Distance Education

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cdie20

Reimagining online grading, assessment, and testing using situated cognition

Daniel Hickey & Tripp Harris

To cite this article: Daniel Hickey & Tripp Harris (2021) Reimagining online grading, assessment, and testing using situated cognition, Distance Education, 42:2, 290-309, DOI: 10.1080/01587919.2021.1911627

To link to this article: https://doi.org/10.1080/01587919.2021.1911627

Published online: 06 May 2021.

DISTANCE EDUCATION
2021, VOL. 42, NO. 2, 290–309
https://doi.org/10.1080/01587919.2021.1911627

Routledge
Taylor & Francis Group

ARTICLE

Reimagining online grading, assessment, and testing using situated cognition

Daniel Hickey and Tripp Harris

Counseling and Educational Psychology, Indiana University, Bloomington, IN, United States of America

ABSTRACT

Increased online learning is helping many appreciate that online grading, formative assessment, and summative testing can cause instructor burnout and leave little time for more productive instructor interactions. We reimagined grading, assessment, and testing in an extended program of design-based research using situative theory to refine online courses in secondary, undergraduate, graduate, and technical contexts. This research minimized private instructor-student interactions (including grading and private formative feedback) while maximizing public interactions. We present 10 assessment design principles, including a new principle concerning diversity and equity. We assume that these principles will be new to many readers and counter-intuitive to some. These principles focus on assessment functions (rather than ostensible purposes) and align learning across increasingly formal levels. We argue that doing so can maximize formative and transformative assessment functions, position students as authors, rather than consumers, reposition minoritized students to empower them, and increase validity and credibility of evidence.

ARTICLE HISTORY
Received 18 January 2021
Accepted 9 March 2021

KEYWORDS
grading; assessment; testing; online learning; situated cognition

The massive shift to online learning in 2020 helped many appreciate that grading, formative assessment, and summative testing have failed to keep up with other changes in educational technology. Many students and educators learned that grading students' contributions to online forums made these clumsy discussions (Thomas, 2002) even less productive for learning (e.g., Peterson, 2020). The difficulty of administering conventional closed-book tests left many educators scrambling for solutions. Some who heeded the many calls for alternative assessments (e.g., Harris, 2020) were overwhelmed by the need to provide detailed feedback and warrants for any deductions, leaving little time to interact with students elsewhere. Others who trusted students to abide by honor codes were sorely disappointed (e.g., Supiano, 2020). Still others discovered that digital proctors are expensive and intrusive (e.g., Flaherty, 2020). Students still find their way to unethical commercial websites (Chin, 2020). More educators now appreciate how the stress and time associated with providing individualized formative feedback and grading student work can contribute to burnout among online educators (e.g., Openo, 2020). Arguably, these developments have led many educators to ask the question that Conrad and Openo (2018) asked in the preface of *Assessment Strategies for Online Learning: Engagement and Authenticity*: "Did traditional forms of assessment *ever* serve us well?" (p. xi, emphasis in

CONTACT Daniel Hickey dthickey@indiana.edu

the original). We certainly believe that the answer is "no." But we also believe that the alternative approaches that are now being widely advocated are potentially problematic and unsustainable in many settings.

This paper presents examples and design principles for a comprehensive approach to grading, assessment, and testing that emerged across a decade of iterative design-based refinements of online courses, consisting of courses that were asynchronous (i.e., self-paced) and semi-synchronous (i.e., cohorts working towards regular deadlines); secondary, undergraduate, graduate, and technical; and credentialed, open, and hybrid. As will be elaborated, this research embraced the theory of learning known as situated cognition, as that theory was embodied in the work of Greeno (1991, 1998) and colleagues (Greeno et al., 1996; Greeno et al., 1993). This contrasts with most contemporary considerations of assessment that embrace cognitive-constructivist perspectives (e.g., Pellegrino et al., 2001) or socio-constructivist perspectives (e.g., Conrad & Openo, 2018). Our initial research has convinced us that this approach can help address the challenges in the previous paragraph regarding online grading, assessment, and testing by reducing the amount of private interaction that instructors have with individual students, maximizing the more efficient public instructor interaction with entire classes, aligning learning across increasingly formal levels of assessments, and increasing the validity and credibility of all assessment outcomes for students, educators, designers, and researchers.

We assume that many readers are familiar with situated learning, Vygotsky's (1978) zone of proximal development, and Dewey's (e.g., 1902/1990) vision of progressive education, and that some are using these perspectives in their teaching and/or research. Nonetheless, we assume that some of this article's suggestions about grading, assessment, and testing will be new to many readers and counterintuitive to some. We believe that the lack of consideration of situative theory in assessment is partly the result of continuing tensions between expository associationist approaches and inquiry-oriented constructivist approaches. In assessment, this tension is manifested in enduring debates over traditional multiple-choice tests and alternative performance and portfolio assessments. We discuss these debates below in the context of our argument that a situative approach offers a way to resolve some of these tensions by pragmatically drawing from both. We further assume that most readers are familiar with the three presences (cognitive, social, and teaching) that make up the influential socio-constructivist community of inquiry framework (CoI; Garrison et al., 2010) and associated approaches to assessment (i.e., Conrad & Openo, 2018). But we share the concerns of skeptics and critics (Maddrell et al., 2017; Rourke & Kanuka, 2009) that scores on the widely used CoI survey have never been convincingly connected to actual (as opposed to perceived) learning outcomes.

A brief summary of situated cognition

Situative theory is a distinctive strand of the broader class of sociocultural theories that gained prominence in the 1990s. With roots in Vygotsky's early Soviet psychology (e.g., 1934/1962), sociocultural theories assume that knowledge is fundamentally connected to the social and cultural context where it is used. Some assessment theorists have employed sociocultural perspectives, primarily to critique existing assessment and testing practices (e.g., Gipps, 1999; Moss, 2005; Shepard, 2000; Torrance, 2012). Sociocultural critiques of

assessment typically argue that individual assessments provide an imperfect and partial representation of individual knowledge while largely ignoring the more social and cultural aspects of knowledge. For example, Torrance argued that "*all* assessment is formative, for student dispositions and self-identities as learners as well as knowledge and understanding, but not necessarily in a positive way" (p. 325, emphasis in the original).

Situative theory is rooted in research carried out from 1986 to 2000 at the Institute for Research on Learning in Palo Alto, CA (IRL, 2020). The IRL was launched by Seely Brown and Greeno, and other influential collaborators including Goldman, Lave, McDermott, Pea, and Wenger. The IRL's commitment to ethnographic studies of cognition and learning resulted in groundbreaking contributions that have since gained wide recognition. Notable early contributions included Lave and Wenger's (1991) theories of communities of practice and legitimate peripheral participation, Brown et al.'s (1989) ideas regarding cognitive apprenticeship, and Greeno's (1991) characterization of learning using Gibson's (1986) theory of constraints and affordances.

A particularly important early contribution from situative research for assessment was Greeno et al.'s (1993) new characterization of the transfer of learning. Assuming that most educational assessments are intended to capture evidence of transfer of learning (from a learning environment) to some subsequent environment (the transfer environment), transfer is essential to understanding and improving assessment. The prevailing cognitive models characterized transfer using a transportation metaphor whereby the learner carries knowledge from the learning environment to the transfer environment. At that time, the prevailing debate was whether the learner carried (i.e., transferred) relatively specific associations (e.g., Singley & Anderson, 1989) or more general schema (e.g., Reed, 1993) from the learning environment to the transfer environment.

In contrast, Greeno et al.'s (1993) model characterized transfer of learning in terms of becoming attuned to the invariant properties of the learning environment. The invariant properties are the ones that are expected to be found in other environments. For learning to transfer, the learning environment needed to feature some of the properties of the transfer environment. From this perspective, failure to learn transferable knowledge can be characterized as a failure to attune learners to the invariant properties of the learning environment. In many school contexts, this likely occurs when students are only attuned to the variant properties of the learning environment. The variant properties, by definition, are unlikely to be present in most conceivable transfer environments. As introduced in Hickey and Pellegrino (2005) and elaborated across the design principles below, this situative theory of transfer has implications for the assessment of learning.

Four levels of online grading, assessment, and testing

This new approach to online grading, assessment, and testing is rooted in four programs of externally funded design-based research that used situative theory to refine formative and summative assessments in multimedia, videogame-based and new media learning environments (Hickey et al., 2013; Hickey et al., 2009; Hickey et al., 2012; Hickey & Zuiker, 2012). Central to this research was the goal of maximizing scores on performance assessments and achievement tests while minimizing what the validity theorist Messick (1995) called "construct-irrelevant easiness" and what most others call "teaching to the test." A related goal in this work was minimizing what Messick called "construct-irrelevant variance." This is when

factors besides what a particular assessment aims to assess introduce variance in student performance. This "noise" makes it difficult to reach strong conclusions about whether iterative refinements have improved student learning outcomes. Minimizing construct-irrelevant easiness and variance was important in this research because learning was iteratively aligned across multiple levels of increasingly formal assessments.

When this earlier program of multimedia assessment research moved into online courses, including both conventional courses (Hickey & Rehak, 2013) and an externally funded open course (Hickey & Uttamchandani, 2017), a generalizable approach emerged for situative grading, assessment, and testing, organized around engagement strategies based on expansive framing (Engle et al., 2012). This article presents this approach to grading, assessment, and testing in the form of ten design principles that aim to address the challenges presented in the introduction by maximizing public instructor and peer-to-peer interaction while minimizing private instructor-student interaction. Like the previous multimedia assessment research, this approach aligns learning across four increasingly formal levels. These four increasingly formal levels are as follows:

Immediate-level ungraded assessment of online discourse via instructor comments

This first level of assessment consists of instructor feedback in the form of comments directly on students' work or contributions to online discussions (e.g., in the form of comments on peer work or posts to discussion forums). These assessments are immediate level because they are so embedded in the learning itself. Such assessment is so informal that some might not consider it assessment at all. Indeed, we worry that labeling these practices as assessments might lead readers to assume that we mean feedback that is evaluative rather than participatory. As elaborated below, situative theories embrace a broad view of learning and therefore a broad view of assessment. We refined this assessment practice in asynchronous courses where students worked at their own pace with regular deadlines.

While there is no grading or formal evaluation involved, these assessments are crucial to fostering student engagement and learning. Because all students can see them and participate in the ensuing threads, they offer an efficient use of instructor time and expertise. Instructors should search for examples where students transfer in relevant prior experiences and make connections with ways that they might use what they are learning in the future. Particularly in early assignments, instructors should highlight examples of the most desired forms of student engagement. As courses get underway, instructor comments should position students as authors, rather than consumers, of disciplinary knowledge (as elaborated below). A particularly efficient strategy is providing relatively extensive feedback to students who post first (who are often the most experienced and ambitious) and encouraging all students to examine and contribute to the comment thread.

As part of this assessment practice, we encourage instructors to search for ideal places to embed more advanced content and ideas within students' work. Relatively advanced concepts (that might overwhelm less experienced learners if they were included in the assignment) can be introduced as useful instructor comments to help other students understand advanced concepts.

Close-level graded assessment of engagement via informal reflections

The second level of assessments in this approach is also informal. After completing each assignment, students respond briefly to several prompts about their prior engagement in the assignment, responses that do not assess expertise or knowledge but are used to assign grades or points for each assignment. They can be public (e.g., on a completed gPortfolio or ePortfolio) or private (submitted via the learning management system [LMS]). In most of our courses, points for engagement comprise the majority of students' grades (typically 60%–70%). These assessments are close-level because they are so closely connected to and contextualized by the curricular routines where the targeted learning occurred.

These reflections are intended to shape engagement proleptically (Cole, 1995), whereby anticipation of the future shapes the present. The assumption is that when students anticipate that they will have to reflect on specific aspects of their engagement, they are more inclined to engage that way after completing the first reflection. For example, in several courses, students complete each assignment by responding to more specific versions of the following reflection prompts:

- Contextual engagement. How well suited was your course goal for this assignment?
- Collaborative engagement. What did you learn from other students and what did others learn from you?
- Conceptual engagement. When you completed the self-assessment, what concepts did you struggle with? Are you are still confused about them?
- Cultural engagement. What are the implications of the ideas and arguments in this assignment for fostering equitable and inclusive education?
- Consequential engagement. How will you use what you learned in this assignment in the future?

To reiterate, close-level assessments do not involve assigning points or grades for the corresponding student artifacts, annotations, or comments for mastery or expertise regarding course content. Therefore, their reflections can be graded very efficiently. Students are awarded full points if they complete the assignment's required elements and their reflections are coherent. Grading typically involves (a) first reviewing student annotations, artifacts, and/or discussions for completeness, (b) adding one or more additional comments as needed (e.g., for students who posted at the last minute), (c) reviewing reflections for coherence, and (d) then giving brief private feedback in the gradebook (e.g., "great job, see my new comment"). If the instructor has been sufficiently engaged in the immediate-level assessment, it can take under a minute to grade each assignment.

Across multiple courses, we have observed that grading reflections becomes laborious only when instructors are compelled to explain why points were deducted for incomplete assignments (e.g., missing elements, failure to annotate or comment) and/or incoherent reflections (i.e., inflated or fraudulent reflections in light of engagement). But such incompleteness and incoherence are usually obvious, undeniable, and rare. While this observation calls for more careful documentation, it seems to us that this is very different than grading artifacts and contributions for evidence of mastery or expertise, where doing so in typical course contexts can require painstaking warrants for every deduction; even so, some students may argue over the deductions. As elaborated below, we assume that

a larger array of assessments yields student credibility toward grades that can minimize this potentially stressful and likely unproductive form of private instructor-student interaction. Time-consuming and stressful, we believe that such interactions can contribute to instructor burnout.

Proximal formative self-assessments

In some of our online courses, students complete a formative self-assessment after they complete each assignment (but before they complete the reflections). These consist of open-ended items covering most or all major topics in the assignment (usually about 5–10 items). These assessments are semiformal in that they present known-answer questions that are decontextualized relative to the assignment. But these assessments are proximal and curriculum-oriented in that they assess only knowledge that was presented in the assignment. They present that knowledge differently from how it was presented in the corresponding assignment, allowing students to self-assess whether they understand each concept in the abstract. Ideally then, students are positioned to use formative feedback from the LMS quiz or test engine (or perhaps Google Forms) to maximize that understanding accordingly.

Sometimes these items are drawn from (or based on) the textbook publisher's item bank, while other times they are created anew. Regardless, each item is accompanied by a correct answer which explains why it is correct and explains away likely incorrect or incomplete answers. The instructions tell students that they (a) should first try to provide an answer from memory, (b) should only then search the text or other resources for additional information as needed, (c) should submit something before seeing the expert answer, and (d) should carefully read the expert answer and make sure they understand it.

We acknowledge that most characterizations of formative assessment assume human feedback (e.g., Black & Wiliam, 2010). The assumption here is that known-answer questions and corrective feedback have relatively modest formative potential beyond what these self-assessments can accomplish. Put differently, in the absence of additional instruction, we assume there is little more to learn from known-answer items than the answer to the item and some cognitive residue of a more abstract representation of the underlying concept that might facilitate subsequent transfer. While this assumption certainly calls for systematic inquiry, our confidence here is bolstered by the assumption that precious instructor time and experience are more effective when used to provide public discursive feedback in immediate-level assessment. Similarly, we recognize that online instructors have long explored peer feedback in formative and summative assessment contexts (Yang & Tsai, 2010). But a situative consideration of typical peer assessments involving known-answer items suggests insufficient formative potential to justify the infrastructure and time required to support it. As it regards both formative and peer assessment, we acknowledge Sadler's (1989) suggestion regarding the need for fuzzy criteria; we agree with Torrance (2012) that this is frequently overlooked. To the extent that this is true, we assume that the infrastructure and time associated with typical formative and peer assessment can be more productively dedicated to informal immediate-level peer assessment in comment posts and threads.

As elaborated below, a situative approach assumes that these proximal assessments simultaneously serve a formative function for subsequent achievement, while serving a summative function for the conceptual understanding that each student took away from the assignment. If these assessments are sufficiently aligned to adjacent assessments, students should be prepared

and motivated to learn from them. Regardless, instructors should occasionally review the answers students are submitting for completeness and refine assessments accordingly to maintain and enhance engagement.

Automated distal summative achievement tests

Many (but not all) of the courses using our approach include automated time-limited multiple-choice tests, typically at the end of longer modules of four or five weekly assignments. Like conventional achievement tests, these assessments are distal and standards-oriented in that they are aligned to targeted standards and include items that require far transfer because they concern topics that were not fully (or even directly) covered in the curriculum or self-assessments.

Of course, many schools and universities require students to sit for conventional closed-book on-campus tests while others employ online proctors (Flaherty, 2020). Online proctors can be effective (Hylton et al., 2016), but they can also be expensive, intrusive, disruptive, and/or thwarted (Cluskey et al., 2011; Flaherty, 2020). We argue that some basic test features and designs can provide valid evidence of achievement for many (but not all) educational contexts while requiring little or no instructor time:

- Include best-answer items whose correct response cannot be located using Google or commercial homework help sites; at minimum, any correct answer items should be structured so that all four or five responses need to be searched. All candidate items should be searched to ensure that they are not already compromised and should be checked occasionally for compromise. In particular, items and answers should not appear directly in Google or at the bottom of the search page under "Searches related to." Readers should note that websites like Chegg (https://www.chegg.com/) sometimes contain entire textbook item banks (see Feldstein, 2021).
- Items can be drawn from textbook item banks and can include items from other textbooks in the same area, although the risk of compromise is greater, particularly with popular texts. Regardless of the source, the items should never directly present course content in the manner it was presented in course readings or other resources.
- Students are typically given 1–2 min per item. This is not enough time for students to search for the correct answer. Rather than requiring dubious honor-code promises to not search Google or readings, we assume that students will search as needed and that doing so can also offer modestly formative functions.
- The difficulty of items and time allotted is refined so that few if any students get a perfect score, the mean score is around 80%, and that students who engage weakly and/or skip the formative assessment perform poorly.
- Items are shown one at a time (to make it hard to print them out) and students are only shown their overall score (and not the correct answer to each item). Item analysis tools in the LMS are used to ensure that items discriminate appropriately and reveal whether items have become compromised.

Generally speaking, such tests should be worth enough points to reward and motivate engagement but not so much that they motivate elaborate cheating strategies; for example, have a summative test for each of three modules, with each worth 10% of the overall grade.

Ten situative design principles for online grading, assessment, and testing

The nature and function of each of the four levels of assessment above evolved as we adapted, studied, and refined the original multi-level assessment framework (i.e., Hickey & Zuiker, 2012) over a decade of design-based research in online courses (summarized in Hickey et al., 2020). This supported the continual revision and expansion of the original design principles, along with an appreciation of the divergence of these principles from prevailing associationist and constructivist approaches to assessment and the need for an additional principle to specifically address issues of diversity, equity, and cultural relevance. Following is a summary of our current design principles.

Embrace situative reconciliation over aggregative reconciliation.

The first principle concerns the use of different forms of individual activity and the reconciliation of individual and social activity. Cognitive approaches reconcile these differences by assigning them different levels of aggregation. Doing so assigns information processing to a lower level of human activity, assigns sense-making and problem-solving to a higher level of human activity, and uses aggregated characterizations of individuals' activities to characterize social activity (Greeno, 1998). This approach to reconciliation is quite widely embraced, often tacitly (Greeno et al., 1996, p. 40).

Situative theory can be distinguished from the broader class of sociocultural theory by its insistence that knowledge resides primarily in the sociocultural context and only secondarily in the minds and actions of individuals. Greeno's (1998) situative synthesis advances an approach to reconciliation that assumes that knowledge represented by the way individuals process information and solve problems is actually a secondary representation of that primarily sociocultural knowledge. This makes it possible to characterize the cognitive associations and/or structures that define how humans process information, solve problems, and make sense of the world as special cases of socially situated activity.

While this principle has proven difficult to convey to broad audiences, it is presented first because it is fundamental to the principles that follow. The situative synthesis makes it possible to consider the entire range of assessment practices from the same theoretical lens. This makes it possible to characterize the knowledge that is captured by individual formative and summative assessments as residue of each student's prior participation in more social learning practices. More specifically, the broad view of knowledge in situative theories offers a single theoretically and pragmatically coherent lens for using the four different forms of assessment introduced above—specifically, using a single theoretical lens to inform and study the following:

- using classroom discourse to assess the design and enactment of curriculum
- using reflections to assess prior engagement in curricular routines
- using curriculum-oriented classroom assessments to assess understanding of targeted concepts
- using standards-oriented tests to assess the impact of an entire course on student achievement or mastery of the broader domain.

In this way, situative assessment goes beyond a sociocultural critique of prevailing assessment and testing practices. Rather, it reframes the entire range of assessment practices as engaged participation (Greeno, 1998). Doing so transcends the enduring debate among cognitive scientists and assessment scholars over whether knowledge is constructed or acquired. The situative synthesis does so by reframing typical classroom assessment as a peculiar form of disciplinary discourse and reframing achievement tests as a potentially bizarre form of discourse—while acknowledging the unique and sometimes necessary role of both in the broader educational enterprise (Hickey, 2015). This in turn offers new ways of thinking about how such practices relate to one another, revealing new roles for familiar practices. While this familiarity facilitates transformation, doing so requires reconsideration of a widely held assumption about those practices, as described next.

Focus on assessment functions rather than purposes.

Most educators are familiar with formative assessment (for learning), summative assessment (of learning), and evaluative assessment (of programs). Our second design principle rejects the cognitivist focus on assessment purposes and the corresponding recommendation that a given assessment not be used for multiple purposes. The central concern is typically that summative purposes undermine formative purposes. This argument was influentially advanced in a report titled "Knowing What Students Know: The Science and Design of Educational Assessment" from the US National Research Council (NRC, 2001). Rather, we draw on situative theories of assessment (Gee, 2003) to focus on assessment functions instead of purposes (Hickey & Pellegrino, 2005). This makes it possible to assign complementary formative and summative functions to assessment practices at each level. Doing so while embracing the broader view of learning associated with situative theories allows each assessment to function as a formative assessment at one level and a summative assessment at the next level. Our second principle means that:

- immediate-level assessment of classroom discourse is a summative assessment of instructors' design of curricular routines but a formative assessment of students' engagement
- close-level assessments are summative for students' prior engagement but formative for their current understanding
- proximal assessments are summative for students' understanding but formative for their subsequent achievement
- distal tests are summative for students' achievement but formative for the overall design of the course (and the impact of that course on the achievement of future students).

In our experience, this idea that the same assessment can serve a formative function for one aspect of learning and a summative function for another aspect of that same learning is counterintuitive for educators and assessment scholars who are rooted in cognitive perspectives of learning. As embodied in the NRC (2001) and Ruiz-Primo et al. (2002), most assessment experts embrace a fundamentally cognitive model of knowing and learning and view all assessment functions through that lens.

In this way, our approach embraces a similar characterization of learning as that of Mislevy (2018) but reaches a rather different conclusion about assessment practice. As a caveat, our approach sidesteps naturalistic cognitive science arguments about what is really happening inside each learner's mind when they are completing the assessment. Of course, the NRC (2001) acknowledged that assessment results are only an estimate of what an individual knows and can do. But the primary title of the report (i.e., "Knowing What Students Know") implies that results are estimates of something that is ultimately knowable and is "based on samples of knowledge and performance drawn from a much larger universe of everything a person knows and can do" (p. 36). Instead, we focus on more pragmatic learning sciences arguments about whether our interpretation of assessment evidence results in new solutions to enduring educational challenges. To the extent that readers agree that the challenges delineated in the introduction of our paper are indeed significant problems for online education, we believe our approach does just that.

Synergize multiple complementary types of interaction.

Our third design principle is rooted in Hall and Rubin's (2013) distinction between different types of interactions among students, peers, and instructors. Their three types of interaction and our examples from online learning are:

- Public interactions directed at all participants in a course. Examples include publicly posted student artifacts and announcements posted by the instructor. In some cases (e.g., using an LMS) these might be visible only to students enrolled in the class and the instructors. In other cases (e.g., using Google Docs), these might be visible to anyone with the link.
- Local interactions carried out in public but directed to an individual. Typical online examples include threaded comments posted on student artifacts by instructors or peers. Discussion forum posts might be local or public, depending on the context in which the post was made.
- Private interactions between individuals. The most obvious example is instructor feedback to students in the course gradebook, for which strict privacy controls are required. Other examples include messages between individuals inside an LMS or invited private chat sessions. Because such interactions involve a single student rather than multiple students, we assume that such interactions represent an inefficient use of instructor expertise and time.

To Hall and Rubin's three types of interactions, we add a fourth type of interaction to convey our suggestions about conventional distal multiple-choice tests: discrete interactions are also private but should be downplayed in importance. Labeling distal tests as discrete highlights our suggestion that they should be positioned as relatively unimportant features of the course, used primarily to validate the success with which the course was designed and enacted, and used for a relatively small proportion of each student's grade that is sufficient to motivate engagement in assignments and formative assessments.

Gains on discrete distal-level tests can undoubtedly provide convincing estimates of impact on subsequent external high-stakes achievement tests targeting the same standards. However, we contend that as long as most of the problems on such tests are

different from problems in the learning environment and formative assessments (i.e., truly distal), these scores are estimates of transfer to the entire range of subsequent settings, including any educational, personal, and professional settings that feature the invariant properties of the curriculum. While sharing Bransford and Schwartz's (1999) enthusiasm for framing transfer as preparation for future learning (PFL), we disagree with their assertion that multiple-choice tests or any sequestered problem-solving assessments (including performance assessments) cannot capture PFL. On the contrary, the validity threat from construct-irrelevant easiness with performance assessments leads us to suggest that that multiple-choice distal tests can be more valid evidence of PFL, so long as they are indeed distal to the curriculum. Put differently, we agree with Bransford and Schwartz that "the ideal assessment from a PFL perspective is to directly explore people's abilities to learn new information and relate their learning to previous experiences" (p. 69). But we believe that such assessment is most appropriate at the immediate and close levels; the effectiveness of this assessment and associated formative feedback can then be objectively evaluated for impact at the proximal and distal levels and refined accordingly.

Thanks to Google Forms and modern LMSs, distal tests can be automatically and securely administered, which makes them ideal in many online course settings. Most importantly for addressing the problems outlined in the introduction, making the best use of each of these four kinds of interactions has significant implications for maximizing instructor and student efficiency while minimizing arguments over grades and instructor burnout. We return to this assumption below.

Use increasingly formal assessments that capture longer timescales of learning.

The fourth design principle is that disciplinary knowledge across levels should be represented at increasingly lengthy timescales (Lemke, 2000). Put differently, increasingly formal assessments capture learning that accrues over longer periods of time:

- Immediate-level assessment is very informal and represents learning across timescales of minutes–hours (e.g., while working on curricular routines).
- Close-level assessment is informal and represents learning across hours–days (e.g., when completing assignments).
- Proximal assessments are semi-formal and capture learning across days–weeks (e.g., the understanding gained in the curriculum).
- Distal tests are formal and capture learning across months–years (e.g., the achievement gained across courses).

This transformation of learning across levels is central to the alignment of learning across levels without compromising the validity of evidence at each level. For example, our proximal assessments are open-ended constructed-response items; students must transfer the learning from formative feedback on such items if they are to correctly answer distal multiple-choice achievement test items targeting the same concept (so long as the proximal assessment does not directly present the actual associations that make up any corresponding distal items). Students must transfer that learning even further to answer distal achievement test items on concepts that were not included in the proximal assessment.

Embrace transformative functions and systemic validity.

In contrast to reliability (a property of a given assessment), validity is a property of evidence considering an argument that assessment results are being used to make. The most obvious arguments concern whether an individual knows something. But these can also be arguments about whether an individual learned something (i.e., they may have already known it before a class) or whether a particular formative assessment is more effective than some other curricular routine. Assessment theorists traditionally focused on evidential validity associated with arguments about summative functions (e.g., Messick, 1988). Increased use of alternative assessment formats in the 1990s drew attention to formative functions and associated concerns with consequential validity (e.g., Messick, 1994). Our fifth design principle concerns transformative functions (Shepard, 2000) represented by the way assessment practices transform the broader culture of classes, schools, and communities. These functions are associated with systemic validity (Frederiksen & Collins, 1989) and validity arguments about these transformations.

Our approach assumes that assessment practices always cause transformations, although such transformations are not always what educators and instructional designers intended. As Torrance (2012) pointed out, assessment practices often serve conformative functions whereby learners' knowledge conforms to relatively specific representations (e.g., on a performance assessment task as intended by many formative assessment practices). But assessments can also serve deformative functions whereby learners' knowledge is deformed by overly narrow assessment representations (e.g., by unwittingly aligning instruction too closely to multiple-choice tests).

While educators can value broader forms of learning in conventional classrooms, in classroom discussions, we contend that assessment plays an outsized role in online education in signaling to learners the value of different forms of learning. As such, extra attention is needed to ensure that assessments support the intended transformations while avoiding unintended negative transformations. One way to accomplish this is by avoiding known answer questions in the public space of the course. Rather these questions are "sequestered" in the private proximal assessments and discrete distal tests.

Position learners as accountable authors.

The sixth design principle concerns a primary goal of our broader approach: engagement and assessment are intended to position (Harré & Moghaddam, 2003) learners as authors and hold them accountable to the discursive norms of the course and discipline. Our efforts in this regard are shaped by Engle and Conant's (2002) four design principles for fostering productive disciplinary engagement (PDE): (1) problematize subject matter from the learner's perspective, (2) give learners authority to address such problems, (3) hold learners accountable to others and to shared disciplinary norms, and (4) provide students with relevant resources for accomplishing these goals. Our efforts are further informed by the design principles for expansive framing (Engle et al., 2012) whereby learners are insistently pushed to establish intercontextuality (Bloome et al., 2009) by finding connections between course topics and people, places, topics, and times beyond the assignment or course.

While our sixth design principle is primarily accomplished in student engagement routines, our grading, assessment, and testing practices are explicitly designed to support these routines. Specifically, engagement routines should avoid known-answer questions in the course's public space, minimizing conversation-killing, initiate-respond-evaluate (Cazden & Beck, 2003) discourse between students and instructors (and instructors' online assignments). This sets the stage for instructors' immediate-level assessment of students' discourse, which helps position students as authors and position both the instructor and students as audience for the insights authored by students. In practice, we often post LMS announcements with hyperlinks to highlight good examples of PDE and expansive framing and support other related forms of positioning. This is further reinforced by the close-level assessments, with the contextual, collaborative, and consequential reflections specifically intended to summatively assess authorial positions in prior engagement and proleptically encourage authorial positions in subsequent engagement.

Reposition minoritized learners for equitable engagement.

We recently extended the previous design principle in response to the burgeoning research in the learning sciences exploring the roles of power and privilege in the pursuit of educational equity. Influential work here includes books (Medin & Bang, 2014), special issues (Bang & Vossoughi, 2016), edited volumes (Esmonde & Booker, 2016), and numerous workshops and symposia. Specifically, we have extended our approach in response to Agarwal and Sengupta-Irving's (2019) critique of Engle and Conant's (2002) PDE design principles. It is important to note that Agarwal and Sengupta-Irving conceded that the four PDE principles may bolster participation among minoritized students, relative to more traditional curricula. Specifically, they agreed that (1) problematizing content through learners' personal perspectives fosters culturally meaningful explanation of that content, (2) allowing students the authority to explore those problems helps them actively construct knowledge, (3) holding learners accountable ensures that their justifications are open to critique, and (4) offering resources that are culturally relevant might draw the interests of diverse learners. However, Agarwal and Sengupta-Irving questioned the extent to which this will actually occur among minoritized students, given the growing body of evidence that minoritized students are routinely positioned out of disciplinary discourse by teachers and more advantaged peers (e.g., Anderson, 2009) and/or positioned as incompetent, lazy, or disruptive (e.g., Lambert, 2015).

Because of this, Agarwal and Sengupta-Irving (2019) argued that implementing the principles of PDE while failing to confront the dynamics of power and privilege limits the potential of PDE for supporting equitable learning opportunities. This is because (1) challenging culturally dominant ways of knowing through problematizing course content can lead to racialized tension, (2) supporting intellectual authority risks ignoring the potentially overwhelming power of social authority, (3) developing authority to share and justify one's ideas is easier than upholding accountability to critique and revise ideas, and (4) students from minoritized groups may have trouble connecting disciplinary concepts with racial and cultural meanings that pervade educational resources.

In response to these concerns, Agarwal and Sengupta-Irving (2019) extended the four PDE design principles with repositioning whereby teachers attend to issues of power and positioning arising in the classroom interactions, and reposition students perceived as low status or minoritized to support their participation. They introduced four new connective and productive disciplinary engagement principles for repositioning students from minoritized groups: (1) use sociopolitical uncertainties to help problematize disciplinary knowledge, (2) keep undue social authority in check, (3) uphold equitable forms of accountability, and (4) treat sociopolitical tensions as resources. Initially, we have found that these goals for student engagement can be effectively accomplished in the immediate-level assessment of classroom discourse, and that this repositioning can be accomplished informally in the existing cultural reflection. In brief, we agree with Neri et al. (2019) that:

> It is not even mildly controversial to suggest that in order to successfully teach nondominant students, all their teachers should possess practical pedagogies that effectively draw on the racial, ethnic, linguistic, and cultural assets that these learners bring to school. (p. 197)

Despite the common-sense appeal of asset-based responses to educational inequities and compelling evidence of effectiveness, Neri et al. (2019) characterized the adoption of CRE approaches as "sporadic and underwhelming." They attribute this to "teacher resistance" that "stems from limited understanding, and belief in the efficacy, of CRE and a lack of know-how needed to execute it" (p. 202).

A recent comparison of two sections of the same online course provides initial support for these ideas. Coding of student work and interactions revealed that deliberate instructor encouragement via immediate-level assessment of student discourse and the introduction of the cultural reflection described above led to a dramatic increase in the proportion of times that students used sociopolitical controversies to problematize (ostensibly nonpolitical) course content, inviting further interpretive and empirical exploration and scrutiny from colleagues who have more expertise in CRE. In short, it appears that immediate-level instructor assessment of student discourse (i.e., in their assignments and discussion of those assignments) is a promising way to reposition minoritized students, encourage all students to transfer in their own relevant cultural experiences, and encourage sociopolitical controversies; it further appears that including a cultural reflection in the close-level assessments is a promising way to support these goals.

Enhance validity of evidence for designers, evaluators, and researchers.

Our eighth design principle concerns the validity of evidence for carrying out iterative design-based refinements of curricula and then documenting the ultimate impact of those refinements on distal standards-oriented tests. This use of multiple assessment levels minimizes the aforementioned construct-irrelevant easiness (Messick, 1995) when formative assessments use problems that are too similar to the curriculum and when summative assessments use items that are too similar to formative assessments. Using problems that are too similar raises scores in ways that are unrelated to the underlying targeted knowledge. Our approach aims to support the validity of observations or scores at each assessment level by ensuring that the representations of disciplinary knowledge are sufficiently transformed from one level to the next.

Although assessments should not be overly similar across levels, we also suggest using assessments that are sufficiently aligned from one level to the next to ensure that the

scores across levels are valid evidence of formative impact of feedback at prior levels. This is most obvious when using distal test results to evaluate and revise course assignments. Such tests are highly sensitive to prior achievement, which is only partly due to learning in a particular assignment. The presence or absence of echoes across assessment levels helps distinguish between random variations in scores and systematic differences due to specific course features introduced in design-based refinements (Hickey & Zuiker, 2012). This is particularly the case when comparing classes that have fewer than 30 students because the realities of the central limit theorem allow outliers to easily undermine tests of statistical significance with smaller sample sizes. In these cases, larger statistically significant differences on proximal assessments (which are more sensitive to interventions) can remove doubt regarding improvements in distal outcomes that do not quite reach conventional levels of statistical significance (i.e., $p < .05$).

Enhance credibility of scores and efficiency for educators.

The ninth design principle extends the intent of the eighth principle to the credibility and efficacy of assessment evidence for educators. Many formative assessment practices can be exhausting for educators because they require so much individualized formative feedback (Bennett, 2011; Hickey, 2015). We contend that much of this burden is because instructors are attempting to accomplish three ambitious goals when providing formative feedback: The first goal is "correcting" each student's conceptual misunderstandings. This can be accomplished quite readily with well-designed formative assessments—as long as instructors accomplish the second goal of motivating learners to use their feedback to improve their understanding. Regardless, accomplishing the first goal is still very time-consuming for online instructors, and accomplishing the second goal can be difficult in all instructional formats.

A situative perspective introduces a third goal of using formative feedback and positions learners to support PDE introduced above. From this perspective, feedback should position learners so that their engagement is both disciplinary and supportive of more generalized learning than solving a specific problem or answering a specific question on the assessment. We contend that this is difficult to accomplish in typical formative assessment contexts because such assessments are focused on what learners know rather than how they gained that knowledge in prior course assignments. Although this assumption certainly calls for more careful experimental validation, we believe that formative self-assessments can accomplish the first goal of formative assessment while making minimal demands on instructors' time and that aligning such self-assessments to achievement tests helps accomplish the second goal. Along with the outcomes of the next principle, these practices should free up instructor time to more efficiently accomplish the third goal via immediate-level public assessment of student discourse.

We recognize that this principle contradicts prevailing beliefs about formative assessment and the value of individualized feedback. It is worth reiterating that our arguments are particularly directed at typical asynchronous instruction. Some may question our instructor efficiency argument, assuming that typical discussion forums can replace the discussions that occur in physical classrooms. Discussion forums can be remarkably incoherent (Thomas, 2002) and hard pressed to support the more efficient just-in-time whole-class educator feedback that can emerge quite naturally from synchronous classroom-based formative assessment. When coupled with educators' concerns over student

privacy and regulation regardings graded student work, online formative assessment often devolves into repeatedly providing the same feedback to multiple students.

Enhance credibility of assessments and grades for learners.

The final principle is that the use of increasingly formal assessments usually enhances the credibility of learners' assessment results as careful alignment helps students see how their engagement at one level impacts their success at the next level. This in turn should help learners regulate their engagement while increasing their trust of assessment results, which can reduce corrosive and unproductive arguments over grades. As outlined above, close engagement reflections and proximal self-assessments provide students with copious opportunities to reflect on and regulate their engagement with course content, making it possible for students with less prior knowledge but who are sufficiently motivated to score well on achievement tests and also helping overstretched students appreciate why their efforts did not warrant the highest scores on the distal tests (Nicol & MacFarlane-Dick, 2006).

Conclusions and next steps

In summary, we have advanced a situative approach to online grading and assessment by presenting examples of four different levels of assessment and ten principles that emerged from a decade of research across many online courses. Many online educational contexts using this approach will minimize private instructor-student interactions and maximize public interactions among instructors, students, and peers.

We believe that this approach can maximize the most productive forms of students' engagement, understanding and achievement of disciplinary concepts; and that this approach will help reduce the instructor burnout that results from the stressful hours required to provide individualized formative feedback, grade students' work, and perhaps argue with them over resulting grades. We believe that this approach will work best when used with engagement routines that embrace situative theories of learning. In our case, these routines build quite directly on the design principles for PDE (Engle & Conant, 2012) and expansive framing (Engle, et al., 2012), pushing students to find personally relevant connections with topics and times beyond the assignment or the course (Hickey et al., 2020). We look forward to collaborations that pair these assessment practices with the increasingly popular social annotation of online educational resources (Kalir, 2020) and the adaptation of this approach to portfolio-based or project-based approaches (Chang & Tseng, 2009; Savin-Baden & Wilkie, 2006) and to other conventional instructional routines.

Acknowledgements

We acknowledge the contributions to this research by Chris Andrews, Tara Kelly, Joshua Quick, Rebecca Itow, and Suraj Uttamchandani. We also acknowledge the contribution to these ideas from our collaborating instructors Jody Duncan, Christine Hitchcock, and Courtney Gaylord.

Disclosure statement

No potential conflict of interest was declared by the authors.

Notes on contributors

Daniel Hickey is a professor of learning sciences at Indiana University.

Tripp Harris is a PhD student in learning sciences at Indiana University.

ORCID

Daniel Hickey http://orcid.org/0000-0001-9146-5089

References

Agarwal, P., & Sengupta-Irving, T. (2019). Integrating power to advance the study of connective and productive disciplinary engagement in mathematics and science. *Cognition and Instruction*, 37(3), 349–366. https://doi.org/10.1080/07370008.2019.1624544

Anderson, K. T. (2009). Applying positioning theory to the analysis of classroom interactions: Mediating micro-identities, macro-kinds, and ideologies of knowing. *Linguistics and Education*, 20(4), 291–310. https://doi.org/10.1016/j.linged.2009.08.001

Bang, M., & Vossoughi, S. (2016). Participatory design research and educational justice: Studying learning and relations within social change making. *Cognition and Instruction*, 34(3), 173–193. https://doi.org/10.1080/07370008.2016.1181879

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

Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 92(1), 81–90. https://doi.org/10.1177/003172171009200119

Bloome, D., Beierle, M., Grigorenko, M., & Goldman, S. (2009). Learning over time: Uses of inter-contextuality, collective memories, and classroom chronotopes in the construction of learning opportunities in a ninth-grade language arts classroom. *Language and Education*, 23(4), 313–334. https://doi.org/10.1080/09500780902954257

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education*, 24(1), 61–100. https://doi.org/10.3102/0091732X024001061

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32–42. https://doi.org/10.3102/0013189X018001032

Cazden, C. B., & Beck, S. W. (2003). Classroom discourse. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of discourse processes* (pp. 165– 197). Erlbaum.

Chang, C. C., & Tseng, K. H. (2009). Use and performances of web-based portfolio assessment. *British Journal of Educational Technology*, 40(2), 358–370. https://doi.org/10.1111/j.1467-8535.2008.00885.x

Chin, M. (2020, April 29). Exam anxiety: How remote test proctoring is creeping students out. *The Verge*. https://www.theverge.com/2020/4/29/21232777/examity-remote-test-proctoring-online-class-education

Cluskey Jr, G. R., Ehlen, C. R., & Raiborn, M. H. (2011). Thwarting online exam cheating without proctor supervision. *Journal of Academic and Business Ethics*, 4(1), 1–7. https://www.aabri.com/manuscripts/11775.pdf

Cole, M. (1995). Culture and cognitive development: From cross-cultural research to creating systems of cultural mediation. *Culture & Psychology*, 1(1), 25–54. https://doi.org/10.1177/1354067X9511003

Conrad, D., & Openo, J. (2018). *Assessment strategies for online learning: Engagement and authenticity*. Athabasca University Press. https://doi.org/10.15215/aupress/9781771992329.01

Dewey, J. (1990). *The school and the society* and *The child and the curriculum*. University of Chicago Press. (Original work published 1902)

Engle, R. A., & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition and Instruction*, 20(4), 399–483. https://doi.org/10.1207/S1532690XCI2004_1

Engle, R. A., Lam, D. P., Meyer, X. S., & Nix, S. E. (2012). How does expansive framing promote transfer? Several proposed explanations and a research agenda for investigating them. *Educational Psychologist*, 47(3), 215–231. https://doi.org/10.1080/00461520.2012.695678

Esmonde, I., & Booker, A. N. (Eds.). (2016). *Power and privilege in the learning sciences: Critical and sociocultural theories of learning*. Taylor & Francis. https://doi.org/10.4324/9781315685762

Feldstein, M. (2021, February 6). The Chegg situation is worse than you think. *eLiterate*. https://eliterate.us/the-chegg-situation-is-worse-than-you-think/

Flaherty, C. (2020, May 11). Big proctor. *Inside Higher Ed*. https://www.insidehighered.com/news/2020/05/11/online-proctoring-surging-during-covid-19

Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, 18(9), 27–32. https://doi.org/10.3102/0013189X018009027

Garrison, D. R., Anderson, T., & Archer, W. (2010). The first decade of the community of inquiry framework: A retrospective. *The Internet and Higher Education*, 13(1–2), 5–9. https://doi.org/10.1016/j.iheduc.2009.10.003

Gee, J. P. (2003). Opportunity to learn: A language-based perspective on assessment. *Assessment in Education: Principles, Policy & Practice*, 10(1), 27–46. https://doi.org/10.1080/09695940301696

Gibson, J. J. (1986). *The ecological approach to visual perceptions*. Lawrence Erlbaum Associates.

Gipps, C. (1999). Socio-cultural aspects of assessment. *Review of Research in Education*, 24(1), 355–392. https://doi.org/10.3102/0091732X024001355

Greeno, J. G. (1991). Number sense as situated knowing in a conceptual domain. *Journal for Research in Mathematics Education*, 170–218. https://doi.org/10.5951/jresematheduc.22.3.0170

Greeno, J. G. (1998). The situativity of knowing, learning, and research. *American Psychologist, 53* (1), Article 5. https://doi.org/10.1037/0003-066X.53.1.5

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

Greeno, J. G., Moore, J. L., & Smith, D. R. (1993). *Transfer of situated learning*. In D. K. Detterman & R. J. Sternberg (Eds.), *Transfer on trial: Intelligence, cognition, and instruction* (pp. 99–167). Ablex.

Hall, R., & Rubin, A. (2013). There's five little notches in here: Dilemmas in teaching and learning the conventional structure of rate. In J. G. Greeno and S. V. Goldman (Eds.), *Thinking practices in mathematics and science learning* (pp. 199–246). Routledge.

Harré, R., & Moghaddam, F. M. (Eds.). (2003). *The self and others: Positioning individuals and groups in personal, political, and cultural contexts*. Greenwood Publishing Group.

Harris, D. (2020, April 29). Online education and authentic assessment. *Inside Higher Ed*. https://www.insidehighered.com/advice/2020/04/29/how-discourage-student-cheating-online-exams-opinion

Hickey, D. T. (2015). A situative response to the conundrum of formative assessment. *Assessment in Education: Principles, Policy & Practice*, 22(2), 202–223. https://doi.org/10.1080/0969594X.2015.1015404

Hickey, D. T., Chartrand, G. T., & Andrews, C. D. (2020). Expansive framing as a pragmatic theory for instructional design. *Educational Technology Research and Development*, 68 (2), 751–782. https://doi.org/10.1007/s11423-020-09759-4

Hickey, D. T., Honeyford, M. A., & McWilliams, J. C. (2013). Participatory assessment in a climate of accountability. In H. Jenkins, W. Kelly, K. Clinton, J. McWilliams, R. Pitts-Wiley, & E. Reilly (Eds), *Reading in a participatory culture: Remixing* Moby-Dick *in the English classroom* (pp. 169–184). Teachers College Press.

Hickey, D. T., Ingram-Goble, A. A., & Jameson, E. M. (2009). Designing assessments and assessing designs in virtual educational environments. *Journal of Science Education and Technology*, 18(2), 187–208. https://doi.org/10.1007/s10956-008-9143-1

Hickey, D. T., & Pellegrino, J. W. (2005). Theory, level, and function: Three dimensions for understanding transfer and student assessment. In J. P. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 251–293). Information Age Publishing.

Hickey, D. T., & Rehak, A. (2013). Wikifolios and participatory assessment for engagement, understanding, and achievement in online courses. *Journal of Educational Media and Hypermedia, 22*(4), 229–263. https://www.learntechlib.org/p/41508/

Hickey, D. T., Taasoobshirazi, G., & Cross, D. (2012). Assessment as learning: Enhancing discourse, understanding, and achievement in innovative science curricula. *Journal of Research in Science Teaching*, 49(10), 1240–1270. https://doi.org/10.1002/tea.21056

Hickey, D. T., & Uttamchandani, S. L. (2017). Beyond hype, hyperbole, myths, and paradoxes: Scaling up participatory learning in a big open online course. In L. Losh (Ed.), *MOOCs and their afterlife: Experiments in scale and access in higher education* (pp. 13–36). The University of Chicago Press. https://doi.org/10.7208/chicago/9780226469591.003.0002

Hickey, D. T., & Zuiker, S. J. (2012). Multilevel assessment for discourse, understanding, and achievement. *Journal of the Learning Sciences*, 21(4), 522–582. https://doi.org/10.1080/10508406.2011.652320

Hylton, K., Levy, Y., & Dringus, L. P. (2016). Utilizing webcam-based proctoring to deter misconduct in online exams. *Computers & Education*, 92, 53–63. https://doi.org/10.1016/j.compedu.2015.10.002

Institute for Research on Learning. (2020, December 13). In *Wikipedia*. https://en.wikipedia.org/wiki/Institute_for_Research_on_Learning

Kalir, J. H. (2020). Social annotation enabling collaboration for open learning. *Distance Education*, 41 (2), 245–260. https://doi.org/10.1080/01587919.2020.1757413

Lambert, R. (2015). Constructing and resisting disability in mathematics classrooms: A case study exploring the impact of different pedagogies. *Educational Studies in Mathematics*, 89(1), 1–18. https://doi.org/10.1007/s10649-014-9587-6

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press. https://doi.org/10.1007/s10649-014-9587-6

Lemke, J. L. (2000). Across the scales of time: Artifacts, activities, and meanings in ecosocial systems. *Mind, Culture, and Activity*, 7(4), 273–290. https://doi.org/10.1207/S15327884MCA0704_03

Maddrell, J. A., Morrison, G. R., & Watson, G. S. (2017). Presence and learning in a community of inquiry. *Distance Education*, 38(2), 245–258. https://doi.org/10.1080/01587919.2017.1322062

Medin, D. L., & Bang, M. (2014). *Who's asking? Native science, Western science, and science education*. MIT Press. https://mitpress.mit.edu/books/whos-asking

Messick, S. (1988). Validity. In R. L. Linn (Ed), *Educational measurement* (3rd ed., pp. 12–103). Macmillan. https://onlinelibrary.wiley.com/doi/pdfdirect/10.1002/j.2330-8516.1987.tb00244.x

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13–23. https://doi.org/10.3102/0013189X023002013

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50 (9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Mislevy, R. J. (2018). *Sociocognitive foundations of educational measurement*. Routledge.

Moss, P. A. (2005). Understanding the other/understanding ourselves: Toward a constructive dialogue about 'principles' in educational research. *Educational Theory*, 55(3), 263–283. https://doi.org/10.1111/j.1741-5446.2005.00003.x

Neri, R. C., Lozano, M., & Gomez, L. M. (2019). (Re)framing resistance to culturally relevant education as a multilevel learning problem. *Review of Research in Education*, 43(1), 197–226. https://doi.org/10.3102/0091732X18821120

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. https://doi.org/10.1080/03075070600572090

Openo, J. (2020). Education's response to the COVID-19 pandemic reveals online education's three enduring challenges. *Canadian Journal of Learning and Technology*, 46(2), 1–12. https://doi.org/10.21432/cjlt27981

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. National Academy Press. https://doi.org/10.17226/10019

Peterson, O. (2020, October 7). Discussion posts: Is there a point? *The Advance-Titan*. https://advancetitan.com/opinion/2020/10/07/discussion-posts-is-there-a-point

Reed, S. K. (1993). A schema-based theory of transfer. In D. K. Detterman & R. J. Sternberg (Eds.), *Transfer on trial: Intelligence, cognition, & instruction* (pp. 39–67). Ablex.

Rourke, L., & Kanuka, H. (2009). Learning in communities of inquiry: A review of the literature. *Journal of Distance Education*, 23(1), 19–48. http://www.ijede.ca/index.php/jde/article/download/474/815

Ruiz-Primo, M. A., Shavelson, R. J., Hamilton, L., & Klein, S. (2002). On the evaluation of systemic science education reform: Searching for instructional sensitivity. *Journal of Research in Science Teaching*, 39(5), 369–393. https://doi.org/10.1002/tea.10027

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144. https://doi.org/10.1007/BF00117714

Savin-Baden, M., & Wilkie, K. (2006). *Problem-based learning online*. McGraw-Hill.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4–14. https://doi.org/10.3102/0013189X029007004

Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill*. Harvard University Press.

Supiano, B. (2020, October 10). Students cheat. How much does it matter? *The Chronicle of Higher Education*. https://www.chronicle.com/article/students-cheat-how-much-does-it-matter?

Thomas, M. J. (2002). Learning within incoherent structures: The space of online discussion forums. *Journal of Computer Assisted Learning*, 18(3), 351–366. https://doi.org/10.1046/j.0266-4909.2002.03800.x

Torrance, H. (2012). Formative assessment at the crossroads: Conformative, deformative and transformative assessment. *Oxford Review of Education*, 38(3), 323–342. https://doi.org/10.1080/03054985.2012.689693

Vygotsky, L. S. (1962). *Thought and language* (E. Hanfmann & G. Vakar, Trans.). MIT Press. (Original work published 1934)

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. Jolm-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Yang, Y. F., & Tsai, C. C. (2010). Conceptions of and approaches to learning through online peer assessment. *Learning and Instruction*, 20(1), 72–83. https://doi.org/10.1016/j.learninstruc.2009.01.003