Practical Assessment, Research, and Evaluation

Volume 27 Article 23

2022

Assumptions Underlying Performance Assessment Reforms Intended to Improve Instructional Practices: A Research-Based Framework

Sarah Wellberg
University of Colorado Boulder

Carla Evans
National Center for the Improvement of Educational Assessment

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 27 Number 23, September 2022 ISSN 1531-7714

Assumptions Underlying Performance Assessment Reforms Intended to Improve Instructional Practices: A Research-Based Framework

Sarah Wellberg, *University of Colorado Boulder*
Carla Evans, *National Center for the Improvement of Educational Assessment*

There is renewed interest around including performance assessments in state and local assessment systems to spur positive changes in classroom instruction and student learning. Previous research has identified the external conditions that mediate the role of assessment in changing instructional practices. We extend that work by focusing on the internal classroom conditions that support improvements in student learning. We identified six key instructional practices from three teacher quality frameworks that may result from policy changes that include complex, performance-based assessments. For each practice, we explored the bidirectional relationships among the instructional core of students, teachers, and content. We argue that altering these relationships requires teachers and students to have both the disposition and the capacity to change, and we identify the assumptions that need to hold in order for those changes to occur in response to the inclusion of performance assessments in state and/or local assessment systems.

Keywords: performance based assessment; instructional improvement; educational policy; elementary secondary education

Introduction

There has been a long-held assumption that educational assessments can serve as both a lever of policy reform and an outcome measure of the extent to which educational reforms are working as intended (Faxon-Mills et al., 2013; Hamilton, 2003). Some advocates for testing reform argue that certain types of assessments incentivize teachers to use instructional practices and techniques that foster the development of deeper learning skills (e.g., critical thinking, problem solving, complex reasoning, and communication) while others encourage rote learning (Conley, 2014; Darling-Hammond et al., 2010; Frederiksen & Collins, 1989; NCEST, 1992; Simmons & Resnick, 1993). Policymakers tend to overestimate the ease with which teachers adopt and implement better instructional approaches in response to assessment changes without attending to the conditions and factors that may mediate those changes (Faxon-Mills et al., 2013), including the job-embedded and on-going professional development, materials, and collaborative supports that are needed to do so (Khattri et al., 1995). Additionally, assessment does not operate in a vacuum. Curriculum, instruction, and assessment should be coherently linked through a common model of learning and alignment to a state's content and performance standards (NRC, 2001). It follows that "we cannot meet the goal of improved teaching through assessment alone" (Marion, 2019, para. 7). Assessments may be a catalyst for change, but they are never sufficient on their own (Faxon-Mills et al., 2013; Hamilton, 2003). If instructional change in support of student learning is the goal, policymakers and

educational leaders in schools and districts must answer the following question: "What types of assessments, situated in what type of system, are likely to support improvements in teaching quality?" (Marion, 2019, para. 15).

Performance-based assessments have long been forwarded as one solution to improve teaching and learning at scale (Conley & Darling-Hammond, 2013; Resnick & Resnick, 1992; Stecher, 2010; Wiggins, 1992). For example, some states attempted to leverage performance-based assessments to improve instructional practices starting in the 1990s, including Kentucky, Vermont, Maine, and Maryland (Firestone et al., 1998; Tung & Stazesky, 2010). While these initiatives had some success, they suffered from inconsistent technical quality and limited resources for things like scoring and professional development (Tung & Stazesky, 2010). The use of large-scale performance assessments was further stifled by the passing of *No Child Left Behind* (NCLB) (Supovitz, 2009). This change in federal policy ushered in an era of multiple-choice tests assessing discrete skills. These tests have been a mainstay in large-scale testing thanks, in part, to the lower cost that comes with the ease of scoring and the ability to reuse items, as they are less memorable than are more involved tasks (Davey et al., 2015). Despite the financial benefits of multiple-choice tests, there has been renewed interest in including performance-based assessments as a signal of deeper learning goals in state-level testing programs by the two consortia – Smarter Balanced Assessment Consortium and PARCC (Faxon-Mills et al., 2013; Herman & Linn, 2013). More recent state testing programs in New Hampshire (NHDOE, 2022) and Hawaii (HIDOE, 2022) share the same policy goals.

Although not all performance assessments require students to demonstrate complex thinking skills because the construct being measured does not require it (e.g., a musical performance), performance assessments typically "allow for the evaluation of both the process used in solving a task and the product" (Lane & Stone, 2006, p. 387). Given the wide variability in performance assessment design, this paper focuses on how complex, performance-based assessments have been forwarded as one lever to promote more desirable instructional practices, especially those relating to deeper learning outcomes (Darling-Hammond et al., 2010; Linn et al., 1991). These types of performance assessments measure how well students apply their knowledge, skills, and abilities to authentic problems (Lane & Stone, 2006). These assessments require a student to produce something (e.g., reports, products, experiments, or performances), which is scored against specific criteria and may be designed to occur over different periods of time (e.g., hours, days, or weeks) depending on the range and complexity of skills to be assessed (Darling-Hammond & Adamson, 2010).

Previous research on the role of assessments in improving instructional quality has emphasized the conditions that mediate the relationship between assessment and instructional practices. Faxon-Mills and colleagues (2013), for example, reviewed the literature on performance assessments in U.S. public schools (among other literature) and found that several conditions mediate the effects of assessment on instruction: attributes of the tests and testing programs; accountability context; educator background, beliefs, and knowledge; school and student characteristics; and district/school policy. This type of research is critical for understanding how previous research has conceptualized studies and findings related to the role of assessment in changing instructional practices. It shows how researchers have focused on various structures and processes that surround the instructional core of classroom practice, but not the core itself. The *instructional core* is the relationships among the level of instructional content, teachers' knowledge and skill, and student engagement (City et al., 2009). As a result, much is known about how individual elements of the core may change in response to the inclusion of complex, performance-based assessments in state- and local systems of assessment (Faxon-Mills et al., 2013) but little is known about how the relationships among those elements are likely to change or about the assumptions that need to hold for teaching practices to shift. Without such conceptualizations, research may continue to focus on external conditions and not also on the internal conditions within classrooms that are necessary for assessments to play a role in promoting more desirable instructional practices.

Therefore, the purpose of this paper is (1) to provide a research-based framework that describes the relational shifts among the content, the teacher, and the students that are needed for desirable changes to the instructional core to occur in response to assessment policy reforms, (2) to identify which

beneficial instructional changes are most likely to occur in response to the inclusion of complex, performance-based assessments in state and/or local assessment systems, and (3) to use our framework to identify the assumptions that need to hold in order for those changes to take place. We begin by establishing a framework centered around the concept of the instructional core of relationships among the student, the teacher, and the content (City et al., 2009) and the idea that teachers (and students) need both the disposition and the capacity to change their practices (McLaughlin, 1990). We then identify six key instructional practices that teachers may adopt in response to the use of complex, performance-based assessments in state and/or local systems of assessment and use our framework to create what we refer to as *observation-assumption triangles*, which elucidate the disposition and capacity assumptions that need to hold for teachers and students to successfully adopt those practices. The paper concludes with implications of this work for research, policy, and future practice. We propose that the observation-assumption triangle framework extends the previous research on the role that assessments can play in changing instructional practices and can be used to anticipate and to monitor instructional changes that are meant to result from educational reforms more broadly.

A Research-Based Framework for Instructional Change

Changes in assessment policy have historically had little success in improving student performance (Alexander et al., 2017; Hanushek & Raymond, 2004). This is likely because these policies tend to result in changes in curriculum but not in instruction (Diamond & Spillane, 2004; Pedulla et al., 2003; Sykes & Wilson, 2016). The alignment of curricular- and tested content is not, on its own, enough to improve student learning outcomes (Polikoff & Porter, 2014). The content being taught is only one aspect of what City and colleagues (2009) refer to as the instructional core: the bi-directional and overlapping relationships among the student, the teacher, and the content within classroom learning systems. In order for a policy to result in positive changes in student learning at scale, City et al. argue that the rigor of the content, the teacher's knowledge and skills, and the students' levels of engagement must all increase. Therefore, policies must impact the relationships among these three aspects of classroom learning systems rather than altering the nature of one piece in isolation or of other aspects of schooling outside of the core in order to effectively promote student learning.

The lack of explicit attention to the instructional core is the main barrier to substantial, policy-based instructional reform because such policies are implementation laden in that they depend primarily upon the actions of teachers (McLaughlin, 1990). If we assume that teachers want what is best for their students, then they are likely already doing their best (Elmore, 2004). If this is the case, then "the [new] policy must be carried out by the very individuals who are regarded as the problem that gave rise to the need for the policy" (Sykes & Wilson, 2016, p. 854). Clearly something must disrupt, and positively inform, how the teacher fosters relationships with their students, selects the content to present, and supports the relationship between the students and the content if instructional change is to take place.

McLaughlin (1990) proposed that teachers need both the *will* (hereafter referred to as *disposition*) and the *capacity* to change their practices. In this framework, teachers must see the value in the proposed changes and must have both the training and the resources needed for successful implementation. If, for example, a teacher believes it is important for students to be actively engaged in learning rather than simply listening to the teacher explain a new concept, but they do not have access to engaging materials or the training to effectively facilitate a student-centered activity, any attempts that they make may not fulfill their instructional goals. Similarly, if a teacher has adequate resources and training but does not think that students benefit from active learning experiences, then they are less likely to put those resources to use.

The success of any educational policy meant to improve instruction is heavily dependent upon the assumptions that teachers have the disposition and the capacity (including access to high-quality resources and training) to improve their decisions about the curriculum, the instructional practices, and the assessments they use in their classrooms. From this perspective, it makes sense that simply incorporating a complex, performance-based task into an assessment system after instruction has occurred has not typically led to the types of formative instructional changes that

policymakers had intended to promote (Firestone et al., 1998; Parke & Lane, 2008; Stecher & Mitchell, 1995). In the next section, we describe six instructional changes that are likely to result from including complex, performance-based assessments in a state or local assessment system, *if* certain dispositional and capacity assumptions hold.

Applying Our Framework to Existing Paradigms of High-Quality Teaching Practices

We examined three existing paradigms for high-quality teaching practices: ambitious teaching (Ball, 2019), Danielson's (2013) framework for teaching, and the National Board for Professional Teaching Standards (NBPTS) framework (2002). We selected the ambitious teaching paradigm because of its widespread use as a theoretical framework for modern education research and teacher preparation and because it helps to support equitable assessment practices (Shepard, 2021). Danielson's framework is widely used for teacher evaluation (Close et al., 2020) and, therefore, reflects the practices that state and/or local officials hope will be used in the classroom. Finally, we selected the NBPTS framework because it is a successful, large-scale program, and certified teachers have been repeatedly shown to be highly effective (see Goldhaber, 2006). We identified six instructional strategies and techniques from these frameworks that teachers may adopt in response to the introduction of a complex, performance-based instruction and assessment system. Table 1 presents these instructional practices along with the sources from which they originated.

In making our selections, we considered why teachers may adopt these instructional practices and what dispositional and capacity assumptions would need to hold in order for these changes to occur as the result of introducing complex, performance-based assessments into an assessment system. Descriptions of the selected practices and these justifications are described in the sections below. Note that underlying all of these practices is also an assumption that the teacher both knows what skills and understandings are being assessed in the complex, performance-based assessments and that they want their students to perform well on these assessments. We would expect that a teacher with those qualities would change their instructional content and processes before the administration of performance assessments in order to better prepare their students to succeed. Whether these changes result in only surface-level test-preparation activities may depend heavily upon the teachers' dispositions and capacities with respect to the instructional core. The ideal is that these assessment and learning activities result in deeper instructional changes and formative feedback cycles that lead to real improvements in students' conceptual understandings and applications of the content.

Table 1. Instructional Practices Possibly Affected by the Implementation of Complex, Performance-Based Assessment

Instructional Practice	Source(s)
Using high-quality questions and prompts	Danielson (2013)
Integrating components of knowledge with habits of thinking	NBPTS (2002)
Actively engaging students in learning	Danielson (2013)
Learning through discussion	Ball (2019); Danielson (2013)
Eliciting and interpreting student thinking	Ball (2019)
Giving students multiple opportunities to showcase their knowledge & abilities	NBPTS (2002)

In the following sections, we explain why each of the selected practices is valuable and why teachers may adopt them in response to the introduction of complex, performance-based assessments. We also examine each practice from the perspective of the instructional core, describing the relationships among the three elements (student-teacher-content) that would be observed if the practices were successfully implemented. Importantly, we also identify the dispositional and the capacity assumptions that would need to hold for those relationships to be observed. We present an observation-assumption triangle diagram, in the format shown in Figure 1, for each instructional practice. These triangles describe the behaviors that one would likely observe if the practice were being implemented with fidelity in the classroom along with the dispositional and the capacity assumptions that support those practices. These

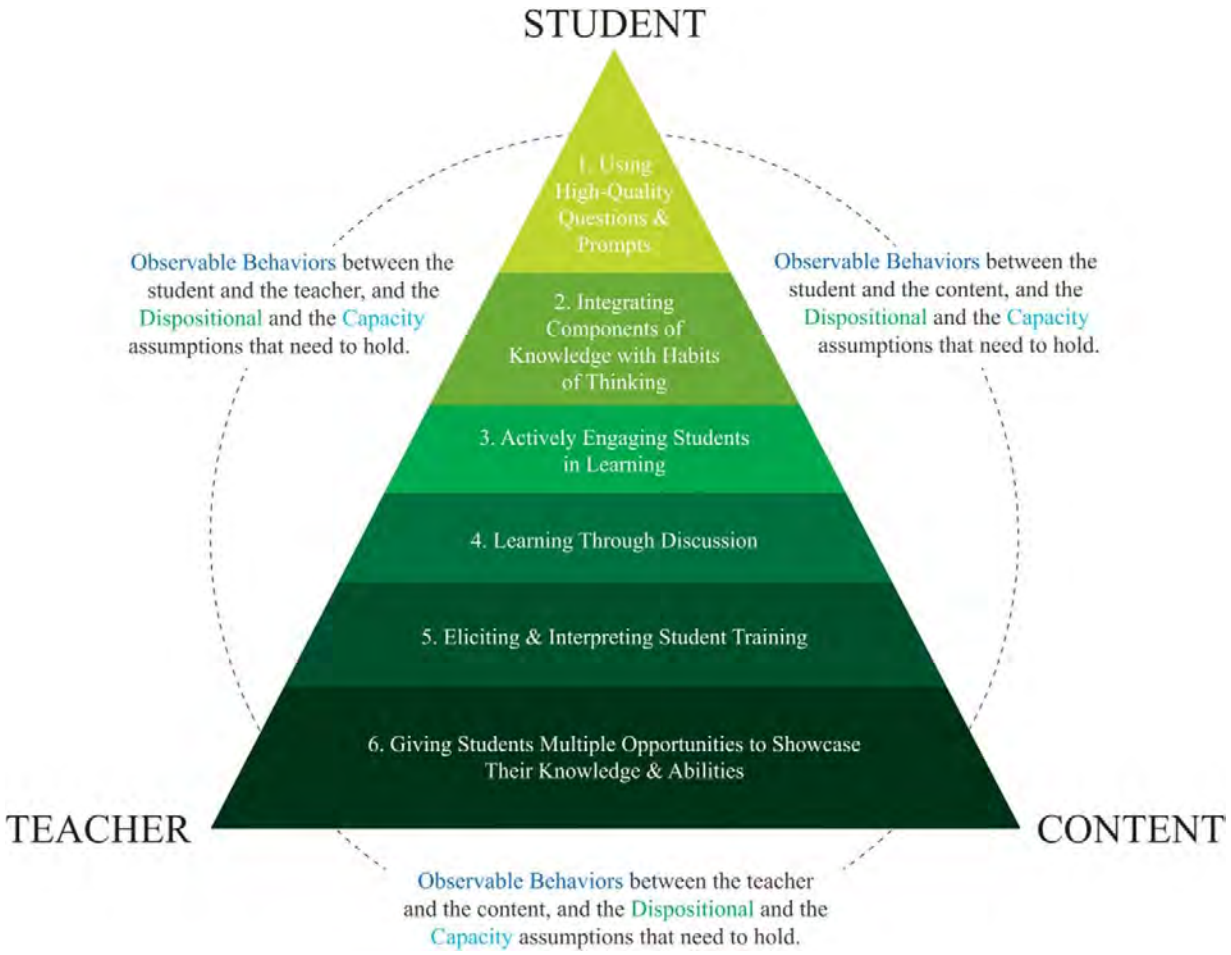
observations and assumptions are listed for each bidirectional relationship within the three elements of the instructional core (student-teacher, student-content, and teacher-content).

Instructional Change 1: Using High-Quality Questions and Prompts

The first thing that we would expect to change with the implementation of performance assessments is the nature and the quality of the tasks that students experience. A task is any classroom activity, related to either instruction or to assessment, that teachers assign to students to help them engage with the content (Tekkumru-Kisa et al., 2020). These tasks are defined by a combination of the products generated by students, how students go about creating those products, and the resources they may use to do so (Doyle, 1983). In some cases, the task may be taking notes while the teacher demonstrates a mathematical algorithm or explains proper use of a semicolon. In other cases, students may be asked to collaboratively solve a real-world, mathematically-based problem or to engage in a debate about a theme in a novel. The nature of these examples is clearly quite different and, unfortunately, students have traditionally been asked to complete lower-level tasks that involve rote learning and recall more frequently than those at a higher level, such as tasks involving application, analysis, or evaluation (Boston & Smith, 2009). This reliance on low-level tasks is problematic because "the work students do…determines how they think about a curriculum domain and come to understand its meaning" (Doyle, 1983, p. 168). If students only engage with low-level tasks, then they are likely to conceive of the subject as something to be memorized and regurgitated rather than as a coherent system of interrelated ideas (Tallman et al., 2016).

One potential benefit of incorporating complex, performance-based assessments into a state, district, and/or classroom assessment system is that in order to

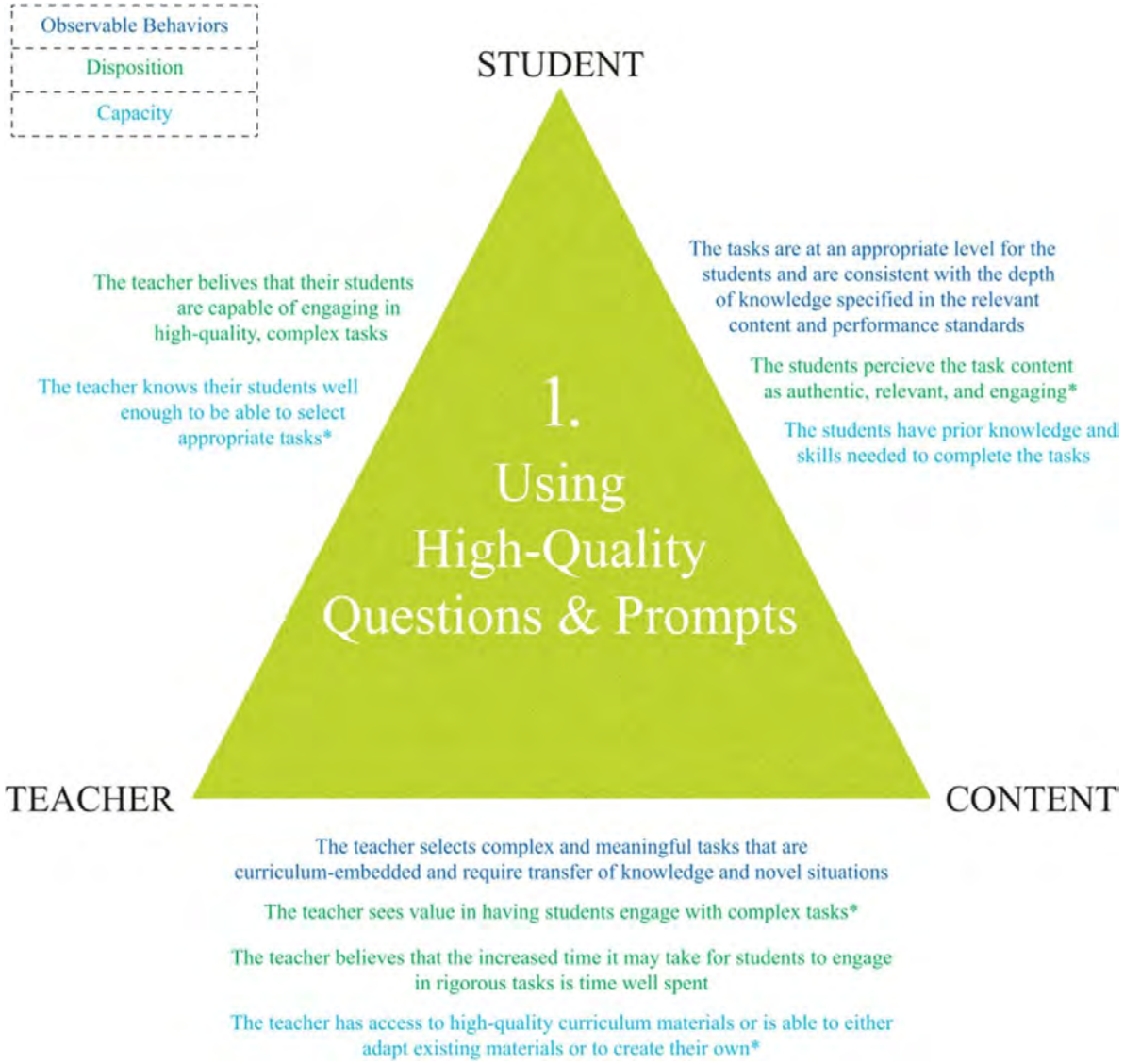
Figure 1. Observation-Assumption Triangle Layout



be successful, students need practice completing higher-order thinking tasks during the formative instructional cycle before they get to the summative assessment. If students are taught seemingly discrete skills and never get to see or to practice how the concepts work together in service of a larger goal, they may not be able to make or to apply those connections on their own in a formal testing setting (Haertel, 1999). We would, therefore, expect teachers to select richer tasks and to model their own thought processes and internal dialogue for their students. This expectation, however, rests upon several assumptions (see Figure 2), starting with the *dispositional* assumptions that the teacher understands that their students need this practice and that they are capable of completing high-quality tasks.

Next, there are the *capacity* assumptions that the teacher has (1) the access to meaningful tasks embedded within high-quality curriculum materials that are designed to elicit student thinking and reasoning, (2) the ability to develop such materials themselves, or (3) the skills to adapt existing materials to increase the level of rigor. Teachers must also know their students well enough to select or to create an

Figure 2. Observation-Assumption Triangle for "Using High-Quality Questions and Prompts"[1]



[1] Note that assumptions marked with * are listed in more than one observation-assumption triangle.

appropriate task, including understanding which contexts the students will find engaging and the prior knowledge that they bring. Given the centrality of tasks in activating student thinking and promoting content-based connections, it is vital that any policy aimed at improving instruction provide teachers with adequate materials and with professional development around the design or selection of rich tasks in support of learning.

Task Implementation. While designing or selecting the right tasks is important, so is maintaining a high level of cognitive demand. *Cognitive demand* is the mental effort that students use while completing a task (Candela, 2016), and it is greatly affected by the way the task is implemented. There are a host of factors that can reduce cognitive demand (Ruk, 2020), and tasks that were designed to elicit high-level thinking from students can easily become procedural, teacher-centered, and over-scaffolded (Smith et al., 2008; Stein et al., 1996). Stein and Lane (1996) address this issue in their task implementation model, which describes three stages of task implementation: task features, task set-up, and task implementation.

The *task features* are the aspects of the task design. Tasks can consist of selected-response items or open-ended questions with multiple acceptable answers. Tasks can require students to use multiple representational forms or be limited to one. Additionally, tasks could ask the student to write a single number or word as their final answer, or they could require students to explain their reasoning. Any of these intended features may change when the teacher introduces a task to the students in the *task set-up*. The way that the teacher describes the task and the things that they ask their students to do can dramatically change the cognitive demand. The transition from the task features to the task set-up is impacted by the teacher's goals, subject matter knowledge, and knowledge of their students (Stein & Lane, 1996).

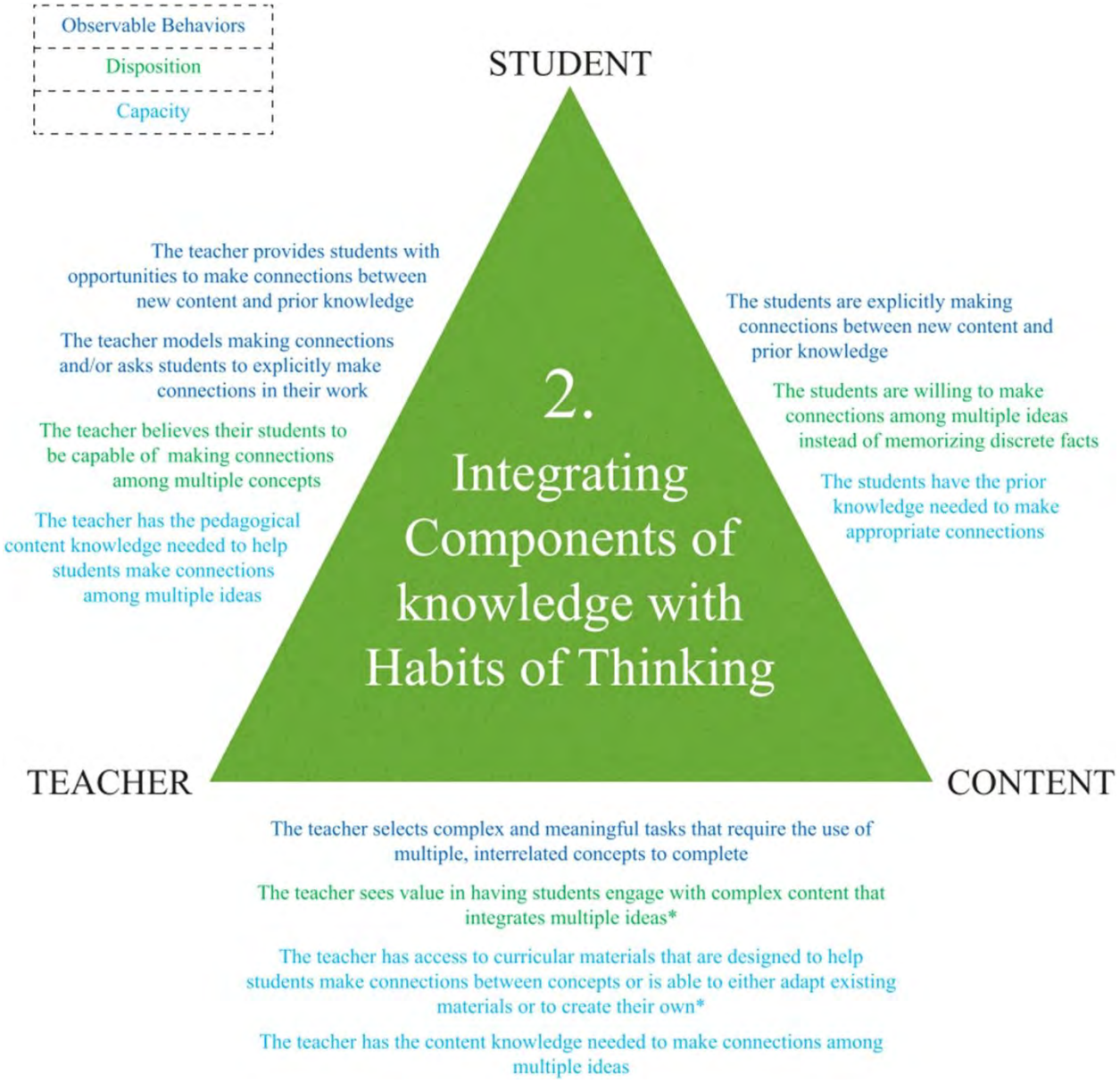
The final stage is *task implementation*, which relates to how the students engage with the task. This stage is affected by both previous stages and by factors such as classroom norms, task conditions, the teacher's instructional dispositions, and the student's learning dispositions. If students are to demonstrate the complex reasoning skills that rigorous tasks are meant to elicit, then teachers must select rigorous tasks that authentically engage students in disciplinary practices, and the intended level of rigor must be maintained. This maintenance occurs via the remaining instructional practices.

Instructional Change 2: Integrating Components of Knowledge with Habits of Thinking

Performance-based assessments typically require students to integrate multiple concepts while solving or responding to a complex problem situated in a novel scenario or context, which requires students to demonstrate higher-order thinking skills. This is a departure from most traditional standardized assessments, in which each item typically measures one content standard at a time (AERA et al., 2014). The well-known limitation of isolated items is that students are not required to apply higher-order thinking skills, such as making connections among multiple concepts, finding and exploiting patterns, or transferring knowledge and skills to new or novel contexts. As stated in the previous section, students cannot make these connections on their own for the first time during a formal assessment situation (Haertel, 1999). Consequently, teachers may increase their focus on building connections among ideas and have students practice making those links explicitly during instruction prior to administering a complex, performance-based assessment. This instructional change, however, depends upon a few assumptions (see Figure 3).

There are the *dispositional* assumptions that the teacher values having students engage with complex content that requires the integration of ideas and that they believe their students are capable of doing this work. There are also the *capacity* assumptions that teachers (1) have access to the types of high-quality tasks embedded within rich curriculum that allow for these connections, (2) have the content knowledge needed to make these connections themselves, and (3) have the pedagogical content knowledge to provide the right type, and amount, of support to help students make these connections themselves. Furthermore, students need to have adequate prior knowledge, as they cannot make connections if they are only vaguely familiar with the content (Kirschner et al., 2006; Vygotsky, 1978). Students must also be willing to make these connections themselves rather than memorizing facts or watching others do this deeper work (Machemer & Crawford, 2007). This requires students to have the disposition and the capacity to be actively

Figure 3. Observation-Assumption Triangle for "Integrating Components of Knowledge with Habits of Thinking"



engaged in the learning process rather than passively receiving information.

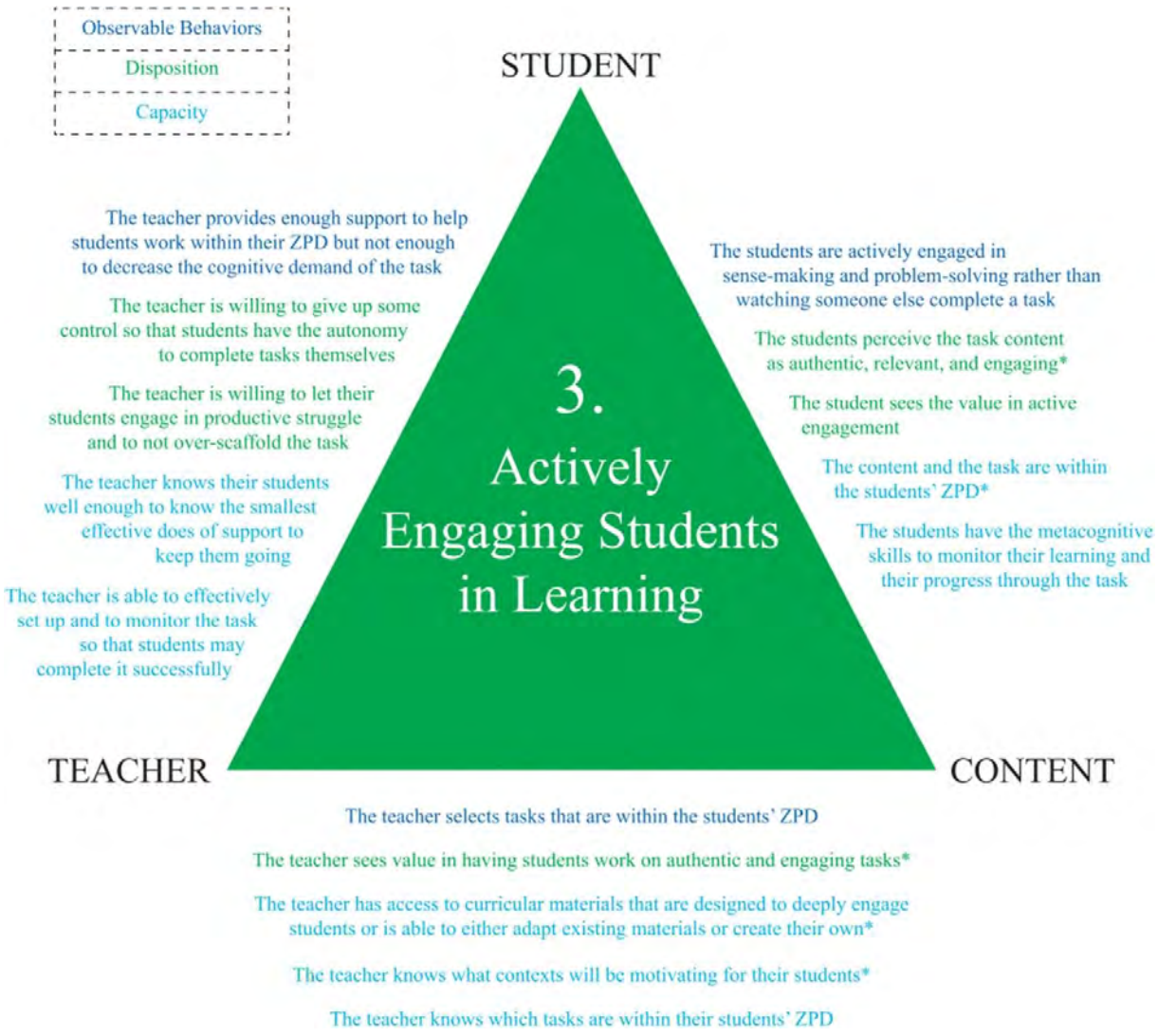
Instructional Change 3: Actively Engaging Students in Learning

It is reasonable to assume that introducing complex, performance-based assessments into a state, district, and/or classroom assessment system may result in more frequent use of active learning techniques. Students need opportunities to develop and to practice the metacognitive skills that these tasks require before the summative performance assessment is administered (Rozencwajg, 2003; Schoenfeld, 2016). Active engagement in learning is associated with increased retention of information (Kvam, 2000; McCarthy & Anderson, 2000), confidence (Townsend et al., 1998), and critical thinking skills (Kim et al., 2013). While active learning can take the form of larger projects and demonstrations of knowledge and skills, such as debates (Hurd, 2000; Oros, 2007) or presentations (Deeley, 2014), it does not need to be a big event to be productive. Cavanagh (2011) found that even including short discussions or other brief periods of reflection in an otherwise traditional lecture can be beneficial. These less-intensive options may be seen as easier to implement, and policymakers may believe that

teachers will take these up in their instructional practices if given the right incentives. While this is an understandable expectation, especially since the benefits of active learning strategies have been known for some time (Biggs, 1988; Boekaerts, 1997; Niemi, 2002), many teachers still do not fully understand how to implement these techniques (Graeff, 2010; Ito & Takeuchi, 2020; O'Grady et al., 2014).

As with the other selected instructional practices, there are several assumptions that underly the effective use of these methods (see Figure 4). The first of which is the *capacity* assumption that teachers have access to, or the ability to create, tasks embedded within high-quality curriculum that will (1) motivate students to actively engage with the content and (2) maintain their attention. Second is the assumption that teachers know their students well enough to select or to create tasks that fall within their students' zones of proximal development (ZPD). The ZPD is the space between what a student can accomplish on their own and what they can do with appropriate support (Vygotsky, 1978). It follows that the tasks teachers select or create need to have the right level of complexity for their students and must include an appropriate amount of risk so as to not overwhelm them (Hurd, 2000). Teachers, therefore, need to be aware of where that space is and of which types of supports will provide just enough of a nudge to help a student keep making progress on a task without reducing the level of cognitive rigor. This involves making individual decisions about how to best support each student depending upon their current level of understanding, which requires the development of individual relationships with each

Figure 4. Observation-Assumption Triangle for "Actively Engaging Students in Learning"



student, student conferrals, and substantial pedagogical content knowledge.

In addition to these capacity assumptions, there is a *dispositional* assumption that the teacher is willing to allow students to engage in productive struggle. Over-scaffolding is one obstacle to students making connections among ideas and transferring knowledge to a new context because it can result in a complex task becoming proceduralized (Stein et al., 1996). High-quality tasks tend to be more complex and less structured than are routine, low-level tasks. This can cause anxiety for students who may not know where to begin, and they may pressure the teacher to be more explicit in the task set-up (Doyle, 1983; Stein et al., 1996). If a teacher is uncomfortable allowing students to struggle (Stein et al., 1996) or believes that students will disengage if they are not provided with the information they need to easily make progress on the problem (Candela, 2016), they are more likely to give away too much information. In doing so, they lower the level of cognitive demand, as they are the ones making the difficult connections while the students attend to the more straight-forward aspects of the task. This is also related to the dispositional assumption that the teacher is willing to give up some of their authority in the classroom (Raney, 2003) and to grant students the autonomy to work on their own with just enough guidance to keep them moving forward (Stefanou et al., 2013).

Once students are given the freedom to take control of their learning, they need to do so by fully engaging in the tasks as set up by the teacher. This can be a big ask, especially for students who have either been successful in traditional classroom settings and do not want to move away from that familiar framework (Machemer & Crawford, 2007) or who have little confidence in their abilities and do not want to be embarrassed in front of their peers (Watkins et al., 2007). Teachers often expect pushback from students who have had negative experiences with collaborative learning methods in the past (Cooper et al., 2000; Finelli et al., 2018). While there are strategies to reduce student resistance, it can be a battle (Finelli et al., 2018; Tino, 2020; Tolman & Kremling, 2017). Students can be especially resistant to instructional activities that require them to interact with one another (Tolman & Kremling, 2017), which is problematic because social interaction is an important part of active learning (Watkins et al., 2007).

Instructional Change 4: Learning Through Discussion

Educational researchers have understood for decades that student-centered classroom discussions are beneficial, leading to higher levels of student engagement, more positive attitudes towards the content, higher achievement, improved communication skills, and opportunities for critical reflection (Carpenter et al., 1989; Delaney, 1991; Helme & Clarke, 2001; Leikin & Zaslavsky, 1997; McKeachie & Kulik, 1975; Wade, 1994). One may expect teachers to begin incorporating more opportunities for discussion into their instruction when performance tasks are introduced into an assessment system. The types of complex tasks that students need practice with may be too large for one student to complete on their own, especially when they are first starting to shift away from traditional modes of learning and towards more active approaches. Performance tasks may also require students to provide some type of explanation or justification, and discussions are one way to have students practice those elaborative skills. Furthermore, discussing different problem-solving methods is known to be an effective way to support students in developing conceptual understanding and key skills such as the ability to critique one another's arguments (Franke et al., 2007), which "helps learners to recognize, clarify, and repair inconsistencies in their own thinking" (Webb et al., 2006, p. 64).

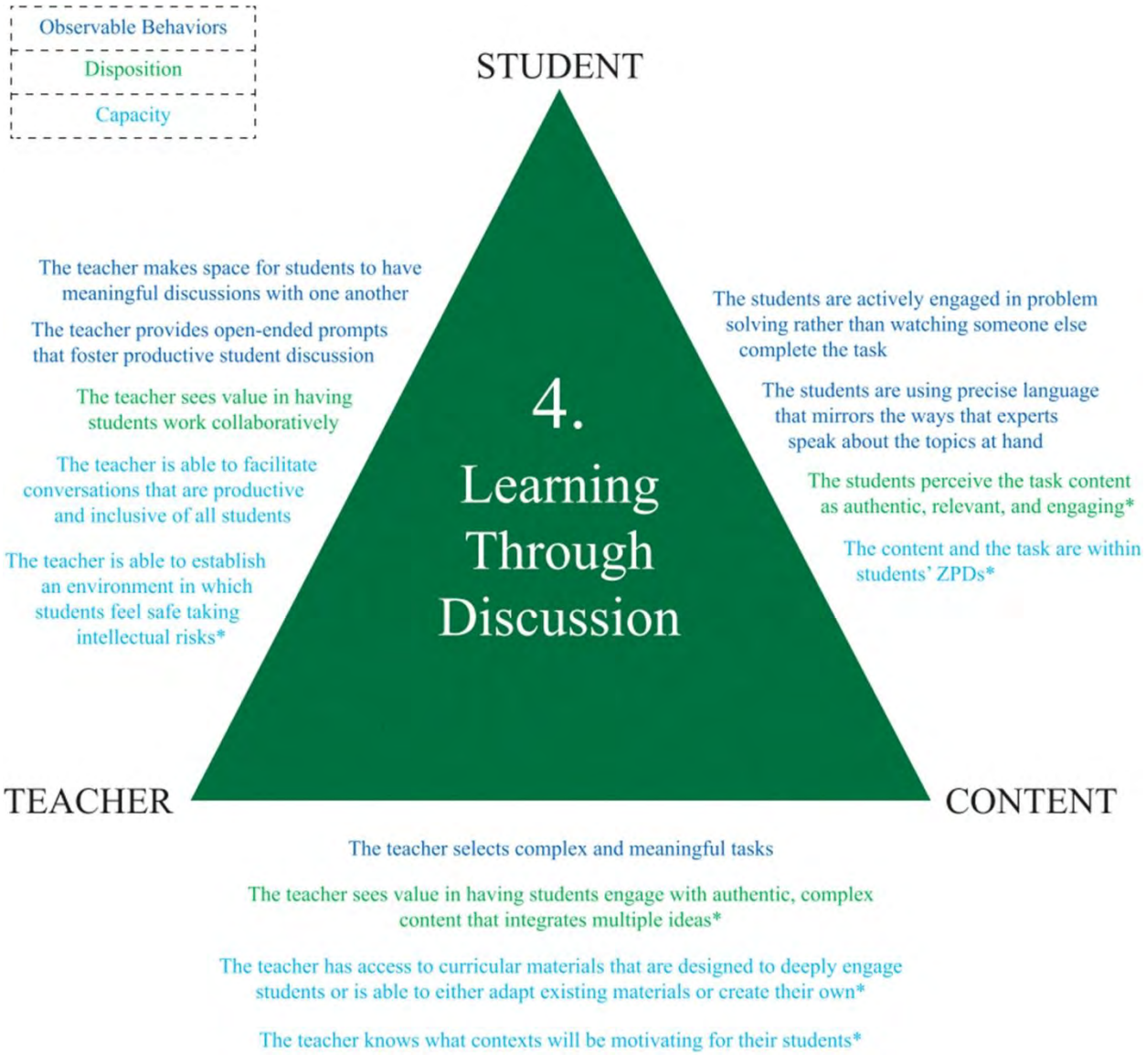
Having students talk to one another, however, does not guarantee these positive outcomes. Classroom conversations need to be structured in a way that maintains students' attention on the task at hand and that makes space for all students to contribute (Franke et al., 2007). Teachers, therefore, cannot simply set their students loose and hope for the best but rather need to continually monitor who is speaking, what ideas are coming up, how students' prior knowledge is interacting with the task, and the attitudes of the students with regard to the task and to one another (Lampert, 2001). Consequently, there are several assumptions that underly successful use of discussion for learning in the classroom (see Figure 5). *Dispositional* assumptions include the teacher seeing the value in learning through discussion and believing that their students are capable of doing so. This works in tandem with the assumption that the teacher understands the benefits of having students engage

with complex, authentic materials that are relevant to their lives and are worth discussing.

Next is the *capacity* assumption that the teacher has access to, and is able to select, tasks that are motivating for their students. This is important, as students may get off track with their conversations (Webb et al., 2006), and they are much more likely to remain engaged with a task if they feel there is something meaningful to talk about (Stein & Lane, 1996). Similarly, the tasks should be within the students' ZPD so that they are able to make progress and not feel overwhelmed by a task that is too far outside of their reach. If the students are not able to make progress on the task, they are more likely to disengage (Greeno et al., 1996; NASEM, 2018). Appropriate task selection, therefore, requires teachers to understand their students' interests and current levels of understanding.

A second *capacity* assumption is that the teacher has developed the previously-mentioned skills that Lampert identified as critical for monitoring productive classroom discussions. It can be intimidating for students to share their ideas, and students from different backgrounds and with different social identities tend to differ in their willingness to engage in classroom discussions (Lubienski, 2002). It is, therefore, up to the teacher to create a classroom environment that is welcoming and supportive and in which students feel safe sharing their work.

Figure 5. Observation-Assumption Triangle for "Learning Through Discussion"



Establishing productive norms for social engagement requires substantially more effort from the teacher than does enforcing more traditional classroom rules that center around students quietly absorbing information (Yackel et al., 1991). If discussion norms are not successfully enacted, then it is reasonable to assume that the work would not be productive, and the teacher may stop trying to use these techniques in favor of more familiar methods.

The final sets of assumptions are about how the students interact with one another and with the content. In order for students to successfully communicate about the content, they need to use precise language so that everyone in the discussion has a clear idea of what the speaker intends to convey (Sfard & Kieran, 2001). If students do not have a shared vocabulary, or if the speaker does not provide enough detail, confusion may arise. It is entirely possible to speak with someone and to have very little understanding of what they are attempting to say, or to think that you understand when you really do not (Trognon, 1993). This disconnect can result in unproductive forms of discussion, such as students speaking mostly to themselves while working on a problem individually rather that reasoning with their partners (Kieran & Dreyfus, 1998). Another common issue is students' unfamiliarity with the types of questions that they may need to ask one another and the types of explanations they need to produce in order to really dig into the content. Webb and colleagues (2006), found that when they tried to introduce collaborative work into a traditional mathematics classroom, "students did not seem to realize that their groupwork interactions could (and should) differ significantly from their traditional classroom interactions" (p. 109). In order for discussions to be productive, both teachers and students need to be trained in how to hold higher-order conversations that promote deeper reasoning.

Instructional Change 5: Eliciting and Interpreting Student Thinking

Teachers who implement complex, performance-based assessments may attempt to get more information about their students' thought processes, since these tasks require far more than recall of memorized facts and procedures. Students typically need to apply and connect their knowledge in new contexts while completing these complex tasks, which may expose misconceptions that they hold. In order to properly prepare students, teachers need to elicit more information about what their students know and can do, so as to address these misconceptions before the assessment. Having students explain their thinking is a common way to get this information. Moreover, such explanations are a common feature of performance tasks that should be practiced and modeled ahead of time. Students need to be engaged in an appropriate task, to have adequate time to think, to see how others approach and solve related tasks, and to be encouraged to explain their thinking if they are to develop the reasoning skills needed to complete complex tasks (Greeno et al., 1996; NASEM, 2018).

One way to encourage students to discuss their thinking is by asking questions about what they have done or what they think they should do. It is essential, however, that teachers, and students engaged in group work, ask the right kinds of questions during task-based discussions, as "the nature of the question has a remarkable impact on the progression of thought in the class" (Dean, 1986, p. 185). The questions that are asked during a task can scaffold students' engagement, create opportunities for students to engage with higher-order ideas, and shape the classroom culture (Kazemi & Stipek, 2001; Smith, 2000). High-quality questions are closely linked to students' experiences, help students develop their reasoning skills, and lead to more creative thinking (Lee & Kinzie, 2012). Regrettably, the majority of questions that are asked in classrooms are short-answer and lower-level (Webb et al., 2006). If students are asked questions, they are often expected to respond with a single answer that is subsequently judged to be either correct or incorrect (McNeill & Pimentel, 2010). This often occurs during what is called an *initiation-response-evaluation interaction pattern*, in which the teacher asks a question, the student answers it, and the teacher evaluates the correctness of the answer, usually with little attention paid to the process that the student used to obtain it (Mehan, 1979). This lack of concern for the thought processes that students use to solve problems can seriously inhibit students' abilities to develop rich understandings of the content by reducing opportunities for them to reflect on and to refine their knowledge, especially if the student produced an incorrect answer (Franke et al., 2009; Shaughnessy et al., 2020). Sequences of probing questions have been found to be the most effective way to follow-up on a

student's response, but teachers do not use this strategy the majority of the time (Franke et al., 2009).

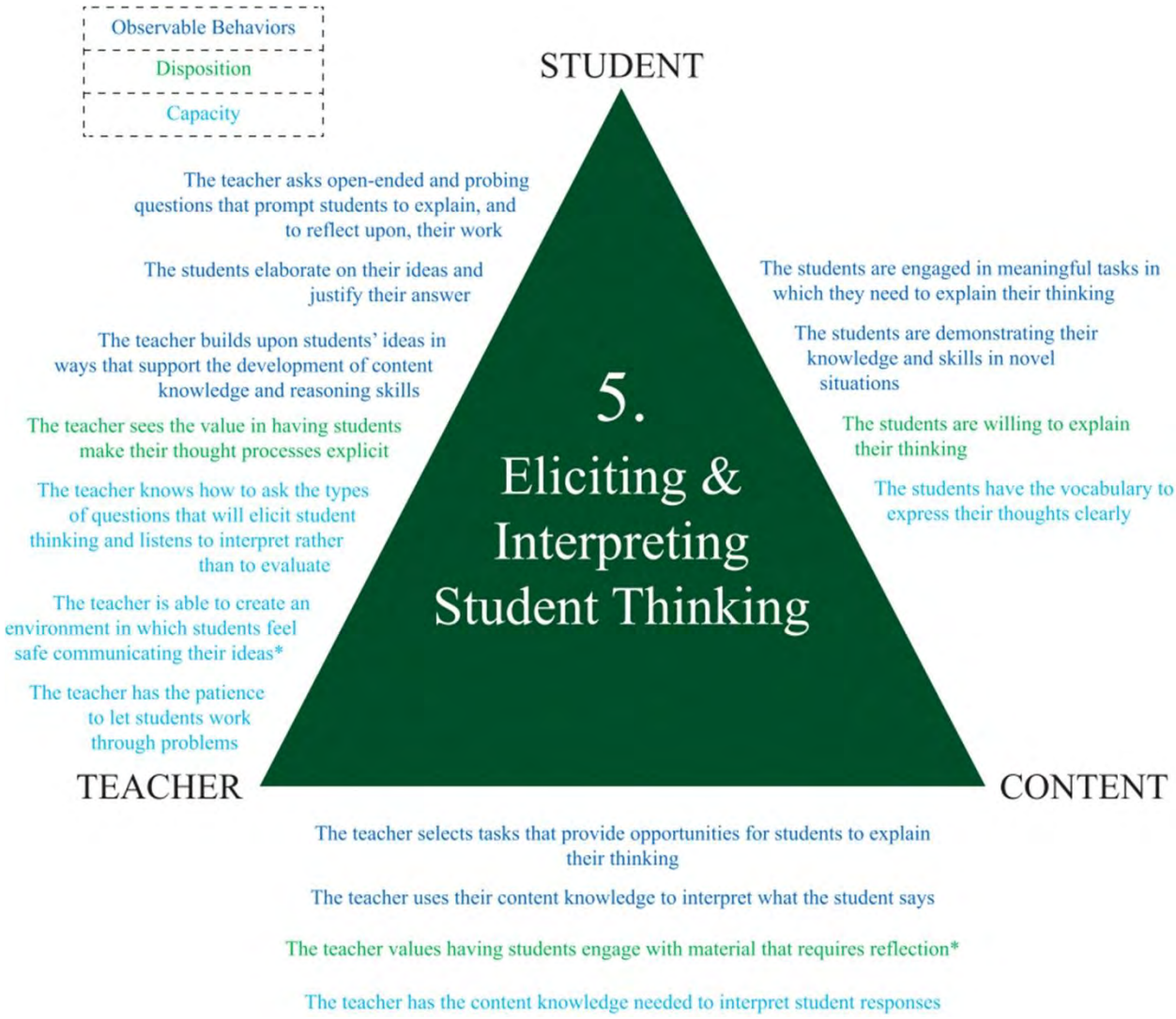
There is ample evidence that probing student thinking has beneficial effects (Jacobs et al., 2007; Sfard & Kieran, 2001; Silver & Stein, 1996), but teachers tend to have difficulty doing so in ways that support learning (Franke et al., 1998; Shaughnessy et al., 2020). In traditional classrooms, teachers tend to ask close-ended questions, which have one correct answer. Kim (2015) found that 78% of the questions asked in traditional classrooms were close-ended, as opposed to 44% in argumentation-based classrooms. Furthermore, many teachers ask close-ended questions in which they fill in students thinking by positing what the student may have done or thought and then having the student either confirm or reject their statement (Shaughnessy et al., 2020). Even if teachers understand the importance of using open-ended questions, it is often difficult for them to alter their practice (Oliveira, 2010; Scott, 1998), possibly because they tend to see student's responses as being either right or wrong rather than containing nuances that need to be explored (Gotwals & Birmingham, 2016). This is concerning because a strict focus on having students produce the "right answer" inhibits conceptual growth by reducing the amount of information the teacher gets about the student's thought process (Shaughnessy et al., 2020). The predominance of close-ended questions not only limits the quality of teacher-student discussions and interactions, but also impacts how students communicate with one another when working in groups. Webb and colleagues (2006) found that students whose teachers asked more open-ended questions were more likely to ask their peers those kinds of questions. It appears that the teacher's questioning style sets the tone for the rest of the class and that "an entrenched culture of low-level questions and explanations is very hard to overcome" (Webb et al., 2006, p. 109).

As with the other practices, certain assumptions must hold for teachers to shift their questioning style (see Figure 6). First is the *dispositional* assumption that the teacher sees value in understanding students' thinking and in having them explain their reasoning. While this is essential for change, it is insufficient on its own (Oliveira, 2010; Scott, 1998). There are also the *capacity* assumptions that the teacher has the content knowledge and the pedagogical content knowledge to be able to ask the types of questions that will uncover students' thinking and to interpret their responses (NRC, 2001). Teachers also need the patience to follow a student through a train of thought until they reach a shared understanding (Matusov & Smith, 2007). Moreover, the teacher needs to be able to establish a classroom culture in which students feel safe elaborating on their ideas (Franke et al., 2007). Students, in turn, need to be engaged in meaningful tasks that call for explanations and justifications. They also need to be willing to share their thoughts and to have the vocabulary required to express those thoughts clearly (Rosebery et al., 2005). If all of these assumptions hold, then teachers will be able to get a deeper understanding of what students know and can do much more frequently than they would in a traditional classroom environment.

Instructional Change 6: Giving Students Multiple Opportunities to Showcase Their Knowledge and Abilities

In an environment in which students mostly listen to lectures and take notes, formal exams or tests may be the only ways that teachers get information about what their students know and can do. This approach has been widely criticized, and the importance of on-going formative assessment processes has been known for some time (Black & Wiliam, 1998, 2009; Crooks, 1988; Sadler, 1998). Nevertheless, if students are not given opportunities to actively engage with the material in some way during a lesson or to share their thinking during student conferrals, it is more difficult for teachers to gain insights into their thinking. If, however, the instructional practices described above are being implemented, there will be ample opportunities for students to communicate what they do and do not yet understand before a formal assessment arrives. The teacher will, therefore, be able to make more informed decisions about how to monitor or adjust their instruction in ways differentiated to student needs. Furthermore, students will also be able to monitor their own learning more effectively, as they will have real-time feedback about how well they understand the concepts they are learning. In this way, these tasks can be seen as *assessment in the service of learning*, in that teaching,

Figure 6. Observation-Assumption Triangle for "Eliciting and Interpreting Student Thinking"



assessment, and learning work together to help students develop deeper understandings (Gordon, 2020). Furthermore, if teachers have adopted instructional tasks that require problem-solving, collaboration, and concept integration, these tasks will allow students to demonstrate their knowledge in a wider variety of ways. These may include portfolios, projects, presentations, essays, capstone projects, small lesson-embedded tasks, and even observations of students as they engage in group work (Maki, 2002).

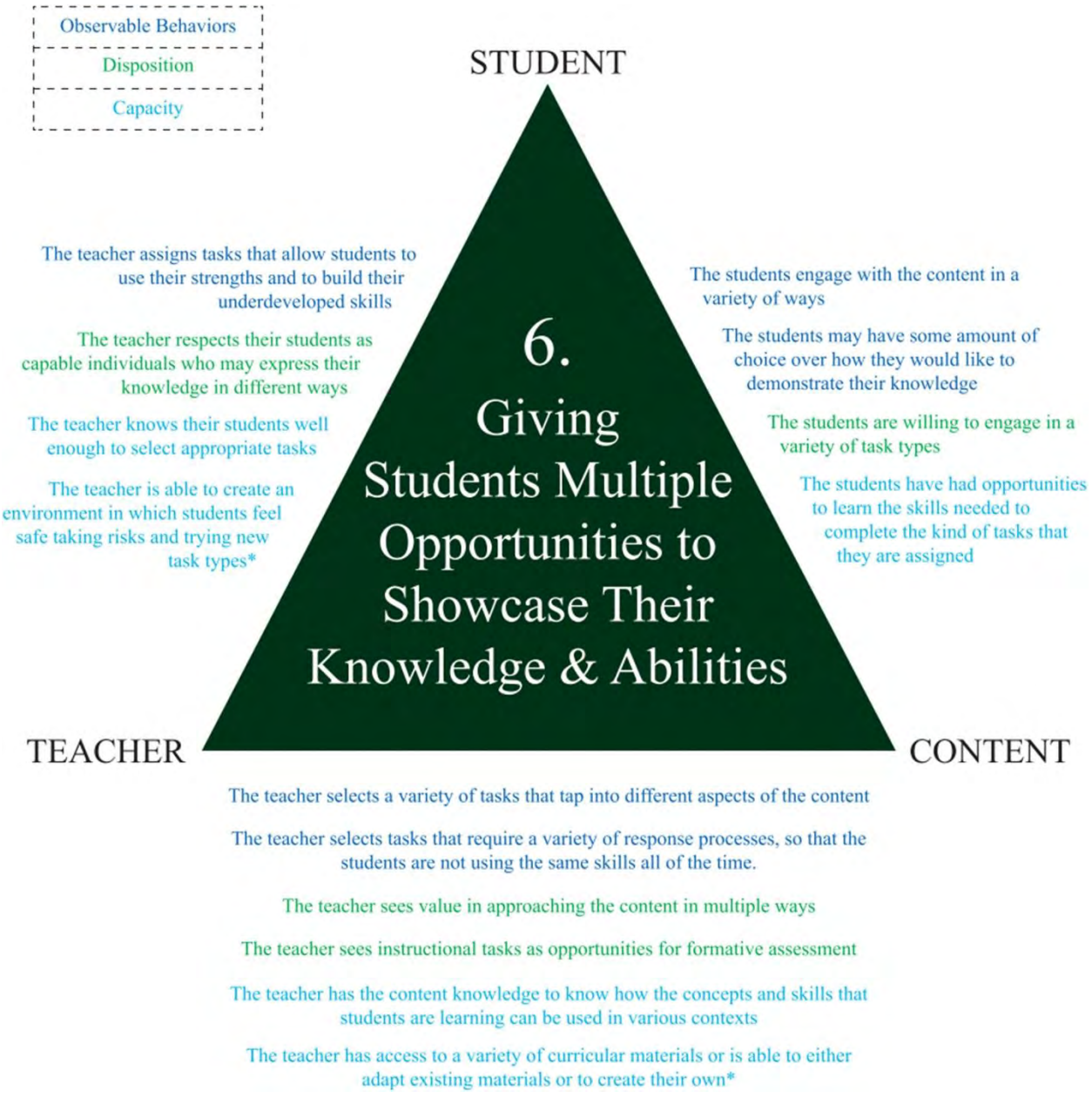
Using the same types of assessments repeatedly puts the same students at a disadvantage each time (Brown, 2005). These students may have the knowledge needed to complete a task but may not be able to express it effectively in the specific way that is asked of them (Stanford, 2003).

Multiple intelligence theory (Gardner, 1983) stresses that there are eight different forms of intelligence and that while everyone has nonzero amounts of each form, some forms may be more well developed than others within an individual. By using a variety of tasks for formative or summative assessment purposes, teachers can help students with different strengths demonstrate what they know and can do by drawing upon various assets including students' habits of mind (Costa & Kallick, 2009) and cultural capital (Yosso, 2005) in ways that allow students to feel confident. Furthermore, complex tasks may help students use their dominant intelligences to build up those that are less well developed (Gardner, 1983). This may also involve some amount of choice on the student's part, as teachers may provide students with options for which task they would like to complete.

The key to this strategy is to ensure that the content representation and the cognitive demand is roughly equivalent among all of the presented options so that students who select one task are not advantaged over those who choose another.

Certain assumptions must hold for students to have a range of opportunities to show what they know (see Figure 7). The first of these is the *capacity* assumption that teachers have access to a variety of task types that use an array of different skills and types of intelligence. Additionally, the teacher must have the content knowledge needed to select which task types are most appropriate for the concepts at hand, and they also need adequate knowledge of their students to select tasks that will help them use their dominant intelligences and assets to demonstrate their knowledge and skills. Furthermore, the teacher needs to be able to create a safe environment in which students feel comfortable trying new task types.

Figure 7. Observation-Assumption Triangle for "Giving Students Multiple Opportunities to Showcase Their Knowledge and Abilities"



Next are the *dispositional* assumptions, starting with the teacher seeing value in providing students with multiple opportunities to demonstrate their knowledge. If a teacher believes that a paper-and-pencil test is the best, or the only, way to assess the subject matter, then they are unlikely to use other types of tasks. Teachers also need to see authentic instructional tasks as valuable sources of formative assessment information that can help inform their instruction and student goal setting. Without this view, teachers are less likely to have students actively engage with the content. Furthermore, the teacher needs to believe their students are capable individuals who are able to develop deep understandings regardless of how they are best able to communicate those understandings. If the teacher does not value non-traditional forms of academic expression and believes that exams are the only valid measure of understanding, it is unlikely that they will value the information that more complex tasks provide.

The final set of assumptions have to do with the students. As discussed in previous sections, many students are resistant to engaging in non-traditional tasks for a variety of reasons. For students to be able to demonstrate their knowledge, they must be willing to engage with the tasks the teacher selects. This is closely linked to the *capacity* assumption that the teacher can create a classroom culture in which the students feel safe trying new things. Finally, there is a capacity assumption that the students have had opportunities to learn, and to receive feedback on, the skills needed to complete the tasks. If a task is being used for instructional purposes and is providing formative information about how students are picking up a new idea, the task needs to be within their ZPD. If the task is a summative assessment, the students need to have had opportunities to not only learn the content but to develop the task-related skills required for a successful performance.

Implications and Future Research

States, districts, or schools that are in the process of developing policies that aim to use performance assessments to affect instructional change would be wise to consider whether the requisite assumptions hold during the design phase. While we have identified several assumptions, there are a handful that are repeated across multiple practices. The two most common assumptions are the *dispositional* assumption that the teacher sees the value in having their students engage with complex, meaningful tasks and the *capacity* assumption that the teacher has access to high-quality materials. These are things that policymakers should consider when designing a new system. Professional development around learning theory and ambitious teaching practices may be needed to help teachers see why students need to actively engage with authentic tasks. Additionally, new policies should include appropriate allocation of resources to ensure that all teachers have access to the types of curricular materials that will facilitate the use of the practices we have identified.

The observation-assumption triangles we have described can benefit researchers and practitioners that are interested in promoting and monitoring the impact of complex, performance-based assessments on teachers' instructional practices. Our fundamental claim is that myriad assumptions need to hold for teachers to effectively adopt the types of practices that would prepare students to successfully complete complex, performance-based assessment tasks. Researchers and practitioners can first explore the extent to which some of the assumptions are in place prior to trying to measure student achievement outcomes. Once there is a reasonable likelihood that teachers are indeed changing instructional practices in ways aligned with the performance assessment reform, then more nuanced information on student academic performance and growth can be collected.

Additionally, the assumptions may also serve as a jumping-off point for school-level teacher evaluation discussions if school administrators do not observe the desired practices being implemented. It is common for administrators to assume that teachers are simply resistant to change (Bantwini & King-McKenzie, 2011). While this may be true in some cases, this assumption about teachers' dispositions needs to be verified before decisions are made on that basis. Bantwini and King-McKenzie (2011) found that teachers who were assumed by administrators to be unwilling to change their practices in response to a new initiative were actually unsure about the policy and did not know what they were meant to be doing. Just as with students, teachers are unlikely to accomplish a task successfully if the instructions are unclear and if they do not receive formative feedback that helps them to improve. This would be something a building

administrator could ask a teacher directly in an observation debrief meeting and future research could explore question such as: Does the teacher understand the objectives being measured by the performance tasks? Do they have the bank of materials and instructional strategies needed to teach towards those objectives? If administrators can uncover the factors that are impeding adoption of preferred instructional strategies, can they then make informed decisions about how to address those issues in a way that promotes the reforms goals?

Conclusion

We have presented a research-based framework for instructional change that is based upon the instructional core of the student, and teacher, and the content (City et al., 2009) and a theoretical argument for why, and how, instruction may beneficially change in response to the inclusion of complex, performance-based assessments in state and/or local assessment systems. Assessment policy changes often result in updated curriculum with very little effect on the instructional practices that teachers use (Au, 2007). We argue that this is likely the result of limitations in teachers' dispositions and in their capacities to implement more effective teaching strategies that change the instructional core of classroom practices. In order for substantive changes in student learning outcomes to occur, teachers need to want, and to be able, to change how they, their students, and the content relate to one another. Identifying the dispositional and the capacity barriers that hinder instructional progress can help stakeholders at all levels better understand why policies that introduce performance assessments are or are not leading to the intended instructional changes. Furthermore, the observation-assumption triangles that we developed are applicable to any instructional practice. Researchers, policymakers, and/or practitioners may examine other practices for their embedded assumptions and observable dimensions in order to monitor the effects of policy reforms.

References

Alexander, N. A., Jang, S. T., & Kankane, S. (2017). The performance cycle: The association between student achievement and state policies tying together teacher performance, student achievement, and accountability. *American Journal of Education*, *123*(3), 413–446.

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

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, *36*(5), 258–267. https://doi.org/10.3102/0013189X07306523

Ball, D. (2019, September). *Teaching ambitiously: What does this mean, and what does it take?* [Keynote Address]. https://static1.squarespace.com/static/577fc4e2440243084a67dc49/t/5d83dc463b6ff512ebe00300/1568922700302/091819_UColorodo.pdf

Bantwini, B. D., & King-McKenzie, E. L. (2011). District officials' assumptions about teacher learning and change: Hindering factors to curriculum reform implementation in South Africa. *International Journal of Education*, *3*(1), 10. https://doi.org/10.5296/ije.v3i1.655

Biggs, J. (1988). The role of metacognition in enhancing learning. *Australian Journal of Education*, *32*(2), 127–138. https://doi.org/10.1177/000494418803200201

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

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (Formerly: Journal of Personnel Evaluation in Education)*, *21*(1), 5–31.

Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction*, *7*(2), 161–186. https://doi.org/10.1016/S0959-4752(96)00015-1

Boston, M. D., & Smith, M. S. (2009). Transforming secondary mathematics teaching: Increasing the cognitive demands of instructional tasks used in teachers' classrooms. *Journal for Research in Mathematics Education*, *40*(2), 119–156.

Brown, S. (2005). Assessment for learning. *Learning and Teaching in Higher Education*, *1*, 81–89.

Candela, A. G. (2016). Mathematics teachers' perspectives on factors affecting the implementation of high cognitive demand tasks. *Proceedings of the 38th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, 315–322.

Carpenter, T. P., Fennema, E., Chiang, C.-P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal*, *26*(4), 499–531.

Cavanagh, M. (2011). Students' experiences of active engagement through cooperative learning activities in lectures. *Active Learning in Higher Education*, *12*(1), 23–33. https://doi.org/10.1177/1469787410387724

City, E. A., Elmore, R. F., Fiarman, S. E., & Teitel, L. (2009). *Instructional rounds in education* (Vol. 30). Cambridge, MA: Harvard Education Press.

Close, K., Amrein-Beardsley, A., & Collins, C. (2020). Putting teacher evaluation systems on the map: An overview of states' teacher evaluation systems post–Every Student Succeeds Act. *Education Policy Analysis Archives*, *28*, 58. https://doi.org/10.14507/epaa.28.5252

Conley, D. T. (2014). *A new era for educational assessment. Students at the center: Deeper learning research series*. Jobs for the Future.

Conley, D. T., & Darling-Hammond, L. (2013). *Creating systems of assessment for deeper learning*. Stanford Center for Opportunity Policy in Education.

Cooper, J. L., MacGregor, J., Smith, K. A., & Robinson, P. (2000). Implementing small-group instruction: Insights from successful practitioners. *New Directions for Teaching and Learning*, *2000*(81), 63–76. https://doi.org/10.1002/tl.8105

Costa, A. L., & Kallick, B. (2009). *Habits of mind across the curriculum: Practical and creative strategies for teachers*. ASCD.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, *58*(4), 438–481.

Danielson, C. (2013). *The framework for teaching: Evaluation instrument*.

Darling-Hammond, L., Adamson, F., & Abedi, J. (2010). *Beyond basic skills: The role of performance assessment in achieving 21st century standards of learning*. Stanford Center for Opportunity Policy in Education.

Davey, T., Ferrara, S., Shavelson, R., Holland, P., Webb, N., & Wise, L. (2015). *Psychometric considerations for the next generation of performance assessment* (pp. 1–100). Center for K–12 Assessment & Performance Management at ETS Educational Testing Service.

Dean, D. (1986). Questioning techniques for teachers: A closer look at the process. *Contemporary Education*, *57*(4), 184–185.

Deeley, S. J. (2014). Summative co-assessment: A deep learning approach to enhancing employability skills and attributes. *Active Learning in Higher Education*, *15*(1), 39–51.

Delaney, E. (1991). Applying geography in the classroom through structured discussions. *Journal of Geography*, *90*(3), 129–133.

Diamond, J. B., & Spillane, J. P. (2004). High-stakes accountability in urban elementary schools: Challenging or reproducing inequality? *Teachers College Record*, *106*(6), 1145–1176. https://doi.org/10.1111/j.1467-9620.2004.00375.x

Doyle, W. (1983). Academic work. *Review of Educational Research*, *53*(2), 159–199.

Elmore, R. F. (2004). Accountability systems. In S. H. Fuhrman & R. F. Elmore (Eds.), *Redesigning accountability systems for education*. Teachers College Pre.

Faxon-Mills, S., Hamilton, L. S., Rudnick, M., & Stecher, B. M. (2013). *New assessments, better instruction? Designing assessment systems to promote instructional improvement*. RAND Corporation.

Finelli, C. J., Nguyen, K., DeMonbrun, M., Borrego, M., Prince, M., Husman, J., Henderson, C., Shekhar, P., & Waters, C. K. (2018). Reducing student resistance to active learning: Strategies for instructors. *Journal of College Science Teaching*, *47*(5).

Firestone, W. A., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis*, *20*(2), 95–113. https://doi.org/10.3102/01623737020002095

Franke, M. L., Carpenter, T., Fennema, E., Ansell, E., & Behrend, J. (1998). Understanding teachers' self-sustaining, generative change in the context of professional development. *Teaching and Teacher Education*, *14*(1), 67–80.

Franke, M. L., Kazemi, E., & Battey, D. (2007). Mathematics teaching and classroom practice. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 1, pp. 225–256). Information Age Publishing.

Franke, M. L., Webb, N. M., Chan, A. G., Ing, M., Freund, D., & Battey, D. (2009). Teacher questioning to elicit students' mathematical thinking in elementary school classrooms. *Journal of Teacher Education*, *60*(4), 380–392. https://doi.org/10.1177/0022487109339906

Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, *18*(9), 27–32.

Gardner, H. (1983). *Frames of mind: The theory of multiple intellegences*. Basic Books.

Goldhaber, D. (2006). National Board teachers are more effective, but are they in the classrooms where they're needed the most? *Education Finance and Policy*, *1*(3), 372–382.

Gordon, E. W. (2020). Toward assessment in the service of learning. *Educational Measurement: Issues and Practice*, *39*(3), 72–78. https://doi.org/10.1111/emip.12370

Gotwals, A. W., & Birmingham, D. (2016). Eliciting, identifying, interpreting, and responding to students' ideas: Teacher candidates' growth in formative assessment practices. *Research in Science Education*, *46*(3), 365–388.

Graeff, T. R. (2010). Strategic teaching for active learning. *Marketing Education Review*, *20*(3), 265–278. https://doi.org/10.2753/MER1052-8008200307

Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). Cognition and learning. In D. Berliner & R. Calfee (Eds.), *Handbook of educational psychology* (pp. 15–46).

Haertel, E. H. (1999). Performance assessment and education reform. *The Phi Delta Kappan*, *80*(9), 662–666.

Hamilton, L. (2003). Chapter 2: Assessment as a policy tool. *Review of Research in Education*, *27*(1), 25–68. https://doi.org/10.3102/0091732X027001025

Hanushek, E. A., & Raymond, M. E. (2004). The effect of school accountability systems on the level and distribution of student achievement. *Journal of the European Economic Association*, *2*(2–3), 406–415.

Hawaii State Department of Education. (2022). *Innovative assessment project*. https://www.hawaiipublicschools.org/TeachingAndLearning/Testing/InnovativeAssessmentProject/Pages/default.aspx

Helme, S., & Clarke, D. (2001). Identifying cognitive engagement in the mathematics classroom. *Mathematics Education Research Journal*, *13*(2), 133–153. https://doi.org/10.1007/BF03217103

Herman, J., & Linn, R. (2013). *On the road to assessing deeper learning: The status of Smarter Balanced and PARCC assessment consortia* (No. 823). Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Hurd, P. D. (2000). Active learning. *Journal of Pharmacy Teaching*, *7*(3–4), 29–47.

Ito, H., & Takeuchi, S. (2020). The demise of active learning even before its implementation? Instructors' understandings and application of this approach within Japanese higher education. *Education Inquiry*, 1–20. https://doi.org/10.1080/20004508.2020.1860283

Jacobs, V. R., Franke, M. L., Carpenter, T. P., Levi, L., & Battey, D. (2007). Professional development focused on children's algebraic reasoning in elementary school. *Journal for Research in Mathematics Education*, *38*(3), 258–288.

Kazemi, E., & Stipek, D. (2001). Promoting conceptual thinking in four upper-elementary mathematics

classrooms. *Journal of Education*, *189*(1–2), 123–137.

Khattri, N., Kane, M., & Reeve, A. L. (1995). How performance assessments affect teaching and learning. *Educational Leadership*, *53*(3), 80–83.

Kieran, C., & Dreyfus, T. (1998). Collaborative versus individual problem solving: Entering another's universe of thought. In A. Olivier & K. Newstead (Eds.), *Proceedings of the Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 112–119).

Kim, K., Sharma, P., Land, S. M., & Furlong, K. P. (2013). Effects of active learning on enhancing student critical thinking in an undergraduate general science course. *Innovative Higher Education*, *38*(3), 223–235.

Kim, S. (2015). *An analysis of teacher question types in inquiry-based classroom and traditional classroom settings* [Doctor of Philosophy, University of Iowa]. https://doi.org/10.17077/etd.ah4k7h3s

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Kvam, P. H. (2000). The effect of active learning methods on student retention in engineering statistics. *The American Statistician*, *54*(2), 136–140.

Lampert, M. (2001). *Teaching problems and the problems of teaching*. Yale University Press.

Lane, S., & Stone, C. A. (2006). Performance assessment. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 387–431). Praeger Publishers.

Lee, Y., & Kinzie, M. B. (2012). Teacher question and student response with regard to cognition and language use. *Instructional Science*, *40*(6), 857–874.

Leikin, R., & Zaslavsky, O. (1997). Facilitating student interactions in mathematics in a cooperative learning setting. *Journal for Research in Mathematics Education*, *28*(3), 331–354.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, *20*(8), 15–21.

Lubienski, S. T. (2002). Research, reform, and equity in U.S. mathematics education. *Mathematical Thinking and Learning*, *4*(2–3), 103–125. https://doi.org/10.1207/S15327833MTL04023_2

Machemer, P. L., & Crawford, P. (2007). Student perceptions of active learning in a large cross-disciplinary classroom. *Active Learning in Higher Education*, *8*(1), 9–30. https://doi.org/10.1177/1469787407074008

Maki, P. (2002). Using multiple assessment methods to explore student learning and development inside and outside of the classroom. *NASPA's NetResults*.

Marion, S. (2019, December 12). Can educational assessment improve teaching?: What would it take to design an assessment system around a goal of improving teachers' knowledge and skills? *CenterLine*. https://www.nciea.org/blog/educational-assessment/can-educational-assessment-improve-teaching

Matusov, E., & Smith, M. P. (2007). Teaching imaginary children: University students' narratives about their Latino practicum children. *Teaching and Teacher Education*, *23*(5), 705–729.

McCarthy, J. P., & Anderson, L. (2000). Active learning techniques versus traditional teaching styles: Two experiments from history and political science. *Innovative Higher Education*, *24*(4), 279–294.

McKeachie, W. J., & Kulik, J. A. (1975). Effective college teaching. *Review of Research in Education*, *3*(1), 165–209.

McLaughlin, M. W. (1990). The Rand Change Agent Study revisited: Macro perspectives and micro realities. *Educational Researcher*, *19*(9), 11–16.

McNeill, K. L., & Pimentel, D. S. (2010). Scientific discourse in three urban classrooms: The role of the teacher in engaging high school students in argumentation. *Science Education*, *94*(2), 203–229.

Mehan, H. (1979). *Learning lessons: Social organization in the classroom*. Harvard University Press.

National Academies of Sciences, Engineering, and Medicine. (2018). *How people learn II: Learners, contexts, and cultures* (p. 24783). National Academies Press. https://doi.org/10.17226/24783

National Board for Professional Teaching Standards. (2002). *What teachers should know and be able to do*.

National Council on Education Standards and Testing. (1992). *Raising standards for American education: A report to Congress, the Secretary of Education, the National Education Goals Panel, and the American people*.

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment* (p. 10019). National Academies Press. https://doi.org/10.17226/10019

New Hampshire Department of Education. (2022). *Performance assessment of competency education*. https://www.education.nh.gov/who-we-are/division-of-learner-support/bureau-of-instructional-support/performance-assessment-for-competency-education

Niemi, H. (2002). Active learning—A cultural change needed in teacher education and schools. *Teaching and Teacher Education*, *18*(7), 763–780. https://doi.org/10.1016/S0742-051X(02)00042-2

O'Grady, A., Simmie, G. M., & Kennedy, T. (2014). Why change to active learning? Pre-service and in-service science teachers' perceptions. *European Journal of Teacher Education*, *37*(1), 35–50. https://doi.org/10.1080/02619768.2013.845163

Oliveira, A. W. (2010). Improving teacher questioning in science inquiry discussions through professional development. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *47*(4), 422–453.

Oros, A. L. (2007). Let's debate: Active learning encourages student participation and critical thinking. *Journal of Political Science Education*, *3*(3), 293–311.

Parke, C. S., & Lane, S. (2008). Examining alignment between state performance assessment and mathematics classroom activities. *The Journal of Educational Research*, *101*(3), 132–147. https://doi.org/10.3200/JOER.101.3.132-147

Pedulla, J. J., Abrams, L. M., Madaus, G. F., Russell, M. K., Ramos, M. A., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers.*

Polikoff, M. S., & Porter, A. C. (2014). Instructional alignment as a measure of teaching quality. *Educational Evaluation and Policy Analysis*, *36*(4), 399–416. https://doi.org/10.3102/0162373714531851

Raney, D. (2003). Whose Authority?: Learning and Active Resistance. *College Teaching*, *51*(3), 86–93.

Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement and instruction* (pp. 37–75). Springer.

Rosebery, A., Warren, B., Ballenger, C., & Ogonowski, M. (2005). The generative potential of students' everyday knowledge in learning science. In T. Romberg, T. Carpenter, & F. Dremock (Eds.), *Understanding mathematics and science matters* (pp. 55–80). Erlbaum Mahwah, NJ.

Rozencwajg, P. (2003). Metacognitive factors in scientific problem-solving strategies. *European Journal of Psychology of Education*, *18*(3), 281–294.

Ruk, J. (2020). A conceptualized framework for assessing factors that maintain and lower cognitive demand during task enactment. In A. I. Sacristán, J. C. Cortés-Zavala, & P. M. Ruiz-Arias (Eds.), *Mathematics Education Across Cultures: Proceedings of the 42nd Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 2351–2352). PME-NA. https://doi.org/10.51272/pmena.42.2020-398

Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 77–84.

Schoenfeld, A. H. (2016). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics (Reprint). *Journal of Education*, *196*(2), 1–38.

Scott, P. (1998). Teacher talk and meaning making in science classrooms: A Vygotskian analysis and review. *Studies in Science Education*, *32*(1), 45–80. https://doi.org/10.1080/03057269808560127

Sfard, A., & Kieran, C. (2001). Cognition as communication: Rethinking learning-by-talking through multi-faceted analysis of students' mathematical interactions. *Mind, Culture, and Activity*, *8*(1), 42–76.

Shaughnessy, M., DeFino, R., Pfaff, E., & Blunk, M. (2020). I think I made a mistake: How do prospective teachers elicit the thinking of a student who has made a mistake? *Journal of Mathematics Teacher Education*. https://doi.org/10.1007/s10857-020-09461-5

Shepard, L. A. (2021). Ambitious Teaching and Equitable Assessment. *American Educator*, *45*(3), 28–48.

Silver, E., & Stein, M. K. (1996). The QUASAR project: The "revolution of the possible" in mathematics instructional reform in urban middle schools. *Urban Education*, *30*(4), 476–521.

Simmons, W., & Resnick, L. (1993). Assessment as the catalyst of school reform. *Educational Leadership*, *50*(5), 11–15.

Smith, M. S. (2000). Balancing old and new: An experienced middle school teacher's learning in the context of mathematics instructional reform. *The Elementary School Journal*, *100*(4), 351–375.

Smith, M. S., Bill, V., & Hughes, E. K. (2008). Thinking through a lesson: Successfully implementing high-level tasks. *Mathematics Teaching in the Middle School*, *14*(3), 132–138.

Stanford, P. (2003). Multiple intelligence for every classroom. *Intervention in School and Clinic*, *39*(2), 80–85.

Stecher, B. M. (2010). *Performance assessment in an era of standards-based educational accountability*. Stanford Center for Opportunity Policy in Education.

Stecher, B. M., & Mitchell, K. J. (1995). *Portfolio-driven reform: Vermont teachers' understanding of mathematical problem solving and related changes in classroom practice* (No. 652302011–001) [Data set]. American Psychological Association. https://doi.org/10.1037/e652302011-001

Stefanou, C., Stolk, J. D., Prince, M., Chen, J. C., & Lord, S. M. (2013). Self-regulation and autonomy in problem-and project-based learning environments. *Active Learning in Higher Education*, *14*(2), 109–122.

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal*, *33*(2), 455–488.

Stein, M. K., & 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), 50–80.

Supovitz, J. (2009). Can high stakes testing leverage educational improvement? Prospects from the last decade of testing and accountability reform. *Journal of Educational Change*, *10*(2–3), 211–227. https://doi.org/10.1007/s10833-009-9105-2

Sykes, G., & Wilson, S. M. (2016). Can policy (re)form instruction? In D. H. Gitomer & C. A. Bell (Eds.), *Handbook of Research on Teaching* (Fifth, pp. 851–916). American Educational Research Association. https://doi.org/10.3102/978-0-935302-48-6_13

Tallman, M. A., Carlson, M. P., Bressoud, D. M., & Pearson, M. (2016). A characterization of Calculus I final exams in U.S. colleges and universities. *International Journal of Research in Undergraduate Mathematics Education*, *2*(1), 105–133. https://doi.org/10.1007/s40753-015-0023-9

Tekkumru-Kisa, M., Stein, M. K., & Doyle, W. (2020). Theory and research on tasks revisited: Task as a context for students' thinking in the era of ambitious reforms in mathematics and science. *Educational Researcher*, *49*(8), 606–617. https://doi.org/10.3102/0013189X20932480

Tino, C. (2020). An integrative interpretation of personal and contextual factors of students' resistance to active learning and teaching strategies. *Andragoška Spoznanja*, *26*(2), 59–74. https://doi.org/10.4312/as.26.2.59-74

Tolman, A. O., & Kremling, J. (Eds.). (2017). *Why students resist learning: A practical model for understanding and helping students*. Stylus Publishing.

Townsend, M. A., Moore, D. W., Tuck, B. F., & Wilton, K. M. (1998). Self-concept and anxiety in university students studying social science statistics within a co-operative learning structure. *Educational Psychology*, *18*(1), 41–54.

Trognon, A. (1993). How does the process of interaction work when two interlocutors try to resolve a logical problem? *Cognition and Instruction*, *11*(3–4), 325–345.

Tung, R., & Stazesky, P. (2010). *Including performance assessments in accountability systems: A review of scale-up efforts*. Center for Collaborative Education.

Vygotsky, L. S. (1978). Interaction between learning and development. In M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (Eds.), *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press.

Wade, R. C. (1994). Teacher education students' views on class discussion: Implications for fostering critical reflection. *Teaching and Teacher Education*, *10*(2), 231–243. https://doi.org/10.1016/0742-051X(94)90015-9

Watkins, C., Carnell, E., & Lodge, C. (2007). *Effective learning in classrooms*. Paul Chapman Educational Publishing.

Webb, N. M., Nemer, K. M., & Ing, M. (2006). Small-group reflections: Parallels between teacher discourse and student behavior in peer-directed groups. *The Journal of the Learning Sciences*, *15*(1), 63–119.

Wiggins, G. (1992). Creating tests worth taking. *Educational Leadership*, *49*(8), 26–33.

Yackel, E., Cobb, P., & Wood, T. (1991). Small-group interactions as a source of learning opportunities in second-grade mathematics. *Journal for Research in Mathematics Education*, *22*(5), 390–408.

Yosso, T. J. (2005). Whose culture has capital? A critical race theory discussion of community cultural wealth. *Race Ethnicity and Education*, *8*(1), 69–91.

Citation:
Wellberg, S., & Evans, C. (2022). Assumptions underlying performance assessment Reforms intended to improve instructional practices: A research-based framework. *Practical Assessment, Research, & Evaluation*, 27(23). Available online: https://scholarworks.umass.edu/pare/vol27/iss1/23/

Corresponding Author:
Sarah Wellberg
University of Colorado Boulder
Boulder, Colorado, USA

Email: sarah.wellberg [at] colorado.edu