

UC Santa Barbara

UC Santa Barbara Previously Published Works

Title

Raising Critical Readers in the 21st Century: A Case of Assessing Fourth-Grade Reading Abilities and Practices

Permalink

<https://escholarship.org/uc/item/35k52008>

Journal

Literacy Research Theory Method and Practice, 71(1)

ISSN

2381-3369

Authors

Arya, Diana J
Sultana, Sabiha
Levine, Somer
[et al.](#)

Publication Date

2022-11-01

DOI

10.1177/23813377221117100

Peer reviewed

Raising Critical Readers in the 21st Century: A Case of Assessing Fourth-Grade Reading Abilities and Practices

Literacy Research: Theory,
Method, and Practice
1-19

© The Author(s) 2022




Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/23813377221117100

journals.sagepub.com/home/lrx



Diana J. Arya¹ , Sabiha Sultana¹,
Sommer Levine¹, Daniel Katz¹,
John Galisky¹, and Honeiah Karimi¹

Abstract

Grounded in the sociocultural nature of literacies and informed of the inherent biases in widely used, English-dominant reading assessments in U.S. schools, this case study traces the planning, development, and pilot administration ($n = 52$) of a culturally inclusive (i.e., participant informed), online reading assessment. The Critical Reading Assessment (CRA) is designed to gauge elementary students' comprehension and critical reasoning (i.e., identifying potential biases or instances of diversity, equity and inclusion) of digital, multimodal texts. Findings from our analysis of recorded pilot sessions with student participants, who are predominantly Spanish/English multilingual learners, suggest (a) the importance of transparency and feedback from multiple stakeholders in the assessment development process; (b) the potential affordance of multiple textual modalities for clarifying comprehension skills and abilities; (c) the potential negative consequences of using established, dominant-English reading tests for determining comprehension abilities; and (d) the need for greater opportunities to practice critical discussions (i.e., questions about perspectives,

¹Education, University of California Santa Barbara, Gevirtz Graduate School of Education, Santa Barbara, CA, USA

Corresponding Author:

Diana J. Arya, Education, University of California Santa Barbara, Gevirtz Graduate School of Education, Santa Barbara, CA 93106, USA.

Email: darya@education.ucsb.edu

representation, and other potential biases) about texts. Implications from this study highlight the need for supporting elementary students and their teachers in dialogic, critical reading practices of multimodal textual information.

Keywords

reading assessments, critical reading practices, elementary students, qualitative assessments

Introduction

Literacy development reflects an increasingly complex landscape of meaning-making and reasoning of various textual resources across sociocultural contexts. Literacy practices are shaped by and help shape an increasingly socioculturally diverse population tasked with navigating digitally expansive, multimodal (print, images, videos, podcasts, etc.) spaces. Yet reading assessments in the U.S. designed to gauge one's *readiness* in reading continue to reflect an antiquated goal of *digging out* main ideas for making sense of print-dominant texts (Arya et al., 2020a) and a decontextualized, monologic definition of literacy (e.g., Gee, 2018). Such assessments are used by the majority of U.S. schools to demonstrate evidence of student progress in English-dominant literacy abilities regardless of potential issues in applying the results to a culturally and linguistically diverse student population (Abedi et al., 2011; Bailey et al., 2007). With these realities in mind, we sought to develop a reading assessment that aligned with the agreed-upon obligations to our long-standing, 20-year partnership with *Ocean Elementary* (pseudonym) located on the central coast of California. Ocean Elementary is an urban public school that serves a predominantly Spanish-speaking community with Latinx/Chicanx roots (i.e., most parents and older relatives have immigrated from Mexico within the past decade). Ocean Elementary is a Title I school and attests that more than 95% of students in grades TK through sixth grade are eligible for free or reduced lunch. Our research team leads a community-based literacy initiative at Ocean Elementary that positions participating students as "co-learners" with undergraduate "college buddies" who attend a nearby Hispanic Serving Institution (HSI). Although the focus of the elementary school is on developing comprehension of English texts, the community-based literacy program is designed to foster the use of *all* linguistic and cultural resources through a critical reading approach, with discussion guides in both Spanish and English to facilitate small-group reading discussions (Arya & Meier, 2020).

The impetus for the development of a new reading assessment, one that elicits a critical eye on textual materials (Souto-Manning et al., 2019), emerged from conversations with school administration, the lead instructor, and teachers in grades 4–6 who expressed general frustration about the district-approved English reading test (STAR; Renaissance Learning, 2022) in terms of understanding particular strengths and needs of students to inform instruction. The fourth-grade teachers were particularly

curious about test reliability in determining reading comprehension levels. As such, we proposed the development of a new, qualitative assessment, situated as a one-on-one interaction between a student and a college buddy around a selected text that the student would read aloud and interpret through discussion. Working within the partnership constraints of English-dominant reading practices, we sought to develop an assessment that would make visible the ways in which participating fourth-grade students ($n = 46$) grapple with reading multimodal information in English that includes charts, videos, and images.

We have framed our exploration as a case study (Stake, 2000), which we believe was most appropriate in this initial phase of a larger longitudinal study that traces the theorizing, planning, development, and revision of the Critical Reading Assessment (CRA). To be clear, this is not a case study on the development of a bi/multilingual assessment; we acknowledge the importance of such work for fully understanding the reading abilities of bi/multilingual students. That being said, we believe that this study may be highly relevant to educators and scholars who work with schools and districts that serve multilingual populations and adhere to English-dominant benchmarks and assessments. The research question guiding this case study is: What theoretical and instructional insights were afforded from the development and administration of an assessment designed to gauge students' engagement, understanding, and critical reading of digital, multimodal texts? Findings from our case study may support similar efforts in developing more culturally inclusive, equitable literacy-related resources.

The *Reading Framework for the 2026 National Assessment of Educational Progress* (NAEP) reflects a new direction, emphasizing reading as a “complex process shaped by many factors” that involve engagement with “multimodal forms” in order to “extract, construct, integrate, critique, and apply meaning in activities across a range of social and cultural contexts” (National Assessment Governing Board, 2021, p. 5). This new national framework highlights the importance of reading assessments for gauging students' understanding and critical reasoning (i.e., discerning relevance, veracity, and ideological influences) of multimodal texts, including podcasts, videos, data representations, and the like. Coupled with this recent national charge is the declaration from the National Academy of Education (NAEd) in February 2021 that literacy assessments would be “as free as possible from racial and cultural biases” (p. 10). To the best of our knowledge, there are no existing assessments or resources available for assessing metacognitive engagement (i.e., making personal or experiential connections with textual media that include critical reasoning) with multimodal texts. Grounded in the understanding that there is no such thing as a socioculturally or ideologically neutral text (Street, 2014), we aimed to develop an approach for gauging students' ability to question and even push back on potential bias in texts, shifting the role of textual resources from an authoritative position to one of multiple potential voices and stances in need of careful, critical review (Jewitt, 2005; Yoon, 2020). Further, we aimed for this new assessment to be more culturally inclusive, noting the constraint of English as the medium of printed content. We

were guided by scholarship on assessment validity issues for multilingual learners (e.g., Ruiz-Primo et al., 2014; Solano-Flores, 2011) who emphasized the importance of understanding the particular knowledge, perspectives, interpretations of textual materials, and experiences of the student population, as well as institutional constraints from the very beginning of assessment development.

The impacts of the COVID-19 pandemic are yet to be fully understood, but it is clear from preliminary research that remote-only access has had negative impacts on our most vulnerable student populations (National Academy of Education, 2021). Reading assessments designed to support instructional decisions have been further compromised due to remote-only pandemic conditions. High-quality, online reading assessments that reflect the new NAEP Reading Framework (National Assessment Governing Board, 2021) are needed now more than ever and will continue to be a valuable asset in the foreseeable, digitally engaged future. Hence, we aimed to design a more culturally inclusive, online-accessible reading assessment designed to gauge students' comprehension and critical reading agency of multimodal texts.

Our work was grounded in theories that reflect a socioculturally situated nature of literacy (Gee, 2018; Street, 2014) and the importance of open-ended, qualitative approaches that elicit interactional engagement from young developing readers who should be encouraged and supported to use all cultural and linguistic resources that they bring to the classroom space (Arya & Maul, 2021; Rojas-Drummond, 2019). Acknowledging ongoing tensions related to what counts as key theoretical and empirical underpinnings within the area of reading studies, our study was built on previous work guided by sociocognitive scholarship that highlights the importance of (a) socio-historical contextual matters that relate to conceptual assertions featured in texts and (b) the active involvement of participating students that are representative of the target population (Arya & Maul, 2021). Such active involvement aligns with research on social-emotional learning and the importance of incorporating students' interests and experiential knowledge in academic contexts (e.g., Fetting et al., 2018). We view such efforts to incorporate relevant, engaging topics as intertwined with the call for raising critical readers of various textual media. Students benefit from explicit support in making textual inferences, identifying potential authorial biases, and questioning stated and implied assumptions in texts that may represent or affirm sociocultural inequities (Arya, 2022; McClung, 2018; McLaughlin & DeVoogd, 2004; Souto-Manning et al., 2019).

Finally, our work incorporates the fact that multimodal textual media such as graphical memes, podcasts, and data representations are key components of communication processes that should be included as part of the textual landscape in school curricular materials and assessments (Arya et al., 2020a; Kress, 2003). As such, we aimed to develop a reading assessment that involves the triangulation of multimodal texts that developing readers are encountering more frequently. In a complementary fashion, our study involved graphical anchors of materials, practices, and findings of our work, hence maximizing the communicative power of multimodality.

The Importance of Studying the Particular, in Reading

The long-standing tensions about the *what* and the *how* of reading comprehension development and ways to assess it stems (at least in part) from a lack of consensus on the theoretical underpinnings of such development (Perfetti & Stafura, 2014). Assessments of reading comprehension are hence a reflection of this lack of consensus (Arya et al., 2020b; Valencia & Pearson, 1986) and as such, warrant closer inspection of the ways that reading assessments are used for making school-based decisions. In California, districts have the freedom to choose reading assessments for informing programmatic and instructional decisions.

The district associated with the present study has adopted the STAR reading test (Renaissance Learning, 2022) for tracking progress in reading development as well as determining readiness for exiting targeted enrichment programs. As such, the STAR reading test is an important marker of academic achievement for participating students in our study. The STAR assessment task used by the participating school involves cloze task items (i.e., passages with missing words) that require the student to read directions and prompts, interpret what is being asked of them, and select the correct answer from a list of options that include distractors crafted to increase the level of challenge. The structural features and linguistic demands of such assessment items have been noted to be problematic, inadequately assessing students' reading abilities across cultural, linguistic, and intellectual groups (Abedi et al., 2011; Bailey et al., 2007). Such assessment items are modern versions of earlier paper-pencil items that have been noted by literacy scholars as "insensitive" for gaining useful information about reading abilities and "virtually useless for making decisions in a school setting" (Valencia & Pearson, 1986, p. 3). Developers of the STAR reading assessment assert that this assessment "is not intended to be used as a 'high-stakes' test" but could be used to "predict performance on high-stakes tests" (Renaissance Learning, 2022, p. 1). Understandably, district leaders associated with this study found the expressed high-stakes association to be a reason to use the STAR for making programmatic decisions to best prepare students for annual testing.

Despite Renaissance's assertions to the contrary, the two fourth-grade teachers associated with this case study expressed that they were unable to use the STAR reading scores to inform instructional practices beyond selecting Lexile-informed leveled texts for their students, the majority (83%, 38 students) of whom being multilingual students with Latinx cultural roots. The practice of navigating forced-choice exams is particularly challenging for multilingual learners who must decode and translate the content of a text, decode and translate the content of a question and all the possible answers, and identify correct responses among distractors while also remembering the text they had originally read. Literacy scholars have long noted the importance of measures matching objectives (i.e., construct relevance) and the negative impacts of forced-item tests in meeting such objectives (e.g., Ferrara & DeMauro, 2006; Hsu & Nitko, 1983). Multiple-choice items of segmented textual

information may be less useful for understanding how students—particularly multilingual students—comprehend, connect, and critically engage with textual information.

The fourth-grade teachers and school leadership (i.e., the principal and instructional coordinator) were interested in learning what information can be gained from the CRA and to what extent the performance on this qualitative assessment aligns with STAR reading performance. We believed that these expressed interests from our partners fit within the aforementioned research question about the particular theoretical and instructional insights gained from developing and administering the CRA. The additional exploration into potential differences between observed performance on the CRA and the STAR for each participating student may clarify the usefulness of the STAR for our partnering teachers and school leaders.

This study focused on a particular case of students within two fourth-grade classrooms ($n = 46$) that participated in a pilot administration of one-on-one qualitative reading assessments (CRA) designed to reflect a natural reading discussion process of multimodal informational texts across disciplinary contexts. Although specific to the participating school context, the expressed practices, experiences, viewpoints, and findings from administered instruments and materials may be instrumental for gaining theoretical insights into critical elementary reading engagement within the digital world; case study research has been noted as an appropriate and potentially powerful way for gaining such insights (Stake, 2000).

Methodology

Study Context

This case study is a two-phase exploration of our growing archive of data sources (fieldnotes, recorded exchanges, produced program materials, and email correspondence) involving multiple layers of stakeholders that included a team of six young participants (ages 7–13) and their parents, eight undergraduate research assistants, two graduate student coordinators, and a faculty program leader. We took up Stake's (2000) account of case study methodology as a starting place for clarifying the complexities involved in how the CRA came into being. As such, we identified and collected all recorded activities, artifacts, and surrounding context factors (prior relationship with the participating school, cultural and linguistic resources, school and community resources, etc.) associated with CRA development and the different layers of actors represented (lead researchers, research assistants, participating teachers, school leadership, young informants, and parental informants). The initial development of CRA materials occurred within our university-housed literacy center designed to foster reading, writing, and other forms of communication predominantly in English through a student-centered, multilingual approach. The university is recognized as a Minority Serving Institution (MSI) with more than 30% of the student population with Latinx/Chicanx roots. University students facilitated instructional sessions at the literacy center and were encouraged to use all linguistic resources (including

Spanish) during their time with the young students. Although textual materials are predominantly English, the center's library also contains multilingual texts, primarily in Spanish, which is the most frequently used language within the surrounding community. Programmatic and research initiatives associated with this center were grounded on a three-dimensional framework of agency, co-learning, and belonging; hence, centering literacy practices that are equitable and involve topics and interests relevant to the local community (Arya, 2022).

The pilot administration of drafted assessment materials took place at Ocean Elementary, which has a 20-year partnership with our university. Recent program efforts from this partnership involved a community-based literacy program that centers on elementary students' interests and curiosities about their surrounding environment and ways to preserve natural spaces. Guided by the same framework as the literacy center, this partnership program was designed for students in fourth through sixth grade to celebrate and incorporate all cultural, linguistic, and disciplinary expertise participating students bring to weekly reading and digital storytelling sessions within their classrooms. Activities associated with this program involved small-group reading discussions and creative projects (e.g., creating videos of new knowledge about locally relevant issues) that were facilitated by undergraduate students from the university. With the exception of three weeks that needed to be virtual due to pandemic conditions, all sessions took place at the elementary school.

As was the case for previous studies, families of participating students in fourth grade (the selected grade level by school leadership) were notified of the present study through an informational session hosted at the school and through messages sent through school-delivered emails. All information was provided in both Spanish and English to ensure accessibility. More than 90% of families of children attending Ocean Elementary are Latinx and speak at least some amount of Spanish at home. Recruited participants represent the majority population of program members and were eligible for free and reduced lunch. With the permission of the partnering school, we invited parents and other caregivers to an informational session at the school prior to the recruitment and administration of the piloted assessment. Attendees were encouraged to ask questions about the project, which we described as a qualitative reading assessment in progress, intended to help teachers understand students' abilities both to make sense of and think critically about interdisciplinary English texts. We also explained the multimodal nature of this reading assessment and showed examples of drafted texts. Questions most commonly reflected curiosities about the STAR and concerns about its usefulness for gauging reading abilities in English. Nearly as common were questions about the ways in which the CRA differed from the STAR and which would be a better tool for gauging English reading ability. The lead researcher (first author) explained that many educational researchers share concerns about the usefulness of the STAR for making educational decisions for students and that we aimed to see what differences exist between students' STAR performance and their performance on the CRA. Sample texts were used to demonstrate how each text included printed English

with an embedded video clip and/or data representation that complemented content relayed in print form (see Figure 1 below).

Members of the research development team include two graduate student coordinators (identified respectively as South West Asian and North America--SWANA and Latina) and a faculty director (SWANA/white). The eight undergraduate assistants represented a range of cultural diversity (one European Chicax American, one South Asian American, two Latinx American, three Chicax American, and one

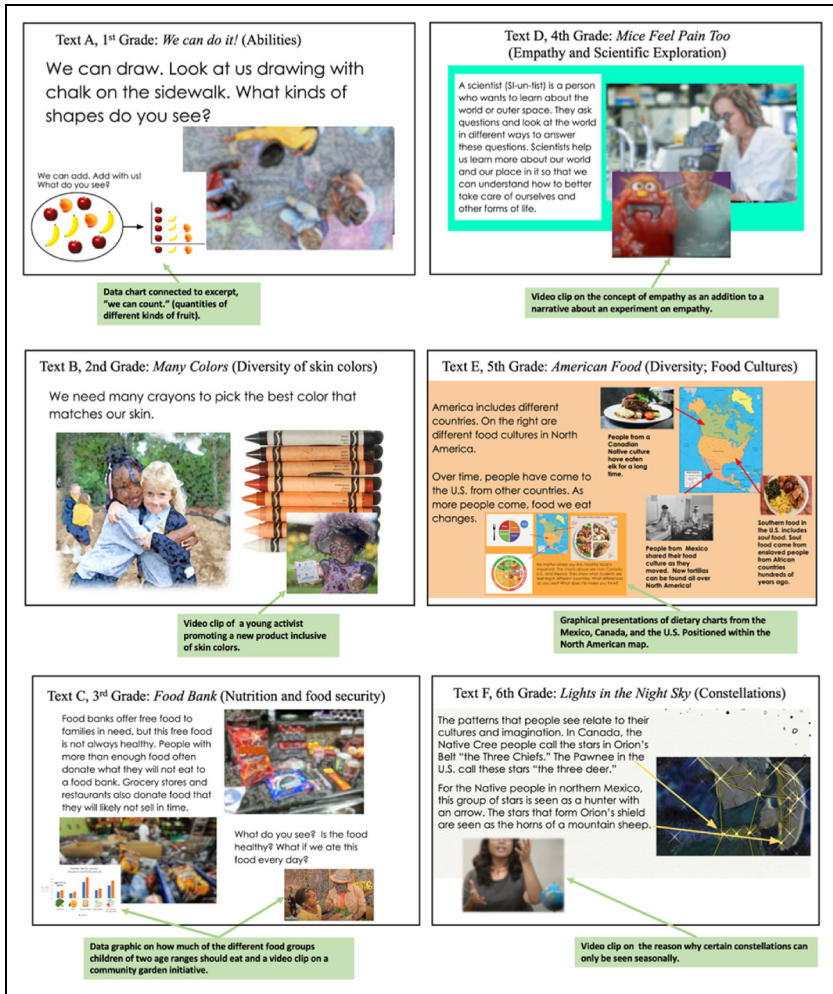


Figure 1. Excerpted text examples in order of complexity, from A to F. Note that multimodal texts described are presented on a separate page/slide.

East Asian) and disciplinary expertise (marine sciences, sociology, history, etc.). Hence, the cultural diversity represented in this study spans all layers of actors including the research team. Mindful of our respective cultural and linguistic resources, we aimed to construct the CRA guide with the lens of affirming and leveraging students' linguistic resources for engaging in reading discussions (García & Kleifgen, 2020). Hence, decision-making throughout this study stemmed from the shared commitment to the development of a reading assessment that would allow students to express their understandings, confusions, interests, and concerns using all available forms of knowledge and expertise. As such, members of the research team took an active, critical stance in reviewing primary sources for adapting texts and associated questions during planning discussions in order to identify and address potential barriers to gaining insights from students. Key factors raised during researching and planning meetings included the presence (and absence) of perspectives represented in primary sources, and cultural representations of featured individuals (Souto-Manning et al., 2019).

Over a five-month period, the university multilevel development team engaged in a series of sessions to support the drafting and revision of assessment tools and materials appropriate for TK-8 readers representing a multicultural, multilingual population. This development reflected a nonlinear, iterative process involving critical analysis of existing formative reading assessments as well as multiple revisions of textual media, associated questions, and the conceptual frameworks for guiding the analysis of student responses.

Participants

Forty-six fourth graders (identified as 22 girls and 24 boys) with parent consent participated in this study. All participants attended Ocean Elementary, which serves young children ranging from preschool age (TK) through sixth grade; more than 95% of enrolled students were eligible for free breakfast and lunch and more than 80% speak at least some Spanish at home. Most participants (38; 83%) have a Latinx/Chicanx background and nearly half of this number (19) indicated that they speak mainly Spanish at home. Other ethnicities represented include four white students, one Black student, and one biracial student (SWANA/white). The two remaining students had undisclosed ethnicities. Thirty-one participants (67%) have been identified as needing language enrichment support as identified by *English Language Proficiency Assessments for California* (ELPAC) with 11 (35%) of these students reclassified into the regular education program. Ten participants (22%) received special education services and five (50%) of these students also received language enrichment support. The diversity represented by our young participants reflects the state-wide trend toward an increasingly diverse student population; more than 50% of K-12 students in California public schools identify as Latinx (California Department of Education, 2021).

CRA Development

Gathered research of widely used and new qualitative assessment tools developed by literacy experts (i.e., Duke, 2020; Leslie & Caldwell, 2017), field notes, correspondence, and planning sessions were collectively used in developing the CRA. The progression of linguistic and syntactic complexity reflected in this collection was guided by previous investigations on the impacts of various elements of textual complexity on the accessibility of information (Arya et al., 2011). Further, we followed earlier work involving text development (i.e., Arya & Maul, 2012), by submitting all drafted texts through the Coh-Metrix software program (McNamara et al., 2005) to determine readability and textual coherence (conceptual similarity of words within a text) of each leveled text. The Coh-Metrix program relies on the same algorithms that produce reading levels used by the STAR test and library systems across the country. Although we acknowledged the problematic nature of using computer-driven formulas for determining text accessibility, we also recognized that our participating school leaders and teachers were supported by district leadership to be dependent on such algorithms for determining growth in reading ability.

Further, our development of this progression of texts followed an iterative effort of (a) conducting think-aloud sessions with young children from our aforementioned six-member panel of young students and (b) debriefing and revising drafted texts according to multiple dimensions of textual complexity as well as sociocultural relevance (Arya & Maul, 2021; Arya et al., 2020b). As suggested earlier, words and concepts included in each leveled text of the CRA aligned with readability and coherence information that allowed for comparisons with STAR test scores (see Online Appendix 1). Embedded graphical representations (e.g., bar charts and geographical maps) and video clips contained information that was complementary to the topic of each text without duplicating the textual information. For example, the sixth-grade text titled *Lights in the Night Sky* (Figure 1) included printed textual descriptions about various cultural interpretations of particular constellations as well as a video clip available on YouTube and designed for young viewers about the reason why people can only see certain constellations during a particular season. Together, the two modalities of text provide a more comprehensive overview of the notion of constellations.

The final draft of this new assessment includes the following components: (1) general student information (e.g., grade, educational status), (2) selected text information (including modalities represented), (3) prior knowledge of the topic, (4) running record analysis for texts read aloud, (5) text-dependent understanding of key information (direct recall and inferential), and (6) critical reasoning/questioning of textual information (identifying biases, missing information, etc.). This assessment was designed to be displayed digitally via electronic device in order to ensure seamless transitions in reading the printed text along with embedded video clips. Further, the digital sources allow for virtual assessment administration, which was a need inspired by the pandemic.

Questions associated with each of the piloted texts included inquiries about key information explicitly or implicitly represented in the presented text. In addition to requests to recall such information, participants were asked metacognitive questions that pertain to what may be interesting, surprising, and inclusive (or exclusive) about the presented text. We intended that such metacognitive questioning be open-ended and encompassing various possible responses from students.

Assessment Administration

Training for administering the CRA followed the general principles of the community-based literacy program; assessors were trained to approach the assessment session as an opportunity to learn about and from students. As such, all instructions were framed with this sentiment (*this is an opportunity for me to learn from you about how you read and what you think is interesting, or if you have any suggestions about how to make the text we will be reading better*). Approaching these sessions through the lens of a colearner helped to establish a comfortable dialogue around the reading and prevent well-documented effects of evaluative framing within academic contexts (e.g., Steele & Aronson, 1995). Assessors elicited information from students beyond general textual understanding, including interests related to the topic and textual information, opinions of what was identified as most important, and views about what might be missing or potentially unfair.

The general structure of administration is roughly aligned with what is typically done with running records or other qualitative reading assessments that are administered individually to student participants (e.g., Leslie & Caldwell, 2017). Hence, we trained 46 undergraduates (one per participant) to administer this assessment through the students' electronic tablets. Each individual assessment began with a brief introduction to familiarize the student with the undergraduate assessor and to establish a comfortable rapport prior to reading. The participating fourth-graders were then guided through a list of words beginning with higher-frequency, easier lexical items (e.g., *cat*, *see*, etc.) and ending with lower-frequency, more complex items (e.g., *establishment*). Performance on this task helped to determine the level of text to be used with the student.

Once a text is selected, students are asked up to two questions to gauge prior knowledge about the topic of interest (e.g., *what do you think the word "empathy" means?*) in order to clarify what the student gained from the textual information. The student is then prompted to read aloud the selected text, which is presented in digital form (i.e., Google slides presented on the student's electronic tablet). Following the reading/viewing of textual information, students were asked a series of questions that elicited the student's general comprehension of textual information (direct recall and inference making). The development of questions aligned with theoretical foundations on textual comprehension and associated empirical evidence (Arya & Maul, 2012; Tierney, 2018). Each assessment ended at the conclusion of student responses to the aforementioned metacognitive questions.

Scoring and Analysis

Recorded sessions and notes for each assessed student (over 15 h of audio footage and approximately 50 pages of notes summarized) were analyzed by three separate researchers that included an undergraduate research assistant, a graduate student with expertise in literacy instruction, and a literacy scholar with strong expertise in literacy assessment and text development. Three analysts separately reviewed each recorded session to determine the depth of understanding and metacognitive engagement based on previous research on comprehension assessment (Arya & Maul, 2012; Arya et al., 2020b) and the use of Wilson's (2004) building blocks framework for building transparent maps of variations in metacognitive engagement. Such mapping efforts (i.e., clarifying a progression or depth of understanding and reasoning) are not to suggest that our constructs of interest (comprehension of textual media and metacognitive engagement) are linear in nature; given the complexities involved in reading and the ways that we synthesize and integrate new information, there are arguably countless maps that could be drawn to trace variations of understanding and perspective-taking. For this study, we focused our observations of comprehension on the extent to which key ideas were fully identified and synthesized similar to previous work (Arya & Maul, 2012). Following the initial construct mapping, we engaged in the second block of text and item design, which involved an iterative process of reading and discussing drafted texts and questions with our six-member panel of young students. Responses from this panel informed our measurement model of the CRA, hence bringing us back to the theoretical maps initially constructed for further exploration and refinement.

We focused our observations of metacognitive engagement informed on aesthetic interactions with multimodal textual media (e.g., Browne et al., 2021), hence exploring the extent to which perspectives reflected a critical stance (e.g., noting potential socio-cultural (mis)representations or other forms of bias). We did not look for predetermined phrasing or examples in such aesthetic responses because we expected a wide range of views based on the varied cultural and experiential knowledge of participating students (Street, 2014). There were no discrepancies between the three separate ratings for each student participant, with full consensus to reassess eight participants who were given a text representing a level that was deemed too high based on responses. Recordings of these reassessments were reviewed by each of the three raters in a similar fashion, resulting in full consensus on the reading abilities and metacognitive engagement of all participants. We organized such ratings by the grade level of the selected text and adjusted labeling of levels typically used by qualitative reading assessments—*emerging*, *developing*, and *independent*. Ratings of metacognitive engagement involved levels that reflected the metaphor of oceanic depth; responses that reflected no such engagement were viewed as *on the shore* while responses that reflected views about equity or bias issues were viewed as *diving deeper* towards criticality. More medial levels included *skimming the shore* (e.g., expressing views on the appeal of images) and *breaking the surface* (e.g., connecting textual information with prior experiences), respectively.

Quantitative Analysis. We regressed the outcome variable—the CRA comprehension score—on the STAR reading scores captured a week earlier. Since both scores use the same scale (grade-level equivalency), we were able to investigate the relationship between performance on the CRA and the performance on the STAR. We hypothesized that based on previous research, students may not be able to fully demonstrate their reading ability on computer-administered assessments like the STAR, and as such, students may score significantly lower on less proximal assessments. However, we also anticipated no significant results given the relatively small sample size of this study.

Case Findings

We were fortunate that none of our participants had previously viewed the publicly accessed video clips embedded within our texts. Further, with the exception of a few instances, most of the content presented across the texts was novel enough to serve as an opportunity to learn something new from the text.

All participants demonstrated at least an emerging level of understanding of the selected grade-level text. Observed levels represent a wide spread of abilities with the largest concentration at the fourth grade, *developing* level (see Online Appendix 2). Similarly, levels of metacognitive engagement also varied by observed depth, with the greatest concentration (17 out of 46 responses) *on the shore*. Although there was a general concentration of participants reading at higher comprehension levels that also demonstrated the deepest (critical) level of metacognitive engagement, not all participants demonstrated such congruity (see Figure 2 below).

Analysis of responses to CRA questions seemed to capture students' ability to understand and synthesize textual information as well as their ability to think metacognitively about such information. Student participants consistently expressed genuine interest in the topics and concepts featured in the CRA texts. Responses to metacognitive questions revealed curiosities about our night sky (*I would add more about why we can't see the same constellations all year*), surprising insights about the history of food cultures (*I thought that this country stole the food from Mexico*), and concerns about the ethics of animal testing (*Scientists should leave the mice alone; the author should find another way to explain [empathy]*). While all students expressed enjoyment in reading with their college buddies, none expressed negative views about the STAR beyond one student stating that *it's boring*.

Results from our comparative analysis revealed a disparity between STAR scores and observed CRA performance that was statistically significant ($p < .05$) with a relatively low standard error (.35). The average difference between STAR and CRA performance scores (both based on a grade-level scale) is approximately eight months with a range of 8–14 months. Simply put, the STAR test seemed to underestimate students' reading abilities as demonstrated by the CRA. The sample size ($n = 46$) precluded us from statistically determining any such disparities for subgroups, namely,

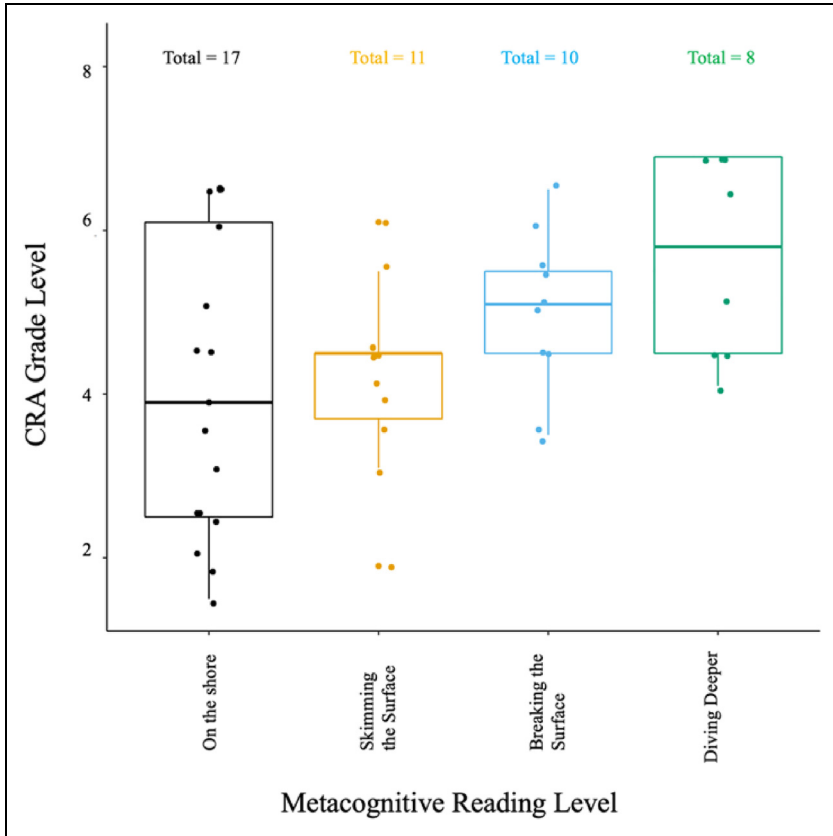


Figure 2. Box plots of observed reading abilities by the depth of metacognitive connections.

those receiving general education services (including reclassified students based on ELPAC performance) and students receiving language and/or special education services.

Discussion

Our case study was guided by our interest in understanding the theoretical and instructional insights afforded by the CRA, which we designed to gauge students' comprehension and critical reasoning of multimodal, interdisciplinary texts. We acknowledge the limitations of case studies for making any generalizations about the various common assessment practices in schools. Further, the relatively small sample size of students precluded our ability to explore potential differences in reading performance across school-identified groups (e.g., special education) within

our sample of participants. We were also unable to determine the potential effects of various textual features and modalities (e.g., embedded video clips) on reading comprehension performance given that we only had a single text for each of the grade levels represented. As such, this initial study is the first of a series of investigations related to improving reading assessment practices in schools.

During our debriefing session with administrative leadership and teachers, there seemed no surprise from our partnering colleagues about the variation in reading abilities, but there was some surprise about the observed abilities of several students, such as one Latinx student receiving both language enrichment and special education services demonstrating grade-level understanding of textual media as well as critical reasoning of information. Hence, the STAR may not be an adequate tool for understanding the reading abilities of students, particularly multilingual learners and those receiving special education services. The observed gap between STAR scores and observed CRA levels gathered within the same timeframe offers a clear indication that students may be systematically underestimated, potentially with harmful consequences (e.g., assigned to pullout remedial programs). Further, students receiving special linguistic and other educational support may not have full access to demonstrate their abilities on standardized assessments (Kohn, 2000), hence the need for more appropriate classroom reading assessments.

We also found that a student's demonstration of critical engagement may not necessarily align with their reading level. For example, fourth-grade participants with the ability to grasp core ideas and terms of sixth-grade texts were not necessarily able to engage metacognitively with such texts. Such a finding suggests a need for greater efforts to practice metacognitive engagement with texts in the early elementary grades. Instructional tools that foster critical discussions about texts (e.g., Arya & Meier, 2020) may be helpful for fostering such practice. Given the increased importance of critical reasoning in reading, we are in need of assessments that can adequately gauge such metacognitive thinking about texts (National Assessment Governing Board, 2021). Preliminary results presented in this study suggest that the CRA could be a valuable tool for schools striving to meet new assessment standards in reading.

The relative ease of administering standardized assessments like the STAR compared to qualitative assessments like the CRA present a challenge for schools that strive to minimize the time taken away from valuable instructional practices. University and community college partners can be valuable sources of support by training and connecting undergraduates and preservice teachers with classrooms to conduct more in-depth, qualitative assessments like the CRA that may provide a more comprehensive, accurate picture of young, developing readers. We believe that our use of the building blocks approach to measurement design (Wilson, 2004) enabled us to clarify our theoretical assumptions while centering on students' interests and curiosities about textual media.

Our partners have welcomed further exploration into this issue to determine any additional group differences across all grade levels. As such, this case is our first of

multiple developmental and explorative efforts in assessing and supporting young developing readers within this digital age. We aim to build on this initial phase of CRA development by continuing to pilot our current text sets while also building new textual sets at each grade level to allow for initial and subsequent assessments for individual students. We also aim to continue our adherence to the iterative approach of measurement design (Wilson, 2004), hence engaging in multiple rounds of evaluation and revision throughout the development process. We hope that our study inspires similar efforts by literacy scholars and educators interested in addressing current issues in reading assessments during this era of standards-driven accountability.


Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Diana J. Arya  <https://orcid.org/0000-0002-1335-3206>

Supplemental Material

Supplemental material for this article is available online.

References

- Abedi, J., Leon, S., Kao, J., Bayley, R., Ewers, N., Herman, J., & Mundhenk, K. (2011). Accessible Reading Assessments for Students with Disabilities: The Role of Cognitive, Grammatical, Lexical, and Textual/Visual Features. *CRESST Report 785. National Center for Research on Evaluation, Standards, and Student Testing (CRESST)*.
- Arya, D. J. (2022). Into the void of discourse. *Linguistics and Education, 68*, 1–10. <https://doi.org/10.1016/j.linged.2021.100964>
- Arya, D. J., Clairmont, A., & Hirsch, S. (2020a). Interpreting and explaining data representations: A comparison across grades 1–7. In T. Phillips & R. Dippre (Eds.), *Approaches to life-span writing research: Generating murmurations towards an actionable coherence* (pp. 177–193). Colorado State University Press & the University Press of Colorado.
- Arya, D. J., Clairmont, A., Katz, D., & Maul, A. (2020b). Measuring reading strategy use [Special issue on validity studies]. *Educational Assessment, 25*(1), 5–30. <https://doi.org/10.1080/10627197.2019.1702464>
- Arya, D. J., Hiebert, E. H., & Pearson, P. D. (2011). The effects of syntactic and lexical complexity on the comprehension of elementary science texts. *International Electronic*

- Journal of Elementary Education*, 4(1), 107–125. Retrieved from <https://www.iejee.com/index.php/IEJEE/article/view/216>
- Arya, D. J., & Maul, A. (2012). The role of the scientific discovery narrative in middle school science education: An experimental study. *Journal of Educational Psychology*, 104(4), 1022–1032. <https://doi.org/10.1037/a0028108>
- Arya, D. J., & Maul, A. (2021). Why sociocultural context matters in the science of reading and the reading of science: Revisiting the science discovery narrative. *Reading Research Quarterly*, 56, S273–S286. <https://doi.org/10.1002/rrq.316>
- Arya, D. J., & Meier, V. (2020). *CRUSH-it! Reading multimodal texts for civic engagement*. University of California, Santa Barbara. Captured from: <https://www.cbleducation.org/cbltoolkit>
- Bailey, A. L., Butler, F. A., Stevens, R., & Lord, C. (2007). Further specifying the language demands of school. In A. L. Bailey (Ed.) *The language demands of school: Putting academic English to the test* (pp. 103–156). Yale University.
- Browne, S., Chen, X., Baroudi, F., & Sevinc, E. (2021). *Reader response theory*. Oxford University Press. <https://doi.org/10.1093/OBO/9780190221911-0107>.
- California Department of Education (2021). *Fingertip facts on education in California*. Author. <https://www.cde.ca.gov/ds/ad/ceffingertipfacts.asp>
- Duke, N. (2020). *Listening to reading—watching while writing protocol (LTR-WWWP)*. Nell K. Duke. <https://www.nellkduke.org/listening-to-reading-protocol>
- Ferrara, S., & DeMauro, G. E. (2006). In R.L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 579–621). Santa Barbara: Greenwood Publishing Group.
- Fettig, A., Cook, A. L., Morizio, L., Gould, K., & Brodsky, L. (2018). Using dialogic reading strategies to promote social-emotional skills for young students: An exploratory case study in an after-school program. *Journal of Early Childhood Research*, 16(4), 436–448. <https://doi.org/10.1177%2F1476718X18804848>
- García, O., & Kleifgen, J. A. (2020). Translanguaging and literacies. *Reading Research Quarterly*, 55(4), 553–571. <https://doi.org/10.1002/rrq.286>
- Gee, J. P. (2018). Reading as situated language: A sociocognitive perspective. In R. Ruddell, M. Sailors, N. Unrau, and D. Alverman (Eds.), *Theoretical models and processes of literacy* (pp. 105–117). Routledge.
- Hsu, T. C., & Nitko, A. J. (1983). Microcomputer testing software teachers can use. *Educational Measurement: Issues and Practice*, 2(4), 15–18. <https://doi.org/10.1111/j.1745-3992.1983.tb00719.x>
- Jewitt, C. (2005). Multimodality, “reading”, and “writing” for the 21st century. *Discourse: Studies in the Cultural Politics of Education*, 26(3), 315–331. <https://doi.org/10.1080/01596300500200011>
- Kohn, A. (2000). *The case against standardized testing: Raising the scores, ruining the schools* (pp. 1–15). Heinemann.
- Kress, G. (2003). *Literacy in the new media age*. Routledge. <https://doi.org/10.4324/9780203299234>.
- Leslie, L., & Caldwell, J. S. (2017). *Qualitative reading inventory, VI*. Scott, Foresman/Little, Brown Higher Education.

- McClung, N. A. (2018). Learning to queer text: Epiphanies from a family critical literacy practice. *The Reading Teacher*, 71(4), 401–410. <https://doi.org/10.1002/trtr.1640>
- McLaughlin, M., & DeVoogd, G. L. (2004). *Critical literacy: Enhancing students' comprehension of text*. Scholastic. <http://doi.org/10.14507/er.v0.491>.
- McNamara, D. S., Louwerse, M. M., Cai, Z., & Graesser, A. (2005). Coh-Metrix (Version 1.4) [Computer software]. Captured from: <http://cohmetrix.memphis.edu>
- National Academy of Education (2021). *Educational assessments in the COVID-19 era and beyond*. Author. <https://naeducation.org/wp-content/uploads/2021/02/Educational-Assessments-in-the-COVID-19-Era-and-Beyond.pdf>
- National Assessment Governing Board (2021). *Reading framework for the 2026 national assessment of educational progress*. U.S. Department of Education. <https://www.nagb.gov/naep-frameworks/reading/2026-reading-framework.html>
- Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, 18(1), 22–37. <https://doi.org/10.1080/10888438.2013.827687>
- Renaissance Learning (2022). *STAR Assessments for reading technical manual*. Author. <https://doi.org/10.1080/10888438.2013.827687>.
- Rojas-Drummond, S. (2019). A dialogic approach to understanding and promoting literacy practices in the primary classroom. In N. Mercer, R. Wegerif, & L. Major (Eds.), *The Routledge international handbook of research on dialogic education* (pp. 306–319). Routledge.
- Ruiz-Primo, M. A., Solano-Flores, G., & Li, M. (2014). Formative assessment as a process of interaction through language. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *Designing assessment for quality learning* (pp. 265–282). Springer. https://doi.org/10.1007/978-94-007-5902-2_17.
- Solano-Flores, G. (2011). Assessing the cultural validity of assessment practices: An introduction. In M. R. Bastera, E. Trumbull, & G. Solano-Flores (Eds.), *Cultural validity in assessment: Addressing linguistic and cultural diversity* (pp. 3–21). Routledge.
- Souto-Manning, M., Rabadi-Raol, A., Robinson, D., & Perez, A. (2019). What stories do my classroom and its materials tell? Preparing early childhood teachers to engage in equitable and inclusive teaching. *Young Exceptional Children*, 22(2), 62–73. <https://doi.org/10.1177%2F1096250618811619>
- Stake, R. E. (2000). Case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 17–25). Sage.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797. <https://psycnet.apa.org/doi/10.1037/0022-3514.69.5.797>.
- Street, B. V. (2014). *Social literacies: Critical approaches to literacy in development, ethnography and education*. Routledge. <https://doi.org/10.4324/9781315844282>.
- Tierney, R. J. (2018). Toward a model of global meaning making. *Journal of Literacy Research*, 50(4), 397–422. <https://doi.org/10.1177%2F1086296X18803134>
- Valencia, S., & Pearson, P. D. (1986). New models for reading assessment. *Reading Education*, 7(1), 1–14. Retrieved from <https://files.eric.ed.gov/fulltext/ED281167.pdf>

- Wilson, M. (2004). *Constructing measures: An item response modeling approach: An item response modeling approach*. Routledge. <https://doi.org/10.4324/9781410611697>.
- Yoon, H. S. (2020). Critically literate citizenship: Moments and movements in second grade. *Journal of Literacy Research*, 52(3), 293–315. <https://doi.org/10.1177%2F1086296X20939557>

Author Biographies

Diana J. Arya is an associate professor and faculty director within the Gevirtz Graduate School of Education at the University of California, Santa Barbara. Their research interests relate to interdisciplinary literacies (researching, communicating, and building knowledge) for a diverse, multilingual student population.

Sabiha Sultana is a PhD student in education at the University of California, Santa Barbara. Her research interests include critical reading assessments and validity issues related to measurement design.

Somer Levine is a PhD student in education at the University of California, Santa Barbara. Her research interests include instructional approaches to literacy and professional development of elementary teachers.

Daniel Katz is a PhD student in education at the University of California, Santa Barbara. His research interests relate to philosophical and analytic aspects of measurement across interdisciplinary contexts.

John Galisky is a PhD student in education at the University of California, Santa Barbara. His research interests relate to science education and related literacy practices and assessments.

Honeiah Karimi is a PhD student in education at the University of California, Santa Barbara. Her research interests relate to sociolinguistics, literacy, and technological applications to educational programming.