state, and that students who graduated high school were not necessarily adequately prepared for post-secondary education or the workplace. Grounded in research about student learning and outcomes, the Partnership for 21st Century Skills (2009) emphasized the importance of "deep understanding rather than shallow skills" and for "thinking skills" such as problem-solving, argumentation, evaluating arguments, critical thinking, and systems thinking as well as content knowledge. Overall rationales for the need for these shifts toward more rigorous expectations were anchored in relatively poor achievement for US students compared with other nations on international assessments, the identification of excessive academic remediation necessary for students entering colleges, and the identification of a lack of (and corresponding need for) workforce preparedness for US graduates in the new global economy (e.g. Schmidt, et al, 1997, CCSS, 2010). Research indicated more positive learning outcomes for students who were provided the opportunity to engage in productive struggle through complex academic work and problem solving compared with students in classrooms that used more traditional and lower complexity tasks (e.g. Kiefer, 1993, Smith, 1996, Stein and Lane, 1996; Schmidt, et al, 1997, Roth, et al., 2006, NRC, 2012). In response to concerns, states collaborated to develop college and career readiness (CCR) standards, emphasizing the critical role that rigorous expectations play in post-secondary success in a global economy and information landscape.

Each resulting new CCR framework or new set of CCR standards differentiated itself from previous ones in relation to the "rigor" of the expectations and connected this attribute to college and career readiness. The Common Core State Standards (CCSS, 2010) purposefully incorporated 21st century skills into the language of the expectations. The CCSS development criteria defined "rigorous standards" as those that "include high-level cognitive demands by asking students to demonstrate deep conceptual understanding through the application of content knowledge and skills to new situations" and further explained these types of high-level cognitive demands involved, for example, "reasoning, justification, synthesis, analysis, and problem-solving." NGSS' Appendix C (2013) also details the "importance of rigorous content for college and career readiness in science," referencing rigor 22 times over 19 pages. Similarly, the C3 framework (2013) states a core objective to "enhance the rigor of the social studies disciplines." Most recent iterations of academic standards incorporate a situative perspective on learning in which the expectations for complex cognitive engagement emphasize participation in aspects of epistemic practices and authentic ways of domain-specific knowing. These shifts were intended to promote "meaningful" learning or "deeper" learning, defined as learning grounded in conceptual understanding and organized such that the learning can be readily accessed and transferred to new situations (Mayer, 2010 NRC, 2013). In general, rigorous expectations are described as including particular types of cognitive engagement with disciplinary content—for example, drawing on underlying conceptual understanding, applying knowledge to new contexts, interweaving ideas, solving problems, reasoning and justifying with evidence, etc.

Unlike earlier iterations of standards, which were criticized for being "horizontal," i.e. focused on the within-grade expectations rather than thinking about learning across grades, the new CCR standards were developed with vertical articulation of expectations, representing K-12 learning sequences (CCSS, 2010, NGSS, 2013, NCSS, 2013; Shepard, et al., 2013). For purposes of this paper, the conceptualization of learning sequences as represented in the CCSS, NGSS, and C3 standards is not assumed to be "right" nor challenged as "wrong" but rather is taken as a "given" component of the system. Consistent with their intent, today's academic standards are treated within this paper as viable models of learning that describe how domain-specific knowledge is developed, structured, and represented. The key point is that today's academic standards emphasize conceptual understanding and other types of complex engagement with disciplinary content, practices, and habits of mind *within all grades* along with increased sophistication of both content and thinking *across the K-12 grade span.*

Why the DOK framework was developed

The need for reliable and useful interpretation of the complexity of the cognitive demands of academic standards was recognized early in the standards reform movement: for standards to have their intended effect of systemic change, alignment was necessary. In other words, all of the different pieces of our education system needed to serve in conjunction with one another to guide the system toward students learning what is expected (Webb, 1997). One of many aspects of systemic alignment is the need for consistency between the complexity of cognitive engagement specified by the standards and the extent to which that complexity plays out in the corresponding curriculum and assessments. This requirement is existential for standards-based reform given the centrality of "rigor" in the overall learning model. Despite broad agreement on overall goals for deeper learning and high expectations, there was an acute need to specify, describe, and define the types of cognitive demands that were detailed within the standards to promote reliable and productive efforts to compare, evaluate, and operationalize the complex cognitive demands within standards.

Prior to the nationwide shift to standard-based instruction, many different classification systems existed to describe the cognitive demands of educational objectives. Although widely known, Bloom's taxonomy (1956) was rejected as a tool for evaluation of cognitive complexity to promote alignment, in large part because the underlying learning model of the taxonomy (e.g., grounded in a behaviorist and hierarchical perspective on learning and knowing) was inconsistent with the generally cognitivist and constructivist perspectives of the times[1]. Additionally, Bloom's taxonomy came from an educational psychology perspective. In contrast, interpreting and operationalizing the expectations within academic standards—a narrative document—required a system that allowed for a *content analysis* of the language of the standards along with a content analysis of curriculum and assessment prompts, tasks, questions, etc. Analyzing the language of a learning expectation 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 ideas and concepts at the intended level(s) of complexity, and adjustments can be made as needed. Thus, to operationalize the goals related to deeper learning, in the context of a coherent or aligned system, a practical tool was needed. This tool needed to be tailored to the particular purpose of content analysis, include the appropriate level of precision, resonate with contemporary learning theory, be useful to practitioners, be possible to use reliably, and be grounded in an underlying conceptualization of complexity that was consistent with the standards.

Through a content analysis of the types of expectations within academic standards across a range of states and grounded in a cognitivist perspective, a new framework and associated tool was developed with input from committees of content experts (Webb, 1997, Webb, 2007, and later revised, again with committee input: Webb and Christopherson, 2014, Webb and

---

[1] Anderson, Krathwohl, and others revised the original Bloom's taxonomy to address several identified issues, including the mismatch of the underlying learning model with the cognitivist perspective on learning. The revised version conceptualizes separate Knowledge and Cognitive Process Dimensions (Anderson, et al, 2001).

Christopherson, 2018). Commonly known as Webb's Depth of Knowledge (DOK), this framework conceptualizes cognitive complexity as related to the degree of processing of concepts and skills required to meet an expectation or complete a task, dependent on the intersection of content with thinking process or practice. In other words, through the lens of DOK, it is not possible to categorize an isolated content topic by complexity (e.g. "fractions") nor is it possible to categorize an isolated thinking process or practice by complexity. For example, the complexity of a process such as "critiquing the reasoning of others" depends on the nature, extent, and context of the reasoning.

The DOK framework was developed by organizing academic expectations by subject area and by 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. Analyzing curriculum and assessment components through the DOK framework promotes coherence by providing a common language to evaluate and communicate about the types of complexity of cognitive engagement within the corresponding learning targets. (Now, as before, many additional classification frameworks continue to be used, many of which largely parallel the DOK categories. For example, the 2009 NAEP reading framework defined three types of "cognitive targets" that generally correspond with the DOK 1, 2, and 3 definitions for reading. Different frameworks target different units of analysis, often focused on one specific piece of the system. For example, some frameworks are focused specifically on task analysis, while others are focused on evaluation of student responses.)

Importantly, the DOK framework asserts that the types of thinking and what constitutes complexity varies by subject area. In other words, while we can sort the types of expectations into similar general categories, what each category "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. (Here, only a very brief summary is provided of some of the general conceptual underpinnings of the DOK framework. A thorough description of the DOK Categories of Engagement for each subject area is beyond the scope of this paper but updated definitions specific to science, social studies, ELA, and mathematics, are available here. DOK can be used consistently by trained reviewers, evidenced by statistical measurements of reviewer agreement over 20+ years of use).

Difficulty vs complexity through the DOK lens

Through the lens of DOK, difficulty and complexity are related-but-distinct attributes of an expectation or task (**Figure 1**). In general, complexity relates to the degree of mental processing required while difficulty can very broadly be considered to relate to effort. The two concepts are not orthogonal, but instead overlap: factors that influence complexity also influence difficulty—due, for example, to the resulting impact on effort and memory burden. While complex tasks are generally difficult because they involve abstraction, generalization, intricacies, and interweaving of interconnected parts, and a high degree of mental processing, difficult tasks are not necessarily complex. For example, tasks such as memorizing the correct spelling of English words (e.g., tough, though, through) or the specific chemical reactions involved in glucose metabolism are difficult but not complex. 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.

Through the DOK lens, "text complexity" is a concept distinct from the complexity of engagement *with* a text. This is another key distinction as the ability to read increasingly complex text is an important component of the conceptualization of learning progression as represented in the standards. CCSS Appendix A highlights research findings that strongly link a high school student's ability to read and make use of complex texts with measurements of post-secondary success. Text complexity depends on both qualitative and quantitative factors such as type of vocabulary, sentence length, frequency of unfamiliar words, the extent to which the text is literal vs figurative, and other aspects. CCSS additionally include factors related to the reader, such as motivation. While the complexity of a text (or, somewhat parallel, of another task context, such as a science phenomenon) may influence the types of student interactions that are viable, it is possible for tasks to require very simple engagement with complex text (e.g. *Use the Table of Contents to find the first page of Chapter 2*; *Name the main character*). Conversely, it is possible for tasks to require very complex engagement with simple texts. For example, an early elementary student could be asked what they think a character from one picture book would think about an event that occurred in real life, referring to the text and graphics for rationale. Differentiating text complexity from complexity of engagement *with* a text can help teachers and content developers ensure questions and prompts are structured to promote meaningful and complex engagement *with* texts at all levels of text complexity. This general concept can be extrapolated to task context for other subject areas as well.

Difficulty that is inherent to an expectation or task can be considered necessary. Tasks may also include difficulty that is unnecessary or irrelevant to the complexity of the task. Sweller et al., 1998 made a somewhat similar distinction between intrinsic cognitive load and extraneous cognitive load when considering the impact of working memory demands on a task, but did not differentiate difficulty from complexity nor relate the task demands to a specific learning or assessment target. In contrast, through the DOK lens, whether difficulty is inherent or extraneous depends on the purpose of a task. Evaluating components of a learning progression to identify necessary difficulty can help teachers to focus and prioritize lesson time and can help to ensure any unnecessary difficulty obstacles are identified and removed to better promote student learning (e.g., Robar and Bryan, 2021).

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

Key factors that influence complexity

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

Key factors that influence difficulty

Low difficulty	**High difficulty**
Low chance for simple errors	High chance for simple errors
Limited effort/perseverance required	Extensive effort/perseverance required
Clear instructions/Language	Unclear instructions/Language
Simple English/lower Lexile text	Dense or /higher Lexile text
Fewer repetitive steps	More repetitive steps
Limited memory burden/recall required	High memory burden/recall required

Factors that influence complexity and difficulty, summarized above, relate to the nature of an expectation, question, prompt, or task, etc. No value judgement is ascribed to the degree of difficulty nor complexity. In other words, it is not “good” or “bad” for an expectation to be low or high complexity nor is it “good” or “bad” for a task to be low or high difficulty. Rather, the purpose of using the DOK lens for content analysis is to 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 high difficulty is necessarily difficult, due to the nature of the assessment target.

The distinction between difficulty and complexity can be made empirically as well as conceptually. For example, **Table 1** maps a set of multiple-choice math and reading/language items by complexity and difficulty. Item difficulty data were provided by a nationally recognized content developer and organized into three categories (low difficulty, medium difficulty, or high difficulty) based on an IRT model and a large population of test takers. DOK was determined by averaging the independent item-level codings of a panel of six content experts (for each of the two subject areas), all of whom had extensive training and experience with DOK and content analysis. These items were considered high quality and did not contain any content-related or technical issues.

As shown in **Table 1**, the largest group of lowest complexity (DOK 1) items were medium difficulty but most DOK 1 items were nearly equally divided between the lowest or highest difficulty categories. Importantly, low complexity did not predict low difficulty (or vice versa).

While ~25% of the lowest complexity items (DOK 1) were also lowest difficulty, the vast majority (~75%) were medium or higher difficulty.

In the item sample used in the example below, most items were DOK 2, and only a small portion of items were DOK 3. If this distribution corresponds to the distribution of complexity within the assessment targets, then the low count of DOK 3 items may be appropriate. If the distribution of complexity within the assessment targets is very different—for example, if most expectations are DOK 3—then the low proportion of DOK 3 items would not be consistent with the complexity of the expectations. Similarly, item difficulty depends on overall intent: here, the distribution may be appropriate if the goal is for items to have a range of difficulty, with most items somewhere in the middle.

Table 1. Number of high school math and reading/language items by complexity and difficulty

		Difficulty		
Complexity		Lowest difficulty	Medium difficulty	Highest difficulty
	DOK 1	26	42	28
	DOK 2	31	100	55
	DOK 3	1	8	7

Traditionally, extraneous item difficulty was considered to be manageable and attributable to poor item quality. However, a wide variety of factors that negatively affect student performance, and are reflected in resulting difficulty, are outside of the control of item design. These factors include language barriers, unmet accessibility needs, opportunity to learn, general familiarity with topics or contexts, item format, test interface, time of day, anxiety, student health, socioeconomic factors, gender identity, prior knowledge, and many more (e.g., , Ercikan, 1998, NRC, 2001, Clarke, et al., 2005, Scherbaum and Goldstein, 2008, Ramirez, et al., 2013, Choe, et al., 2019, Lin and Steedle, 2020). Aspects of affect, including attitudes about one's own capacity to learn, perspectives of self-efficacy, confidence, and more can also influence difficulty (e.g., Lubienski, 2013, Snipes and Tran, 2017). Even small details and contextual factors that may seem trivial, such as semantic structure or the color of an object, have been observed to affect a student's success with a task (e.g. NRC, 2001, Özdemir and Clark, 2007). Because a wide range of extraneous factors can interfere with students' opportunity to demonstrate what they know, efforts are made to limit the effect of these factors, such as through application of Universal Design for Learning (UDL) in the development process and DIF analysis in field testing and item evaluation. For items that successfully pass through extensive quality design processes, the resulting item difficulty is sometimes used as a proxy for complexity. However, as shown in **Table 1**, item difficulty does not predict complexity. Using DOK as well as other frameworks intended for differentiating item type, Schneider, et al. (2013) similarly found that item difficulty did not predict complexity.

The empirical findings shown in **Table 1** are consistent with the conceptual underpinnings of the DOK framework: While complex tasks are generally difficult, difficult tasks are not necessarily complex. As part of their exploration of the relevance of learning progressions for

NAEP, Shepard, et al. (2013) noted that decisions about ordering items by complexity and/or difficulty constitute a “critical conceptual decision.” One reason this decision is so critical is 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. Similarly, ideas about the relationships of difficulty and complexity with learning progressions can consequentially influence how teachers interpret and use test results, and therefore, what type(s) of learning opportunities students will be provided. For example, if a student is found to be on the low end of a learning progression, does that mean the student should be presented with low complexity tasks? Easy tasks? Will that student be provided access to the types of relevant, high complexity opportunities that we know motivate learning? From a qualitative perspective, when concepts of difficulty and complexity are not differentiated, we tend to see items that are tricky or very difficult misclassified as complex, and false positive decisions about content coherence or alignment between assessment tasks and academic standards. Because teachers bring their own personal epistemologies to the table, they may perceive that a student who cannot complete a particular task is therefore “not ready” for learning opportunities that are more difficult and/or complex, resulting in harmful gatekeeping. When working with teachers on DOK-related concepts, sorting out ideas of difficulty vs. complexity is often reported to be an “aha moment.” For example, in evaluation responses after participation in professional development about DOK, the most common theme within teacher comments is related to the value they found in differentiating the ideas of difficulty and complexity (WCEPS, 2022).

Implications related to the distinct-but-intertwined concepts of difficulty and complexity—and how they relate to the learning expectations within academic standards—extend beyond the assessment of learning progressions. Specific to the context of learning progressions, use of a practical tool such as DOK can help both test developers and teachers differentiate difficulty and complexity to support their interpretation and use of assessment results, connecting these attributes to a model of learning with greater system coherence. While some progression of difficulty may be inherent to learning progressions, underlying progression of difficulty is not a rule, particularly at the micro-level (e.g. Shepard, 2013, OECD, 2019, Robar and Bryan, 2021). For example, Robar and Bryan (2021) identified components of learning progressions that they termed “Trip Steps.” Similar to the conclusions of Shepard, et al, (2013), they note that these “disproportionately difficult” components of learning progressions are important puzzle pieces, necessary to support students as they work toward the goals within the standards. Understanding the factors that influence difficulty and complexity can help support both content developers and teachers to work with greater intentionality toward shared goals of student learning within a coherent education system.

Implications of low-to-high difficulty and/or low-to-high complexity conceptualization of learning:

While ideas of difficulty and complexity come into play when attempting to operationalize learning progression-based curriculum and/or assessments, the nature of learning is typically described in the literature as a trajectory of more “sophisticated” thinking (e.g., Smith, et al., 2006, NRC, 2007, Deane, et al., 2012). This increase in sophistication has been described as “the transformation of naive understanding into more complete and accurate comprehension" as well as a process of modifying the organization of knowledge which may require additions or extensions as well as "radical reconstruction" of the knowledge structures (Glaser and Baxter, 1999, NRC, 2001). Learning progressions may be thought of as a movement toward expertise and are often qualified by the grain size and amount of time involved (e.g. Duschl, et al., 2007, Heritage, 2008). Overall, learning tends to be considered a qualitative change but it is sometimes represented as including a change in amount, scope, or quantity. For example, CCSS math standards for Kindergarten expect addition and subtraction within 10 but more numbers are included for grade 1 (up to 20) and as noted by Bennett (2015), CCSS grade 6 RL.1 expects students to cite textual evidence but in grade 7 RL.1 the expectation expands to include *several* pieces of textual evidence, etc. (CCSS, 2010). It is generally agreed that learning occurs in myriad ways and is not necessarily directional at the finer grain, but that learning becomes more thorough or complete over time and reasoning and problem-solving strategies generally develop over time (e.g., NRC, 2007, Alonzo and Steedle, 2009, Steedle and Shavelson, 2009, Gotwals and Songer, 2010, Gotwals and Songer, 2012, Gotwals and Songer, 2013). Instead of as a "progression," learning has also been envisioned as a non-linear "conceptual ecology" (originally Posner, et al., 1982) 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 is (*sic*) shifting from a "static, linear learning-progression model" to a "non-linear, dynamic model." Whatever the conceptualization, Gotwals and Songer's (2013) description of the "messy middle" of a learning progression resonates widely and is often quoted. Within this middle, qualitative changes in thinking may include inaccurate ideas (e.g. Posner, et al, 1982, Furtak, 2009; Alonzo, 2017).

While broad trends toward increased difficulty and complexity are consistent with most conceptualizations of increased sophistication of thinking, some key considerations are necessary to fit that idea with a standards-based learning model. Within a standards-based model, expectations are carefully described to define a specific learning or performance expectation as relates to interactions with academic concepts and are unequal in terms of the difficulty and complexity. For example, consider the expectation to "Use an apostrophe according to the rules of standard English." Once this expectation is mastered, there is no particular further learning trajectory related to this specific goal—no use of an apostrophe in more difficult nor in more complex ways. Any "progression" of difficulty or complexity would require an adjustment to the scope or target of learning, for example: "Write explanatory text with tone appropriate for audience and punctuation according to the rules of standard English." Conversely, the progression "toward" a highly complex learning outcome may involve engagement in difficult as well as easy tasks and complex as well as simple tasks. One rationale for the use of teaching strategies such as the 5E model, as well as problem-, project-, and phenomenon-based learning is that they help disrupt the tendency to "start easy" or to "start simple." From a conceptual perspective, the observation of tasks becoming easier may be evidence of learning. For example, as a student builds fluency with reading or the use of standard math algorithms, these tasks become easier. Similarly, the novice-to-expert conceptualization of learning, including in the context of high complexity expectations, suggests that as learners progress toward expertise, they organize ideas in ways that make tasks easier, although still complex. This applies broadly: For example, consider the experience of a researcher submitting their first grant application (novice) compared with the experience of a researcher submitting their 50^{th} grant application (expert). Both are engaged in a highly complex task, involving synthesizing ideas to build an argument, iterative processes requiring metacognitive thinking, etc. However, the task is likely to be harder for the novice than for the expert because the expert has developed strategies and structures that help to complete the work. As noted previously, observations such as Robar and Bryan's (2021) Trip Steps reinforce the assertion that there is no low-to-high difficulty rule as relates to learning progression.

Perhaps more importantly, due to the implications for equitable opportunity to learn, it is important to consider the implications of applying a low-to-high complexity interpretation of learning progression—a perspective very commonly held by teachers. With the caveat that the term "complexity" is not necessarily used in the same way as it is defined in this paper, learning is often conceptualized as movement from low-to-high complexity thinking. For example, Furtak (2009) describes examples of learning progressions about science ideas that were developed with the conceptualization of ideas unfolding "in a linear fashion, progressing from simple to complex." However, if learning were to progress from simple to complex, then we would expect to see lowest complexity expectations in the academic standards for lower grades moving up to highest complexity expectations in highest grades. Instead, and for all subject areas, we see both lower and higher complexity expectations across all grades, typically with a greater proportion of higher complexity expectations in higher grades (Sato, et al. (2011), WebbAlign, n.d.). In some contexts, (e.g. standard algorithms or common equations, such as $l \times w = a$) academic standards introduce more complex conceptual work before expecting fluency and rote work with the same ideas. The underlying learning model of today's standards-based education system does not reflect an assumption that learning progresses, as a rule, from low-to-high complexity. A low-to-high complexity trajectory also conflicts with empirical observations in multiple content areas. For example, children working in street markets have been observed to do complex math in context (e.g. use ratio reasoning, apply a variety of strategies to work with large numbers) even if they were not able to solve more simple math problems out of context (e.g., Carraher, T.N., et al, 1985; Saxe, G.B, 1988). Conversely, observations that even students who successfully completed extensive science education still could not convey basic science facts (e.g. Posner, et al, 1982, Harvard-Smithsonian Center for Astrophysics, 1987) prompted extensive rethinking about assumed learning progressions within science education. A core concern with a representation of learning as progressing along a continuum of complexity is that "postponing more complex reasoning about subject matter would be antithetical to the intentions of both the CCSS and learning progressions research" (Shepard, et al., 2013). Similarly, teachers bringing underlying assumptions about a low-to-high complexity progression could misinterpret or misuse the results of assessments intended to provide information about student learning.

Because academic standards emphasize *complexity* of engagement, learning progressions are understood to reflect increased *sophistication* of thinking, and practical implementation of assessment relies on metrics of *difficulty*, attention to how increased sophistication of thinking relates to both difficulty and complexity may be fruitful for the development and assessment of learning progressions as well as for the interpretation and use of test results by teachers. Test developers must strike a balance between providing clear, precise results to teachers and communicating the much messier reality of learning as it occurs in the wild. Greater coherence within underlying learning models could enhance the capacity of assessments to provide instructionally useful guidance as relates to learning progressions in the context of rigorous academic standards. Differentiating difficulty and complexity from changes in sophistication along learning progressions is important to support the equity-focused goals of a standards-based education system. Equipped with an understanding that learning progressions are not dogmatically linear trajectories of low-to-high difficulty and/or complexity, teachers can work to ensure all students have access to engaging, relevant, and complex learning opportunities—the types of opportunities that we know promote learning—no matter where they are along a learning progression.

How we envision learning progressions influences how we measure them and how teachers interpret and use the results of assessments. Additional research on the relationship between difficulty and complexity and how assessment task difficulty and complexity relate to the inferences we make about students and learning processes is warranted.

Alonzo, A.C & Elby, A. (2019) Beyond Empirical Adequacy: Learning Progressions as Models and Their Value for Teachers, Cognition and Instruction, 37:1, 1-37, DOI: 10.1080/07370008.2018.1539735

Alonzo, A.C (2017) Tracing the Assessment Triangle for Formative Assessment: Not All Learning Progressions Are Created Equal, Measurement: Interdisciplinary Research and Perspectives, 15:3-4, 163-167, DOI: 10.1080/15366367.2017.1404369

Anderson, L.W. (Ed.), Krathwohl, D.R. (Ed.), Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J., & Wittrock, M.C. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives (Complete edition). New York: Longman.

Badrinarayan, A., Christopherson, S. and Harris, C. (2017) The Alignment Challenge. Reidy Interactive Lecture Series: Assessing Student Learning of the Next Generation Science Standards. The National Center for the Improvement of Educational Assessment

Badrinarayan, A, Christopherson, S., Gong, B., McCrae, A. (2018) Developing a Common Language to Understand Content Complexity for Alignment Studies of the NGSS. National Conference on Student Assessment

Badrinarayan, A., Christopherson, S., Davis-Becker, S., Everson, H., and Forte, E. (2019). Representing Cognitive Complexity in Test Design and Evaluation. ATP Innovations in Testing
Blank, R.K., Porter, A. & Smithson, J (2001). New Tools for Analyzing Teaching, Curriculum and Standards in Mathematics & Science. Results from the Survey of Enacted Curriculum project, Final Report. Washington, DC: CCSSO.

Baxter, G. P. and Glaser, R. (1998). Investigating the Cognitive Complexity of Science Assessments. Educational measurement, issues and practice 17(3) p. 37-45

Bloom, B., Engelhart, M., Furst, E., Hill, W., & Krathwohl, D. (Eds.) (1956). Taxonomy of Educational Objectives: The Classification of Educational Goals, Handbook I: Cognitive Domain. New York: David McKay.

Carraher, T.N, Carraher, D.W, and Schliemann, A.D (1985) Mathematics in the streets and schools. British Journal of Developmental Psychology 3:21-29. Great Britain: The British Psychological Society

Clarke, T., Ayres, P., and Sweller, J. (2005) The Impact of Sequencing and Prior Knowledge on Learning Mathematics through Spreadsheet Applications. Educational technology research and development 53(3) p. 15-24 DOI:
10.1007/BF02504794

Confrey, J. (2019). A Synthesis of research on learning trajectories/progressions in mathematics(EDU/EDPC(2018)44/ANN3; Future of Education and Skills 2030: Curriculum Analysis. OECD.

Choe, K.W., Jenifer, J.B, Rozek, C.S, Berman, M.G, and Beilock, S.L. (2019) Calculated avoidance: Math anxiety predicts math avoidance in effort-based decision-making. Science Advances 5(11). DOI: 10.1126/sciadv.aay1062

Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (Eds.). (2007). Taking science to school: Learning and teaching science in grades K–8. Washington, DC: National Academies Press.Ercikan (1998)

Ertmer, P. A., & Newby, T. J. (2013). Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. Performance Improvement Quarterly, 26(2), 43-71.

Harvard-Smithsonian Centero for Astrophysics (1987) A Private Universe Available as of March, 2023 at https://www.learner.org/series/a-private-universe/

Heritage, M. (2008) LEARNING PROGRESSIONS: SUPPORTING INSTRUCTION AND FORMATIVE ASSESSMENT. Council of Chief State School Officers, Washington, DC

Later, Freedle and Kostin (1997)

Lin, Qiao and Steele, Jeffrey (2020). Investigating Assessment Conditions Potentially Associated with Differential Item Functioning. ACT Research & Policy Technical Brief

Lubienski, S.T., Robinson, J. P., Crane, C.C., & Ganley, C. M. (2013). Girls' and Boys' Mathematics Achievement, Affect, and Experiences: Findings from ECLS-K. Journal for Research in Mathematics Education, 44(4), 634–645. https://doi.org/10.5951/jresematheduc.44.4.0634

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. 2012. A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas. Washington, DC: The National Academies Press. https://doi.org/10.17226/13165.

National Council for the Social Studies (NCSS), The College, Career, and Civic Life (C3) Framework for Social Studies State Standards: Guidance for Enhancing the Rigor of K-12 Civics, Economics, Geography, and History (Silver Spring, MD: NCSS, 2013).

National Research Council. (2012). Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century. Committee on Defining Deeper Learning and 21st Century Skills, J.W. Pellegrino and M.L. Hilton, Editors. Board on Testing and Assessment and Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

NGSS Lead States. 2013. Next Generation Science Standards: For States, By States. 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. doi: 10.1186/s40468-022-00163-8. Epub 2022 Jun 15. PMCID: PMC9197553.

OECD (2019) Future of Education and Skills 2030L: OECD Learning Compass 2030. A Series of Concept Notes. https://www.oecd.org/education/2030-project/teaching-and-learning/learning/learning-compass-2030/OECD_Learning_Compass_2030_Concept_Note_Series.pdf

Özdemir, D. and Clark, D.B. (2007) Eurasia Journal of Mathematics, Science & Technology Education, 3(4), 351-361

Plass, J., Moreno, R., and Brünken, R. (Eds.). (2010). Cognitive load theory. New York: Cambridge University Press.

Polikoff, M. S., Porter, A. C., & Smithson, J. (2011) How well aligned are state assessments of student achievement with state content standards? American Educational Research Journal, 48(4), 965–995.

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

Ramirez, G. Gunderson, E.A., Levine, S.C & Beilock, S.L. (2013) Math Anxiety, Working Memory, and Math Achievement in Early Elementary School, Journal of Cognition and Development, 14:2, 187-202, DOI: 10.1080/15248372.2012.664593

Robar, J. and Bryan, J (2021). The critital role of Trip Steps in math recovery. Renaissance Learning Retrieved from https://www.renaissance.com/2021/09/24/blog-the-critical-role-of-trip-steps-in-math-recovery/

Roth, et al (2006) Teaching Science in Five Countries: Results From the TIMSS 1999 Video Study. National Center for Education Statistics

Sato, E., Lagunoff, R., and Worth, P. (2011) SMARTER Balanced Assessment Consortium Common Core State Standards Analysis: Eligible Content for the Summative Assessment Final Report West Ed
Saxe, G. B. (1988). The Mathematics of Child Street Vendors. Child Development, 59(5), 1415–1425. https://doi.org/10.2307/1130503

Schmidt W. H. McKnight C. C. Raizen S. A. & Third International Mathematics and Science Study. (1997). A splintered vision : an investigation of u.s. science and mathematics education. Kluwer Academic.

Scherbaum, C.A. and Goldstein. H.W. (2008) Examining the Relationship Between Race-Based Differential Item Functioning and Item Difficulty Educational and Psychological Measurement 2008 68:4, 537-553

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 Acheivement-Level Descriptors. Educational Assessment 18:99-121. Taylor& Francis Group, LLC

Shepard, L., Daro, P., Stancavage, F.B (2013) The Relevance of Learning Progressions for NAEP. NAEP Validity Studies Panel; National Center for Education Statistics

Smith . J.B. (1996) Does an Extra Year Make Any Difference? The Impact of Early Access to Algebra on Long-Term Gains in Mathematics Attainment. Educational Evaluation and Policy Analysis 18(2) pp. 141-153 American Educational Research Association doi: 10.2307/1164553

Snipes, J., & Tran, L. (2017). *Growth mindset, performance avoidance, and academic behaviors in Clark County School District* (REL 2017–226). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory West. Retrieved from http://ies. ed.gov/ncee/edlabs.

Stein. M.K. and Lane, S. (1996) Instructional Tasks and the Development of Student Capacity to Think and Reason: An Analysis of the Relationship between Teaching and Learning in a Reform Mathematics Project. Educational Research and Evaluation 2(1) doi.org/10.1080/1380361960020103

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. Educational Psychology Review, 10(3), 251–296. https://doi.org/10.1023/A:1022193728205

U.S. Department of Education, National Center for Education Statistics (2003). Overview and Inventory of State Education Reforms: 1990 to 2000, NCES 2003–020, by David Hurst, Alexandra Tan, Anne Meek, and Jason Sellers. Project Officer: Edith McArthur. Washington, DC

U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), What Does the NAEP Reading Assessment Measure? Accessed March, 2022
https://nces.ed.gov/nationsreportcard/reading/whatmeasure.asp

Valencia, S. W., & Wixson, K. K. (2000). Policy-oriented research on literary standards and assessment. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), Handbook of reading research: Vol. III. Mahwah, NJ: Lawrence Erlbaum.

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

Weir, C.J. and Porter, J. (1994) The Multi-Divisible or Unitary Nature of Reading: The Language Tester Between Scylla and Charybdis. Reading in a Foreign Language 10(2)

Wixson, K. K., Fisk, M. C., Dutro, E., & McDaniel, J. (2002). The alignment of state standards and assessments in elementary reading. Ann Arbor, MI: Center for the Improvement of Early Reading Achievement.

Yi, S., Deane, P., Graf, E. A., and van Rijn, P. (2013). Using Argumentation Learning Progressions to Support Teaching and Assessments of English Language Arts CCSS CBAL EDC. ETS R&D Connections, n22