Applied Measurement in Education

ISSN: 0895-7347 (Print) 1532-4818 (Online) Journal homepage: https://www.tandfonline.com/loi/hame20

Learning progressions as tools for assessment and learning

Lorrie A. Shepard

To cite this article: Lorrie A. Shepard (2018) Learning progressions as tools for assessment and learning, Applied Measurement in Education, 31:2, 165-174, DOI: 10.1080/08957347.2017.1408628

To link to this article: https://doi.org/10.1080/08957347.2017.1408628

Accepted author version posted online: 28 Nov 2017.
Published online: 23 Feb 2018.

APPLIED MEASUREMENT IN EDUCATION
2018, VOL. 31, NO. 2, 165–174
https://doi.org/10.1080/08957347.2017.1408628

Learning progressions as tools for assessment and learning

Lorrie A. Shepard

School of Education, University of Colorado Boulder, USA

ABSTRACT
This article addresses the teaching and learning side of the learning progressions literature, calling out for measurement specialists the knowledge most needed when collaborating with subject-matter experts in the development of learning progressions. Learning progressions are one of the strongest instantiations of principles from *Knowing What Students Know*, requiring that assessments be based on an underlying model of learning. To support student learning, *quantitative* continua must also be represented *substantively*, describing in words and with examples what it looks like to improve in an area of learning. For formative purposes, in fact, *qualitative insights* are more important than *scores*. By definition, learning progressions require iterative cycles of development so as to build in horizontal coherence among curriculum, instruction, and assessment. Learning progressions are also an important resource for teacher learning. With accompanying professional development and institutional supports, they can help teachers engage their students in richer and more equitable learning experiences. Examples are cited whereby learning progressions can be used to help teachers improve their skills in setting learning goals, interpreting student ideas in relation to a progression, and responding to student ideas with specific interventions that serve to move learning forward.

The relatively new, 21st-century idea of learning progressions gained its impetus from *Knowing What Students Know* (National Research Council, 2001), a report authored by a National Research Council committee specifically constituted to bring together advances in the cognitive and measurement sciences. This effort to integrate what is known about learning with the machinery of psychometric modeling remains critical to our understanding of research on learning progressions today. The most important integrative idea is that *quantitative* continua must also be represented *substantively*, describing in "words and examples what it means to make progress or to improve in an area of learning" (National Research Council, 2001, p. 137).

Learning progressions are defined as "descriptions of successively more sophisticated ways of reasoning within a content domain" (Smith, Wiser, Anderson, & Krajcik, 2006, p. 1). They are different from either item-anchored test-score scales or curricular scope and sequence strands, because they require *both* conceptual analysis and empirical testing of hypothesized developmental pathways (Corcoran, Mosher, & Rogat, 2009). [1] (Item anchoring is an empirical method used, after

CONTACT Lorrie A. Shepard Lorrie.Shepard@Colorado.edu Campus Box 249, School of Education, University of Colorado Boulder, Boulder, Colorado 80309

[1]As many authors have recognized, definitions of learning progressions vary (Alonzo & Steedle, 2009), and not all examples using this label have satisfied the requirements that hypothesized sequences be empirically tested and linked to instruction. In this commentary, I rely on the seminal work *Knowing What Students Know* (National Research Council, 2001)—which called for cognitive models that document the means for furthering learning, not just a sequence of end-points. In the same vein, the Center on Continuous Instructional Improvement Science Panel's consensus attributes of learning progressions note that they are "crucially dependent on instructional practices" (Corcoran et al., 2009, p. 38). Smith et al. (2006) also agree with these requirements, calling their model for matter and atomic-molecular theory a *possible* or *proposed* learning progression rather than an *actual* progression because it lacks empirical testing derived by following students provided with concomitantly designed instruction.

the fact, to try to give substantive meaning to locations on a test score scale; historically, for example, items with a 65% probability of being answered correctly were used to anchor or describe the meaning of National Assessment of Educational Progress (NAEP) scores at 250, 300, 350, and so forth. Curricular sequences are devised conceptually by experts but are not tested to see if hypothesized orderings work in practice).

Measurement experts may hear in the definition of learning progressions a resonance with Glaser's (1963) arguments more than 50 years ago for "criterion-referenced measures," which he distinguished from norm-referenced measures that convey only students' relative standing on a performance continuum. According to Glaser, "criterion-referenced measures indicate the content of the behavioral repertory, and the correspondence between what an individual does and the underlying continuum of achievement" (p. 520). This century's learning progressions are more than this, however. They go beyond sequential status descriptions and also account for how and by what instructional means progress toward more sophisticated understandings is achieved.

This commentary is written for a measurement audience, yet the articles in this special issue of *Applied Measurement in Education* are focused not on measurement but on teacher learning (i.e., teacher learning about the use of learning progressions to support student learning). Thus, it helps to be reminded of the properties of learning progressions that integrate theories of learning with measurement models. Before proceeding to the topic of teacher learning, I wish to back up and first say a bit more about what measurement specialists should know about the development of learning progressions, then consider learning progressions as tools to support student learning, and lastly return to implications for teacher professional development.

Defining and developing learning progressions

Learning progressions benefit student learning by illuminating student thinking and by providing insights about instructional next steps. For these instructional claims to hold true, however, learning progressions have to be designed and then tested to verify that intended inferences are warranted. Learning progressions (or more specifically, assessment tasks intended to elicit evidence of progress) cannot be developed using traditional approaches to test construction. More sophisticated reasoning and adept knowledge use are not represented well by composite-score scales nor by heterogeneous item sets covering a domain map or test blueprint. Developmental continua require more than simplistic ordering of items by difficulty. Tests developed based on a well-conceptualized, unidimensional construct and Rasch (1960) item calibration come the closest to modeling learning progressions. Indeed, prior work by Masters, Adams, and Wilson (1990) and Wilson, Draney, and Kennedy (2001) using the Rasch model informed *Knowing What Students Know* (National Research Council, 2001) arguments in favor of developmental assessment. Rasch analyses alone, however, are often insufficient for building learning progressions because they do not give adequate attention to the learning side of things. For example, even carefully crafted construct definitions have not typically attended to the preliminary understandings and intuitions that students hold when they begin with a new topic, nor has attention been paid to particular learning experiences that help to transform these ideas.

Quibbling over the definition of learning progressions and the adequacy of traditional test construction methods is worthwhile because in the past many good ideas about testing have been corrupted by superficial understandings and a tendency to continue old habits under new labels. Glaser (1963) wrote about substantive, content-focused criterion-referenced measurement from the perspective of a cognitive theorist. But his ideas were quickly corrupted by measurement experts to mean quantitative cut scores; and the testing industry just as quickly re-sorted items from norm-referenced tests into objective-referenced subtest score reports, a practice that has continued to oversell the diagnostic properties of standardized tests. In our present-day context, Common Core State Standards (CCSS) were initially said to represent learning progressions. Yes, CCSS were designed conceptually by content experts to "progress" across grades to arrive at college and career readiness by 12th grade, but they lacked the empirical research and development work required to build real learning progressions. The premature efforts that followed—to build more challenging large-scale assessments without accompanying instructional research—have done very little to support ambitious teaching practices in schools with the greatest needs. New large-scale tests still promote breath over depth with no new insights about what instructional moves might actually bring tougher standards within reach for more students.

Elsewhere, my colleagues and I have argued that it is better to build learning progressions from the bottom up, focusing on local jurisdictions or curricular projects, where it is more likely to be possible to design for coherence among curriculum, instruction, assessment, and teacher learning (Shepard, Penuel, & Davidson, 2017). *Knowing What Students Know* authors (National Research Council, 2001) called this *horizontal coherence*, which requires that these elements be co-constructed to represent the same underlying model or theory of learning. Learning progressions are one type of cognitive model; but note that there are a number of other ways of modeling how thinking and mature practice develop in a domain, such as facets-based or knowledge-in-pieces approaches (diSessa & Minstrell, 1998; Hammer, 1996). *Knowing What Students Know* (National Research Council, 2001) also called for *vertical coherence* between classroom assessments and large-scale assessments, but this is nearly impossible to accomplish across districts with different curricula that do not share the same underlying model of learning. In addition to political and cost constraints that likely weigh against building large-scale assessments from learning progressions, the current research base also argues for developing, testing, and implementing learning progressions for smaller scale domains rather than for broad subject matter areas such as science, mathematics, or English language arts.

Funded primarily by the National Science Foundation, construct names from the existing research base help to illustrate appropriate grain sizes for learning progression development. Titles include "scientific modeling," "matter and atomic-molecular theory," "carbon cycling in socio-ecological systems," "scientific argumentation," "modern genetics," "complex reasoning about bio-diversity," "energy," "evolution," and "celestial motion." [2] Such development efforts typically span multiple grade levels, but they are "narrow" in the sense that they focus on only one or two big ideas in science or a single scientific practice. Consistent with present-day research on learning, however, these learning progressions do not treat content domains separately from scientific processes (National Research Council, 2007). Rather, for any given big idea, developers have selected specific practices—but not all possible scientific practices—to engage and deepen. To develop young children's understandings of matter and atomic-molecular theory, for example, Smith et al. (2006) engaged children in practices that included "identifying and classifying, ordering and measuring, representing data, and posing questions" (p. 28). Similarly, when the construct was a scientific practice such as scientific modeling (Schwarz et al., 2009), learning progression developers integrated increasingly complex examples of these practices within specific content units on condensation and evaporation at the elementary level, and on the physics of light and seeing and the chemistry of smell for middle school students. A look at the deeply supportive and demanding learning activities within these research projects makes it clear why it would be difficult to use performance tasks generated in these studies for broader scale summative purposes with students who have not had access to these curricular experiences.

A final point, worthy of note for a measurement audience, is that a certain amount of validation research is built into the recursive process of developing learning progressions. Starting with a conceptual analysis of intermediate steps toward envisioned learning goals, curriculum, instructional activities, and formative assessment tasks must be jointly constructed. Then assessment tasks are empirically tested, embedded within instructional activities. The focus is on student thinking as test developers might more typically examine in separate cognitive lab studies. Teachers and researchers

[2]References associated with these constructs are as follows: Berland and McNeill (2010); Catley, Lehrer, and Reiser (2005); Duncan, Rogat, and Yarden (2009); Mohan, Chen, and Anderson (2009); Neumann, Viering, Boone, and Fischer (2013); Plummer and Krajcik (2010); Schwarz et al. (2009); Smith et al. (2006); Songer, Kelcey, and Gotwals (2009).

then propose and try out subsequent instructional activities that might extend students' knowledge, skills, and reasoning. The adequacy of these follow-up activities is evaluated in turn. Assessment tasks, instructional activities, and the progression itself are all subject to revision during this process. Learning progressions do not assume that all children proceed lockstep through a specified sequence, but they are the most useful when there is a research base for identifying frequent orderings (e.g., understanding weight before density, or perceiving that sugar dissolves in water before accepting that there are invisible particles called molecules) as well as frequent conceptual challenges. Of course, the validity argument for a given learning progression does not rest solely on data collected during the development cycle. Validation should continue—as is the case with the Furtak Heredia, and Circi article in this issue—to seek evidence of learning gains using independent measures of the intended learning outcomes.

Learning progressions as tools to support student learning

In arguing for vertical coherence between the classroom level and policy-level, large-scale assessments, *Knowing What Students Know* (National Research Council, 2001) authors acknowledged that a model of learning shared between the two levels would necessarily have to be specified at very different grain sizes. This idea that classroom-level assessments would require much more finely elaborated models of cognition was consistent with recognized differences in assessment purposes and corresponding tradeoffs in assessment design. Measurement specialists are nonetheless still likely to underestimate the extent of development needed to enact learning progressions at the classroom level—unless they are willing to go beyond an image of the typical end-of-unit test and think instead about curriculum development projects.

One or two examples may also help to illustrate the extensive research and development and detailed materials needed to create learning progressions that directly support student learning. In mathematics, learning progressions are more often called learning trajectories, and one of the most widely respected sets of examples is the *Building Blocks* early childhood curriculum (Clements & Sarama, 2007). *Building Blocks* activities have a game-like quality, focusing on patterns, spatial and geometric competencies, one-to-one correspondence, counting, equality, classifying, sorting, and sequencing, and so forth. In addition to staying faithful to naturalistic contexts of play, each learning trajectory explicitly extends what students can do initially, engages mathematical thinking and ways of talking, and, when appropriate, connects concrete manipulated objects with technological representations. Clements and Sarama conceptualized their curriculum based on decades of research on young children's mathematics learning, but they also conducted a number of formative research studies, as well as outcome efficacy studies that have demonstrated a positive impact on student learning. In a randomized controlled trial with 35 teachers and 252 students, researchers found that students in classrooms where *Building Blocks* was implemented performed better on a test of early mathematics than students in comparison classrooms where a different mathematics curriculum was used (Clements & Sarama, 2008).

Another example is provided by the School Water Pathways Unit and assessment resources (Caplan, Gunckel, Warnock, & Cano, 2012) upon which the Covitt, Gunckel, Syswerda, and Caplan teacher study in this issue is based. The School Water Pathways Unit is, in turn, based on the conceptual learning progression for Water in Socio-Ecological Systems and accompanying formative assessments developed by Gunckel, Covitt, Salinas, and Anderson (2012). In addition to making the point about the level of detail required, both the *Building Blocks* curriculum and School Water Pathways Unit help to illustrate how much curriculum development work it takes to begin to enable more ambitious teaching practices. In their conceptualization of student learning about water systems, Gunckel et al. (2012) are aware that very often existing curricula make it possible for students to memorize elements in the water cycle without ever being able to engage in model-based reasoning about how the movement of water through the system is affected by man-made structures, steep or flat terrains, or the permeability of soils. The School Water Pathways Unit is designed around the learning progression and specifically addresses typical "misconceptions." For example, students may have the idea that water moves from one place to another, but they may not account for water that is hidden, in ground water or as water vapor. The School Water Pathways Unit instructional activities help students develop new understandings about the features of their school yard (area, slope, and surface type) and then engages them in making summary representations and model-based predictions about runoff. The unit continues with additional investigations calling on students to measure evaporation, measure transpiration (and relate this to what they know about different states of water), and then measure infiltration to learn how gravity and soil structure drive and constrain water movement in the ground.

These examples illustrate the *substantive* depth needed for learning progressions to be effective in supporting student learning. An important and related point—worth emphasizing for a measurement audience—is that *scores are not needed* for either the formative assessment or instructional purposes of learning progressions. Scoring could even harm rather than enhance learning. I share this concern with measurement colleagues because in our research we have seen examples in which focusing on scores has detracted from substantive insights made possible by rich curricular tasks (Shepard, Davidson, & Bowman, 2011). It is also true that the research-based claims about the benefits of formative assessment and assessments built from models of cognition rely on substance. In Kluger and DeNisi's (1996) long-ago meta-analysis of feedback studies, the one third of studies that showed negative effects more often gave feedback that focused on the person rather than features of the task. This pattern in research findings has persisted in the ensuing 20 years. It does little good to tell a fourth grader that his essay is a "2" or that he needs three more items correct to reach mastery on a numeracy test. Feedback is more likely to be beneficial when it helps students see *how* to improve, and this invariably calls for information that is qualitative rather than quantitative. Talking to students about their scores—and implicitly representing learning goals as numbers—has motivational consequences as well. It leads to the commodification of learning, promoting extrinsic over intrinsic motivation, which is antithetical to what Sadler (1989) had in mind when he argued that formative assessment should foster development of expertise by helping students to internalize the features of good work.

When measurement specialists partner with subject-matter experts in the development of learning progressions, it might be wise to think about scores and test-score scales as being most important to ensure vertical coherence between classroom and large-scale assessments—assuming that curricular materials have been jointly developed to enact next-generation standards and more ambitious teaching practices at either the district or state level. Scores are useful in the background to make sure that a "3" in a classroom rubric corresponds to a "3" on an external exam exercise, but at the classroom level, scores per se are of limited value for guiding teaching and learning.

Teacher learning

Bennett (2011) has made a compelling argument that effective interventions—such as formative assessment—must specify a theory of action explaining how key processes and components are expected to work together to produce intended outcomes. In our recent handbook review of research on formative assessment, Bill Penuel and I (Penuel & Shepard, 2016) built on this idea, arguing that to support learning, formative assessment practices must be grounded in a theory of learning. As part of our analysis of several theories of action underlying various approaches to formative assessment, we attended to learning theory plus the material resources, activities necessary to prepare participants for new roles, and social and institutional conditions for success. Learning progressions are examples of what we identified as sociocognitive formative assessment interventions. The learning goals aimed for with learning progressions, and the ambitious expectation that understandings of core ideas and disciplinary practices will be jointly developed, means that the learning progress represented is not found routinely in today's classrooms. It follows then that an adequate theory of action must attend to the supports needed to help teachers learn to teach in these new ways.

The work described in the three articles in this special issue may be thought of primarily as demonstration projects, illustrating the kinds of research questions that need to be addressed as part of learning progression development, with special attention to the adequacy of supports for teacher learning. Although they do not provide definitive proofs, these studies also illustrate the kinds of data collection methods that may be useful in the formative development of learning progressions and subsequently in evaluating their efficacy.

Importantly, the Covitt et al. study reported in this issue is part of a larger curriculum project and illustrates well the types of material resources that can help teachers engage their students in richer learning experiences. Here they focus specifically on their goals for teacher learning about learning progressions, laying out a framework specifying degrees of proficiency on three sequential pedagogical skills: "setting learning goals aligned with a LP, interpreting student ideas with respect to a LP, and responding to student ideas with productive instruction that supports students in moving to higher LP achievement levels" (p. ##). Measurement specialists will be pleased to see the close correspondence between their intended learning goals and the design of the subsequent assessment tasks completed by teachers participating in their environmental literacy professional development project.

Covitt, Gunckel, Syswerda, and Caplan (2018) acknowledge that teacher learning outcomes are not so strong as might have been hoped, but such findings are typical of most research intended to transform teaching practices. The majority of teachers set learning goals "reflective of didactic instruction, focusing on facts rather than model-based reasoning" (p. ##): they tended to focus on students' ideas being right or wrong rather than considering the nature of what students were thinking, and their instructional responses to student errors were more often telling students the correct facts rather than engaging in activities designed to help students develop deeper understandings. The case study of Laurie is a promising example of the kinds of more ambitious teaching practices that are possible with well-crafted curricula and embedded formative assessments tied to learning progressions. The study was not designed to determine whether the project was the reason that Laurie was able to use a three-dimensional watershed model and water dropper to challenge students' incomplete understandings about water flow, but the case is still a powerful example of what the upper end might look like of continua representing teachers' learning progression-based formative assessment practices. Given their important focus on teacher learning, the authors might also want to consider for the future what kinds of substantive feedback might be given to teachers who are "2"s about how to improve their formative assessment practices. For example, what if the best examples (coded as "3"s) from the research were shared and analyzed as part of the next year's professional development?

The von Aufschnaiter and Alonzo study (2018) is innovative in several respects. Its focus on the use of learning progressions to support the learning of *preservice* teachers is rare, possibly even unique. As has already been said, learning progressions provide discipline-specific models for attending to student thinking and for appreciating how detailed aspects of partial understandings might mature into more sophisticated conceptions. In the case of force and motion, for example, students might have further-along conceptions of each before they think about the relationship between the two. In my experience, teaching methods courses have often focused on vibrant inquiry-oriented instructional practices but have set aside the topic of assessment. As a result, novice teachers adopt traditional testing routines once they have their own classrooms. This should be no surprise given their lifelong experience with standardized tests and lack of opportunity to learn about forms of assessment that are directly linked to ambitious teaching practices.

Video-based evidence of student thinking also solves the problem of scale needed to provide preservice teachers with extensive practice at interpreting student thinking. The examples in this study are only a starting point. Of course, preservice teaching methods courses cannot provide candidates with practice relevant to every curricular unit they are likely to encounter. But, as with student learning, the goal should be to determine how much practice—with carefully scaffolded analyses of student thinking and expert-devised instructional adaptations—is necessary before teacher candidates are able themselves to think of productive ways to build on student thinking. Preservice teachers in this study had not had prior experience with learning progressions. Similar to a next question that might be asked in the Covitt et al. (2018) project, I would ask here whether learning progressions could be used instructionally to help preservice teachers gain more detailed experience with incomplete or inaccurate conceptions and perhaps brainstorm together about instructional next steps. Such teacher preparation efforts should pay special attention to the idea of using qualitative descriptions of student understandings and should consider how it might be that practice with assignment of students to numeric levels might shift attention away from attending to the particulars of student thinking, as sometimes occurred in this study. This is again the old problem of focusing on scores instead of meaning (necessary for large-scale assessments but not helpful here where preservice teachers should be trying to figure out what to do instructionally).

The Furtak et al. study (2018) takes the final step in a complete program of research on learning progressions, examining the effects of professional development on teacher-devised formative assessment tasks and the effect of these, in turn, on student learning. Teachers who participated in the project were able to develop their own formative assessment tasks that elicited evidence of student thinking for most, but not all, of the natural selection learning progressions. Moreover, there was some correspondence (although not perfect alignment) between the subtopics assessed by teachers—such as random mutations, variation and heritability, differential survival and reproduction, and deep time—and areas where the learning progression project showed greater pre–post learning gains than had occurred at the outset of the study.

An important feature of the Furtak, Heredia, and Circi (2018) study is their use of an independent achievement measure, the Daphne Assessment of Natural Selection, to evaluate learning gains from the learning progression project separate from participating teachers' own course exams. In addition to comparing mean pre–post achievement gains in project years to those in a baseline year, Furtak et al. also took the novel approach of analyzing gains on individual test items. To do this, they reversed the usual assumption in applications of the Item Response Theory Partial Credit Model (holding item parameters constant and estimating changes in examinee ability, θ values) and instead held examinee θ values constant and estimated shifts in item threshold values from pre- to posttest. This analysis allowed the authors to identify the two-item set addressing Transformationist ideas, which did not improve with instruction. Despite gains in students' increasing understandings of deep time, heritability, and selective reproduction, Lamarkian ideas about short-term adaptations and the ability to pass these on to offspring continued to have intuitive appeal for students. This example illustrates the kinds of insights that learning progression data can provide for subsequent rounds of curriculum revision. The authors were also able to use results from the PCM to distinguish between learning progressions for which students were able to reach the highest level of the continuum versus those for which gains occurred at lower thresholds without moving significant numbers of students into the highest level. This too provides insight about where instructional improvements are needed.

Conclusion

The studies in this special issue of *Applied Measurement in Education* (AME) clearly address the teaching and learning side of learning progressions, rather than measurement per se. Why then are such studies appropriate for a measurement journal? The first and most obvious reason has to do with the knowledge needed to collaborate with subject-matter specialists in the development of learning progressions. In the 2008 special issue of AME, Shavelson (2008) referred to this type of collaboration as a "*romance*—of assessment developer and curriculum developer" (emphasis in original, p. 294) who are together seeking to create embedded formative assessments for an existing curriculum. A deeper and more compelling version of this same reasoning has to do with what measurement specialists need to know about the theory of action underlying the development of learning progressions for learning purposes.

Coming originally from social and organizational theory, theories of action describe how the elements of a social policy are expected to work together to produce a desired outcome. A theory of action for a test or testing practice is akin to validity and interpretive arguments (Kane, 2006) needed to guide both test development and validity evaluations (Shepard, 2016). The theory of action framework, laid out by Penuel and Shepard (2016) for formative assessment interventions, has learning theory at its center. To be effective, formative assessment must be grounded in a theory of learning—which requires explicit attention to the value of learning goals; sufficient evidence regarding means to goals; consideration of motivation and identity development, as well as cognitive goals; and a weighing of consequences for equity. Learning progressions are promising as tools to foster enactment of disciplinary practices (as the science examples in this issue illustrate) and the attainment of more ambitious learning goals. They are constructed to enact both "big" and "little" learning theories, meaning that they reflect both broad sociocognitive conceptions of learning and development arising through social processes and more fine-grained learning theories detailing how understandings of force and motion, the water cycle, or natural selection are developed.

Learning progressions have a certain amount of validity "built in," not just because of their initial conceptual theorizing, but because they are developed recursively with explicit testing of proposed orderings and effectiveness of specific instructional supports. Therefore, for learning purposes, measurement specialists will want to begin validity investigations by examining whether the learning model works as intended, but they should also investigate known limitations of learning progressions. For example, Penuel and Shepard (2016) noted that sociocognitive approaches to formative assessment may not support equitable instruction if they do not allow for multiple pathways into participation. Another potential limitation, well known to measurement specialists, is that present-day learning goals, requiring the integration of disciplinary practices with content knowledge dimensions, are not likely to be modeled well with simple unidimensional methods.

Understanding the classroom learning purposes of learning progressions is foundational to other uses of learning progressions as well. Evidence from research and development cycles helps us to understand what it is that "progresses" and what it is that might be aggregated for summative purposes. Summative and large-scale uses will, of course, require new validity investigations. Given that learning progressions build in instructional supports, a key validity question will be whether tests scores can have the same meaning in large-scale contexts where these supports are not available but where other combinations of knowledge and practices are fostered.

References

Alonzo, A. C., & Steedle, J. T. (2009). Developing and assessing a force and motion learning progression. *Science Education*, *93*(3), 389–421. doi:10.1002/sce.v93:3

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy, and Practices*, *18*(1), 5–25. doi:10.1080/0969594X.2010.513678

Berland, L. K., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. *Science Education*, *94*(5), 765–793. doi:10.1002/sce.v94:5

Caplan, B., Gunckel, K. L., Warnock, A., & Cano, A. (2012). Investigating water pathways in schoolyards. *Green Teacher*, *98*, 28–33.

Catley, K., Lehrer, R., & Reiser, B. (2005). *Tracing a prospective learning progression for developing understanding of evolution*. Paper Commissioned by the National Academies Committee on Test Design for K-12 Science Achievement. Washington, DC: National Academies of Sciences.

Clements, D. H., & Sarama, J. (2007). Effects of a preschool mathematics curriculum: Summative research on the building blocks project. *Journal of Research in Mathematics Education*, *38*, 136–163.

Clements, D. H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal*, *45*(2), 443–494. doi:10.3102/0002831207312908

Corcoran, T., Mosher, F. A., & Rogat, A. (2009). *Learning progressions in science: An evidence-based approach to reform* (CPRE Research Report #RR-63). Philadelphia, PA: Consortium for Policy Research in Education.

Covitt, B., Gunckel, K. L., Syswerda, S., & Caplan, B. (2018). Teachers' use of learning progression-based formative assessment in water instruction. *Applied Measurement in Education*.

Di Sessa, A. A., & Minstrell, J. (1998). Cultivating conceptual change with benchmark lessons. In J. G. Greeno & S. V. Goldman (Eds.), *Thinking practices in mathematics and science learning* (pp. 155–187). Mahwah, NJ: Lawrence Erlbaum.

Duncan, R. G., Rogat, A. D., & Yarden, A. (2009). A learning progression for deepening students' understandings of modern genetics across the 5th–10th grades. *Journal of Research in Science Teaching*, *46*(6), 655–674. doi:10.1002/tea.v46:6

Furtak, E. M., Heredia, S. C., & Circi, R. K. (2018). Exploring alignment among learning progressions, teacher-designed formative assessment tasks, and student growth: Results of a four-year study. *Applied Measurement in Education*.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes. *American Psychologist*, *18*, 519–521. doi:10.1037/h0049294

Gunckel, K. L., Covitt, B. A., Salinas, I., & Anderson, C. W. (2012). A learning progression for water in socio-ecological systems. *Journal of Research in Science Teaching*, *49*(7), 843–868. doi:10.1002/tea.v49.7

Hammer, D. (1996). Misconceptions or p-prims: How many alternative perspectives of cognitive structure influence instructional perceptions and intentions? *Journal of the Learning Sciences*, *5*(2), 97–127. doi:10.1207/s15327809jls0502_1

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport, CT: American Council on Education and Praeger Publishers.

Kluger, A. N., & DeNisi, A. (1996). The effect of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*, 254–284. doi:10.1037/0033-2909.119.2.254

Masters, G. N., Adams, R. A., & Wilson, M. (1990). Charting of student progress. In T. Husen & T. N. Postlethwaite (Eds.), *International encyclopedia of education research and studies. Supplementary volume 2* (pp. 628–634). Oxford, England: Pergamon Press.

Mohan, L., Chen, J., & Anderson, C. W. (2009). Developing a multi-year learning progression for carbon cycling in socio-ecological systems. *Journal of Research in Science Teaching*, *46*(6), 675–698. doi:10.1002/tea.v46:6

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment.* Committee on the foundations of assessment. In J. W. Pellegrino, N. Chudowsky, & R. Glaser (Eds.), *Board on Testing and Assessment, Center for Education. Division of Behavioral and Social Sciences and Education.* Washington, DC: National Academies Press.

National Research Council. (2007). Taking science to school: Learning and teaching science in grades K–8. Committee on Science Learning, Kindergarten Through Eighth Grade. In R. A. Duschl, H. A. Schweingruber, & A. W. Shouse (Eds.), *Board on Science Education, Center for Education, Division of Behavioral and Social Sciences and Education.* Washington, DC: The National Academies Press.

Neumann, K., Viering, T., Boone, W. J., & Fischer, H. E. (2013). Towards a learning progression of energy. *Journal of Research in Science Teaching*, *50*(2), 162–188. doi:10.1002/tea.v50.2

Penuel, W. R., & Shepard, L. A. (2016). Assessment and teaching. In D. H. Gitomer, & C. A. Bell (Eds.), *Handbook of research on teaching*. Washington, DC: AERA.

Plummer, J. D., & Krajcik, J. (2010). Building a learning progression for celestial motion: Elementary levels from an earth-based perspective. *Journal of Research in Science Teaching*, *47*(7), 768–787. doi:10.1002/tea.20355

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Danmarks Paedogogiske Institut. (Chicago: University of Chicago Press, 1980)

Sadler, R. (1989). Formative assessment and the design of instructional assessments. *Instructional Science*, *18*, 119–144. doi:10.1007/BF00117714

Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Acher, A., Fortus, D., … Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, *46*(6), 632–654. doi:10.1002/tea.20311

Shavelson, R. J. (2008). Guest editor's introduction. *Applied Measurement in Education*, *21*, 293–394. doi:10.1080/08957340802347613

Shepard, L. A. (2016). Evaluating test validity: Reprise and progress. *Assessment in Education: Principles, Policy & Practice*, *23*(2), 268–280. doi:10.1080/0969594X.2016.1141168

Shepard, L. A., Davidson, K. L., & Bowman, R. (2011). *How middle school mathematics teachers use interim and benchmark assessment data*. CSE Technical Report 807. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Shepard, L. A., Penuel, W. R., & Davidson, K. L. (2017). Design principles for new systems of assessment. *Phi Delta Kappan*, *98*(6), 47–52. doi:10.1177/0031721717696478

Smith, C. L., Wiser, M., Anderson, C. W., & Krajcik, J. (2006). Implications of research on children's learning for standards and assessment: A proposed learning progression for matter and the atomic-molecular theory. *Measurement: Interdisciplinary Research & Perspective*, *14*(1&2), 1–98.

Songer, N. B., Kelcey, B., & Gotwals. (2009). How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning about biodiversity. *Journal of Research in Science Teaching*, *46*(6), 610–631. doi:10.1002/tea.20313

von Aufschnaiter, C., & Alonzo, A. C. (2018). Foundations of formative assessment: Introducing a learning progression to guide pre-service physics teachers' video-based interpretations of student thinking. *Applied Measurement in Education*.

Wilson, M., Draney, K., & Kennedy, C. (2001). *GradeMap [computer program]*. Berkeley: BEAR Center, University of California.