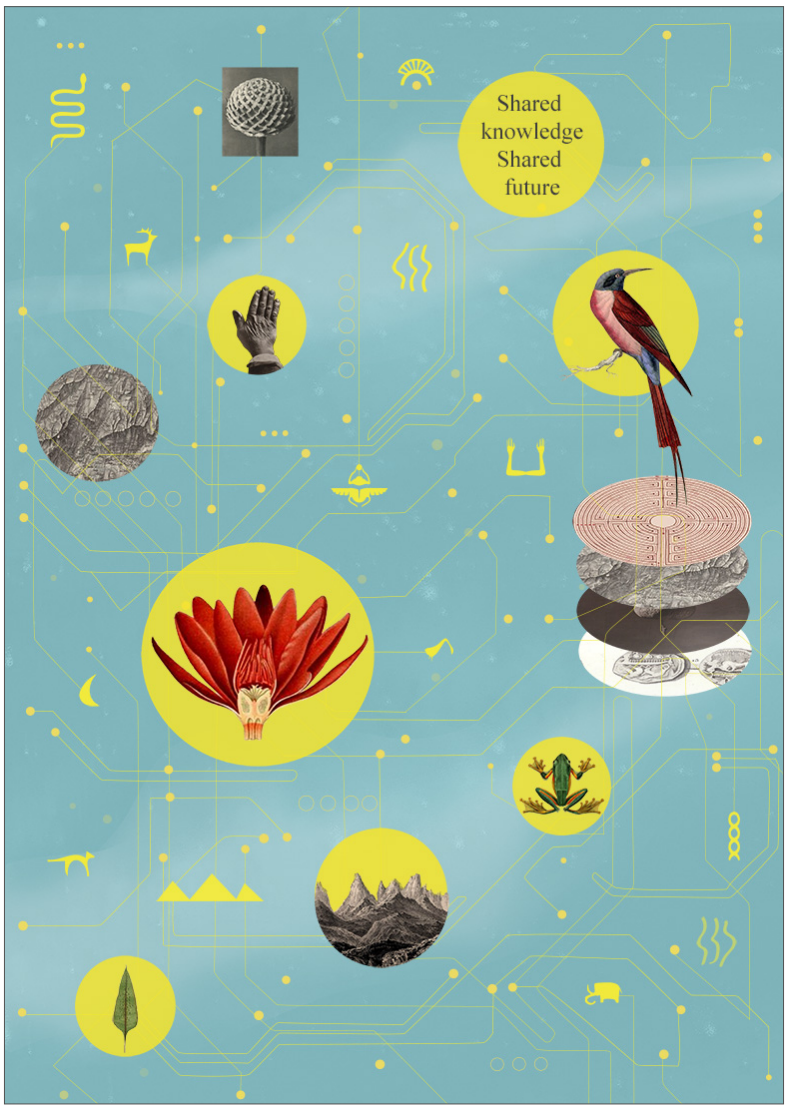
Generative Artificial Intelligence, Writing Placement, and Principled Decision Making in U.S. Postsecondary Contexts

A White Paper

Christie Toth, Jessica Nastal, Tiffany Buckingham Barney, Kris Messer, and Jason Godfrey



January 15, 2026

Commissioned by the editors of *The Journal of Writing Analytics*

Authors

Christie Toth, Ph.D. is associate professor of writing and rhetoric studies at the University of Utah, where she leads the development of placement assessment for first-year writing in addition to coordinating undergraduate transfer initiatives and community college partnerships. Her research on writing placement has been published in *Assessing Writing*, *Journal of Writing Assessment*, *Writing Program Administration*, and the collection *Writing Assessment, Social Justice, and the Advancement of Opportunity* (Poe, Inoue, & Elliot, eds.). With Jessica Nastal and Mya Poe, she co-edited *Writing Placement in Two-Year Colleges: The Pursuit of Equity in Postsecondary Education*, which received the Council of Writing Program Administrators Best Book Award for 2022. Toth chairs her department's AI working group and is the primary facilitator for its new faculty micro-credential in GAI-aware writing pedagogy.

Jessica Nastal, Ph.D. is assistant professor of English at College of DuPage and co-leads the Two-Year College English Association's writing placement network with Kris Messer. Her research on writing assessment has been published in *ETS Research Reports*, *Intersection*, *Journal of Response to Writing*, *Journal of Writing Assessment*, *Pedagogy*, *Teaching English in the Two-Year College*, and in edited collections. With Mya Poe and Christie Toth, she co-edited *Writing Placement in Two-Year Colleges: The Pursuit of Equity in Postsecondary Education*, which received the Council of Writing Program Administrators Best Book Award for 2022. Nastal is incoming co-editor of *Teaching English in the Two-Year College*.

Tiffany Buckingham Barney is a Ph.D. candidate in the Writing and Rhetoric Studies Program at the University of Utah and a former adjunct faculty member at Salt Lake Community College. Her research is on computational models and the affordances and limitations offered by GAI in writing assessment and writing placement. Barney's recent scholarly work focuses on guiding students in rhetorical analysis and prompt engineering as well as offering faculty support regarding GAI pedagogy and assessment.

Kris Messer, Ph.D. is associate professor of English at the Community College of Baltimore County and co-leads the Two-Year College English Association's writing placement network with Jessica Nastal. Her teaching work is focused on literacy praxis, community formation, identity performance, and community-based and socially engaged learning. Her writing assessment scholarship has been published in *Teaching English in the Two-Year College*, *Journal of Writing Assessment*, and in edited collections. Messer is chair of the 2026 TYCA national convention.

Jason Godfrey, Ph.D. is the Managing Director of Data and Information Systems at Accelerate, where he leads cross-state cohorts build integrated education data systems, oversee large-scale tutoring grants, and pass supportive legislation. With expertise in mixed-methods research, he focuses on evidence-based practice to advance equity aims in educational settings. Godfrey's research on writing assessment has been published in *Journal of Writing Assessment*, *Assessing Writing*, and *Writing Program Administration*.

Cover Image: "Shared Knowledge, Shared Future" by Luisa Brando for Creative Commons, licensed under CC-BY-SA 4.0 International.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ADF:	anticipatory design frameworks
A*f*L:	assessment for learning
AWE:	automated writing evaluation
CoT:	chain of thought
CIP:	Classification of Instructional Programs
CR:	constructed response
CSP:	culturally sustaining pedagogy
ECD:	evidence-centered design
FYWA:	First-Year Writing Assessment
GAI:	generative artificial intelligence
GPT:	generative pre-trained transformer
ICL:	in-context learning
IPEDS:	Integrated Postsecondary Education Data System
IT:	information technology
IUA:	interpretation and use argument
JWA:	*The Journal of Writing Analytics*
LLM:	large language model
NLP:	natural language processing
PII:	personally identifying information
RAG:	retrieval-augmented generation
SEI:	social and environmental impact
SLM:	small language model
SLO:	student learning outcomes
SSP:	student self-placement
SWE:	standardized written English
ToA:	theory of action
TYCA:	Two-Year College English Association

Preface & Acknowledgements

This white paper emerged from an urgent need in a specific context at a particular historical moment. In April 2025, co-author Christie Toth, coordinator of the University of Utah's first-year writing placement process since 2021, was tasked by her department and college leadership to work with the institution's Artificial Intelligence (AI) Initiative team to incorporate generative artificial intelligence (GAI) into the writing placement scoring process. Initially, those leaders were pushing to begin using GAI-assessed placement with incoming students by July. Christie's previously scheduled editorial meeting about a potential special issue of *The Journal of Writing Analytics* with co-authors Jessica Nastal and Kris Messer turned into a much-needed brainstorming session about how she should respond to this sudden demand on a dangerously compressed timeline. Norbert Elliot and Jessica tapped their national network of writing analytics colleagues to provide references and expert input as a starting point, but it quickly became clear that emerging scholarship on GAI and writing assessment had barely begun to grapple with how these technologies might be incorporated into placement. Norbert proposed this white paper to help colleagues get ahead of what seemed likely to become a new horizon for writing assessment scholarship and practice. For Christie, the project was a frantic effort to build the plane mid-air.

Fortunately, Christie's college leadership decided to slow down the proposed implementation process. As of January 2026, the University of Utah's GAI-assessed placement initiative is on an indefinite pause. Christie and colleagues are using the likely temporary reprieve to develop the theory of action (ToA), evidence-centered design (ECD), and interpretation and use argument (IUA) that this document calls for (§4). So far, the experience has been a case study in how institutional urgency surrounding GAI can create placement "emergencies" that threaten faculty shared governance grounded in disciplinary expertise and understanding of local context. It is also an example of how our multidisciplinary community can respond to this moment in ways that extend our knowledge; uphold our values; and center opportunity, equity, and justice for students as we imagine transformative possibilities for writing placement at this moment of deep technological and political uncertainty.

The co-authors of this white paper extend their deepest thanks to Norbert Elliot for his extraordinary mentorship, advocacy, and intellectual generosity through every stage of the document's development. Norbert's contributions are particularly evident in Section 4: Toward a Theory of Writing Placement and GAI, which reviewers noted could be an article in its own right. Thank you also to Alaina Tackitt, Alex Rudniy, and David Eubanks for formative conversations about GAI, writing assessment, and policy. Many thanks to the stellar group of reviewers who provided invaluable feedback on the

initial draft of this manuscript and agreed to be acknowledged by name: Laura Aull, Brian Gogan, Anuj Gupta, Sarah Z. Johnson, Michael Laudenbach, Jung-Hsien Lin, Jens Lloyd, Ben Naismith, Jennifer Burke Reifman, Zhaozhe Wang, and Carl Whithaus, in addition to one anonymous reviewer. Thank you to Lindsey Harding for excellent copyediting. We would also like to thank Matthew Osborn for his elegant design of this document.

No GAI platforms were used to generate the text of this white paper. Some members occasionally used Microsoft Copilot along with non-AI digital search tools to locate sources. Members of the University of Utah's AI Initiative lab used ChatGPT's deep research function to generate a report about writing placement and GAI that they presented to Christie during early institutional conversations about developing a pilot GAI-assisted placement process. Discussion of that document did foreground some key considerations (and some misgivings about ChatGPT's ability to synthesize writing assessment scholarship) that influenced the development of this white paper.

Executive Summary

This white paper documents the thinking of one small constellation of scholars coming together to articulate principles and practices we hope can help guide decision making when colleagues at a range of institution types encounter—with varying degrees of volition and enthusiasm—the prospect of incorporating generative artificial intelligence (GAI) into writing placement. Our positions reflect our individual and collective lived experiences, beliefs, and institutional positions, as well as the many thoughtful resources, suggestions, and critiques our peer reviewers volunteered. We offer the resulting white paper to try to make sense of our current moment and make meaning together going forward.

Audience: This white paper is designed for communities of literacy program coordinators (i.e., writing program administrators and others with coordination roles in postsecondary literacy programs) who are considering incorporating GAI into writing placement processes at their postsecondary educational sites in the United States. Related audiences include those stakeholders internal to institutions (instructors, students, professional staff, administrators) and those external to institutions (community members, psychometricians, legislators, and members of accrediting bodies).

Section 1: The Landscape of Generative Artificial Intelligence, Writing Placement, and Principled Decision Making: The authors advocate transparent, principled decision-making strategies to address the unique challenge of using GAI for writing placement. The authors establish a groundwork for principles and strategies by describing the present uses of GAI for assessing student writing samples and the significance of heterogeneity in the United States as it pertains to the language models used to evaluate these samples. The authors also discuss four potentially intersecting paths forward for GAI-assisted writing placement (opting out, slowing down, selectively implementing, reimagining) and provide an overview of decision-making principles that shape these paths.

Section 2: Principles to Guide Decision Making About Generative Artificial Intelligence and Writing Placement: Central to this white paper, the authors provide six principles that serve as a touchstone for decision making regarding GAI-assisted writing placement: combining disciplinary integrity with multidisciplinary collaboration, prioritizing logic of evidence, aligning with programmatic values and commitments, respecting individuals and communities, attending to emerging technologies and relationships, and being open to transformative effects. Special attention is paid to place-making through writing assessment—using assessments of incoming students to continuously (re)make institutional and programmatic places to serve students

equitably, honor students' dignity, and advance social justice.

Section 3: GAI, Writing Analytics, and Writing Assessment: This technical section examines the current state of GAI writing assessment, identifying many areas in which further research is needed to inform the design of GAI-assisted placement processes with strong evidence of fairness, validity, and reliability. The authors provide an overview of significant text-based corpora, discuss the evolution of formative and summative automated writing evaluation (AWE) in relation to recent developments in GAI, and present considerations for various approaches to GAI training. The authors identify potential biases and social and environmental impacts of GAI use, as well as promising possibilities for GAI-assisted writing placement when instruments are trained using local corpora of human-evaluated student writing.

Section 4: Toward a Theory of Writing Placement and GAI: This theoretical section includes information on anticipatory design frameworks (ADF), score interpretation and use arguments (IUA), and theories of action (ToA) emphasizing consequences of GAI-assisted writing placement. The authors position the white paper in a tradition of evidence-centered design (ECD) and in a program of research focused on principled writing assessment. Advancing the idea that (G)AI may be best understood as a human creation situated in social processes, the authors provide three models that can be used to promote reflection, collaboration, and deliberative action through attention to student backgrounds and needs at local postsecondary educational sites.

Section 5: Possible Uses of GAI in Writing Placement: In this conceptual section, the authors examine possible approaches to GAI-assisted writing placement, consider the likely competence of GAI-based assessment instruments, discuss GAI scoring, and seek transformative uses for GAI in writing placement beyond scoring. The authors apply the principles and theory articulated earlier in the white paper to consider how current writing placement processes might incorporate GAI. Focused on potential, the authors contend GAI in writing placement could afford additional opportunities for local postsecondary educational sites to support students' agency and achievement.

Section 6: Next Steps: The authors conclude with a discussion on the role of peer review and collaboration in development of this white paper, as well as 2025–2026 plans to foster opportunities for students and colleagues across and beyond the United States to discuss uses of GAI in writing placement and contribute to iterative revision/reimagining of this initial document.

Section 1 • The Landscape of General Artificial Intelligence, Writing Placement, and Principled Decision Making

1.0 Overview

This white paper advocates transparent, principled decision-making strategies to address the unique challenge of incorporating generative artificial intelligence (GAI) into writing placement in U.S. postsecondary institutions. In this introduction, we lay the groundwork for these principles and strategies by briefly describing a) the present uses of GAI for assessing student writing, and b) the significance of heterogeneity in the United States as it pertains to these writing assessment processes. We also discuss four potentially intersecting paths forward for GAI-assisted writing placement and provide an overview of the six principles that inform our discussion.

The primary audience for this white paper is literacy program coordinators who are interested in —and/or being encouraged, pressured, or required —to incorporate GAI into writing placement processes at their institutions. We follow Two-Year College English Association (TYCA) chair Joanne Baird Giordano in using the capacious term "literacy program coordinators" (Giordano et al., 2025) because a) not all colleges have formal writing programs or writing program administrators, b) coordination roles are often distributed across multiple faculty and staff, and c) reading and writing instruction are often integrated, particularly at access-oriented institutions. We recognize our audience might include scholars whose primary areas of expertise are outside writing assessment, writing analytics, or computers and writing. As part of our commitment to accessibility, we have sought to provide necessary context for these readers, including basic exposition of theories and specialized terminologies from these fields; we invite colleagues who are already familiar with these concepts to navigate the document strategically. Because this document is the first white paper of its kind, we have engaged in extensive pre-publication peer review. We are committed to inviting broad public input on the initial publication and ultimately revising the work presented here in a planned 2.0 version to be published in fall 2026.

Our advocacy for principled decision making regarding uses of GAI in writing placement—decision making that is multidisciplinary, evidence-based, culturally and ethically sustaining, and technically and consequentially alert—aims to enhance the fairness, validity, and reliability of next-generation writing assessment practices. As teachers and scholars, we are dedicated to sociocultural assessment, with special attention to campus localization. Situated within educational measurement, the basic premise of sociocultural assessment "is that students' learning and performance is inextricably tied to the social, cultural, and linguistic contexts within which they

live and develop knowledge" (Bennett et al., 2025, p. 1). There is no one standard by which writing should be evaluated, and any innovation in writing placement—including incorporating GAI into local processes—must create space for multiple ways of knowing, being, and doing through language.

1.1 GAI in Writing Assessment

Since OpenAI released its ChatGPT chatbot to the public in November 2022, GAI has been transforming practices across professional, political, and academic contexts, with particular implications for literacy instruction. In ChatGPT's initial months, teachers described the proliferation of GAI in their classrooms as "intensely negative" (Klopfer et al., 2024) as more than half of students aged 14 to 22 began to use these platforms (Center for Digital Thriving, 2024). Over the last three years, public discourse regarding GAI has cycled through rhetorics of hype, hope, doom, and disappointment (Martin, 2023; Sloane et al., 2024), exacerbated by emerging concerns about developers' disregard for intellectual property and data privacy (Kuru, 2024), fears about worker displacement (Occhipinti et al., 2025), and reckoning with the energy and water demands of these technologies and their environmental impacts on communities where GAI facilities and power sources are housed (Jiang et al., 2024). News outlets regularly report on GAI's capacity to support academic dishonesty or circumvent critical reading and thinking, contributing to what Johnson (2023), Byrd (2023), Basgier (2025), and others have called an *AI literacy crisis*. As the field centrally concerned with providing introductory college literacy instruction, writing studies has been at the forefront of efforts to understand and respond pedagogically to this technological upheaval.

Amid the crisis discourses emanating from within and beyond the academy, it is important to recognize that AI use in writing—and writing assessment—is not new. Teacher-scholars and practitioners have used a number of different computational methods and techniques over the years to analyze and evaluate writing. In writing assessment, automated writing evaluation (AWE) technologies designed to mimic human rater behavior using AI have existed for decades (Page, 1966). The earliest versions included lexical evaluation using word counts, word predictions, text parsing, and classification. Later versions added evaluation of semantic, syntactic, and discourse features using natural language processing (NLP) techniques (Shermis et al., 2013). These instruments were trained to match human responses to the text, scoring some language features higher than others on a designated scale or scales, and then used to provide scores on writing samples. For instance, many of us who applied to U.S. graduate programs after 1999 had our Graduate Record Exam (GRE) essays scored by both humans and e-Rater, an AWE application (see Burstein, 2003). AWE systems have been the subject of extensive disciplinary analysis (Shermis & Wilson, 2024) and critique (Ericsson & Haswell, 2006;

Herrington & Moran, 2001; Perelman, 2012, 2014), and we hope that this kind of deliberation will continue as new writing assessment technologies emerge.

While rapid, widespread uptake of GAI is a recent phenomenon, the technology itself is not. AI that generates newly calculated output, rather than pre-written output, had previously been used in task-specific software—including AWE and digital writing technologies with text prediction—for several years before GAI chatbots were released. The latest iterations of GAI include newly developed large language models (LLMs), which generate large amounts of novel and coherent output in response to open-domain user text input. These models rely on deep neural networks to mimic not only human behavior but, to some degree, human *learning*. LLMs such as OpenAI's generative pre-trained transformer (GPT) and Google's Pathways Language Model (PaLM) are a type of GAI pre-trained using very large sets of textual data; they power various GAI platforms including OpenAI's ChatGPT and Google's original chatbot, Bard. LLMs can be finetuned for various contexts and tasks, altering the capabilities of their applications using several methods, which we discuss further in §3.2.1, §3.2.2, and §3.2.3; these methods, among others, train models to create the current multimodal GAI platforms (sometimes referred to as *agents*) that generate not only new and unique text but also images, audio, and video.

LLMs are currently powering several GAI platforms such as ChatGPT, Anthropic's Claude, Microsoft's Copilot, Google's Gemini, and DeepSeek's DS-R1, all open for public use. These platforms are each trained for a variety of generation tasks and are continually adding capabilities. In terms of writing assessment, they can offer formative feedback toward revision that some students and teachers find useful, particularly when combined with human peer review (Dai et al., 2024; Guo, 2024; Saini et al., 2024; Sperber et al., 2025). Researchers are currently exploring their possibilities for summative writing evaluation (Mansour et al., 2024; Mizumoto & Eguchi, 2023; Naismith et al., 2023; Pack et al., 2024; Yancey et al., 2023). So far, however, such research has been conducted primarily by faculty and developers outside of writing studies.

To date, the field of writing studies has focused its energies on understanding the ever-shifting ways students are using GAI for college writing; grappling with the implications of GAI for classroom instruction, assignment design, curriculum development, and programmatic/institutional plagiarism policies; and debating disciplinary positioning in relation to GAI, including various possibilities for ethical engagement and principled refusal (e.g., Adisa et al., 2024a; Adisa et al., 2024b; Byrd et al., 2023; McIntyre, 2024; Sano-Franchini et al., 2025; Vee et al., 2023). Disciplinary conversations regarding GAI and writing *assessment* have centered on uses (and possible misuses) of GAI for facilitating peer feedback and formative and summative assessment

in course instruction (e.g., Gegg-Harrison & Shapiro, 2025; Sperber et al., 2025). However, initial organizational and disciplinary responses to GAI have not substantively addressed its possibilities and perils for writing placement. Drawing on critical perspectives within and beyond writing studies, this white paper responds to that gap.

As we respond, we must contend with the reality that scholarly and pedagogical debates surrounding GAI have unfolded across an increasingly dire political context for higher education. A growing number of postsecondary institutions are facing new waves of austerity under present fiscal crises and projected enrollment cliffs, as well as state and federal higher education policies aimed at dismantling diversity, equity, and inclusion efforts. Some of these funding cuts terminate postsecondary grants in the arts, sciences, and humanities, disproportionately affecting Historically Black Colleges and Universities, Tribally-Controlled Colleges and Universities, and two-year colleges (Bedekovics & Ragland, 2025). At some institutions, leadership is exerting pressure on writing/literacy programs to demonstrate "efficiency gains" through GAI adoption, sometimes combined with increased course enrollment caps and faculty teaching loads, as is happening in Utah. This white paper aims for a rigorous examination of GAI's current abilities, limitations, and costs in relation to writing placement, with full recognition that this examination cannot be disentangled from the layers of political struggle surrounding GAI, locally and nationally.

1.2 Heterogeneity in U.S. Society and Higher Education

Any consideration of the future of writing placement—and the broader political contexts that shape it—must account for the growing heterogeneity of the United States and its potential college-going population. The 2020 decennial census recorded a total U.S. population of 331,449,281, a 7.4% increase from the 2010 census. There was an even split in gender identities between individuals who identified as female (49.1%) and individuals who identified as male (50.9%), with no response option available to those who identify with neither or both of these gender categories. While population growth slowed for individuals under 45, they still represented the largest age groups in the country: under 18 years of age represented 22.1% of the population and those 18–44 represented 35.7%. Those between 45 and 64 years of age represented 25.4% and over 65 represented 16.8%. Turning to race and ethnicity, the largest group was White/Caucasian alone at 57.8%, which is a decrease from 67.8% in 2010. The next largest was Hispanic/Latinx at 18.7%, Black and African American alone at 12.1%, and Asian alone at 6%. This was the first census that introduced Middle Eastern/North African as an option, and 3.5 million people identified as part of this group. These shifts mark the nation's ongoing movement toward minority-majority status: by 2045,

demographers project that more than 50% of the U.S. population will hold non-White racial identities, and those groups will be younger, on average, than the White population.

Turning to the most recent data in the *Digest of Education Statistics*, published by the National Council of Education Statistics (NCES, 2024), more than fifteen million students were enrolled in postsecondary undergraduate education in Fall 2023, about 40% of whom were enrolled at two-year colleges (Community College Research Center, 2026). In terms of race and ethnicity, 46.3% of students identified as being Black, Indigenous, People of Color (BIPOC), 48.4% of students identified as Caucasian/White, and 5.3% were international students (NCES, 2024, Table 306.10). Most communities saw enrollment growth from Fall 2022, with strongest growth in Hispanic/Latinx, African American/Black, and international student enrollment; Hispanic/Latinx and Asian American student enrollment has grown most since Fall 2019, while enrollment has declined for Native American, Pacific Islander, African American/Black, and Caucasian/White students since the COVID-19 pandemic (NCES, 2024, Table 306.10). Nearly three-quarters (71.4%) of undergraduate students received some form of financial aid across all institution types (NCES, 2022, Table 331.10), with 31.6% of first-time, full-time students receiving Pell grants in 2019–2020 (NCES, 2026).

Examining similar data, Oliveri, Mislevy, and Elliot (2020) concluded that the key to equitable postsecondary admissions and placement "may not rest in single solutions such as remediation but, rather, in understanding obstacles comprehensively and designing strategies to overcome them in more exact ways" (p. 350). They proposed assessment frameworks such as those found in §4 that simultaneously consider both the meaning of score-based interpretations and consequences of test use. As U.S. and global demographics continue to shift, educators could face rising demands to assess even as they encounter increasingly diverse student populations. These changes require those who design, assess, and sustain writing placement processes—regardless of whether those processes incorporate GAI—to engage in ongoing, concurrent consideration of anticipatory design frameworks (§4.0), interpretation and use arguments (§4.3), and theories of action (§4.1). This white paper offers theories, principles, and concrete strategies for such engagement in relation to emerging GAI technologies.

1.3 The Current State of Writing Placement

The term *writing placement* refers to a range of practices postsecondary institutions implement with the goal of matching incoming students with learning contexts that align with their current writing abilities and support needs relative to the specific local literacy curriculum. Placement is sometimes framed as a matter of gauging perceptions of students' "preparedness" for college reading and writing, a way to determine which admitted students are likely to succeed and which

are "at-risk" of failing (see Messer et al., 2022). Historically, writing placement has often functioned as a binaristic choice between two courses: 1) an introductory college-credit-bearing writing course, typically a general education requirement for all students, and 2) a non-college-credit-bearing writing course (often called *developmental*, *basic*, or *remedial*; see Rose, 1983), intended to provide preparatory instruction for students deemed unlikely to succeed in the required college-credit course. Writing placement, however, can and often does encompass a wider range of possible course and support options beyond this two-course "sorting" model, with considerable variation across institutions and institution types (Cauoette, 2019; Klausman et al., 2016; Toth et al., 2019).

Until recently (e.g., Maloy et al., 2024; Ran & Lee, 2024; Sullivan & Adams, 2023), many colleges and universities—particularly those that have open or nonselective admissions—have required large numbers of students to enroll in developmental literacy coursework, and that requirement has usually been determined by some kind of writing placement process. Institutions have typically offered sequences of developmental reading and writing courses that range from as few as one to as many as seven courses students could be placed into, often with additional courses for multilingual students (which may or may not be classified as developmental). The last 15 years have seen pushes to accelerate developmental education; that is, to reduce the number of credits and amount of time students spend in developmental courses before gaining access to college-credit-bearing courses. These initiatives have led to a number of interrelated changes: the integration of separate reading and writing courses into single courses; the shortening of developmental course sequences; and, in some cases, a large-scale reduction or elimination of developmental courses in favor of co-requisite support models that provide more intensive and individualized instruction alongside the required college composition course (Hassel et al., 2015; Hodges et al., 2020; Toth et al., 2019). Meanwhile, institutions with widely varying selectivity, including many that have no developmental writing courses, have pursued forms of advanced placement and competency-based credit that enable students to "place out" of introductory college writing courses and enter directly into more advanced courses in required composition sequences (e.g., Toth et al., 2024; Zanders & Wilson, 2019).

Just as curricular and support options vary, institutions also use a range of possible approaches to writing placement: automatic placement based on high school GPA or externally administered standardized tests (e.g., SAT, ACT, AP), proprietary testing products (e.g., ACCUPLACER), locally developed exams, portfolio assessment, student self-placement (SSP) processes, and multiple measures (e.g., a combination of any of the preceding and/or other student records data). Many of these approaches encompass multiple possible assessment practices, such as multiple-choice tests of grammar, usage,

mechanics, and/or reading comprehension; timed, impromptu essay exams; or opportunities to revise a text. They may or may not involve direct assessment or scoring of students' writing. These approaches all present different possibilities and potential perils for GAI integration, to which we return in §5.1 and Table 5. We further articulate possible consequences of GAI integration into writing placement in our theory of action framework (§4.1).

Over the last century, the field of writing assessment has seen broad, if uneven, movement from writing placement based on decontextualized multiple-choice tests and large-scale purchased exams toward an emphasis on contextualized assessment of student writing through locally administered scoring of student essays, portfolios, and wide-ranging iterations of multiple measures assessment and SSP (Huot, 1996; Pantelides & Whittig, 2024; Yancey, 1999). Multiple measures and SSP have become particularly prominent in the last decade, as questions of racial equity and additional ethical dimensions of fairness have come to the fore and as the COVID-19 pandemic disrupted many longstanding placement practices (Gross, 2024; Nastal et al., 2022b). The TYCA white paper on writing placement reform (Klausman et al., 2016), two dedicated special issues and a special section in the Journal of Writing Assessment (Toth et al., 2019; Nastal & Messer, 2025; Pantelides & Whittig, 2024), one edited collection (Nastal et al., 2022b), and dozens of individual articles in the *Journal of Writing Assessment*, *Assessing Writing*, *College Composition and Communication*, *Writing Program Administration*, *Journal of Basic Writing*, and *Teaching English in the Two-Year College* demonstrate that comprehensive, ethically-minded writing placement reform has become a central concern in the field.

We may be at yet another inflection point for writing placement. The latter half of the 2010s saw a national movement toward test-optional college admissions policies in which students were not required to submit standardized test scores (e.g., ACT, SAT, GRE). This shift, initially grounded in concerns about racial equity, was accelerated by the logistical challenges of the COVID-19 pandemic (Cruz & Webb, 2025). These changes, alongside reforms to developmental education aimed at improving equity, created fertile ground for advances in writing placement. Today, however, policies designed to increase equity and fairness in college admissions are under scrutiny by the federal government. While more than 80% of four-year institutions in the United States did not require fall 2025 applicants to submit SAT or ACT scores, some very selective institutions are beginning to reinstate admissions testing requirements (Wood, 2024). In this climate, Pantelides and Whittig (2025) have expressed "fear" that "there is a turning away from [approaches like] SSP and, for a variety of reasons (state mandates, ease), a return to single measure standardized tests for placement" (p. 2).

Literacy program coordinators have always had varying degrees of input, oversight, and control

over writing placement (Huot, 1994), with two-year college faculty often facing the hardest fight to assert disciplinary authority in their institutions' placement processes (Klausman et al., 2016; Ostman, 2013; Toth et al., 2019). Now, a growing number of public institutions are subject to additional state or systemwide policies that mandate or constrain what approaches to placement are possible (Hodges et al., 2020; Miller et al., 2017; Ran & Lee, 2024; Toth et al., 2024). The advent of GAI further disrupts institutional processes already in upheaval amid politically motivated reductions of federal and often state-level support for higher education, particularly when conceived as more than narrow workforce training. Ongoing legislative attacks on academic freedom and shared governance can further erode faculty authority regarding placement processes. These intertwined technological and political disruptions could precipitate ethical retrenchment, or they could be an opportunity to develop new approaches to writing placement that, in Nastal and Messer's (2025) words, "showcase [our] deepest values about respect for student experiences and backgrounds and the variegated joys of writing," to "redress historic wrongs by deliberately creating processes for students to assert their interests, goals, and experiences and to highlight their assets." This white paper is our collective effort to help ourselves, our colleagues, and the field navigate these crossroads, guided by an awareness that new developments always and forever offer avenues for the past to continue to write its wrongs on our futures. We can also use new platforms and technologies to build different futures, this time with voices and resources that might have gone unnoticed or underutilized without the insights these developments make possible.

1.4 Four (Potentially Crossing or Recursive) Paths Forward

The emergence of GAI and its influence on a variety of institutional domains is creating systemic shifts that require intentional redirection. Sano-Franchini, McIntyre, and Fernandes (2025), authors of the *Refusing Generative AI in Writing Studies* project, have asserted the right of literacy faculty, programs, and students to "consciously and intentionally choose to refuse GAI *use*, when and where we are able to do so." Rooted in rhetorical theory and an understanding of "the relationship between language, power, and persuasion," (Sano-Franchini et al., 2025), their arguments for refusal center on concerns about GAI's tendency to erase linguistic diversity and advance White language supremacy, its theft of intellectual property and threats to data privacy, the unfair faculty labor demands of incorporating GAI into curricula and potential long-term negative impacts on workers across many industries, and the technology's demonstrable environmental harms. Sano-Franchini and colleagues (2025; see also McIntyre et al., 2025) have offered ethical grounds for opting out of GAI use in writing courses and programs that many GAI-skeptical scholars in writing studies—including some of the authors of this white paper—find compelling.

Figure 1. *Four Possible Paths Related to Generative Artificial Intelligence and Writing Placement*

1. *Opting Out.* These principles might, in some situations, provide grounds for opting out of GAI integration into writing placement. The principles reflect the reality that there are many reasons to be deeply concerned about the ethics of using existing GAI platforms for large-scale writing placement. For some literacy programs, opting out might be the most ethical option. Indeed, some programs may choose to opt out of writing placement altogether, embracing the principle that "an admitted student is a qualified student" (Poe et al., 2019; see also Tinkle et al., 2024); such programs might instead marshal resources for comprehensive academic support structures that enable all students to thrive in and beyond the required college credit-bearing writing course.
2. Slowing Down. These principles can support a case for intentionally slowing down GAI placement initiatives. Slowing down might be a particularly useful strategy when initiatives are driven by unwarranted urgency rooted in GAI hype or administrative pressures to integrate GAI for its own sake, often at the cost of students' opportunities to learn and fair labor practices for faculty and staff. Ensuring that a particular use of GAI within a specific writing placement process in a given local context is ethical requires careful planning, rich data collection and analysis, and ongoing assessment, none of which can or should be implemented in a rush. These principles can help literacy programs make pragmatic and ethical arguments for proceeding with integration at an appropriate and just pace.
3. Selective Implementation. These principles offer literacy programs actively interested in incorporating GAI into their writing placement processes a strong theoretical foundation for proceeding in ways that advance equity, honor student dignity, foster learning through human interaction, and expand opportunities for institutions to listen to students and communities. In these cases, GAI will not fully replace human readers or advisors. Rather, it will enable processes that facilitate human connection and provide support where students need it most.
4. Reimagining. These principles invite colleagues to take this moment of GAI-generated crisis to reimagine not just the mechanics and technologies but also the underlying purposes of longstanding writing assessment processes that have been called *placement*. We suggest that a range of writing analytics methods and technologies, including current or future iterations of AI, might be used to shift from the de facto sorting logic of writing placement assessment to writing assessment for *place-making*. In this reimagining, the purpose of assessing writing with incoming students would be to engage those students in reflective conversations about writing, and to gather information that we could use to continuously (re)make our institutions and our literacy programs so every student has the support that enables them to thrive.

Literacy program coordinators inclined toward principled refusal might find, however, as some of us have, that "when and where we are able to do so" is increasingly limited terrain. For those facing direct mandates from their administration or governing body, refusing GAI in writing placement might not be possible. In some cases, refusing to cooperate with GAI-assisted assessment could result in literacy programs losing control over, or even input in, writing placement processes altogether. Indeed, some independent writing studies departments risk being reorganized out of existence if they cannot demonstrate their contributions to institutional AI initiatives. Administrators who may have only recently turned to faculty to oversee placement during the COVID-19 pandemic might decide just as quickly to turn writing placement over to campus AI developers or purchase an externally developed AI-based assessment product with little or no input from literacy faculty. In such situations, cooperating knowledgeably and critically with efforts to incorporate GAI into writing placement —and assert whatever degree of local disciplinary authority is possible—might be a harm reduction strategy.

We also recognize that many in our field question or reject the GAI refusal position. Some literacy program coordinators—including several authors of this white paper—might take up possible uses of GAI in writing placement voluntarily and/or with enthusiasm. Those familiar with writing assessment's long history of engagement with machine learning and NLP (Shermis et al., 2013) might see GAI's potential to provide insight into students' literacy work through theoretically sound, justice-oriented placement processes. Colleagues in under-resourced institutions, where a single faculty member could be the only individual tasked with managing placement, might wish to use GAI to support their work, extending the ways many of us use learning analytics technologies to deepen our understanding of teaching and learning (e.g., Palmquist, 2020).

Recognizing the range of informed, good faith dispositions toward GAI literacy program coordinators might bring, as well as wide-ranging institutional conditions and needs, the principles in this white paper (§2) support at least four potential responses to the possibilities of GAI-assisted writing placement (described in Figure 1): opting out, slowing down, selectively implementing, and reimagining. These paths forward are not necessarily linear or mutually exclusive; rather, one path might loop back, lead to, or converge with others as technologies, local conditions, and student literacy practices continue to evolve.

1.5 Principles for Decision Making about GAI and Writing Placement

If the coming years are characterized by continued financial, demographic, and political pressures on U.S. postsecondary institutions, we think it is likely that many literacy program coordinators will be asked (or told) to incorporate GAI into writing placement. These disruptions threaten to undercut

the last decade of placement reform efforts grounded in principles of equity, fairness, and justice emerging from writing assessment's ethical turn. Some institutions have already begun to explore GAI's potential for large-scale writing placement, although these early adopters have not necessarily centered disciplinary expertise in writing assessment or writing analytics as they develop their models. Administrators without such disciplinary grounding might prioritize instruments that can be implemented quickly and at low cost to the institution to appease external stakeholders concerned about the bottom line—that is, perceived *efficiency*, which has long undermined efforts to prioritize validity and fairness in writing assessment (see Williamson, 1994). Whether literacy programs seek to incorporate GAI into writing placement of their own initiative or in response to external pressures, *The Journal of Writing Analytics* (JWA) recognizes that the field needs a clear initial articulation of principles to promote theoretical and ethical integrity in writing placement, with particular attention to sociocultural elements amid rapid technological, demographic, political, and institutional change. This white paper offers key principles and practical advice to literacy program coordinators at the full range of postsecondary institution types that are encountering opportunities, pressures, or mandates to incorporate GAI into writing placement. These principles and practices attend to what Moss (1998) observed decades ago: Assessment is "a social practice with social origins and social consequences" (p. 111). With the TYCA White Paper on Placement Reform (Klausman et al., 2016) as a foundational influence, our six principles are as follows: combine disciplinary integrity with multidisciplinary collaboration, prioritize logic of evidence, align with literacy programs' values and commitments, respect individuals and communities, attend to emerging technologies and relationships, and be open to transformative effects.

These principles, which we introduce in Figure 2 and discuss at length in §2, can inform any and all of the potential paths forward identified in Figure 1. Although the many challenges facing postsecondary education and the attendant labor demands can feel isolating, no literacy program needs to chart these placement paths alone.

Figure 2. *Six Principles for Generative Artificial Intelligence, Writing Placement, and Decision Making*

Principle 1: Combine disciplinary integrity with multidisciplinary collaboration. Any use of GAI in writing placement must be grounded in multidisciplinary expertise—local, national, and international—regarding writing evaluation in the 21st century.
Principle 2: Prioritize logic of evidence. Any use of GAI in writing placement must be supported with ongoing evidence regarding fairness, validity, and reliability.
Principle 3: Align Generative Artificial Intelligence with literacy programs' values and commitments. Any use of GAI in writing placement should advance literacy programs' values and commitments, which can include but are not limited to the following: fair labor practices, shared faculty governance practices, and the program's stated efforts to be accessible, culturally sustaining, linguistically just, and responsible to land, water, air, and other-than-human life.
Principle 4: Respect individuals and communities. Any use of GAI in writing placement must be ethical, transparent, and consensual for students and communities.
Principal 5: Attend to emerging technologies and relationships. Understanding emerging relationships between humans, GAI, and writing placement helps facilitate principled human-computer relations in teaching and assessing writing.
Principle 6: Be open to transformative effects. Because GAI could (and maybe should, and hopefully will) transform conceptualizations of the purposes and processes of writing placement and interrelated pedagogies, we should look for and amplify effects, anticipated or otherwise, which have positive value for students.

Section 2 • Principles to Guide Decision Making About GAI and Writing Placement

2.0 Overview

We offer the following principles to literacy faculty and program coordinators who are navigating mandates, pressures, or welcomed opportunities to incorporate GAI into writing placement. These principles provide support for multiple potential courses of action across four paths colleagues may be contemplating or moving between: opting out, slowing down, selectively implementing, and reimagining placement (§1.4). We envision them as heuristics as well as potential arguments for whatever ways forward best meet the needs and possibilities of specific institutional contexts.

Literacy program coordinators considering GAI-assisted writing placement may feel ill-equipped to advocate for disciplinary values if they are not secure in their own understanding of writing assessment, writing analytics, or GAI technologies. Sections §3, §4, and §5 provide in-depth examinations of how GAI works, what GAI can and cannot do at this stage, anticipatory design frameworks (ADFs) for developing well-theorized placement processes, and imagined possibilities for incorporating GAI into writing placement. Finally, in §6, we invite readers to collaborate with us throughout the 2025–2026 academic year as an extension of the peer-review process. In fall 2026, we plan to publish a revised 2.0 version of this white paper based on these ongoing conversations.

Principle 1: Combine disciplinary integrity with multidisciplinary collaboration.

Any use of GAI in writing placement must be grounded in multidisciplinary expertise—local, national, and international—regarding writing evaluation in the 21st century.

2.1 Principle 1: Combine Disciplinary Integrity with Multidisciplinary Collaboration

2.1.1 Writing Assessment Disciplinary Expertise

Any effort to incorporate GAI into writing placement must be grounded in the theoretical principles, methodologies, and empirical evidence generated within writing assessment and writing analytics over the last century. Neither technical understanding of GAI, nor generalized knowledge of prompt engineering (a fundamentally rhetorical and domain-specific act; see Graham, 2023), nor a simple review of writing placement processes at other institutions can substitute for deep theoretical understanding of language and literacy pedagogies, educational measurement, and the ethical dimensions inherent in evaluations of writing.

That said, the field of writing assessment is inherently multidisciplinary. We find the Classification of Instructional Programs (CIP)

codes helpful for tracing this multidisciplinarity. Writing assessment draws on scholarship and methods from a range of fields, including writing/rhetoric/composition studies (23.13 and 09.0101); educational assessment, evaluation, and research (13.06); writing program administration (23.1304); two-year college studies (13.0407); second-language writing (16.1701); applied linguistics (16.0105); ethics (38.0103); philosophy (38.0101); the scholarship of teaching and learning (SoTL) (13.0101); and critical race, ethnic, gender, disability, and Indigenous studies (05.01, 05.02, and 30.2601). Writing analytics is similarly multidisciplinary (see Moxley et al., 2017), drawing on fields that include writing/rhetoric/composition studies (23.13 and 09.0101); educational assessment, evaluation, and research (13.06); computer and information sciences (11.01); linguistics (16.0102, 16.0105, and 30.4801); computers and writing (23.1303); digital humanities (30.5202); and cognitive psychology (42.2701). Taken together, this multidisciplinary expertise provides people, programs, institutions, and the field with the knowledge base to engage with GAI and writing placement in ways that advance shared aims for postsecondary literacy education.

As scholars who have participated in writing assessment's ethical turn (e.g., Messer, 2024; Nastal, 2019; Toth, 2018), several of us have observed how judgments about writing are enacted within and have the potential to reinscribe systemic inequities and persistent social injustices based on a range of interrelated identities and experiences: race, ethnicity, nationality, documentation status, Indigeneity, language background, socioeconomic status, college-going generation, gender, sexual identities, religion, disability, neurodivergence, trauma, incarceration, and educational path, among others. Writing assessment can be violence (Lederman & Warwick, 2018), and without strong theoretical and ethical grounding, it can constrain or deny rather than enable and expand opportunities to learn for the least advantaged (Elliot, 2016). We discuss specific disciplinary values regarding sociocultural frameworks for assessment and central commitments to fairness in §2.2.

Efforts to incorporate GAI into writing placement not grounded in deep engagement with writing assessment scholarship risk harming students, faculty, and communities. Without that disciplinary foundation, we risk further disenfranchisement—further restrictions of educational access—potentially on a much larger scale. Institutional staff hired by upper administration to support the development of localized AI instruments are often in information technology (IT) departments, specialize in managing information systems and data warehouses, or operate as part of interdisciplinary design teams. Like many across postsecondary education, they rarely have a background in writing assessment other than what they themselves experienced as students or what they have been asked to create for other departments that use writing primarily as a means to assess content knowledge. Chances are they, like many,

operate from the kinds of biases (see Hood & Hopson, 2008; Randall, 2021) and uninterrogated standard language ideologies (see Flores & Rosa, 2015; Suh et al., 2021) that writing assessment scholars like Inoue (2019) have identified as sources of inequity in most institutions' assessment ecologies. Designers engaged in *responsible* development of GAI instruments rely on their professional grounding in ethics of intellectual property and data privacy. They should know that they need to bring their good intentions and technical expertise into conversation with the disciplinary and local knowledge of literacy instructors. We call upon our institutions to uphold fundamentals of shared governance and to create spaces for multidisciplinary teams to collaborate on all GAI-assisted writing placement initiatives.

2.1.2 21st Century Writing Assessment Requires Multidisciplinary Collaboration

At present, many literacy educators are actively and continuously deliberating on what their GAI syllabus policies and class practices should entail. Some are examining—and asking students to consider—the impact of GAI on linguistic diversity, water consumption, and communities in the Global South, where workers are paid poverty wages to label traumatizing images and videos to train the models (Ali et al., 2024). Some teachers are using a range of GAI platforms to aid in their course or meal planning, communications with a child's school principal, or managing medical conditions and care. Most of us are considering the ways in which students are interacting with or thinking about GAI and trying to understand more about how this technology is affecting the ways we communicate and circulate—or diminish—ways of thinking, knowing, and being.

Some of us have years of experience coordinating writing assessment programs where we create course or program learning outcomes, accompanying curricula and professional development, norming sessions to improve inter- and intra-rater reliability, and analyses of assessment results. However, we are just beginning to gain experience designing learning outcomes, curricula, and assessments in a world where students are using GAI in their writing processes. Most of us do not have the expertise required to build or train a GAI instrument that eliminates bias (see Dai et al., 2024, and their companion GitHub repository for more: https://github.com/KID-22/LLM-IR-Bias-Fairness-Survey). As we have noted, our colleagues in the IT department do not often have extensive training in writing assessment or in literacy pedagogy. Just as we need their expertise in database administration and information security, they rely on us to discover how to cultivate student self-efficacy and metacognition through writing. We need to collaborate to understand how to curate and analyze a local corpus of writing to, for instance, become better informed about our students.

The current moment calls for ways of conceiving disciplinary integrity in writing assessment that differ from what Huot (1996) and Klausman and

colleagues (2016) envisioned even ten years ago. First, we posit that a *multidisciplinary* approach to principled decision making in GAI and writing placement is necessary. It can bring together the wisdom of administrators focused on long-range planning; library, computer, and information specialists who have a keen understanding of information architecture, data privacy, and software engineering; educators expert in human development, literacy, language, educational assessment, cultural studies, and ethics; and environmental scientists. Second, *collaboration* is necessary. It is impractical to require writing studies faculty to immediately become expert in educational measurement and writing assessment as well as in machine learning, and to apply that expertise to two rapidly evolving enterprises: writing placement and GAI. It is also short-sighted to do so. Collaboration is vital to advancing knowledge and expanding our own capacities. It is vital to writing, to the social acts of communication, and to working (see Cox & Riedner, 2023).

To build an appropriately multidisciplinary team, we suggest the following, inspired by principles of community and labor organizing (e.g., Combahee River Collective, 1981) and shared governance (e.g., Cole et al., 2017):

- **Take stock of your own disciplinary and institutional knowledge.** Identify what gaps exist in your expertise when considering the path(s) you'll take (§1.4, Figure 1)—opting out, slowing down, selectively implementing, and reimagining—toward incorporating GAI into writing placement. Refer to the CIP codes that may be relevant to your work; in addition to those listed in 2.1.1, you may wish to consider artificial intelligence (11.0102), library and information science (25.0101), computational science (30.3001), human computer interaction (30.3101).

- **Consider potential campus partners in your writing placement work**; for instance, faculty in communications, humanities, STEM, and computer science; staff in institutional research, institutional effectiveness, and information technologies; and faculty or academic staff in writing centers, centers for teaching and learning, literacy centers, libraries, disability services, community outreach, and other programs that support students from historically underrepresented and structurally disadvantaged groups.

- **Consider consulting with people beyond your institution for insight.** These might include experts in writing assessment, writing analytics, computers and writing, and educational measurement. They might also include partners at the high schools, two-year colleges, universities, and libraries in your district. They might even be people in your communities outside your professional role, like your friends who work in computer programming, software engineering, and database warehouse administration.

While we recognize that not all individuals tasked with writing placement have the same resources and ability to assemble formal institutional teams, what has become apparent today is that many in writing studies, even those of us with expertise in writing assessment, are in new territory. Even if collaborations begin as informal consultations with colleagues who hold different roles with different expertise, that groundwork can be important for gathering the resources and perspectives that will inform the systems created and undergird advocacy. We do not recommend proceeding alone, and we offer this white paper as a starting point for building and finding multidisciplinary communities of support.

Principle 2: Prioritize logic of evidence.

Any use of GAI in writing placement must be supported with ongoing evidence regarding fairness, validity, and reliability.

2.2 Principle 2: Prioritize Logic of Evidence

While blue books have had their place in many of our educational experiences, and timed, impromptu tests may offer insight into students' written fluency, the best academic writing often emerges through recursive practices that often involve drafting, conversation, feedback, revision, and editing. Creating opportunities for students to compose writing in circumstances that permit such recursivity allows them to attend to all four domains Aull (2023), building on White and colleagues (2016), has articulated for a robust writing construct:

- **The cognitive or discursive domain:** related to reasoning, memory, and content knowledge (including, we surmise, knowledge of particular language and genre conventions). This domain has historically been the focus of most writing assessments, while the other three domains have been neglected.
- **The interpersonal domain:** related to collaborating with others, including how to ask for information or feedback and brainstorm ideas with others.
- **The intrapersonal domain:** related to reflecting and self-moderating, including adaptable strategies for responding to an unfamiliar writing task.
- **The health domain:** related to well-being, particularly the need for sufficient food, rest, and safety as preconditions for doing one's best possible writing (Aull, 2023, pp. 55–56).

Designing writing tasks with each of these domains in mind requires us, for instance, to implement principles of accessibility and trauma-informed teaching (Tayles, 2021): We know that people cannot perform at their best potential when they are unsafe, unable to access necessary materials, or unsupported in their efforts to learn and practice (Hammond, 2014; see more in §2.3). Writing tasks that incorporate peer and instructor feedback and writerly reflection on choices offer

unmatched evidence for researchers, practitioners, and students interested in understanding more about teaching, learning, and writing.

In an introduction to the Tools & Technology Forum in the journal *Assessing Writing*, Hartwell and Aull (2024) asserted that emerging writing assessment technologies like GAI "raise important implications regarding validity, fairness, and equity—three pillars that must undergird any writing assessment tool or tech." These pillars offer a theoretical frame for evaluating the specific uses of particular GAI technologies in local contexts for writing placement. Recognizing that many readers of this white paper might be new to writing assessment theory and/or charged with explaining these concepts to stakeholders outside the field, we provide an overview of these terms in this section. We offer further discussion of specific implications for fairness, validity, and reliability/precision when incorporating GAI into writing placement in Section 4. The GAI-Anticipatory Design Framework (GAI-ADF) (§4.0 and Table 3) may be of particular interest to readers concerned with logic of evidence.

2.2.1 Evidence of Fairness

Fairness, as Elliot (2016) has proposed, "is the first virtue of writing assessment." It requires faculty to structure opportunities "to achieve an ethical outcome" (Elliot, 2016), and it hinges on the social consequences of an assessment process (J. Hammond, 2019). Fairness demands interrogating "cultural and linguistic biases" (Hartwell & Aull, 2024, p. 2) in assessment, including "Eurocentric ways of knowing and processing information" (Randall et al., 2024) that have historically overlooked "the diverse cultural and linguistic resources" of marginalized communities (Hartwell & Aull, 2024, p. 2), as well as issues of access related to disability, neurodiversity, and varying material resources, including time (see Kryger & Zimmerman, 2020). In the literature, fairness has been discussed as the elimination of bias, the pursuit of validity, acknowledgement of social context, legal responsibility, and ethical obligation (Poe & Elliot, 2019). It encompasses ensuring "whether every student has a chance to demonstrate their learning in a given assessment, free from bias or discrimination" (Hartwell & Aull, 2024, p. 2) and requires practitioners to consider "race, ethnicity, age, gender, socioeconomic status, disability, or language or cultural background . . . throughout all stages of development, administration, scoring, interpretation, and use so that barriers to fair assessment can be reduced" (American Educational Research Association (AERA) et al., 2014, p. 50).

Fairness is closely tied to validity and reliability because its goal "is to maximize, to the extent possible, the opportunity for test takers to demonstrate their standing on the construct(s) the test is intended to measure" (AERA et al., 2014, p. 51). Assessment practitioners must create circumstances that facilitate students' ability to demonstrate this standing. When considering the fairness of GAI-assisted writing placement, practitioners can ask questions like the following:

Which communities' language practices, literacy practices, and ways of knowing and being are advantaged or devalued by a given use of a specific GAI platform trained in a particular way? To what extent does this use enable students with the full human range of embodiments and neurotypes to demonstrate their learning? What patterns emerge when we disaggregate scores or placement outcomes by student community, and what do those patterns tell us about whether our assessment practices are perpetuating, reinforcing, or challenging social inequities?

2.2.2 Evidence of Validity

Validity is a central concern across the decades of writing assessment literature, with scores of accounts tracing the history of the term, how it has been used in educational measurement for decades, and how it has been misconstrued to maintain social hierarchies (Cushman, 2016). We ground our work in the *Standards for Educational and Psychological Testing* (2014): "Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (AERA et al., 2014, p. 11). Practitioners are charged with ensuring results are used appropriately (Hartwell & Aull, 2024; Messick, 1989), which necessarily relies on identifying and using robust categories of evidence to make the validity argument (Cronbach, 1988; Kelly-Riley & Elliot, 2014). Validity does not reside within an assessment instrument; rather, it is an argument people make about the results of that assessment (see Cronbach, 1988; Kane, 2006, 2013) and how they will use those results to make decisions about students, curricula, and programs (see Nastal, 2019; Solano-Flores, 2008). Assessment, therefore, is about consequences.

To begin to validate a writing placement practice, practitioners start by asking questions like: How well does the writing construct of this assessment relate to the literacy program's learning outcomes and curricula? How well does it align with other forms of writing? What role does student insight play in the process? What claims are we making based on the results about what a student knows and can do (see §4.3)? What evidence can we use to inform those claims? How does the placement practice correspond to other institutional measures, such as course grades, credit acquisition, persistence, and credentialing?

2.2.3 Evidence of Reliability

Reliability generally refers to the consistency of scores (AERA et al., 2014). While not all placement processes involve scoring student writing, any writing assessment that generates a score or placement recommendation must be piloted, tested, and refined a number of times to ensure it can dependably elicit reasonable and consistent responses from its users. In settings where student writing is read and evaluated, it also takes time to train raters, whether those raters are computer systems or groups of faculty coming together to learn more about the writing in their program. Reliability requires practitioners to determine how and when they will adjudicate

scores, and when they will turn to a third reader, expert reader, or human reader to decide about a complex scenario. As we discuss in §3 and §5, the capabilities of current GAI platforms raise particular concerns about reliability in scoring.

To understand the reliability of a writing placement practice that involves scoring, practitioners will ask these kinds of questions: Does a student receive a similar score upon retaking the placement test? How consistently do raters score across tests and test-takers? Do different raters score similarly? Do test-takers or raters perform differently depending on context or modality? Are there any scores that require more attention? That is, is there more variance in ratings within a certain context or in a certain score range? These questions apply whether raters are human, GAI, or some combination of both.

Principle 3: Generative artificial intelligence align with literacy programs' values and commitments.

Any use of GAI in writing placement should advance literacy programs' values and commitments, which can include but are not limited to the following: fair labor practices, shared faculty governance practices, and the program's stated efforts to be accessible, culturally sustaining, linguistically just, and responsible to land, water, air, and other-than-human life.

2.3 Principle 3: Generative Artificial Intelligence Align with Literacy Programs' Values and Commitments

2.3.1 The End of Isolation (Again)

Drawing upon the breadth of their teaching and research, Poe, Inoue, and Elliot (2018) introduced the edited collection *Writing Assessment, Social Justice, and the Advancement of Opportunity* by insisting "that writing assessment must be understood within an ecological framework" (p. 4), "reject[ing] the disciplinary isolation of assessment from the social justice orientation of Writing Studies" (p. 12). We amplify their goals: If writing placement is to be meaningful, it must align with individual, programmatic, departmental, disciplinary, and broader educational values. It must also involve critical consideration of and engagement with the complex, competing, and sometimes contested values of the communities we serve. Further, it must be aligned with the writing construct, constructed response task and scoring rubric (if applicable), and student learning outcomes (SLOs; see §4.0 and Table 3), as well as curricula, syllabi, pedagogical actions, and other assessments (e.g., classroom, exit, and program; as identified in §4.1, and Figures 4, 5, and 6). If writing placement processes are disconnected from or at odds with this ecological context, we risk replicating or deepening the social inequities writing studies has worked to understand and redress (see Sternglass, 1997).

Throughout the decades-long shift in writing placement away from decontextualized, standardized purchased exams and toward more locally contextualized methods (§1.3), literacy program coordinators have documented efforts to better align placement processes with their institutional locations and programmatic values (e.g., Kelly-Riley & Elliot, 2020; Nastal et al., 2022a; Sweeney & Colombini, 2024; Toth et al., 2024). This is, in part, due to increasing recognition that "placement testing . . . has significant consequences for students and is their first encounter with college writing instruction" (Harrington, 2005, p. 11). As Harrington (2005) has observed, placement is a fundamentally communicative, rhetorical act: Our placement processes teach students what college literacy practices are and what we value in and through them. Students determine what is important to our literacy programs based on what we show them we care enough to assess. Thus, any effort to incorporate GAI into writing placement must consider how such processes align with and present the local construct of writing—not just as a matter of validity (§2.2.2), but as a matter of communicating our literacy programs' values (see Kelly-Riley & Elliot, 2020) to the students we are greeting at the door (Gilman et al., 2019).

SLOs can both articulate and help enact key dimensions of the local construct of writing. Writing placement processes that include student writing samples, written in response to constructed tasks aligned with target SLOs, afford literacy programs the opportunity to highlight their deepest values and most exciting innovations. Those values are introduced in this initial act of writing assessment and sustained through syllabi, classroom practice, professional development, and programmatic and general education learning outcomes. They indicate what is important to a program. SLOs provide structure: to a sequence of assignments in a class or a sequence of courses in a program; to the feedback provided by peers, instructors, writing coaches, and technologies; to the reflections a student embarks on to inform their revisions, improve their writerly habits, and deepen their learning (see Messina et al., 2023).

Of course, the process of clearly articulating the local construct of writing to develop valid, hospitable placement processes often reveals ways our programs, practices, and pedagogies fall short of the values we espouse (Gilman et al., 2019; Toth & Aull, 2014). At its best, then, the ongoing work of developing, assessing, and revising writing placement processes can be an opportunity to critically interrogate our local writing construct. Building on Cushman's (2016) call to decolonize the concept of validity, Gilman and colleagues (2019) wrote, "Even assessment programs that appear to be theoretically sound and exhaustively validated will continue to yield inequitable and unjust outcomes if the underlying construct of writing enshrines universalist/colonial ideologies that fail to recognize the pluralities of rhetorical resources in students' communities." Channeling the values expressed through position statements, policies, and scholarship in writing studies, we suggest that all literacy programs should work

toward iteratively developing writing placement processes—with or without GAI integration—and corresponding institutional assessment ecologies that further the following values-driven disciplinary commitments: fair labor practices; shared governance; accessibility; culturally sustaining pedagogies; linguistic justice; and responsibility to land, air, water, and other-than-human life. We acknowledge that not all programs share (or, in some states, are permitted to officially express) these commitments. They are, ultimately, *our* values as teachers, researchers, and writers in writing studies who are engaged in the inevitably political work of literacy instruction (Sullivan, 2015; Vieira et al., 2019).

2.3.2 Fair Labor Practices

All writing placement requires labor: student labor; the labor of staff and administrators; and the labor of literacy program faculty involved in designing, maintaining, coordinating, evaluating, and revising assessment practices. Consideration of whether and how to incorporate GAI into writing placement must account for what kinds of labor—whose labor, and at what cost to other labor those people might be doing under what existing reward structures—will be involved at every stage: implementation, ongoing assessment activities, and continuous evaluation and improvement of the process.

For students, the labor of writing placement must be reasonable in terms of time and effort and, ideally, constitute a meaningful learning experience in and of itself. Students involved in the design, piloting, and revision of placement processes should be fairly compensated for that labor, either financially or through forms of course credit. For faculty, staff, and literacy program coordinators, regardless of rank or contingent status, the labor of designing, evaluating, and sustaining writing placement must be adequately compensated and recognized in institutional reward structures. Compensation includes additional pay, release time from other duties, and meaningful support to sustain other aspects of their work required for retention, promotion, tenure, and/or merit-based pay increases. Likewise, faculty and staff whose labor is displaced by the incorporation of GAI into writing placement should receive support finding other meaningful, comparably compensated opportunities. Whenever possible, unions representing faculty and staff should be involved in negotiating labor policies surrounding all aspects of GAI use, including the incorporation of GAI into placement processes.

2.3.3 Shared Governance

The principle and existing structures of shared faculty governance are under direct attack in many states. The hype, crisis rhetorics, and manufactured urgency surrounding GAI in postsecondary education risk becoming a pretext for circumventing faculty's rightful role in institutional deliberation and decision making, particularly in matters that affect curriculum and instruction in their programs. Decisions about whether and how to incorporate GAI into writing placement must

be made through transparent, equitable processes of shared faculty governance that enable all instructors, regardless of rank or contingency status, to participate (Paris, et al., 2025). That principle begins at the program and department level and extends to the academic senate or comparable faculty governance structures at the institution level. Any pressures from within or beyond the department to circumvent these deliberative processes in the name of expedience should be strenuously resisted.

2.3.4 Accessibility

As CCCC's (2020) Disability Studies in Composition: Position Statement on Policy and Best Practices stated, "Disability should be considered in all components of program administration," a principle that extends to writing placement processes, including any possible incorporation of GAI. There is evidence of ableist algorithmic bias in at least some GAI platforms (e.g., Gadiraju et al., 2023), a possibility with which literacy programs must contend in their decision making. No use of GAI in writing placement is acceptable if it creates barriers to access or disadvantages students based on ableist norms (see Kryger & Zimmerman, 2020). Likewise, no use of GAI in writing placement is acceptable if it fails to account for students' differential access to necessary material resources and supports, including time and technology (Hubrig, et al., 2022; Wood, 2017). Accessibility is a core principle in designing any assessment (§2.2), and it applies to access constraints related to factors like socioeconomic status, geographical location, college-going generation, and documentation status, as well as disability.

Literacy programs need affirmative evidence that their approach to GAI-assisted assessment *improves* the accessibility of their writing placement processes. Indeed, contemplating GAI incorporation can be an opportunity to interrogate whether and to what extent existing writing placement processes are accessible and to whom. Whichever path(s) forward a literacy program decides to take (§1.4), the process of considering GAI should advance the program's commitment to developing accessible, anti-ableist assessment ecologies.

2.3.5 Culturally Sustaining Pedagogies

Following Alim and colleagues (2020), Paris (2012), and Paris and Alim (2014, 2017), many literacy programs—particularly those at access-oriented institutions—have sought to develop culturally sustaining curricula and pedagogical practices. Culturally sustaining pedagogy (CSP) is "a critical framework for centering and sustaining Indigenous, Black, Latinx, Asian and Pacific Islander communities as these memberships necessarily intersect with gender and sexuality, dis/ability, class, language, land, and more," and begins by "decentering whiteness and recentering communities" (Alim et al., 2020, p. 261). CSP is "necessarily and fundamentally a critical, anti-racist, anti-colonial framework that rejects the White settler capitalist gaze and the kindred

cisheteropatriarchal, English-monolingual, ableist, classist, xenophobic and other hegemonic gazes"; it "reimagines education not only within the context of centuries of oppression and domination, but critically, draws strength and wisdom from centuries of intergenerational revitalization, resistance and the revolutionary love of our communities in the face of such brutality" (Alim et al., 2020, p. 262). As Alim and colleagues (2020) have asserted, "CSP, at its core, proposes educational contexts as sites for sustaining the lifeways of communities rather than eradicating them" (p. 262).

We quote Alim, Paris, and Wong (2020) at length because we believe they set an important and challenging bar for what our institutional literacy assessment ecologies can be, should be, and very rarely (maybe never) are. Writing studies writ large has begun to imagine what culturally sustaining postsecondary writing assessment, including writing placement, might look like in local contexts (Leggett, 2023; Randall, 2021; Sassi, 2018; Slomp et al., 2025; Tremain et al., 2023). GAI has been extensively critiqued for its algorithmic biases relating to race, gender, and disability, and for the ways it advances White supremacy, settler colonialism, and capitalism (see Sano-Franchini et al., 2025). However, experiments and early studies also show promising possibilities for "using prompt engineering as a means to reprogram these biases for more equitable AI outputs" (Gupta, 2024). In the medical field, new algorithms are being created to correct biases present in the corpora used to train LLMs (Chang & Wiens, 2024).

Fundamentally, CSP cannot be separated from the pursuit of disability, linguistic, and environmental justice, and GAI presents particular potential threats to these pursuits (§2.3.6, §2.3.7). As with accessibility (§2.3.4), we suggest that the process of considering GAI-assisted writing placement should prompt literacy programs to interrogate their entire institutional assessment ecology through the framework of CSP. Any use of GAI in writing placement must advance the program's commitments to CSP—again, a very high bar. Whatever path(s) the program takes (§1.4), contemplating GAI-assisted writing assessment can and, in our view, should move that program toward more culturally sustaining assessment ecologies.

2.3.6 Linguistic Justice

An essential component of CSP is the ongoing pursuit of linguistic justice for oppressed, marginalized, and colonized communities in the face of longstanding institutionalized White language supremacy (Baker-Bell, 2020; Baker-Bell, et al., 2020). Likewise, much of the ethical turn in writing assessment (§2.2.1) has focused on naming and challenging the ways writing assessment has often functioned as a technology of middle-class White language supremacy (Inoue, 2009, 2012, 2015; Inoue & Poe, 2012; Poe et al., 2018; Randall, 2021). Arguably the most significant movement in writing studies over the last fifteen

years has been the ongoing effort to fulfill the long unmet promise of CCCC's statement on *Students' Right to Their Own Language* (1974) and to center linguistic justice as a core disciplinary commitment. In their examination of the relationships between linguistic and algorithmic justice, Fernandes and McIntyre (2025) summarized this commitment as follows:

[A]s literacy educators, we have a specific set of responsibilities to our students, including the responsibility to help them develop the tools that allow them to achieve their own rhetorical goals and express themselves in the forms and in conversation with the communities that are meaningful to them. We have a responsibility to understand the language we teach, including how its varieties and dialects function, how that language is acquired, and what it means to support the linguistic expressions of all the students whom we teach. (p. 67)

The most prominent GAI products, which enact default standard English ideologies and often erase or mischaracterize the diversity of U.S. and global Englishes (Bender et al., 2021; Byrd, 2023; de Roock, 2024; Fernandes & McIntyre, 2025; Owusu-Ansah, 2023; Sano-Franchini et al., 2025; see also §3.3.1), undercut this commitment. Sano-Franchini and colleagues (2025) have observed that, although the field has often failed to live up to its commitment to linguistic justice, "Writing studies as a discipline stands against linguistic homogenization, which is accelerated and advanced by GAI." Any consideration of GAI-assisted writing placement must confront this threat head-on.

Again, the prospect of incorporating GAI into placement should prompt literacy programs to interrogate the extent to which their local construct of writing and current assessment ecologies do (and do not yet) challenge White language supremacy, resist linguistic homogenization, and advance linguistic justice (an aspect of transformative effects; §2.6). As Gilman and colleagues (2019) have suggested, "If the valued local construct of writing inherently privileges middle-class White languaging and literacies, then efforts to ensure fairer writing placement hinge on changing and pluralizing that local construct." The design, application, and ongoing evaluation of particular GAI instruments for writing placement must center programs' commitment to linguistic justice. No approach to GAI for writing placement is acceptable if it advances White language supremacy or devalues the diverse linguistic repertoires of students and communities.

2.3.7 Responsibility to Land, Water, Air, and Other-Than-Human Life

Many discussions of GAI literacy treat its well-established environmental harms—its demands for energy and water, its damaging effects on communities already subject to generations of environmental racism and (settler) colonial extraction (Bender et al., 2021; Crawford, 2021; Edwards, 2020; Sano-Franchini et al., 2025;

Valdivia, 2024)—as a marginal or temporary concern. We argue that analysis of environmental impact should be central to any consideration of GAI-assisted writing placement (see §3.3.2).

Literacy programs have a responsibility to understand how specific GAI models, products, and training methods affect land, air, water, and other-than-human life (see Price & Chao, 2023, for a discussion of this and other critical terms for discussing planetary lifeforms beyond the human), as well as the human communities who live in interdependence and kinship with them. The energy and water demands and sociogeographical siting of GAI facilities can vary considerably by model and platform, and some GAI products are more transparent about their environmental impact than others. Thus, environmental considerations can shape not just whether a program chooses to incorporate GAI into writing placement, but which combinations of platforms, models, and prompting methods it adopts. Such decisions include whether to use locally-developed small language models (SLMs) rather than more energy-intensive general LLMs (Sauer, 2025); whether to forego using products that decline to provide transparency about carbon emissions and other environmental impacts (TED, 2023); and how to approach model training (§3.2). On the other hand, at least one study found that total carbon emissions for GAI-generated text are lower than those for comparable human-produced writing (Tomlinson et al., 2024). This finding suggests that programs should account for the potential environmental costs of human writers *and raters* in particular placement processes and weigh those costs in relation to the outcomes and potential benefits—personal, pedagogical, and programmatic—of various forms of human participation in the assessment process.

One reviewer of this white paper suggested we were subjecting GAI to environmental scrutiny that other now-ubiquitous digital technologies (e.g., video-hosting platforms) have not faced. In our view, such observations are not grounds for ignoring the environmental harms of GAI, but rather a wake-up call to critically examine the environmental costs of all our technologies (Berthelot et al., 2025), digital and otherwise, as applied in (and beyond) our assessment processes. We argue that environmental impact—fundamentally, a matter of *consequences* for human and other-than-human communities—should be a central component of validity evidence for *any* assessment, and we need more frameworks and tools for making these determinations (see Hosseini et al., 2025; §2.2.2; §4.3).

Principle 4: Respect individuals and communities.

Any use of GAI in writing placement must be ethical, transparent, and consensual for students and communities.

2.4 Principle 4: Respect Individuals and Communities

As literacy educators, we believe our communities sustain us, and we contribute to sustaining our communities. Our commitment to educational access includes helping students understand and explore ways they do and can shape meaning. This work must be grounded in a deep respect for individuals and communities; otherwise, we risk extending the colonial practice of imposing our will upon others (Villanueva, 1997). As Iris Marion Young (2011) asserted, "those who participate in the production and reproduction of structural processes with unjust consequences share a responsibility to organize collective action to transform those structures" (p. 184). Centering respect for students as individuals and respect for students' communities in any consideration of GAI-assisted writing placement is one way we can transform institutional structures. Respect can and, we argue, should include direct collaboration with students and communities in the development and ongoing assessment of any writing placement process.

2.4.1 Ethical Research Guidelines

As any institutional researcher, psychometrician, or literacy program coordinator knows, assessment—including writing placement—is fundamentally an inquiry-driven research project. We begin with a question (e.g., How well are students learning the stated outcomes for a course?), gather evidence, draw conclusions, and communicate the results (§5.0; Figure 8). Researchers are bound to uphold both compliance standards and ethical conduct (see Israel & Hay, 2006), including the following principles: The Nuremberg Code (1947); The Declaration of Helsinki (2024); The Belmont Report (1979), and The Common Rule (US 45 CFR 46, 2020). The Belmont Report promotes respect for persons, beneficence, and justice as its three central tenets to protect the rights of people involved in research (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979). Application of these tenets in informed consent, assessment of risks and benefits, and subject selection is essential to ethical research conduct.

Writing placement matters in part because of its scope: First-year writing classes are required for nearly all postsecondary degree programs and are among the most highly attended courses at most institutions. Even modest changes in writing placement practices can have seismic local implications, including affecting prerequisites for other classes in a general education curriculum or enrollment at the institution as a whole. As the community college faculty co-authoring this white

paper have witnessed, some prospective students will *avoid enrolling at a college altogether* when community members share negative experiences with the writing placement process. It is, therefore, essential to research the impact of a practice as well as any changes to that practice.

Literacy program coordinators are often required to report on the assessment of course and program learning outcomes at regular intervals to meet accreditation requirements; however, writing placement inquiry may not have a routinized schedule. Ethical research must be designed to capture both immediate and long-term effects of placement decisions. Regular monitoring of the practice throughout a term—with annual research into fairness, validity, and reliability—is one important step. Placement outcomes should be studied in relation to broader measures of student success, including course completion, concurrent and subsequent GPA correlations, persistence, and credential attainment rates. Coordinators may also find insight when investigating Integrated Postsecondary Education Data System (IPEDS) cohorts, looking into three- and six-year graduation rates at two- and four-year institutions. Findings must be shared transparently with students, faculty, and institutional leaders in accessible formats so evidence of validity, reliability/precision, fairness, and equity impacts is widely available. Such distribution is not only a matter of scholarly rigor but also a matter of accountability to the communities most directly affected by placement practices. Potential consequences of GAI in writing placement are explored in detail in §4.1 and Figures 4, 5, and 6.

Ethical research further requires careful handling of student data. All personally identifying information (PII) should be removed from student writing in any placement process: doing so is essential to protecting students' privacy and mitigating bias. GAI raises additional concerns regarding student information because of the risk that identifiable user input will be appropriated to train language models. If a literacy program is adopting GAI-assisted writing placement, they must ensure all GAI platforms are FERPA compliant with licensed guarantees that input will not be used to train LLMs. Those involved in the placement process must ensure student PII is removed from student writing, especially before inputting that writing into any GAI platform for local training (§3.2) or placement assessment purposes. One possible practice is to run named-entity recognition algorithms that redact PII (Rudniy, 2018) prior to any interaction with large or small language models.

2.4.2 Transparency and Informed Consent

The Nuremberg Code, the Declaration of Helsinki, the Belmont Report, and the Common Rule center the importance of informed consent in human subject research, drawing attention to how people have been abused, harmed, and killed in the name of research. Individuals must know they are participating in research studies—and affirmatively agree to that participation—and

researchers must inform individuals of the risks and benefits of their participation as part of the consent process. The era of Big Data and now GAI present specific challenges to transparency and informed consent in the realms of biometric privacy, medicine, and intellectual property (e.g., Andreotta et al., 2022; Cohen, 2019; Pletz, 2025). Notably, Anthropic recently agreed to pay $1.5 billion to authors in settlement of a copyright lawsuit (Metz, 2025) because it did not pursue their consent before using their books to train its GAI platform Claude.

In writing studies, norms and expectations surrounding GAI use disclosure are in flux. At present, scholarly journals and presses typically require authors to identify whether and how they have used GAI in manuscript preparation. Some require reviewers to agree they will not interact with GAI or input an anonymous author's work into a platform. Others have used GAI to assist with copyediting, sometimes with no disclosure to authors. Individual instructors, programs, and postsecondary institutions might have varying expectations for disclosure surrounding student writing. The MLA–CCCC Joint Task Force on AI (Adisa et al., 2024a; Adisa et al., 2024b; Byrd et al., 2023; Flores et al., 2025) has asserted that students have an ethical responsibility to disclose whether and how they have used GAI in the process of writing their assignments. Reciprocally, if we are to construct a learning space that respects students as partners in shaping their education and places for learning, then we must respect students enough to offer them transparency and choice in how they interact with GAI. Literacy programs have a responsibility to disclose whether and how they use GAI to assess student writing, including in the context of writing placement.

By principles of transparency and consent, any GAI-assisted writing placement process must clearly explain to students:

- That GAI is part of the process.
- How GAI is being used, including what evaluation criteria are included or explicitly excluded from GAI-assisted assessment.
- What evidence the institution has gathered and continues to collect regarding the fairness, validity, and reliability of the placement process, including any evidence of equity gaps and explicit documentation of environmental impact based on the specific GAI platforms and training methods used.
- Whether and how the institution will use student writing and any other data students provide, including whether student writing will be used to train local GAI instruments.
- What measures are in place to protect student writing and any other data students provide from being used by tech companies to train LLMs or profit in other ways.

Transparency measures like those outlined above are essential for students to consent meaningfully to GAI-assisted writing placement processes. Furthermore, we argue students should be given

the explicit opportunity to opt out of GAI-assisted placement and request an assessment option that is fully human-evaluated.

2.4.3 Student Use of GAI

Any incorporation of GAI into writing placement must consider how students might use GAI during their writing processes. If students employ GAI to draft, revise, or enhance writing they submit for placement assessment, institutions must examine how these practices shape the writing samples being assessed, the nature of any formative feedback, and the fairness of any GAI-assisted evaluation (e.g., whether GAI assessment instruments systematically prefer writing with features most similar to the text they produce; see §3.3.1). All placement processes, whether or not they incorporate GAI, should be designed with an awareness of how students might be using GAI so those assessments reflect the complexities of students' writerly development, including their critical GAI literacies (e.g., Kim & Chon, 2025). Ongoing validity studies must also account for how evolving student and faculty uses of GAI, including the introduction of new GAI-aware SLOs, might be reshaping the local writing construct in ways that the placement process should reflect. Recognizing and addressing this reality is essential for ensuring placement decisions continue to be aligned with the goals and values of writing programs. Further discussion of academic integrity related to GAI is discussed in §3.3.3.

Principle 5. Attend to emerging technologies and relationships.

Understanding emerging relationships between humans, GAI, and writing placement helps facilitate principled human-computer relations in teaching and assessing writing.

2.5 Attend to Emerging Technologies and Relationships

GAI technologies are proliferating and changing, and it is uncertain whether and when their abilities might plateau. There are relatively few definitive statements about GAI capabilities, risks, and harms that we can be sure will hold constant in the coming years. Thus, scholarship and practice surrounding GAI in writing placement must proceed with the recognition that GAI and other AI technologies continue to emerge and evolve. They might change in ways that directly affect other principles—for example, programmatic values relating to fair labor, accessibility, culturally sustaining pedagogy, linguistic justice, and environmental justice—and these effects, in turn, might reshape local arguments regarding fairness, validity, and reliability in particular use cases.

Principled incorporation of GAI in writing placement requires investing dedicated resources into ongoing efforts to stay current with GAI technological developments as well as emerging

GAI-related scholarship. Such resources include, but are not limited to:

- Compensation for the time and energies of literacy program coordinators and faculty who must continuously learn about these technologies alongside their other teaching, service, and research responsibilities
- Strategic new faculty hires
- Ongoing professional development opportunities
- Extensive institutional technology support

It is unsustainable and unjust to expect literacy program coordinators to shoulder this new, never-ending labor as an individual, uncompensated responsibility. Recognizing that GAI technologies are still young and evolving is also reason to proceed carefully when considering the possibilities for GAI-assisted writing placement. If efficiency is a strong institutional motivator, everyone involved must be aware that it is inefficient to put extensive resources (time, labor, and money) into designing and implementing a large-scale GAI-assisted placement process that could quickly become obsolete. Likewise, the resources required to develop and sustain a principled GAI-assisted placement process might outweigh whatever efficiencies institutions thought they would gain by incorporating GAI.

Finally, we suggest that the emerging human-computer relationships at the heart of GAI-assisted writing placement must be tended consciously and ethically at every stage of development and implementation. Calibrating GAI's capacity to assess depends on an iterative, rhetorical process of human feedback, what Graham (2023) described as "dialogic engagement with the LLM and its outputs" (p. 166). Through such dialogue, humans and GAI are co-constituting our written texts, our writing assessments, and each other. The field is only beginning to imagine what responsibilities such relations might entail. Turning to Indigenous theories and praxis lends insight. As Lewis, Arista, Pechawis, and Kite (2018) have explained, a core principle of Indigenous epistemologies is that "man is neither height nor centre of creation"; rather, "Indigenous communities worldwide have retained the languages and protocols that enable us to engage in dialogue with our non-human kin, creating mutually intelligible discourses across differences in material, vibrancy, and genealogy" (Lewis et al., 2018). Anticipating this moment when AI is permeating many of our communicative acts, Lewis and colleagues (2018) have called on readers to consider how specific Indigenous epistemologies rooted in relationality can contribute to "conceptual frameworks that conceive our computational creations as kin and acknowledge our responsibility to find a place for them in our circle of relationships" (p. 4). We do not have to anthropomorphize GAI to reach toward ethics of "respect and reciprocity" (Lewis et al., 2018, p. 13) in our emerging relationships with these machines—nor do we have to sacrifice our humanness or the interconnected nature of our being.

Principle 6. Be open to transformative effects.

Because GAI could (and maybe should, and hopefully will) transform conceptualizations of the purposes and processes of writing placement and interrelated pedagogies, we should look for and amplify effects, anticipated or otherwise, which have positive value for students.

2.6 Be Open to Transformative Effects

Despite well-intended reform efforts by generations of scholars, writing placement nearly always functions to sort student writers into institutional places that have been categorically determined for them. This sorting still has a "diagnostic" logic, seeking to identify what ails a student linguistically and prescribing a remedy through a course sequence that ultimately pronounces them "college-ready." In our view, such patriarchal and binaristic systems have never served our students, our writing programs, our institutions, or even our society very well. These persistent disappointments might be because longstanding approaches to placement have prioritized assessment *of* learning (see §2.6.1) according to one restrictive set of linguistic practices and values: i.e., standardized written English (SWE) and normative constructs of academic writing "conventions." When students are invited into the writing placement process, we begin to honor their experience, knowledge, and beliefs. We create an opportunity to learn together about the gifts our students bring, how the institution might support and develop them, and how students might help us transform.

2.6.1 Assessment for *Learning*

Rather than exclusively prioritizing assessment *of* learning, we strive to enact assessment *of*, *as*, and *for* learning. We view assessment as a public articulation of individual, programmatic, institutional, and disciplinary values (§2.3.1). We concur with Heritage and Wylie (2018) that assessment's purpose "is to inform learning, as it is taking place, in order to advance the learning from its current status" (p. 730). They identified four features of assessment *for* learning (A*f*L): communicating learning goals to students, intentionally obtaining evidence of learning through a variety of means while learning is taking place, student self-assessment, and timely feedback from teachers and peers that supports students in taking action to advance their learning toward the desired goals (Heritage & Wylie, 2018, pp. 730–731). Across these features, assessment activities are in themselves meaningful learning opportunities for students; they are also opportunities for dialogue with teachers and mutual adaptation. Formative assessments allow students to consider what, how, and how well they may be understanding a concept; summative assessments may do the same *and* provide instructors with information to adapt to in response. We invite readers to imagine how writing

placement—with or without GAI—can better enact these principles, encourage student self-reflection and dialogue, and embrace the potentially transformative programmatic effects of these reconsiderations.

2.6.2 From Placement to Place-Making

Eminent educational measurement scholar Edmund W. Gordon offers a foundational premise we find applicable to writing assessment research. Reflecting at the age of 102 on his own legacy and on the field, Gordon asserted, "Measurements ought not only be about measuring developed ability. Measurement also has the potential to contribute to the cultivation and development of ability. . . . educational measurement should inform and improve learning, *inform and improve learning*" (Mark et al., 2023, p. 184; emphasis in original). Yes, we are required to report on how well students are achieving the course and program learning outcomes our departments have identified, but we do writing assessment because it offers us insight into the teaching and learning taking place in our local contexts. At the heart of writing assessment is this recognition: Students are offering us insight into the ways in which they use written communication. Curiosity about students' literate lives and practices should inform work to improve the teaching and learning taking place in our environments. As we contemplate the emergence of GAI and the possibilities it presents for writing placement, we should ask whether and how this technology can create new opportunities to learn about our students' writing in the context of their lives, their communities, and their hopes for the future.

As we have grappled together with the magnitude of the challenges GAI presents to existing writing placement processes, we have increasingly found ourselves asking bigger questions about the assumptions and purposes that authorize such practices at all. We turn, again, to Lewis and colleagues (2018) and take up their charge "to open up our imaginations and dream widely and radically about what our relationships to AI might be" (p. 13). If GAI has generated (yet another) existential crisis in writing studies, we invite colleagues to take this opportunity to consider shifting from *writing assessment for placement* to what we are calling *place-making through writing assessment*.

Several coauthors of this white paper have spent the bulk of our careers examining how students experience our postsecondary educational literacy sites and experimenting with methods to understand how we might transform them to become more relational instead of hierarchical, more responsive to the individual learners with whom we work and the communities in which we are situated. Inspired by community practice and ways of knowing, as well as sociocultural and political frameworks, we understand place-making to be about how "people experience the concrete, material dimensions of place affectively and cognitively (Entrikin, 2003), but these understandings of place are also constructed, and especially communicated, through social

negotiation, including conflict and difference" (Pierce et al., 2011, p. 60). *Place-making through writing assessment* means moving away from sorting logics and instead creating assessments with incoming students that inform efforts to continuously (re)make our institutional and programmatic places to serve those students equitably, honor their dignity, and advance social justice. Through this framework, the disruption precipitated by GAI can become an opening for change, helping students and instructors rethink the purposes of placement in writing instruction and student success. Instead of a verdict, writing placement can become a moment of reflection, a conversation, and a catalyst for meaningful growth —for students, but also for teachers and programs. We explore these possibilities further in §5.4 Beyond Scoring: Transformative Uses for GAI in Writing Placement.

Section 3 • GAI, Writing Analytics, and Writing Assessment

3.0 Overview

Every fall, millions of students participate in writing placement processes, sometimes through standardized tests like ACCUPLACER (Isaacs, 2018; Scott-Clayton, 2012), sometimes through locally developed student self-placement (SSP) processes that require extensive writerly self-reflection (Messer et al., 2022; Toth, 2018; Toth et al., 2024), sometimes through high school transcript evaluations requiring no active student participation at all. Each placement process carries distinct affordances and constraints. Even the most carefully constructed standardized exams often account for less than 10 percent of the variation in students' actual writing performance (Godfrey, 2024). For instance, using natural language processing (NLP) modeling techniques, Godfrey (2024) compared nearly 50,000 student essays across a ten-year period at the University of Michigan with their test scores on SAT, ACT, and AP exams and found only weak relationships, recalling Nastal and colleague's (2022b) conclusions about the COMPASS exam's limited predictive ability.

Writing well involves more than mastering mechanics or surviving the stress of a timed test; it emerges from an iterative process of reflection, conversation, and growth (Badenhorst et al., 2015; Beigman Klebanov et al., 2019; Lillis & Scott, 2007). As colleges experiment with GAI's possibilities for assessment, it is wise to recognize that it too has affordances and constraints. It can be wielded reductively as a blunt instrument of judgment, or it can be incorporated as a formative, reflective support that helps students better understand their own writing and makes placement a transformative starting point rather than a fixed verdict. Making informed critical decisions about whether and how to incorporate GAI into writing placement in ways that align with programmatic values requires understanding how this technology works and what its current capabilities are.

We begin this section with background information about the corpora and datasets used to train LLMs, offering information for the reader to make connections about ways in which LLM pre-training may affect GAI performance. We then move into basic information about how GAI platforms can be trained to perform more specialized tasks. Finally, we identify additional considerations that might impact decision making regarding the use and training of these technologies for writing placement.

3.1 Corpora and Datasets Used for Training

3.1.1 Text Corpora

The act of collecting, duplicating, and distributing corpora of texts has taken place since at least the 1440s with the Hebrew Bible Concordances (Ben-Shalom, 2011). Individual corpora have been

emerging over more than 500 years and have been digitized over the last five decades. In the 1960s, larger corpora designed for linguistic research became available, including the Brown Corpus (Kučera & Francis, 1967) and its British counterpart, the Lancaster-Oslo-Bergen (LOB) Corpus (Leech & Smith, 2005). Many corpora have been assembled since, including linguistically annotated corpora and a growing number of multilingual translated corpora (Palmer & Xue, 2010; Zanettin, 2014). Annotated corpora, multilingual translated corpora, and unaltered corpora can be useful in prompting existing GAI platforms (§3.2.1), integrating retrieval-augmented generation (RAG) systems (§3.2.2), and finetuning LLMs (§3.2.3). To pre-train a model, however, a very large number of corpora are needed to serve as examples of language that can be analyzed for calculating word prediction (§3.2.4).

Table 1 shows six of the many digital English-language corpora currently available. As of 2025, four of these corpora—Project Gutenberg, Enron Email Corpora, BookCorpus, and BookCorpus3—have been used to train some mainstream LLMs (Gao et al., 2020).

This range of existing corpora used—and not used—to train widely used LLMs demonstrates the very human, culturally-mediated selection process involved in constructing these platforms (§2.5), which is inevitably shaped by ideologies and asymmetries of power. Importantly, it is unlikely the corpora in Table 1 include scored or graded essays written by students in formal educational settings.

While LLMs rely on language patterns rather than grades for pre-training, using GAI to evaluate writing requires assigning value to various language patterns in full texts or traits of texts. Without detailed guidance, the GAI instrument will likely assign the highest value based on the texts with which it was pre-trained, and, as Wataoka and colleagues (2025) noted, favor its own writing style. We cannot overstate the importance of understanding model pre-training corpora and datasets when considering incorporating GAI into writing placement: LLMs pre-trained primarily on texts produced through professional editorial processes have a high likelihood of encoding middle-class White language supremacy. To counteract this tendency, it is important to prompt, finetune, and otherwise inform LLMs using texts that exemplify the range of potential traits and values the GAI placement instrument will assess (see Byrd, 2023). Below, we discuss some of the corpora and datasets that have been used to pre-train LLMs, some of which do include small corpora of graded essays.

Table 1. *Digital Corpora of English-Language Texts*

Year	Corpus Title	Number of Items	Reference
1971	Project Gutenberg	75,000+ books	Project Gutenberg, 2026
2000	Enron Email Corpora	619,446 messages	Klimt & Yang, 2004
2004	American National Corpus	22 million words	Reppen et al., 2005
2005	Corpus of Late Modern English Texts-Extended Version (CLMETEV)	176 texts from 1710–1920	Diller et al., 2011
2015	BookCorpus (from Smashwords–free books not yet published)	11,038 books More than 20k words; 16 genres	Zhu et al., 2015
2020	*Books3	190,000+ books	Rowberry, 2025

*This corpus has been under scrutiny for violating copyright in its LLM pre-training methods (Rowberry, 2025).

3.1.2 Large Datasets

In machine learning, *corpora* are considered to be text-based sets of information and bodies of knowledge, typically used for linguistic analysis. They may be large quantities of texts that have been annotated, translated, and otherwise cleaned data used to inform language models, and they may also include raw or untreated texts. *Datasets* used for LLM pre-training consist of various corpora but also other types of data, such as numbers, audio, video, and images. Some GAI developers have been transparent about their selection of datasets for pre-training, while others have been less so, and some of their decisions are working their way through the legal system as we prepare this white paper. Notably, some developers have used The Pile, a very transparent dataset established in 2020, for training many LLMs (Gao et al., 2020). The Pile consists of 22 datasets covering the areas of webpage language, including a Common Crawl dataset from 2008 to 2020, which performed monthly crawls of raw webpages, metadata, and text extractions (commoncrawl.org). The Pile additionally includes Wikipedia entries; internet talk; academic writing primarily in the fields of math, computer science, physics, and philosophy; medical writing; legalese; creative writing/stories and other books; multilingual translations; and examples of math and coding. The Pile offers a significant amount of data for language models to use for generating writing, coding, and mathematical problems and solutions.

GAI performance depends on the size and types of samples in the pre-training datasets. For instance, Gao and colleagues (2020) have explained that when creating The Pile, they included Deep Mind Mathematics—which includes problems, processes, and answers—because previous iterations of LLMs had performed poorly on mathematical tasks. Importantly, although Gao and colleagues have not disclosed details, we do know The Pile consists of some datasets derived from GitHub.com. GitHub includes several corpora of graded writing including the Automated Student Assessment Prize (ASAP) dataset used for the Hewlett Foundation's 2012 AWE competition, the English Language Learner Insight (ELLIPSE) dataset, PERSUADE, and others; however, these corpora—which are often produced in K-12 settings and/or include short-answer responses rather than essays or are scored for primarily grammatical issues of English language learners—are exceedingly small compared to the overall amounts of data used to pre-train LLMs. Consequently, the ability of any LLM pre-trained on The Pile to consistently and accurately score college essays on specific features demands close scrutiny.

3.1.3 Small Datasets

GAI developers used very large datasets to pre-train LLMs in their efforts to create models that *generate* realistic conversation. For LLMs or small language models (SLMs, §3.2.4) to *evaluate* writing, the model needs some way of valuing various features of the writing; for instance,

particular genre features or specific kinds of rhetorical moves valued within the local construct of writing. GAI might mirror human communication, but these platforms do not think or reason; they use mathematical calculations and parameters set by humans to generate text that aligns with the texts humans chose to pre-train models. Without additional training or prompting, GAI would need to use the writing included in its pre-training dataset as a standard of "good" writing. LLMs have thus far not been pre-trained with very large sets of writing that identify, for instance, the ways in which texts do or do *not* effectively contextualize quotes or qualify claims or demonstrate metacognitive practices deemed to be "good" by many writing instructors in postsecondary U.S. contexts. Such qualities need to be explicitly identified and valued for the LLM to recognize them. Small datasets that present or exemplify value attached to specific features could be used to pre-train SLMs, to finetune existing LLMs, to integrate a RAG system into a GAI instrument, or to prompt existing GAI platforms (§3.2).

3.1.4 Local Corpora and Datasets

The TYCA White Paper for Placement Reform (Klausman et al., 2016) asserted the importance of local contexts in developing and validating writing placement practices. This insight extends to using local corpora and datasets to train GAI-assisted writing placement instruments. Doing so helps ensure that an instrument is calibrated toward features that align with the local construct of writing with enacted traits representative of the local student population. For writing placement processes that involve scoring student writing, for instance, actual local example texts of all the different scoring points for each trait, rubrics explaining the expectations for each score, or both might inform the instrument through any of the training methods outlined below. These scored samples give the GAI platform items with which to compare the texts it is evaluating. It is also possible, however, that training language models with both writing samples and rubrics could contribute to machine confusion; further research on these practices is needed (see Rudniy, 2024).

At present, we cannot say with any certainty whether informing an instrument with locally-developed corpora and datasets will yield fair, valid, and reliable results. Biases can be evident in any writing evaluation practice, whether designed locally or externally, and whether evaluated by humans, machines, or a combination of both. Fairness depends, in part, on the language usage and language ideologies (see §3.3.1) embedded in the local construct of writing in addition to the literacy program's sociocultural attunement. Further, compiling strong evidence of validity and reliability depends in part on consistency within a literacy program—whether instructors within the program are in alignment regarding their evaluation practices—so that the scored samples accurately represent the ways writing is evaluated across the program.

3.2 Understanding GAI Options

We can envision different ways GAI could be incorporated into writing placement assessment: formatively, summatively, or both (see Table 5; §5), all of which may offer rich research opportunities. The development of AI approaches to writing assessment has typically been controlled with very narrow and specific focus, whereas using "off-the-shelf" GAI for these purposes is not, primarily because they have been developed by for-profit entities with far different goals than nonprofit educational organizations centered on improving teaching, learning, and educational access.

A number of longstanding AI platforms have been designed to provide *formative* feedback and are available as open-source or paid products (e.g., eRevise, Docuscope), many of which are now incorporating GAI (e.g., Docuscope's My Prose, ETS's eRater, Grammarly). Several recently introduced platforms are specifically designed to provide GAI-based formative feedback to students (e.g., Kahn Writing Coach, Winnow). These products are designed to provide writing feedback; they are not open-source GAI chatbots or multimodal agent applications, though some may well be in development. Current open-source GAI platforms, such as Google Gemini, can achieve some of the goals of these specialized formative assessment applications, particularly when incorporated into peer review (Sperber et al., 2025) or used for scaffolding writing assignments (Barney, 2023). Preliminary research also offers some promise for using open-source GAI platforms for formative feedback to students in the classroom (Dai et al., 2024; Guo, 2024; Huang et al., 2025a; Sperber et al., 2025). These studies suggest formative assessment might be a viable use for GAI and could contribute to reconceptualizing writing placement as place-making (§2.6.2; §5.4).

AWE instruments trained with AI to provide *summative* feedback have a long history, from Ellis Batton Page's Project Essay Grade (PEG) in 1966 to contemporary work from different entities supporting AWE for large-scale administered writing tests in post-secondary educational admissions, like ACT, GMAT, and TOEFL (Dikli, 2006). Due to its recent development, GAI for summative writing assessment remains underresearched. Following the direction of formative feedback products, existing summative writing assessment products could be augmented with some of the capabilities of LLMs. At this stage, however, researchers are studying the capabilities of existing LLMs and their various GAI interfaces to evaluate and score writing (e.g., Flodén, 2025; Lan et al., 2025; Li, et al, 2025; Mansour et al., 2024; Naismith et al., 2023; Pack et al., 2024).

GAI could be trained to evaluate essays in several different ways, four of which we explore below (§3.2.1-4). In this context, *training* is a complex term with sometimes overlapping meanings; in this white paper, we are using it to refer to the following approaches: *prompting* GAI platforms (§3.2.1), *integrating* RAG systems deep within the

prompt (§3.2.2), *finetuning* large or small language models (§3.2.3), and *pre-training* large or small language models (§3.2.4). Prompting gives words and contexts for applications to use within a single session. Integrating a RAG system inserts applicable dataset sections into the prompt for an LLM to use for response each time, and finetuning and pre-training both change the long-term functions of the GAI platform and large or small language model. Pre-training, finetuning, or integrating a RAG include introducing datasets (§3.1); prompting can also include introducing datasets if desired. After offering a brief comparative consideration of the potential demands and capabilities of these options, we present simplified definitions of each kind of training and explore some of their possible uses in GAI-assisted writing placement.

The technical expertise, cost, and capability of each training option are important considerations for literacy program coordinators and multidisciplinary teams when making decisions in their local contexts. Table 2 illustrates the rising level of technical expertise correlated with the higher cost (Cottier et al., 2024) of each option and the potentially lower validity associated with options that require less technical expertise. Each of these options would draw on datasets tailored to local constructs.

Table 2. *Technical Expertise, Financial Cost, and Anticipated Reliability of LLMs by Training Method*

Training Method	Technical Expertise	Financial Cost	Anticipated Reliability of Output
Pre-training an LLM	Very high	Very high	Low
Pre-training an SLM	High	High	High
Finetuning an off-the-shelf GAI	Medium	Medium	Medium
Integrating a RAG with an existing LLM	Medium	Medium	Medium
Prompting an off-the-shelf GAI platform	Low	Low	Low

As noted above, more research is needed to determine how to most effectively calibrate a GAI-assisted writing placement instrument to provide feedback, scores, or evaluations that have strong evidence of fairness, validity, and reliability within a local literacy program. While scored writing samples or detailed rubrics may work well enough, they also present distinct risks that compound decades of research detailing the harms done when assessing students' writing. For instance, without proper training or context, sets of scored example essays might prompt the GAI instrument to inaccurately identify features like essay length, spelling, or punctuation as indicators for writing placement. Detailed rubrics themselves might become the writing sample the GAI instrument measures student essays against. It is worth exploring various methods of combining these two forms of training data, along with others. Additionally, practitioners will need to explicitly train GAI instruments *not* to value or devalue features of the writing that could otherwise contribute to bias—for example, linguistic markers of marginalized English varieties, or student-volunteered information like disclosure of a disability or reference to attending a prestigious high school.

As Table 2 indicates, with the exception of pre-training LLMs, required technical expertise and financial costs increase in relation to the likely reliability of output. Pre-training an LLM for writing placement is noted as "low" anticipated reliability of output to indicate it as an outlier here because the number of student texts, rubrics, and other contextual information from the literacy program needed to pre-train an LLM would be close to impossible to collect: terabytes of data or trillions of artifacts (§3.3.1). Instead, if a postsecondary educational institution chooses to incorporate GAI into writing placement, using a smaller dataset with one of the other methods listed—pre-training an SLM, integrating a RAG, finetuning an off-the-shelf GAI, or prompting an off-the-shelf GAI—to provide the instrument with information about the valued traits of the local writing construct might be more feasible. Those with higher budgets and larger multidisciplinary teams—with experts, for instance, in sociocultural studies, writing pedagogy, writing assessment, psychometrics, cognitive science, machine learning, computational linguistics, and cloud computing—might want to attempt pre-training an SLM or finetuning an LLM. Those with very little collaboration or support might want to work only with prompt engineering or decide to opt out of GAI-assisted placement. Those with some support or who are interested in learning technical practices might want to explore creating a script for prompting or integrating a RAG system with a dataset of local student texts, instructor assignments and scoring guides, course and program learning outcomes, and annual assessment reports. When using any of these options, however, it is critical to keep a human-in-the-loop (§5.3.4); that is, to incorporate human judgment or feedback at key points in complex GAI-assisted processes to improve performance and mitigate bias (von Davier & Burstein, 2024),

and to engage iteratively in the assessment cycle (§5.0).

3.2.1 Prompting Existing GAI Platforms

When prompting GAI for writing placement, several important considerations arise. The rapid changes to open-source GAI platforms are likely to continue changing their functionality, potentially with no or little notice to the user. These changes can radically alter how prompting open-source GAI platforms can or should be used for writing placement and necessitates frequent review (§2.5). Current LLMs used to power open-source GAI platforms have been pre-trained using corpora that included few evaluated essays or rubrics for assessing writing (§3.1.2) but do reflect the linguistic, stylistic, and cultural biases of their creators and developers. The practice of relying on these models to evaluate writing features beyond SWE (modeled as best practice by most datasets used for pre-training) or mood/tone (modeled for best practice by large amounts of narrative work used for pre-training) therefore deserves careful oversight and consideration in any educational setting and particularly in writing placement scenarios. Furthermore, given GAI's capacity to generate new language combinations, and thus possibly new scores, with each output, there is not yet adequate research regarding its ability to assess writing consistently. Practitioners must commit to analyzing GAI output for intra-rater and intra-platform reliability (§5.2.2). LLMs are designed as language generators, not evaluators; therefore, it is vital to maintain human participation in the process (§5.3.4; see also Burstein & LaFlair, 2024; von Davier & Burstein, 2024).

Prompt engineering is receiving research attention across the curriculum (Bansal, 2024; Lee & Palmer, 2025; Micus et al., 2024; Ranade et al., 2024; Yoshida, 2024). More research is needed, however, on prompt engineering as a rhetorical activity (Aguilar, 2024; Graham, 2023; Gupta & Shivers-McNair, 2024), as well as on how to *effectively* prompt LLMs to evaluate writing (Stahl et al., 2024). Current research is exploring the use of prompt engineering as a method for assessing student essays (Li, et al, 2025; Liu et al., 2025; Mansour et al., 2024; Naismith et al., 2023; Pack et al., 2024), with patterns beginning to emerge. The bullet points below list general names for these types of prompts (MIT Sloan Teaching & Learning Technologies, 2025). The parentheticals indicate ways researchers have applied these types to assessments of student writing:

- Contextual prompting (introducing the assignment and the task)
- Role-based prompting (assigning the role of instructor or evaluator)
- Instructional prompting (sharing the rubric and giving scoring instructions)
- Few-shot prompting (providing multiple local writing samples representing scores) (§3.1.4)

Ranade and colleagues (2024) have suggested that all elements of the rhetorical situation need to be present for any prompting task concerning writing

to produce usable output. Well-designed prompts can lead to better output in each session, but it is important to recognize that prompting does not train the LLM. Instead, it provides *in-context learning* (see Gong et al., 2024) by giving information to be used within an individual session, mirroring some of the ways humans learn. For writing placement applications, this need for in-context learning means all of the prompting information needs to be reentered each time an essay is evaluated by a GAI instrument. Such information includes not only contextual prompting regarding the constructed response (CR) task (i.e., the writing prompt students are responding to), but also the local, contextualizing information the GAI-assisted writing placement instrument needs to know to be able to produce usable output, such explicit articulations of programmatic values along with few-shot prompting (using exemplar texts), instructional prompting (including scoring guides and rubrics), and role-based prompting. When inputting large amounts of prompts and information, the process must be monitored by members of the multidisciplinary team for machine confusion, which is when GAI starts responding to incorrect sections of the data.

3.2.2 Integrating a Retrieval-Augmented Generation (RAG) System

Integrating a RAG system into an existing GAI instrument allows the LLM powering that instrument to access a specialized dataset to answer prompts. In the case of GAI-assisted writing placement, that dataset may include scored example texts, trait-based scoring guides with explanations, or both, depending on the local setting and literacy program. Integrating a RAG does not change the core functionality of the LLM; instead, it gives the LLM more data to use to inform the prompt input and the response. In so doing, the LLM maintains its functionality as the producer of the language and the RAG gives the LLM more accurate information to inform its response.

A RAG, at its very basic level, first indexes the dataset by breaking the text down into smaller segments. When the GAI is prompted, the RAG matches the prompting information to the dataset segments that are most analogous. The RAG then integrates the identified segments with the prompt for added context. More advanced RAGs take more steps to enhance segment retrieval and optimize prompts, thus improving output (see Gao et al., 2024, for more discussion of RAG components, functionality, and challenges). Creating a RAG system to integrate into GAI and inform its LLM could prove to be an effective option for building GAI-assisted writing placement processes. Because of its ability to segment the dataset and select the most appropriate segment, a RAG system is likely to produce better output than prompting alone (§3.2.1).

3.2.3 Finetuning LLMs

Another option for incorporating GAI into writing placement is to finetune the LLM that powers the GAI platform and then prompt that newly altered instrument using some of the methods in §3.2.1. Finetuning makes good use of the LLM's capabilities by adding datasets and training for certain tasks. Finetuning can involve several different methods or combinations of methods, including, but not limited to:

- Instruction Tuning: adding an instruction-formatted dataset and/or Chain of Thought (CoT) examples that show the LLM step-by-step how to perform a task atop the language model (Naveed et al., 2024; Won Chung et al., 2022)
- Alignment: freezing layers of the pre-trained LLM to preserve their functionality while adding task-specific training and then giving a higher rating to the task-specific layers (Naveed et al., 2024; Srinivasan et al., 2024)
- Reinforcement learning from human feedback: adding a layer of human preferences for a specific task (Parthasarathy et al., 2024)

Naveed et al. (2024) found that finetuning with additional datasets can improve performance on a specific task but may cause the model to "forget" some of its pre-training (§3.2.4). Linking a smaller dataset to several pre-training samples or using a smaller dataset for finetuning may lead to better performance (Naveed et al., 2024). Finetuning for writing evaluation tasks, however, is still in very early exploratory phases (Lan et al., 2025; Naismith et al., 2023; Yamashita, 2025), and more research is needed to determine what finetuning methods might be most effective for developing writing placement instruments.

3.2.4 Pre-training LLMs and SLMs

LLMs are pre-trained on selected text corpora and datasets, calibrated for meaningful output, and trained for safety (see Ali et al., 2024, for a critique). Some LLMs are pre-trained for specific purposes, including coding, scientific knowledge, music, audio, and finance (Naveed et al., 2024; Zhou et al., 2024). The amount of resources and computing power needed to pre-train a new LLM can be cost-prohibitive (Hu et al., 2024; Strubell et al., 2019), and LLMs may impact social-environmental health significantly (Hosseini et al., 2025; Schick & Schütze, 2021; Strubell et al., 2019; §3.3.2). As discussed in §2.3 and further elaborated in the theory of action (§4.1; Figures 4, 5, & 6), these concerns must be part of the decision making and validation process for GAI-assisted writing placement.

Small language models (SLMs), or "mini-giants," that mitigate some of these negative consequences are on the rise (Hu et al., 2024; Lepagnol et al., 2024). Zhou and colleagues (2024) asserted, "Relying solely on prompt engineering"—required for working with LLMs—"often falls short. Therefore, smaller language models can bring forward great benefits to these industries, offering the much-needed flexibility for adaptation,

allowing them to full leverage the power of AI and thus catalyzing innovation within them." (Zhou et al., 2024, p. 1). Because LLMs are now reasonably functional, developers have been able to decipher which LLM elements require more or less focus while still producing fairly good results, leading to the ability to create SLMs. It is possible to train an SLM on a subset of an LLM (Hu et al., 2024; Sanyal et al., 2024). Yet another option is to pre-train SLMs (Zhou et al., 2024) from scratch (Hu et al., 2024; Lepagnol et al., 2024). Lepagnol and colleagues (2024) showed SLMs can perform comparatively and sometimes better than LLMs on text classification tasks, a capacity that might extend to valuing segments of text or entire essays. Smaller datasets might also be more tailored to domain-specific purposes (Schick & Schütze, 2021; Zhou et al., 2024), which aligns with the need for GAI-assisted writing placement instruments to be able to evaluate writing according to domain-specific criteria (§3.2.3). Again, however, more research is needed to determine the evaluative capabilities of SLMs.

3.3 Other Considerations for Implementing GAI-Assisted Writing Placement

As our discussion of principles in §2 suggests, contemplating the incorporation of GAI into writing placement both amplifies long-standing ethical considerations in writing assessment and presents new considerations, which we discuss below. We offer a detailed heuristic for considering immediate and long-term consequences of GAI-assisted writing placement in §4.

3.3.1 LLMs and Language Bias

One limitation of current LLMs is potential linguistic bias (§2.3.6). Bias is traditionally defined in educational measurement as tasks (such as writing tasks used for assessment) and their related measurement (such as scoring rubrics) that differentially impact the performance of different groups of test takers, resulting in consequences that call into question the fairness, validity, and reliability of the interpretation and use argument (IUA) that justifies score use (§4.3). In statistics, bias is described as systematic error, whether intentional or unintentional, in a test score (AERA et al., 2014, p. 216). If bias is detected—that is, if scores or other feedback on the measurement are shown to be "subject to systematic error that is related to examinee characteristics (such as race/ethnicity), these indices should not be used unless they are modified in a way that removes the bias" (AERA et al., 2014, p. 211). These two definitions and suggested actions are relevant in discussing bias in LLMs, which can manifest in several different forms, including performance and standard bias, representation bias, and training bias.

In fields interested in the development and performance of LLMs, using "LLM-as-a-judge" to evaluate the dialogic output of other LLMs is a common practice (Chen et al., 2025; Wataoka et al., 2025; Zheng et al., 2023). In an analysis of eight separate LLMs, Wataoka and colleagues (2025) found all had a self-preference bias in their evaluations, meaning they preferred language that

was similar to their own capabilities. The models even gave higher rating preferences to language with lower-perplexity similar to their own over more complex and sophisticated human-written text. These findings match earlier biases noted for machine learning (Deutsch et al., 2022), suggesting they are persisting and that we might be building models upon a biased system.

Next, languages like English, Chinese, French, and German, among others, have enough textual data on the internet to pre-train LLMs, which requires "hundreds of gigabytes or even terabytes of data" (Cassano et al., 2024). Other languages, however, such as Hindi, Irish, Navajo, or Swahili, do not have a large enough linguistic representation on the internet to significantly impact pre-training. Similarly, some dialects of English, such as African American Englishes, Appalachian English, or Cockney English, as well as other World Englishes or Entangled Englishes (Lee & Rüdiger, 2025), such as Spanish-English, Vietnamese-English, or Philippine-English, do not have a large enough linguistic representation on the internet to significantly affect model pre-training (Li et al., 2024). Models finetuned for these languages, dialects, and varieties are usually created on top of models pre-trained primarily in the high-frequency language. That practice might leave grammatical artifacts of the pre-training language in the model performance of the target languages (Papadimitriou et al., 2023). In short, LLMs tend to be biased toward languages used in pre-training. Tran et al. (2025) illustrated the disparity between English, which has a higher linguistic representation digitally, and Irish, which has a lower linguistic representation digitally, in two closed-source LLMs (GPT 4.1 and Gemini 2.0 flash) and in two open-source LLMs (Llama-4 and Scout-Instruct) for evaluation using LLM-as-a-judge. All models performed with lower accuracy when evaluating Irish than English. Such studies demonstrate LLM favoritism toward SWE, which can result in inequitable outcomes for students and communities involved in writing placement.

Because LLMs are pre-trained on languages with greater digital representation and value language similar to their own more highly, they do not give output in any other language or dialect without being prompted to do so. This bias might persist even if the LLM has been finetuned with smaller datasets to generate text in these dialects and languages. However, we do want to note the potential of ongoing projects like Gupta's Translinguo, which endeavor to mitigate these challenges via prompt training to support World Englishes such as Hinglish (Gupta, 2024).

Finally, LLMs may be biased based on data in the pre-training set, including published articles that explicitly discuss bias in writing evaluation toward specific groups of students. That is, articles in the training set that document the existence of White language supremacy—many of which were written by scholars in writing studies and related fields—might inadvertently train GAI platforms to enact racialized language judgments that align with those findings. For instance, Yamashita (2025) evaluated OpenAI's generative pre-trained transformer, GPT

4, by finetuning it specifically to the task of writing evaluation. They found significant bias in the LLM related to race/ethnicity but not related to gender or socio-economic status; GPT 4 assigned consistently higher scores to Pacific/Asian Islanders and lower scores to Hispanic/Latinx writers. Chen et al. (2025) found similar bias but also found that better prompting reduced these biases to some extent.

If researchers identify linguistic bias in local GAI-assisted assessment output, further training along with more human-in-the-loop power (von Davier & Burstein, 2024; §5.3.4) could help mitigate this issue. These unresolved risks of GAI bias—and the need to develop processes to identify and mitigate them—might be reason to slow down or opt out of incorporating GAI into writing placement processes.

3.3.2 Social-Environmental Impact

Although GAI applications can support efforts to address environmental impact issues such as climate change, resource management, and clean energy concerns (Shafik, 2024), the potential use of GAI in writing placement processes raises significant concerns about social and environmental impact (SEIs) (§2.3.7). Hosseini et al. (2025) proposed an SEI framework for estimating the ramifications of a technology designed and implemented with limited consideration for consequences (see §4.1 and Figures 4, 5, & 6 for further discussion of consequences). As we observe in §2.3.7, such considerations are not unique to GAI and should extend to our considerations of all assessment technologies and processes.

That said, examples of the particular adverse impacts of GAI hardware (the tech stack) and LLM training abound. Abilene, Texas (pop. 130,501) is experiencing drought conditions so severe residents are asked to restrict their showers while data centers continue to consume billions of gallons of water (Seamon, 2025). In Chandler, Arizona (pop. 280,000, where electricity and space are affordable), "city officials have limited the construction of new data centers because they do not create quality jobs for local communities and produce disturbing levels of noise" (Hosseini et al., 2025 p. 1). In addition to social consequences, "training a model like GPT3 (with 175 billion parameters) typically consumes about 1287 MWh of electricity and emits approximately 552 tons of CO_2" (Hosseini et al., 2025, p. 1). In comparative terms, that amount of electricity would power a community of about 120 homes for a year, emissions equal to a year's worth of pollutants from 120 cars in that neighborhood. Amazon's planned data center in northern Indiana is forecast to use up to 17.7 megawatts of energy per year; nearly half of what all Indiana residents used in 2024 (Pete, 2025). To address such SEIs, we urge inclusive efforts and targeted accountabilities—such as investigation of impacts for affected communities—to mitigate social impacts. Selecting training options that do not require creating an entirely new LLM or updating all of

the LLM's parameters can help mitigate some SEIs (§3.2).

Continuing GAI innovation could reduce SEIs. The DeepSeek open-source suite, for instance, "achieves high performance through algorithmic and architectural innovations, enabling comparable performance with fewer high-end servers and lower energy consumption" (Wang et al., 2025, p. 1). As a result, the platform reduces reliance on large, centralized data centers such as those in Chandler and alleviates the overall electricity demand involved in training a model like GPT3. Additionally, SLMs (§3.2.4) are shown to have similar capabilities to LLMs with lower SEIs (Hu et al., 2024, Lepagnol et al., 2024). We concur with Wang and colleagues (2025): "In this wave of the democratization of computing power, it is crucial to explore the integration of efficient AI technologies with existing energy infrastructures" (p. 2) and prioritize environmental justice in our decisions about GAI platforms, training methods, and whether to incorporate GAI into writing placement at all (§2.3.7).

3.3.3 Academic Integrity

A critical issue when considering GAI-assisted writing placement is academic integrity: on the part of the institution, the department, the program, and the student. Any writing task students are asked to undertake for assessment must have clearly articulated parameters about whether and how students can use GAI in the creation of their text (§2.4.3). While not all students use GAI, or use it in the same ways, we should assume that incoming students, at least for now and possibly into the future, need explicit instruction on and clear examples of the ways in which GAI can be useful to support academic writing tasks—and when its use is not appropriate. Such instruction includes helping students understand the expectations of their readers: the teacher in the classroom and, if applicable, the GAI-assisted writing placement instrument. When a literacy program incorporates GAI into writing placement, they are communicating to all stakeholders—including students—that GAI can have appropriate uses in academic writing scenarios. They are modeling one example of accepted GAI use at a postsecondary institution.

In some writing placement processes, it may be necessary to redesign or reimagine the kinds of writing tasks students undertake. A new GAI-assisted writing placement process might invite students to reflect critically on their own choices about whether and how to use GAI to complete that writing task. If effectively communicated, GAI-assisted writing placement can teach students about ethical GAI use within the institution's expectations and values. Such reflective tasks can also invite students to be transparent about their use of GAI in writing situations and to critique GAI output and its rhetorical value. Parameters for acceptable GAI use in the writing placement process should align with the local construct of writing as well as institutional policies regarding academic honesty and integrity.

One final aspect of academic integrity includes the institution's responsibility to ensure privacy of each student's personally identifying information (PII) and intellectual property. Literacy programs can encourage students to understand that their writing as their own and educate students about their right to make informed choices about whether and how their writing is used as training data for GAI. Ensuring the privacy and protection of students' writing is essential to preserving integrity on both sides of the equation. Ethical frameworks for privacy, consent, and transparency play a critical role in guiding these design decisions (§2.4).

Section 4 • Toward a Theory of Writing Placement and GAI

4.0 Overview

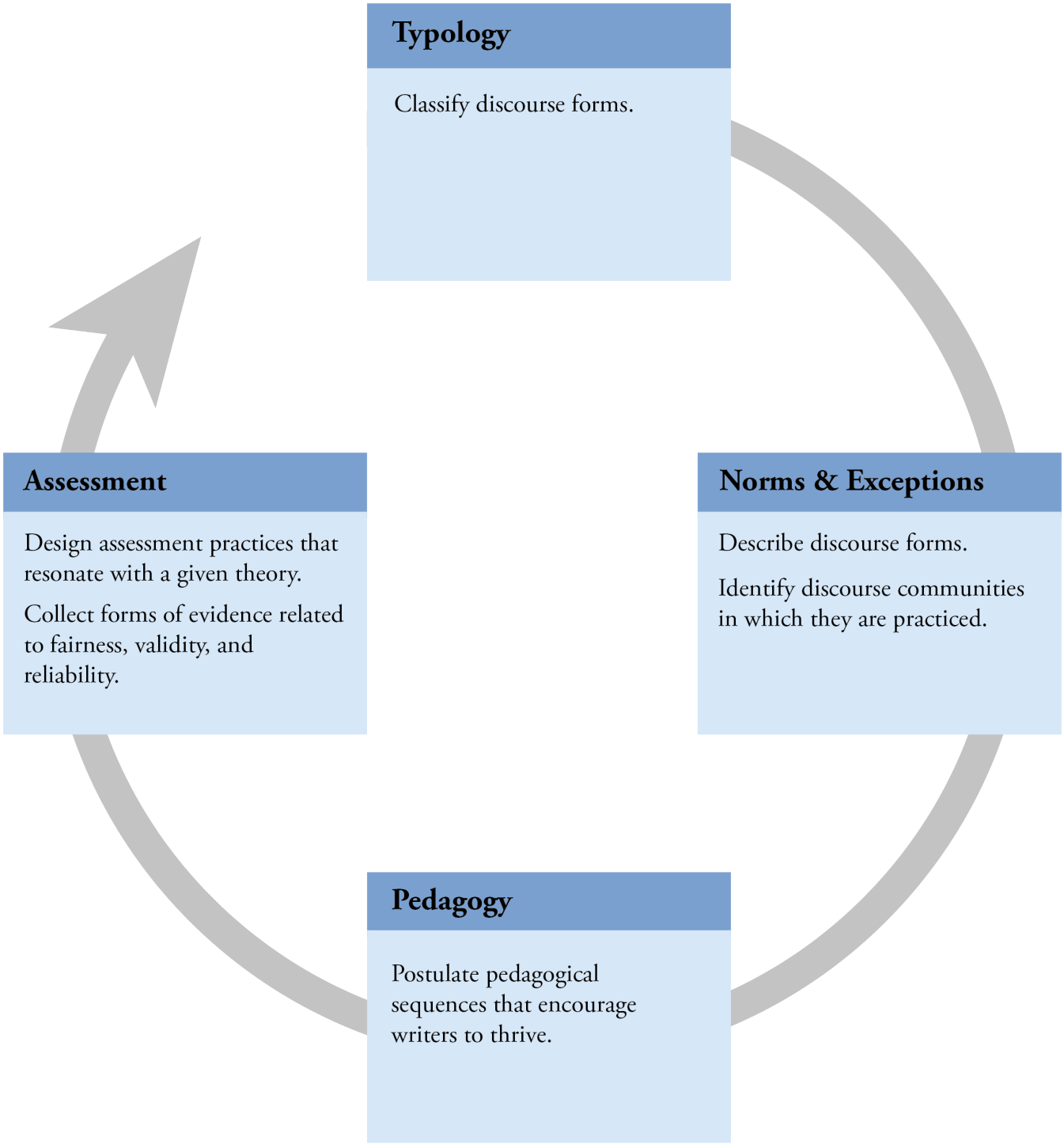
In the immediate aftermath of GAI's arrival, literacy educators were grappling with how to make sense of the technology *right now*, often in the absence of institutional policies and frameworks, and facing the following questions: How will GAI affect students' writing? How will it affect the work of the class? How will it affect academic integrity policies? After three years, we now have the benefit of time. We can reflect on some of the ways education and praxis have shifted, and we can review research by leaders in the multidisciplinary enterprise of GAI literacy. We can also return to our foundations in writing assessment to consider: What must a theory of GAI-assisted writing placement actually do? As is the case with theories of teaching and assessing writing more broadly, a theory of GAI-assisted writing placement should achieve the following (visually represented in Figure 3):

- Establish a typology which classifies discourse forms
- Describe the norms (and exceptions) of these forms and the discourse communities in which they are practiced
- Postulate pedagogical sequences in ways that will encourage writers to thrive
- Design assessment practices that resonate with a given theory and collect forms of evidence related to fairness, validity, and reliability (§2.2; §4.2)

This is a heavy lift under any circumstance, and especially heavy when we consider GAI's potential uses in writing placement, and the possibilities that those uses will reinscribe systemic injustices and disenfranchise communities.

As we have noted (§1.1), GAI, a subset of AI, is a growing technological field that has been in development since the 1960s (Shermis et al., 2013; Johnson, 2023). Forged primarily in computer science, AI is created through rules and algorithms that, in the case of written communication, identify and classify lexical, grammatical, and syntactic language patterns. Recent rapid advancements in GAI have been made through the invention of attention-based architecture, increasing computational power, parallel computing, and continuous parameter adjusting, capturing massive amounts of data from the internet and other sources to use for computing likely word combinations (§3.1.2). Driven by this tech stack, AI is always and everywhere in the process of becoming. Traditional AWE applications have evolved at fast but trackable rates, but if we catch a glimpse of a GAI platform one day, it may change the next in ability and function.

Figure 3. *Theory of Generative Artificial Intelligence Loop*



To understand possible roles of GAI in writing placement, the first step is to understand *contextualization*. "Why," Bruno Latour (1999) asked, "is it so difficult to measure, with any precision, the mediating role of techniques? Because the action that we are trying to measure is subject to 'blackboxing,' a process that makes the joint production of actors and artifacts entirely opaque" (p. 183). Invoking the Labyrinth of Crete, Latour (1994) noted that technology is "shrouded in secrecy. Can we open the labyrinth," he asked, "and count what is inside?" (p. 36). We cannot: "The depth of our ignorance about techniques is unfathomable" (Latour, 1994, p. 37).

What to do? Latour suggested we follow sociologists, particularly the field of ethnomethodology—the study of how humans produce order through interactions, particularly in local settings. To paraphrase Latour, AI, our MacBook Air laptops, our copies of *Failing Sideways: Queer Possibilities for Writing Assessment* (West-Puckett et al., 2023), the WAC Clearinghouse, and TYCA, as well as our membership on municipal committees and in book clubs or as coaches for our children's teams, all organize, shape, define, and limit our interactions—even when we think about GAI and writing placement. While Latour (1999) continued this line of thought to a level of complexity beyond the scope of this white paper, it is worth remembering his conclusion: Technological artifacts do not necessarily mediate human actions. Rather, since we have created them, they are us. As Byrd has stated, "Our literacies are its literacies" (quoted in Johnson, 2023, p. 170).

Most members of our discipline have had little direct involvement with creating these technologies. All of us, however, have played a role in literacy education, the very enterprise necessary to create the corpora used to train GAI, read its output, and think about its implications. Some have contributed important research related to corpus analysis, AWE, and machine learning. And we share concerns about whose language practices have been privileged in the development of these technologies, as well as most other writing assessment technologies and practices over the decades (Aull, 2023, §3.3). As equity-minded literacy teachers grappling with GAI, we are primed to respond to this moment. Carrying Latour's concept forward—AI may be best understood as a human creation situated in social processes—we offer a series of interrelated theoretical frameworks for GAI-assisted writing placement processes that involve analysis or evaluation of student writing. Doing so, we believe, provides literacy educators with the opportunity to carefully deliberate on possible, principled courses of action.

These frameworks all fall under the broad category of anticipatory design frameworks (ADFs)—that is, extended thought experiments articulated *before* an assessment is enacted. Various approaches to ADF, including the options we present here, can help facilitate a cohesive assessment process in which each principle is identified and then

modeled. As Mislevy (2018) has noted, modeling provides ways to conceptualize "reasoning across many unique real-world situations, in each case abstracting salient aspects of the situation, and then going beyond it in terms of mechanisms, causal relationships, or implications at different scales or time points that are not apparent on the surface" (p. 9). Modeling is one important step toward putting theory into practice. As we will demonstrate, the modeling processes offered by ADF provide specifications that identify important elements, establish forms of evidence, and ask key assessment questions.

In this section, we present three particularly constructive approaches to ADF, each of which facilitates thoughtful anticipatory design for GAI-assisted writing placement from different lenses and angles of focus: theory of action (ToA), evidence-centered design (ECD), and interpretation and use arguments (IUA). Each of these approaches can be taken up independently or they can work together to offer layered attention to interrelated potential consequences, forms of evidence, and reasoning for decisions made based on assessment results. Notably, these three ADF models are accessible. They do not require specialized training, tools, or extensive resources. They can be drafted individually and refined collaboratively through a series of regularly scheduled committee meetings. We encourage literacy programs considering GAI-assisted writing placement to engage with the anticipatory design affordances of each approach.

4.1 Theory of Action (ToA)

ToA frameworks support far-reaching considerations of the consequences of a particular assessment practice in context. They have been used with success at a range of institutions; for instance, at one two-year college, Nastal and Elliot articulated a ToA with colleagues (Hazelton et al., 2021) for an AWE application integrated into a composition course and an adult literacy class in a workforce development program for labor union members. Built on Lewin's (1946) conceptions of action research (Adelman, 1993), Argyris's (1997) emphasis on theory-in-use, and careful attention to material conditions, ToA frameworks advance reflection, collaboration, and deliberative action through attention to individual experiences in specific locations. A ToA should be developed during the initial stages of GAI writing assessment research, so that fundamental knowledge of the innovation includes identification of, and relationships among, situated instructional components, stakeholders, pedagogies, and consequences. The ToA offers a useful starting point for theorizing an institution-specific use of GAI in writing placement. Engaging in this extensive consequential mapping at the outset can inform decisions about a program's path or paths forward: opting out, slowing down, implementing selectively, or reimagining writing placement altogether (§1.4).

Our ToA model is based on twelve features. The first four features are (1) identification of stakeholders, (2) components of a specific GAI, (3)

hypothesized pedagogical actions, and (4) hypothesized programmatic actions. The next eight features—the consequence model—are intended to encourage reflection about positive and negative consequences of actions over time: (5) intended positive intermediate consequences (midway desired gains), (6) unintended positive intermediate consequences (midway unexpected gains), (7) intended positive long-term consequences (enduring gains), (8) unintended positive long-term consequences (enduring unexpected gains), (9) intended negative intermediate consequences (midway expected challenges), (10) unintended negative intermediate consequences (midway unexpected challenges), (11) intended negative long-term consequences (long-term expected challenges), and (12) unintended negative long-term consequences (long-term unexpected challenges).

Figure 4 illustrates how fundamental knowledge of GAI-assisted placement assessment can be generally understood and applied in specific assessment situations. It begins with comprehensive identification of stakeholders involved, which emphasizes the reach of this one act of writing assessment. Modeling can be done at the site level that exclusively focuses on the most relevant stakeholders (see Nastal, 2024, for an example situated in a two-year college). Components refer to how programs can incorporate GAI into writing placement in ways that align with individual, programmatic, and disciplinary goals. Focusing on components requires the multidisciplinary team to articulate how the process will operate and function at the institution, how it will advance the team's values and rhetorical theory, and how it will help students learn more about writing. Hypothesized pedagogical and programmatic actions identify the goals of incorporating GAI in writing placement. For instance, in Figure 4 we are committed to improving students' experiences in writing placement and at the institution in support of their educational goals. We are cognizant of the messaging writing placement communicates to various stakeholders, including students, and we consider how a goal of advancing writing studies principles can be achieved through GAI-assisted writing placement.

Figure 4. *Theory of Action for Generative Artificial Intelligence, Writing Placement, and Principled Decision Making: Stakeholders, Components, and Actions*

Stakeholders	Generative Artificial Intelligence Components	Pedagogical Actions: Hypothesized	Programmatic Actions: Hypothesized
Internal to the institution • Students • Instructors • Labor union members • Professional staff • Administrators • Boards of trustees **External to the institution** • Community members • Local K–12 partners • Psychometricians • Multidisciplinary specialists • Legislators • Accrediting bodies • Developers	• Component 1: Integration within a literacy program where writing placement, feature-based construct models, curricula, student learning objectives, and methods of assessment are aligned • Component 2: GAI chatbot chosen by multidisciplinary collaborative team, reflective of institutional priorities and partnerships • Component 3: GAI chatbot training for writing placement overseen by multidisciplinary team • Component 4: Gen AI program of research reviewed by IRB, with emphasis on respect for individuals • Component 5: GAI scores comparable with human scores in terms of evidence based on fairness, validity, and reliability • Component 6: Beyond scores, GAI feedback to supports self-regulated writing processes aligned with feature-based construct models	• Objective 1: Close loop between assessment and instruction • Objective 2: Assure GAI chatbot creates opportunities to learn for all students • Objective 3: Encourage metacognition about writing and revising in digital environments • Objective 4: Improve GAI literacy • Objective 5: Enhance student self-efficacy through formative feedback • Objective 6: Increase students' awareness of writing and the literacy program • Objective 7: Increase students' awareness of college resources and support services	• Objective 1: Use GAI scoring and pedagogical insights to enhance curricula, syllabi, course and program learning objectives, professional development offerings, and programs of research • Objective 2. Use longitudinal GAI scoring and pedagogical information to provide insight into students' writing and dispositions to advance opportunities to learn

Once the scope is articulated, Figures 5 and 6 explicate short- and long-term positive and negative consequences. Those who choose to incorporate GAI into writing placement might have ambitious long-term objectives shared by many literacy educators: to support students as they work toward their individual goals, which can increase their well-being and satisfaction. Wise administrators know that when students feel a sense of belonging, feel intellectually challenged, and feel celebrated in their educational experiences, institutional retention and enrollment can increase. More immediately, as this white paper has articulated, stakeholders might learn more about writing, gain more experience in computer-mediated writing scenarios, and increase their metacognition and self-efficacy. These changes can lead to more ethical and more collaborative institutional research opportunities. As is the case with any choice, however, there are risks involved with GAI writing placement. Introducing the technology into students' first experiences with the writing program might not offer students new insights but instead be perceived as permission to use GAI in all writing scenarios without ethical consideration. Given what is known about how much energy and water current GAI platforms use (§3.3.2) and how LLMs have been built, stakeholders might perceive GAI-assisted writing placement as evidence of the compromised state of higher education today and diminish their views of writing, the writing program, and the institution.

Figure 5. *Theory of Action for Generative Artificial Intelligence, Writing Placement, and Principled Decision Making: Positive Consequences*

Intended Positive Intermediate Consequences	Unintended Positive Intermediate Consequences	Intended Positive Long-Term Consequences	Unintended Positive Long-Term Consequences
• Consequence 1: Increased course-level success for students • Consequence 2: Stakeholders gain more insight into students' writing processes, practices, and dispositions • Consequence 3: Refreshed alignment among writing constructs, CR tasks, scoring rubrics, and student learning outcomes • Consequence 4: Creative experiences with writing and revising in digital environments • Consequence 5: Novel application of score interpretation and use and formative feedback for students • Consequence 6: Increased emphasis on, and opportunities for, multidisciplinary research collaboration and expanded programs of research • Consequence 7: Efficacy in scoring and pedagogical use	• Consequence 1: Enhanced awareness across the curriculum of value of alignment among writing constructs, CR tasks, scoring rubrics, and student learning outcomes • Consequence 2: Greater attention to ethical research practices, with special attention to respect for individuals • Consequence 3: Increased stakeholder training on GAI applications, data warehousing, writing assessment, and rhetorical traditions • Consequence 4: Improved human-computer interactions	• Consequence 1 Increased awareness of benefits of GAI for teaching and assessing writing • Consequence 2: Targeted pedagogical supports identified for wide range of students, with attention to individual differences • Consequence 3: Increased awareness of value of construct models • Consequence 4: Increased student self-efficacy • Consequence 5: Increased number of credits earned, persistence, and completion for students • Consequence 6: Shift from placement to place-making, enhancing students' sense of belonging • Consequence 7: Increased participation by writing studies faculty in various assessment genres • Consequence 8: Increased attention on fairness, validity, and reliability as forms of evidence • Consequence 9: Increased emphasis on linguistic diversity • Consequence 10: Stakeholders in under-resourced institutions gain equitable access to information about incoming students' acquired literacies	• Consequence 1: Improved well-being and happiness for students and stakeholders • Consequence 2: Increased student earnings and socioeconomic mobility • Consequence 3: Increased institutional enrollment • Consequence 4: Contributions to writing pedagogy and writing assessment research and development • Consequence 4: Contributions to GAI research and development

Figure 6. *Theory of Action for Generative Artificial Intelligence, Writing Placement, and Principled Decision Making: Negative Consequences*

Intended Negative Intermediate Consequences	Unintended Negative Intermediate Consequences	Intended Negative Long-Term Consequences	Unintended Negative Long-Term Consequences
• Consequence 1: Student use of GAI rapidly shifts literacy programs and perceptions of writing, creating curricular disjuncture • Consequence 2: Disjuncture in formative feedback between GAI and expert human scores • Consequence 3: Increased demand on institutional resources • Consequence 4: Failure of disciplinary consensus on best practices for GAI • Consequence 5: Increased environmental demands • Consequence 6: Amplification of linguistic and cultural biases via LLMs • Consequence 7: Institutional and ethical research delays in GAI scoring and pedagogical applications (e.g., when a college does not have an IRB or has an under-resourced Institutional Research office)	• Consequence 1: Diminishes student voice and self-confidence • Consequence 2: Severe construct underrepresentation • Consequence 3: Reward of formulaic writing • Consequence 4: Large-scale opting out by students • Consequence 5: Large-scale failure of academic integrity • Consequence 6: Diminished self-efficacy for faculty • Consequence 7: Institutional commercialization leading to costs passed to students	• Consequence 1: Writing placement by humans viewed as obsolete • Consequence 2: GAI methods of teaching and assessing writing used to decrease funding and staffing and increase course loads and caps • Consequence 3: Increased technocratic control over education • Consequence 4: Promotion of addictive technologies with evidence of leading to psychosis	• Consequence 1: Permanent flattening of linguistic diversity • Consequence 2: Diminished perception of higher education • Consequence 3: Catastrophic increased demands on land, water, air, and other-than-human life • Consequence 4: Long-term stress and related comorbidity on workers • Consequence 5: Contribution to bifurcation of society

Finally, the ToA is a living model that can and should be continuously revisited as the placement process is implemented, evaluated, and revised. As part of the validation process, a team can refine their ToA to include *demonstrated* pedagogical and programmatic actions and create more hypotheses. Similarly, they can further identify unintended consequences that might not have been imaginable in the anticipatory stage. Writing assessment is about understanding consequences. Before incorporating any technology or implementing any assessment practice, it is necessary to identify potential impacts on the stakeholders involved. Special attention must be devoted to those most vulnerable in the system, the people who will be affected most by the results of the assessment. In most cases, those people are students.

4.2 Evidence-Centered Design (ECD)

Evidence-Centered Design (ECD) is the earliest model for ADF in educational measurement (Mislevy et al., 2003). This model gave rise to *Expanded ECD* (Arieli-Attali et al., 2019) and the *Justice-Oriented, Antiracist Validity Framework* (Randall et al., 2024). In writing assessment, *Design for Assessment* (DFA) (White et al., 2015), the *Integrated Design and Appraisal Framework* (Forzani et al., 2024; Slomp, 2016), and *Category of Evidence* modeling (Haswell & Elliot, 2019) are all indebted to this ECD tradition. If the ToA framework is helpful for anticipating a range of potential consequences of a writing assessment for different stakeholders, variations on ECD are particularly useful for thinking through the forms of evidence literacy program coordinators already have and will need to determine how well their GAI-assisted writing placement process is enacting its stated principles in their local context. In Table 3, we provide an ECD model adapted to GAI-assisted writing placement. The GAI-ADF integrates the six principles for GAI and writing placement (§2.1-§2.6) into the modeling traditions identified above. This ECD model emphasizes forms of evidence and key questions related to these principles to facilitate the thoughtful planning of writing placement processes literacy program coordinators embark upon.

Table 3. *Generative Artificial Intelligence Anticipatory Design Framework (GAI-ADF)*

Principle	Key Elements	Forms of Evidence	Key Question
1. Combine disciplinary integrity with multidisciplinary collaboration	Assembling strategic expertise in service of principled assessment	Team member documented experience in pedagogy, practice, research, and scholarship; written, publicly available program of research; established budget; planned sustainability	How may the community assemble the expertise necessary to conduct the assessment so that each GAI-ADF principle is followed?
2. Prioritize logic of evidence	Defined interpretation and use argument (IUA)	Evidence of fairness, validity, and reliability	How can an evidence-gathering plan be designed to advance the site-specific IUA?
3. Align with literacy programs' values and commitments	Resonance among writing construct, constructed response (CR) task, scoring rubric, and student learning outcomes (SLOs)	Construct visualization; structured CR tasks; trait and holistic score rubric design; syllabus-based SLOs	How may explicit alignment be established among the construct, CR task, scoring rubric, SLOs, and program's values and commitments?
	Attention to fair labor practices	Cost analysis involving human labor, with attention to sustainability	What kinds of labor are involved in GAI implementation, ongoing assessment activities, and continuous evaluation and improvement of the process?

Principle	Key Elements	Forms of Evidence	Key Question
	Attention to shared faculty governance practices	Statement establishing equitable processes of shared faculty governance regardless of rank or contingency status	What principles are in place to ensure faculty's role in institutional deliberation and decision making and in matters that directly affect curriculum and instruction in their programs?
	Responsivity to accessibility best practices and concerns	Statement establishing that all aspects of the GAI implementation follow Section 508 of the Rehabilitation Act providing federal requirement for digital accessibility	Does the GAI implementation adhere to CCCC's (2020) Disability Studies in Composition: Position Statement on Policy and Best Practices?
	Responsibility to culturally sustaining practices	Commitment to GAI prompt engineering to reprogram LLM biases	Does a written, critical framework exist for centering and sustaining diverse communities as these memberships intersect with gender and sexuality, dis/ability, class, language, land, and other sociocultural elements?

Principle	Key Elements	Forms of Evidence	Key Question
	Responsibility to linguistic justice	Commitment to GAI applications that are developed with conscious attention to language diversity	Does a written statement exist establishing that all students have equitable access to opportunity structures regardless of the language varieties they speak and write?
	Responsibility to land, water, air, and other-than-human life	Environmental impact estimates for GAI training and development	How may GAI applications be identified with attention to their environmental impact as a key selection criterion?
	Attention to assessment *of*, *for*, and *as* learning	Transformation of emphasis *of* learning to and assessment *for* learning about students' assessment *as* a form of learning	How does GAI implementation close the loop between assessment and instruction?
4. Respect individuals and communities	Transparency to students, anonymization, and IRB processes	Written, publicly available GAI disclosure statements to students; removal of student identification information before upload; IRB waiver or approval	How may the community be made aware of information about GAI placement use, its alternatives, and its impact?

Principle	Key Elements	Forms of Evidence	Key Question
5. Attunement to emerging technologies and relationships	Relationship articulation among writing placement principles, GAI development, LLMs, and campus-specific corpora	Written, publicly available statements identifying GAI assessment and pedagogical paths; established prompting processes; language model and data set selection; investigation of language bias; awareness of social and environmental impact	How may the community be aware of the technical aspects of GAI used for writing placement in ways accessible to a wide variety of audiences?
6. Be open to transformative effects	Careful, staged implementation, with emphasis on place-making; consequence planning for short term, intermediate, and long-term impact	Written, publicly available theory of action (ToA) model	How can the community enact place-making through writing assessment to continuously (re)make institutional and programmatic ways to advance student learning?

4.2.1 GAI-ADF Case Study

We will take the University of Utah's exploration of GAI-assisted writing placement as an example for the GAI-ADF model (Table 3). Utah's statewide general education requirements include a two-course first-year writing sequence, WR1 and WR2. The University's placement process, called the First-Year Writing Assessment (FYWA) and currently hosted through the online survey platform Qualtrics, introduces students to key learning outcomes and pedagogical orientations for both courses. It then invites students who believe they would be best served by entering directly into WR2 to write a 500–750-word essay that makes their case for advanced placement by (1) demonstrating they understand the decision they are making by articulating the key similarities and differences between WR1 and WR2 based on the materials provided in the assessment instrument, (2) discussing how their prior reading and writing experiences relate to what those courses offer and expect, and (3) reflecting on their own writing-related strengths and areas for growth in relation to the WR1 and WR2 curricula. A team of experienced first-year writing instructors assesses the essays using a rubric with three trait scores and a holistic score, a process grounded in long-standing traditions in writing assessment that value the pedagogical expertise of instructors (e.g., Haswell & Wyche-Smith, 1994).

The goal of the FYWA is to identify students who accurately understand their course options, have had opportunities to achieve the student learning outcomes for WR1, have developed meaningful capacities for writerly self-reflection, and want to enroll directly in WR2. The FYWA is a relatively new process developed in Spring 2024, after the Utah System of Higher Education mandated that the university shift from its existing informed self-placement process to an institutionally developed "challenge exam" compliant with statewide policy (see Toth et al., 2024). In its first two years, the FYWA has been scored by human raters in ways that have allowed evidence to be gathered regarding fairness, validity, and reliability.

The proposed shift to some form of GAI-assisted scoring for the FYWA requires the University of Utah's writing program to retain some of the processes and principles for gathering this evidence: for example, group analysis for disparate impact, analysis of present and next course success, and analysis of inter-rater agreement and reliability. These processes and principles will also need to be expanded in several ways:

1. New members will need to be added to the research team (§2.1), and the team will need to be assembled to advance the stated principles of the assessment. As part of the GIA-ADF, forms of expertise will need to be established and documented, a budget will need to be expanded to provide support for the GAI program of research, and a plan for long-term sustainability must be in place.
2. While a logic of evidence (fairness, validity, and reliability) model (§2.2) may be established,

these processes need to be maintained—and expanded—to account for evidence relating to GAI use. For example, do the scores produced by the GAI reveal evidence of language bias, inadequate construct coverage leading to course failure, or scores with inadequate reliability relationships to human scores? New areas of qualification—such as environmental impact (§3.3.2)—will also now be considered.

3. Resonance among the FYWA's targeted writing construct, writing task, scoring rubric, and course learning outcomes is always important, and coherent documentation and mapping becomes even more important in GAI scoring. As discussed in §5.3, scoring is determined by very specific one-shot or multi-shot prompting of the GAI. The risks are high if there is disjuncture or contradiction among the stated values and commitments in the documentation of the FYWA those determining the scoring might give incorrect prompting to the machine, resulting in scoring that does not capture key curricular elements.

4. In an era dominated by efficiency discourse, it takes focused effort to remind ourselves of the ethical principles and IRB processes noted in §2.4.1, with universal emphasis on respect, benevolence, and justice (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979). Some institutional stakeholders seeking to move quickly might express doubt about whether it is necessary to consult IRB before experimenting with training GAI using student writing generated through the FYWA in its first two years. It is helpful to list the reasons for investigating GAI assistance in the writing placement process, then reference the Declaration of Helsinki (1964): "These purposes can never take precedence over the rights and interests of individual research participants" (World Medical Association, 2024, General Principle 7). Because the consequences of writing placement are always about students and are so closely related to whether a student gains access to higher education, their rights and interests must always be at the forefront of our work. Informed consent and transparency remain essential (§2.4.2).

5. As Willingham (1974) observed, college writing placement involves complicated processes that are both social and technological, and a multidisciplinary team (§2.1) is needed to follow the emerging GAI technologies discussed in §3.0. The GAI-ADF calls for written statements available to institutional and public stakeholders (§2.4.2) that identify and discuss developments related to the use of LLMs and their variants to score student writing samples. To uphold this standard, the University of Utah would need to maintain a live page on its website to present, update, and create avenues for

community input on the GAI-assisted FYWA.

No matter what GAI path an institution is on at present (§1.4), there are principles and elements, forms of evidence, and key questions about them that are useful to advance student learning and engage communities in decisions about the applications of these technologies. What, one might justifiably ask, if only some—or perhaps none—of the forms of evidence identified in the ADF are present on a campus? We suggest embracing the shift from *writing placement assessment* to *place-making through writing assessment* (§2.6.1; §5.4). Responding to the theoretical and ethical demands of GAI-assisted writing placement can be a moment of reckoning with heretofore under-theorized and inadequately evidenced assessment practices that might have been in place for some time. This technological change can be an exigency for developing a stronger evidentiary foundation for students' first encounters with writing assessment at our postsecondary institutions.

Practitioners can coordinate with offices of institutional effectiveness or institutional research to pull together evidence of fairness, validity, and reliability. They can turn to colleagues in information technologies, computer sciences, and information sciences for insight on data privacy and data governance. Their colleagues in education and psychology might have resources devoted to construct response and learning outcomes development and assessment. It is even possible that a literacy program coordinator's colleagues across the college or university might have access to different budgets and funding streams to support this ongoing work.

4.3 Interpretation and Use Arguments (IUA)

The landscape of educational assessment and writing assessment is persistently fraught because the consequences are so significant. The decisions we make as a result of a student's or a group of students' performance on an assessment can shape their entire postsecondary trajectory. As Nastal (2019) wrote,

> Test score use relates not only to one specific action but to an array of implications, judgments, and social consequences based on labels, connotations, actions, and subsequent lines of inquiry pursued. In placement, for instance, a test score is used to determine the first writing course a student is eligible for. It may also result in a label designation, institutional tracking, and/or subsequent academic opportunities offered or withheld.

What we do with assessment results—the decisions and inferences we make—proceed from a range of assumptions. IUA frameworks offer options for thinking through the meaning a program makes of the results yielded through a particular assessment practice and the decisions it makes based on that interpretation.

Like the ToA, which is created *before* an assessment is enacted, a hypothesized IUA allows anticipation

of the argument that will be used to evaluate the claims made to justify or decide against specific uses of GAI in writing placement. Figure 7 provides the basic elements of one such approach. Like Kane (2006, 2013), we return to Toulmin's (2003) model for argumentative logic; however, we note that Toulmin's model is one among many possible models of argumentation. We have added three contexts to understand how the argument might shift across non-academic, academic, and assessment settings. In addition, following recent sociocultural models advanced by Mislevy, Oliveri, Slomp, Crop Eared Wolf, and Elliot (2025), we have added alternative explanations to qualifications.

Figure 7. *Generative Artificial Intelligence Interpretation and Use Argument*

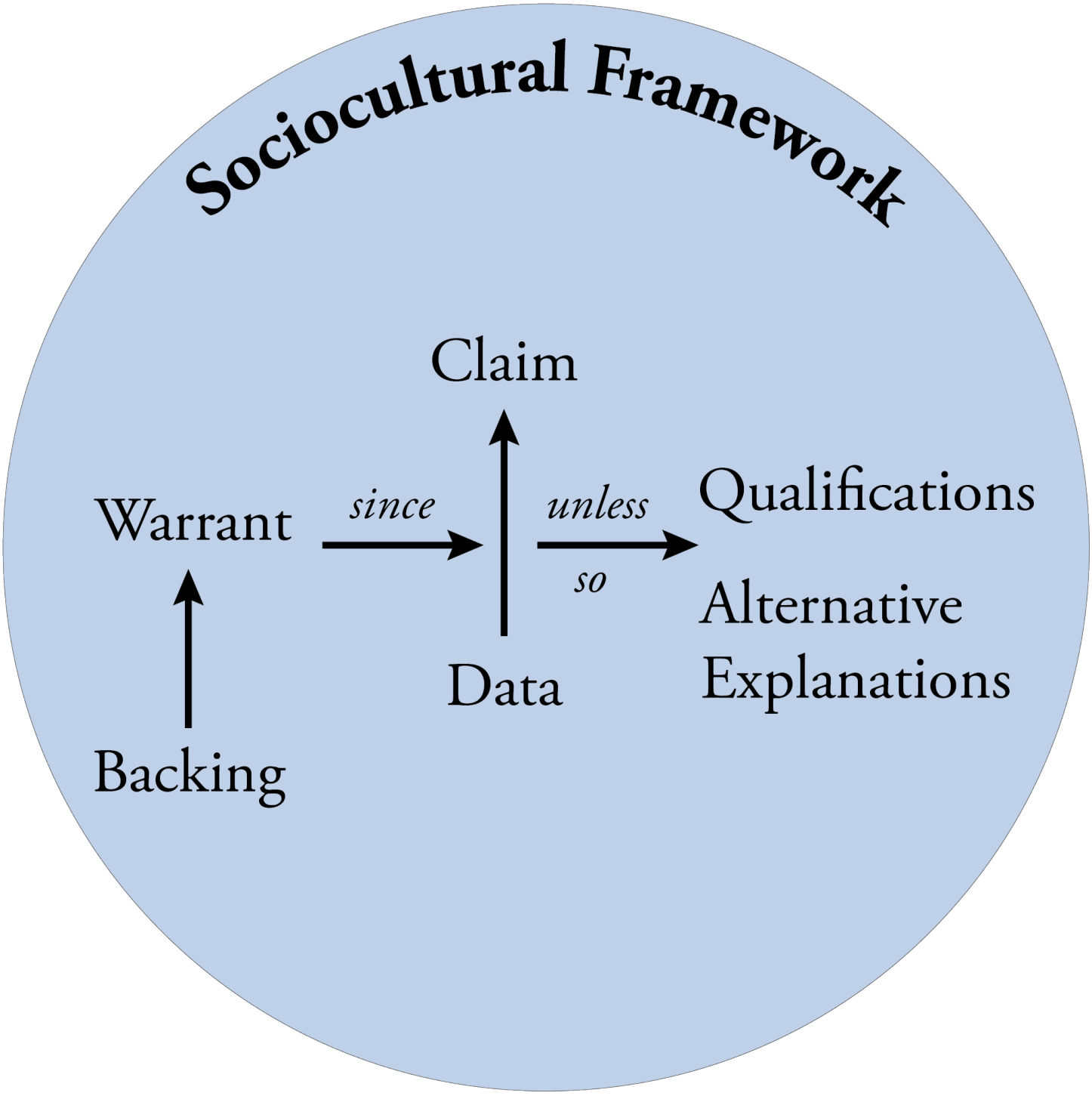


Figure 7 advances the traditional Claim-Data-Qualification structures associated with IUA in scientific measurement. We use IUA as defined by Kane (2013) to include all claims based on the scores. If we return to the University of Utah case study, a central claim is that the results of the FYWA indicate whether a student is likely to succeed if awarded advanced placement into WR2. The validity of a proposed interpretation or use of test scores in a given assessment can be defined in terms of the plausibility and appropriateness of the proposed interpretation and use at that specific time. A proposed interpretation or use can be considered valid to the extent that a) the IUA is coherent and complete (in the sense that it fully represents the proposed interpretation or use) and b) its assumptions are either highly plausible *a priori* or are adequately supported by evidence. The kinds of evidence required for validation are determined by the claims being made, and more ambitious claims require more evidence than less ambitious claims. We hold that a given IUA model for GAI-assisted writing placement must exist within sociocultural frameworks central to knowledge-making in writing studies.

The three contexts—nonacademic, academic, and assessment settings—allow us to tailor construct-based claims about students' capabilities. The addition of alternative explanations allows us to articulate evidence that may, in some cases, so weaken the argument that it cannot go forward. In the case of GAI-assisted scoring for the FYWA, we might articulate the IUA as follows:

- **Claim:** When the decision is made to incorporate GAI into writing placement, a program is making two central claims: a) that we can determine whether a student is a good candidate for a specified class and b) that we are confident the GAI instrument is effective. With a corpus including tasks identical or very similar to tasks on a given placement episode—in this case, FYWA essays written by students and scored by humans between 2024 and 2025—a well prompted GAI instrument can be used in assessment for advanced placement into WR2. Specifically, the application can be prompted with a small training corpus combined with a human-in-the loop scoring process while adequate data is collected on the fairness, validity, and reliability of the application. Important to this process is the availability of a human-scored corpus that can be used to train the GAI instrument and for comparison with GAI-generated scores.
- **Data:** Data collected on the following can be used to compare descriptive and inferential evidence (§5.3) for this specific scoring application using methods in line with current AI writing assessment research:
  - *Fairness evidence* (§5.2.3) from collecting and examining overall scores and scores of all relevant groups identified by the IPEDS in terms of gender identity, race, ethnicity, and socio-economic status, as well as communities pertinent to local

institutions, with particular attention to the presence of disparate impact.

  - *Validity evidence* (§5.2.1) from examining construct relationships among writing task, scoring criteria, and established student learning outcomes for WR1 and WR2; concurrent evidence of correlations with current semester writing course grade and GPA; and predictive evidence of next writing course grade and next semester GPA.

  - *Reliability evidence* (§5.2.2) including inter-rater agreement between GAI and human scores, intra-rater reliability between GAI and its own scores, and precision comparing GAI and human mid-range scores with human scores as the determining factor.

- **Backing:** A systematic literature review can be undertaken to determine the most current GAI prompting methods (§3.2.1), corpus creation (§3.1.4), and evidential forms (§2.2). Thresholds using descriptive and inferential statistics, including but not limited to correlations and regression analyses (§5.2), can be set to establish standards for the local assessment.

- **Warrant:** An explicit system demonstrating that the above-mentioned literature review is relevant to the local assessment situation—in this case, the University of Utah's locally developed GAI-assisted writing assessment process.

- **Qualifications:** Exceptions to the analyses demonstrating that, for some groups or for some scores, the GAI instrument's scores may not be used in isolation.

  - For example, the FYWA rubric uses a six-point scale for each trait and holistic score. Analyses might demonstrate that for some or all traits, GAI scores align closely with human raters on essays scoring in the high (5–6) and low (1–2) ends of the scale, where essays are clearly passing or not passing, but are less reliable in the mid-range (3–4), a highly consequential part of the scale where the difference between a 3 or 4 determines whether a student passes the assessment. In this case, essays receiving scores in the 3–4 range from the GAI instrument should not be used in isolation and must be referred to human raters.

  - In our IUA, environmental impact of relevant GAI tech stacks is also a qualification: Some (or potentially all) proposed tech stacks might be evaluated as too environmentally harmful to use for GAI-assisted writing placement.

- **Alternative Explanations:** Examination of the IUA to indicate that some aspect of the assessment might limit score use or be so severely misaligned (in terms of intra-GAI

scoring agreement or agreement with human scores) that the scores should not be used for placement. For instance, if a student's placement result on the First-Year Writing Assessment indicates WR1 rather than WR2, one claim that could be made is that the student is not qualified to succeed at the kinds of literacy work that WR2 requires (§1.3). An alternative explanation—"unless"—might be that the student was tired or hungry or was not able to devote enough time to the placement process, or that the GAI model has been trained insufficiently or is replicating the biases embedded within its human-informed algorithms. Careful consideration of possible interpretations and reasoning is essential.

This IUA is one possible model among many. Thinking through its constituent parts can help practitioners articulate the assumptions they hold about what the writing placement practice means within their literacy programs. Explicating the model can also help individuals identify how they can evaluate the claims made to support incorporating—or potentially opting out of—specific uses of GAI in writing placement.

4.4 Toward Transformative Effects

Ultimately, we believe reflecting on existing writing placement processes and potential approaches to incorporate GAI might lead programs toward more radical reimagining of the processes and purposes of such assessment, particularly as student and faculty uses of GAI technologies might be changing the local writing construct itself. It is through such reflective praxis that we begin to realize the transformative potential of place-making through writing assessment.

An example might help illustrate this claim. The College of DuPage, in suburban Chicago, reflects the heterogeneity of the United States (§1.2) and has a growing number of Hispanic/Latinx and Middle Eastern and North African students, many of whom speak English in addition to Spanish and Arabic. If the college were to incorporate GAI into a new student self-placement (SSP) process (§5.4.2), one intended positive consequence could be to learn more about students' linguistic repertoires or the sophisticated ways students communicate in their communities. Coordinators could share these insights with instructors and offer supportive professional development opportunities so faculty could modify their curricula to reflect and engage students' knowledge and experiences—movement toward locally responsive and culturally sustaining pedagogy (§2.3.5). They could examine how GAI advances or undercuts linguistic justice, for instance (§2.3.6), deliberating on Byrd's (2023) argument that we now "have an opportunity to partner with their students on creating digital content that the next iteration of LLMs will one day scrape. To this end, we are positioned to launch critical inquiries on corpus texts" (p. 138). Students could be invited to inform the college about who they are, and literacy program coordinators could draw on

that information to validate their processes and refine their framework.

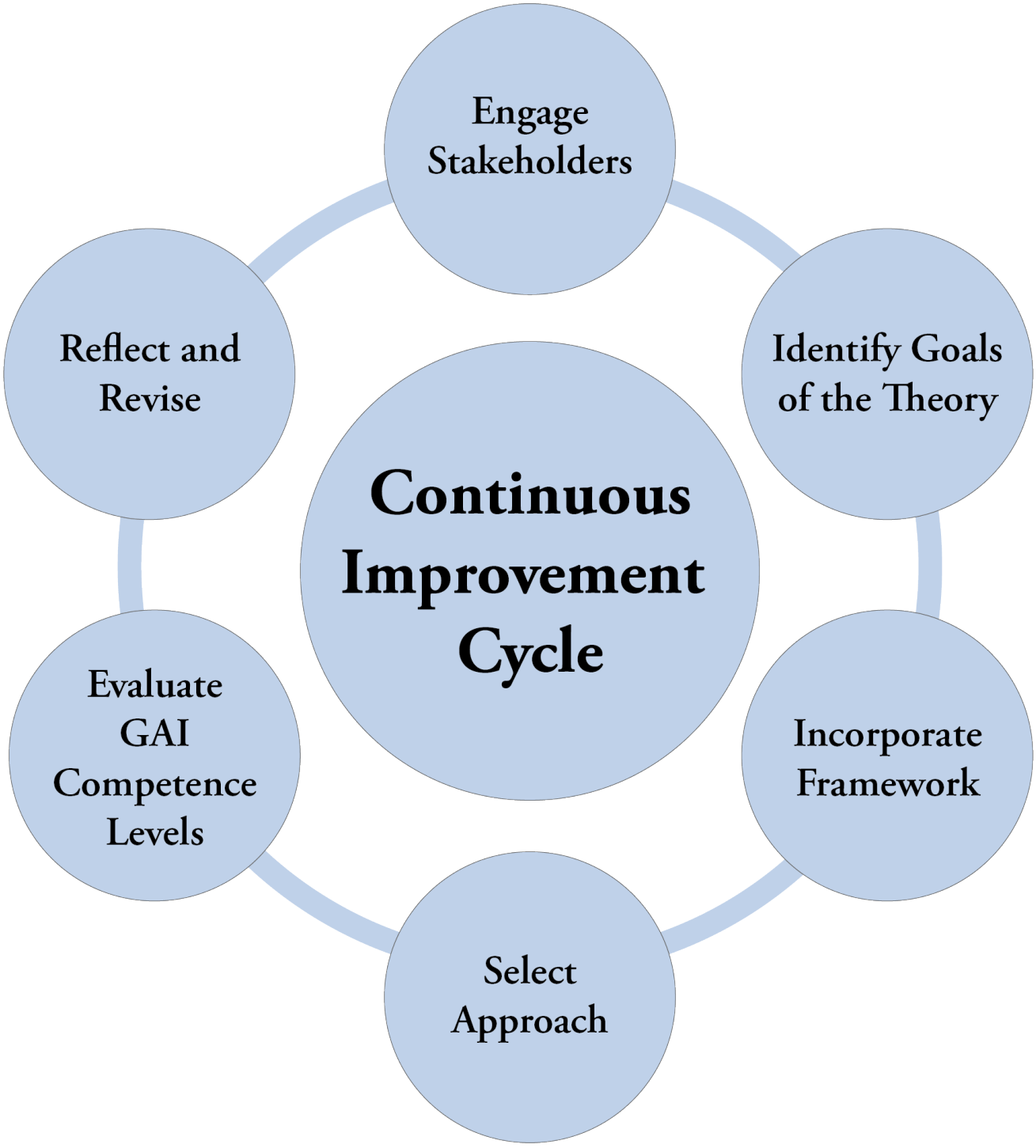
The ToA, ECD, and IUA frameworks we have offered here are preliminary, though they are built upon our collective bodies of work undertaken in a range of geographic areas, institutions, and positions where we have focused on deliberative equity-centered assessment practices. We believe these frameworks offer useful, actionable starting points for conceptualizing GAI in writing placement. Articulated *before* an assessment occurs, the decisions to adopt these or other ADFs will often support the need to *slow down* GAI-assisted placement initiatives in order to ask questions, insist on clearly articulated theories and strategies, acquire and analyze historical and current assessment data, and define the process with all stakeholders.

Section 5 • Possible Uses of GAI in Writing Placement

5.0 Overview

In §4.0 (Figure 3), we introduced a loop that represents the goals a theory of GAI-assisted writing placement should achieve. Within that ADF loop, the left-most square suggests literacy program coordinators ought to enact practices that resonate with theories and pedagogies relevant in their local contexts and collect forms of evidence related to fairness, validity, and reliability (§2.2). To continually evaluate such ADF work after initial implementation, we recommend using a second cycle common in education assessment: the improvement cycle. Applied to GAI-assisted writing placement, continuous improvement cycles can be used as frameworks to consider how using GAI will align with a college's mission, vision, and values, and how it might affect a literacy program's interactions with incoming students (see Figure 8).

Figure 8. *Generative Artificial Intelligence Writing Placement Improvement Cycle*



The GAI writing placement improvement cycle begins and comes full circle with engaging stakeholders. Table 4, below, identifies some of the many ways various stakeholders can be involved in a GAI-assisted writing placement process. As articulated in the ToA framework (§4.1), not all stakeholders carry the same weight when considering the effects of GAI-assisted writing placement. We must prioritize attention to those who have the most at stake in the process: students.

We strongly encourage literacy programs to involve compensated student collaborators in the early conceptualization and design phases of any writing placement process, including processes incorporating GAI. Students should also have multiple layered opportunities to provide feedback on the process during and after placement, including an easy-to-locate option to provide anonymous comments through the program's website. Kryger, Mitchum, and Higgins (2024) have made a compelling case for adopting user experience (UX) design principles and practices—including usability studies, equity-centered UX/participatory design methods, and opportunities for user reflection embedded in the placement instrument—to create and continuously refine accessible and localized placement processes. Engaging with students means interacting with them in a variety of ways to understand their experiences of the writing placement process and refine that process to achieve the literacy program's goals.

Table 4. *Methods to Incorporate Local Stakeholder Feedback into Generative Artificial Intelligence-Assisted Writing Placement*

Stakeholders			
Students	**Faculty**	**Institutional colleagues**	**External to the institution**
• Participation in placement conceptualization and design • Feedback on placement practices (e.g., ongoing UX studies, embedded reflection, surveys) • Reflections in-class • Conversations with advisees • Tracking number of opt-outs	• Participation in placement conceptualization and design • Feedback on placement practices (e.g., surveys) • Using placement materials to guide department meetings • Tracking number of students moving between courses at start of term	• Feedback on placement practices (e.g., surveys, focus groups) • Participating on strategic long-range planning and accreditation committees • Meeting with labor union leadership and faculty senate • Collaborating with administration	• Feedback on placement practices (e.g., online comment form, focus groups) • Participation at community and board of trustee meetings • Conversations with dual credit and concurrent enrollment partners • Conversations with local literacy partners

Moving through the continuous improvement loop includes identifying goals of the theory (§4.0), incorporating an ADF framework (§4.1, §4.2, & §4.3) to conduct inquiry into how the GAI-assisted writing placement model functions, selecting an approach (§5.1; §5.4) to enact GAI-assisted writing placement, and evaluating GAI competence (§5.2). Closing the assessment cycle loop includes reflecting and revising (returning to §4.1, §4.2, and §4.3) and then again engaging stakeholders. Repeating the improvement cycle helps ensure placement processes are being assessed for evidence of fairness, validity, and reliability (§2.2) and are in alignment with programmatic goals. It also keeps GAI-assisted writing placement processes current—a primary consideration given the quickly evolving AI landscape (§2.5).

5.1 Possible Approaches to GAI-Assisted Writing Placement

There are many approaches to writing placement used at institutions throughout the US. Some involve scoring student writing, and some do not (see Table 5). We can envision ways GAI might be useful to support a range of writing placement processes, even as we remain cognizant of the limitations of technology designed to *generate*—not evaluate—language (§3). Table 5 presents our initial thought-mapping regarding how GAI could be incorporated into various writing placement approaches. These are ideas, not recommendations, and they are preliminary: They have not been tried or tested and would need to be subject to evaluation techniques (§5.2) for evidence of fairness, validity, and reliability in local contexts. If deemed usable, they could be incorporated into the improvement cycle, receiving repeated evaluation over time as GAI evolves.

Table 5. *Approaches to Writing Placement and Possible Generative Artificial Intelligence Uses*

Placement Approach	Description	Possible GAI Uses
High school GPA	Course placement based on high school grade point average (Belfield & Crosta, 2012; Koretz et al., 2014).	GAI could assist with demographic and geographic data analysis outside the scope of this paper.
Externally administered standardized tests	Course placement based on scores on standardized tests that students complete prior to admissions (e.g., state common core exams, SAT, ACT).	GAI uses depend on the components of the exam. AWE platforms are currently used to assess written components of these exams and may soon leverage some LLM benefits or new GAI applications may soon be developed. Primary uses: • Score student writing • Offer student feedback • Prompt student reflection • Large-scale corpus analysis of student writing
Proprietary testing products	Course placement based on scores on proprietary testing products purchased by the institution (e.g., ACCUPLACER). Typically include multiple choice tests of grammar, usage, and mechanics and might include a writing task scored either internally or externally (Isaacs, 2018; Nastal, 2019).	GAI uses depend on the components of the exam. Where AWE platforms are currently used to assess written components of these exams, GAI applications may soon be developed. Primary uses: • Score student writing • Offer student feedback • Prompt student reflection • Large-scale corpus analysis of student writing

Placement Approach	Description	Possible GAI Uses
Locally developed exams	Course placement based on locally designed and scored exams, often including student writing composed in response to a prompt (often a reading-to-write task). Sometimes written in controlled testing conditions, sometimes completed in the time and place of students' choosing (Nastal et al., 2022a).	GAI uses depend on the components of the exam. Primary uses: • Score student writing • Offer student feedback • Prompt student reflection • Refer students to resources/support • Rapid identification of time-sensitive student concerns expressed in writing • Large-scale corpus analysis of student writing • Large-scale analysis of growth in student writing
Portfolio assessment	Course placement based on evaluation of portfolios of writing students have produced in other contexts, often with additional reflective writing that contextualizes the portfolio contents (e.g., Huot, 2003; Willard-Traub et al., 1999).	Where portfolio evaluation has traditionally been done by human readers, GAI may offer a much more rapid assessment. Consider providing simple rubrics to help with validity and reliability. Primary uses: • Score student writing • Refer students to resources/support • Rapid identification of time-sensitive student concerns expressed in writing • Large-scale corpus analysis of student writing samples and/or reflective essays • Large-scale analysis of growth in student writing

Placement Approach	Description	Possible GAI Uses
Student self-placement (SSP)	Processes in which students have partial or total control over their own course placement. SSP processes typically include learning about the local writing curriculum, often include filling out questionnaires, and might involve completing a reflective writing task. In some versions of SSP, students' placement decisions may be subject to reviewer approval (see Pantelides & Whittig, 2024; Reifman et al., 2025; Tinkle et al., 2022, 2024; Toth, 2019).	GAI uses depend on the structure of the SSP process. They may be able to extend current corpus analysis methods. Primary uses: • Score student writing in SSP processes that require evaluation of student placement decisions (e.g., Reifman et al., 2025; Wang, 2020) • Offer student feedback • Prompt student reflection • Refer students to resources/support • Rapid identification of time-sensitive student concerns expressed in writing • Large-scale corpus analysis of student reflective writing • Large-scale analysis of growth in student writing
Multiple measures	Processes that draw on multiple sources of evidence regarding student writing ability to make a course placement determination, often combining several of the kinds of data from the approaches listed above (Bahr et al., 2019; Klausman et al., 2016; Mokher et al., 2024).	GAI uses depend on which measures are used. Primary uses: • Score student essay exams, portfolio contents, or reflective writing tasks • Refer students to resources/support • Rapid identification of time-sensitive student concerns expressed in writing • Large-scale analysis of student reflective writing

We encourage literacy program coordinators to use this table as a starting point for informed consideration of which GAI-assisted writing placement paths are most appropriate in their local context (§1.4). Regardless of the path(s) chosen, this table can be a catalyst for reimagining the ways students are being placed and thinking through ways of place-making through writing assessment (§2.6.2, §5.4).

5.2 Evaluating GAI Competence

At present, we do not know whether GAIs pre-trained to generate language will also be able to provide evaluations of student writing with consistent and robust evidence of fairness, validity, and reliability. Indeed, creating a human-evaluated writing placement approach that fulfills these criteria is challenging. Throughout this white paper, we have traced the concatenation of fairness, validity, and reliability (§2.2, §4). For validity studies, a good starting place is examining the extent to which particular GAI platforms are able to evaluate the wide range of potential writing characteristics valued in the local writing construct. Evaluating for reliability traditionally includes inter-rater reliability and intra-rater reliability measures; with GAI, such evidence must include both consistency with human evaluation *and the GAI instrument's consistency with itself*. Finally, if we take Elliot's (2016) claim that "fairness is the first virtue of writing assessment" as our guide, we must turn attention to "forms of evidence that are social, legal, and ethical" (Poe & Elliot, 2019) to attend to fairness. In practice, this means addressing the established linguistic and cultural biases in GAI (§3.3.1), as well as examining socioeconomic, gender identity, and disability-related bias. Furthermore, by incorporating heuristics such as the justice-oriented validation framework advanced by Randall and colleagues (e.g., 2021, 2023, 2024), we can begin to create not only fairer methods of assessment, but also philosophies and praxis that encompass more just logics. Multidisciplinary teams can begin to understand more deeply the consequences of all actions involved with incorporating GAI into writing placement.

5.2.1 Evaluating for Validity

Evaluating a GAI instrument's capabilities to assess specific writing tasks can help create a cohesive assessment process with strong evidence of validity. Practitioners can begin this process by using the guiding questions included in §2.2.2: How well does the writing construct of this assessment relate to the literacy program's learning outcomes and curricula? How well does it align with other forms of writing? What role does student insight play in the process? What claims are we making based on the results about what a student knows and can do (see §4.3)? What evidence can we use to inform those claims? How does the placement practice correspond to other institutional measures, such as course grades, credit acquisition, persistence, and credentialing?

Here, we amplify von Davier and Burstein's (2024) *human-centered AI* principle: There should always

be a human-in-the-loop to make judgement calls and decisions (§5.3.4), both in the initial ADF phases and throughout the iterative and cyclic improvement process. Because writing placement has such significant consequences for students, we emphasize that an individual who is an expert in writing studies and has deep local knowledge of the institution's literacy program must be involved in the decision-making process.

Current GAI platforms might fare well when evaluating grammatical, lexical, and syntactical patterns in a writing sample, especially when that text is being evaluated for its adherence to SWE, as has been found in AWE scholarship over the decades (e.g., Shermis et al., 2013). If literacy program coordinators seek to develop a GAI-assisted writing placement instrument that evaluates texts for more complex traits, such as paragraph development or integration and use of sources, more complex and defined prompting or RAG integration will likely be necessary (see §3.2.1, §3.2.2). Those looking for a GAI-assisted writing placement instrument to evaluate texts on aspects of rhetorical performance—for example, attending to the needs of particular discourse communities (academic or non-academic); effectively integrating culturally grounded forms of storytelling; or using what Faison (2022) identified as Afrocentric rhetorical patterns such as call out, call and response, or repetition (see also Carter-Tod, 2021; Young & Robinson, 2018)—might consider using local datasets with appropriate representation to finetune an existing LLM or pre-train an SLM (§3.2.3, 3.2.4) and prioritize local human evaluation.

When training GAI to evaluate writing, identifying the norms and exceptions of each trait is useful. Language models are pre-trained using different datasets (§3.1.2), most of which do not include rated student writing, assignment prompts, or rubrics, and LLMs regularize patterns that are biased toward both their dataset and toward high-frequency languages from their internet samples (§3.3.1). To account for these biases, GAI-writing placement teams should evaluate the tool's performance related to the range of traits and discourse forms the specific literacy program prioritizes. Performing this type of evaluation can contribute to informed discussions (Figure 8) and decisions about which path(s) a literacy program will take (§1.4). Once a program incorporates GAI into its writing placement process, it should make regularly evaluating the instrument's capabilities part of the improvement cycle. Such ongoing evaluation can help literacy programs consider whether to continue with a particular instrument or move to another, an option that might be possible as more GAI platforms become available to the public or as institutions develop their local datasets and GAI instruments.

5.2.2 Evaluating for Reliability

Reliability is a key concern in any instance of evaluating or responding to student writing and has been an area of inquiry in writing studies for

more than a century (e.g., Starch, 1913). The importance of reliability is why instructors use scoring guides, and why course management systems like Canvas or Blackboard offer the ability to review student submissions anonymously or by timestamp rather than alphabetical order. AWE scholarship has long focused on machine scoring reliability, often in comparison to human scoring (see Elliot & Klobucar, 2013; Foltz et al., 2013; White et al., 2015); in part, this is because practitioners can achieve greater consistency within a literacy program or large-scale scoring scenario, such as for the SAT or COMPASS exam, when raters can focus explicitly on a trait. Foundational studies by Huddleston (1954) and Diederich (1974) offer further discussion of scoring approaches and reader variability. Human-to-AWE scoring research has been repeated in recent human-to-GAI scoring work (see Li, et al, 2025; Liu et al., 2025; Mansour et al., 2024; Mizumoto & Eguchi, 2023; Naismith et al., 2023). After human-to-human reliability has been established, human-to-GAI reliability can be measured to inform choices about the technology's use for writing assessment in specific contexts. We additionally recommend performing GAI instrument-to-GAI instrument (machine to itself) reliability testing. GAI presents a new challenge in reliability testing: Whereas past iterations of AWE technologies were designed to give consistent scores, GAI is designed to give different output with each iteration. When incorporating GAI into writing assessment, context expands from location, population, staffing, writing task, and other familiar variables to consider available GAI platforms and measure their reliability against each other.

In Appendix A, we have provided a simplified overview of standard writing assessment evaluation methods that can be used to measure reliability for processes that involve scoring. This appendix includes information regarding sampling plans, score distributions, and scoring discrepancies. All of these analyses can be conducted manually or with readily available computer programs (e.g., Excel, STATA, SPSS, R, Python). (While we hope the Appendix A overview is useful for literacy program coordinators who might have limited experience with statistical analysis, we emphasize the value of collaborating with a multidisciplinary team or a colleague grounded in quantitative research methods.) The findings of these evaluations should be discussed with stakeholders, and multidisciplinary teams should use these methods to regularly reevaluate the placement instrument—both the GAI platform selected and the methods used to train it for localized assessment.

5.2.3 Evaluating for Fairness

To attend to fairness, multidisciplinary teams must consider social, legal, and ethical impacts of any writing placement decision (assisted by GAI or not). Practitioners must design culturally sustaining literacy curricula and corresponding placement processes so all students have the opportunity to demonstrate their learning. In

other words, the entire assessment ecology must be designed from a fairness-first standpoint. Once designs have been implemented, teams can conduct evaluations for fairness in GAI-assisted writing placement with descriptive or inferential statistics. Evaluating for fairness necessarily involves reviewing all traits, scores (when applicable), and other evaluation results once an assessment is complete to determine whether there is any over- or under-representation of student communities; practitioners may use IPEDS identity categories (e.g., race, ethnicity, gender identity, Pell-grant status, veteran status), as well any locally pertinent communities. The linguistic biases already documented in datasets used for pre-training LLMs (§3.1.2) and LLMs' preference for their own linguistic markers (§3.3.1) make language bias a particularly prominent fairness concern.

Descriptive statistics (counts and percentages) can give simple and reader-friendly ways of gauging fairness (for examples related to race and writing assessment, see Inoue, 2015; Inoue & Poe, 2012). More specifically, the disaggregated counts of students taking a placement exam; placing into the courses offered; and registering, enrolling, and succeeding in a selected course offer revealing information about whether any student community is over- or under-represented, which may indicate potential biases or unfair practices (see the Center for Urban Education's *Racial Equity Toolkit*, 2020). Examining scoring differences between humans and a GAI instrument might also be revealing. We provide more information about scoring inquiries in Appendix B.

Inferential statistics can be another valuable tool to inform the evaluation of fairness in GAI-assisted writing placement. Inferential statistics at a student level can offer more specific information for interpretation by showing associations, for instance, between individual demographics and placement results. This format differs from most readily available institutional-level data, which typically reports the demographics of an entire group of students in aggregate, with no individual descriptors of any one test-taker. If student-level data linking individual demographics and results are available, these data can be used to calculate correlations for analysis.

Several inferential statistical methods can be used for analyzing correlations between individual demographics and test scores (e.g., ANOVA, structural equation modeling, t-tests). Regression models might be particularly useful for an analysis of GAI-assisted writing placement processes and can offer correlation information between demographic traits and scores. To achieve fairness, generally no relationship should exist between a student's identity and evaluations of the writing they produce in a GAI-assisted writing placement process. We offer an overview of regression modeling in Appendix B.

5.3 Scoring Student Writing With GAI

Again, not all writing placement processes involve scoring student writing, but many do—in some cases because the literacy program chooses such an approach and in others because some form of scoring is mandated by the institution or statewide education policy (e.g., Toth et al., 2024). As they develop frameworks for theory building (§4) and work through the GAI-assisted writing placement improvement cycle (§5.0), literacy program coordinators using GAI-assisted scoring in their placement processes will need to design, revise, and possibly realign evaluation practices. Identifying a scoring typology, describing the norms and exceptions, ensuring scoring accountability, and modeling a well-founded theory of scoring all play an integral role in such design. Below, we define and present key considerations for each of these aspects of scoring (re)alignment.

It is important to note that when using GAI in writing placement, an institution's multidisciplinary team will need to provide context to the instrument in each scoring session. This context must include the scoring typology, scoring criteria, and norms and expectations. Additionally, context could include the institutional mission, vision, and values; programmatic student learning outcomes; and program-level guiding principles or curricular emphases.

5.3.1 Scoring Typology

As in any assessment process, writing placement that incorporates GAI-assisted scoring requires literacy program coordinators to identify a) the discourse forms that will be evaluated in a given assessment, and b) the scoring criteria used to evaluate those discourse forms. A scoring typology can include analytic, holistic, and primary trait scoring, or a combination thereof. An example of an analytic scoring system is a traditional rubric, where various criteria or characteristics are identified, along with descriptors of levels of achievement for each criterion. Holistic scores refer to evaluating a text as a whole. Raters might evaluate based on a descriptive scale (often from A–F) or on a binary (ready/not ready for Composition I). Primary trait scoring, developed specifically for writing assessment (Lloyd-Jones, 1977), identifies and explicates key features of a text for raters to evaluate independently; trait scores are often supplemented with a holistic score. This method allows for greater flexibility than a traditional analytic scoring guide. *The Journal of Writing Analytics*, for instance, implements such a model in its peer-review process.

Take, for example, the University of Utah's FYWA, where students learn about the first and second courses in the two-semester first-year composition sequence (WR1 and WR2), decide whether they want to try to enter directly into WR2, and then write an essay to make their case for advanced placement (§4.2-3). Here, an interactive GAI instrument could provide students with formative

feedback on their draft response prior to submission, feedback that aligns with the local construct of writing and the four domains of writing (intrapersonal, interpersonal, cognitive, and health; Aull, 2023) discussed in §2.2. Students could consider the instrument's feedback and choose to revise and edit their writing, prior to submitting it for review. The text they submit could then be scored according to specific traits (e.g., the traits currently scored by human raters: curricular understanding, evidence of prior learning, and writerly self-reflection) as well as holistically (e.g., an overall score that includes both the traits and writing features not covered by the traits). The scoring typology would thus include exposition, evidence-based argument, and reflection, among other discourse forms.

5.3.2 Scoring Norms

Once the scoring typology for the identified discourse forms is determined, literacy program coordinators must describe their norms and exceptions, as well as the discourse communities in which the forms are practiced. Scoring norms and exceptions must be understood as fundamentally rhetorical, and therefore sociocultural, with all the attendant possibilities for advantage, disadvantage, and inequity. As discussed in §3.2, recent research on in-context learning (ICL) holds promise for GAI applications in writing placement. For scoring specifically, ICL could be used to "warmup" (Dong et al., 2024) the placement instrument once it has been pre-trained and fine-tuned. If literacy program coordinators create a rhetorical précis to accompany the scoring guide that provides critical context and rationale for the discourse forms, norms, and exceptions, including potential implications for fairness, the GAI instrument might then be able to apply the précis to the scoring instance. Additionally, as Wei and colleagues (2023) suggested, coordinators could provide a few exemplar models with a key of "input-label mappings" (p. 1), identifying the kinds of language patterns that are to be treated positively or negatively. The GAI instrument can scan the exemplars for the input-labels and then apply those insights to the writing to be scored. If those labels match the model's prior semantic understanding, the scoring accuracy ought to be relatively high (around 90% for some models); if those labels do not match the model's prior semantic understanding, the scoring accuracy has been shown to be relatively high for LLMs and closer to random guessing for SLMs (Wei et al., 2023). Importantly, Wei and colleagues (2023) found that GPT3 showed performance closer to random guessing in all instances.

A literacy program's writing placement prompt or constructed response (CR) task must be designed in concert with the scoring typology and identified genre conventions. As in any scoring scenario, expected discourse features need to be identified, as well as the various traits that are prioritized in a local literacy program and how they may emerge in student writing. For GAI-assisted writing placement, the multidisciplinary team must additionally identify which traits are likely to be immediately apparent, which traits might be less

obvious, and which might not be present at all while still being a successful piece of writing for that particular setting.

To attend to fairness, literacy program coordinators and multidisciplinary GAI-assisted writing placement teams must carefully consider which students might be advantaged or disadvantaged when aspects of scoring typology are left implicit. Additionally, imagining how GAI might be incorporated into writing placement offers literacy program coordinators an opportunity to determine alternative means for students to fulfill a writing task and demonstrate their literacies. The CR task and the scored student writing resulting from it can be used to train future iterations of the GAI-assisted placement instrument to help programs and institutions respond better to local student populations and to work their way toward place-making through writing assessment. We discuss ethical student data collection in GAI-assisted writing placement in §2.4.2.

5.3.3 Scoring Accountability

As faculty and program coordinators, we are entrusted not only with literacy education in classroom contexts but also policy setting—including in gate-keeping scenarios such as writing placement, which often determines access to college credit-bearing writing courses and affects the cost (in terms of time and money) of completing degree requirements. We thus have an ethical responsibility to our stakeholders, particularly our students. The accountability agenda in education has been widely critiqued, but we can remain critical of the neoliberal ideologies undergirding that movement while recognizing the vital responsibility we hold for our students' literacy education and access to opportunities to learn (see Warnke & Higgins, 2018; Higgins & Warnke, 2020). This responsibility is particularly keen when considering GAI-assisted writing placement.

Accountability measures are essential for any literacy program incorporating GAI into writing placement: GAI has a history of hallucination, irrelevant output, and machine confusion (e.g., Sun et al., 2024), in addition to issues related to bias (§3.3.1), social and environmental impact (§3.3.2), respect for intellectual property, and academic integrity (§3.3.3). The risks are simply too high to leave decisions related to students' education to chance. At a minimum, if a literacy program proceeds with GAI-assisted writing placement, we recommend prompting the instrument to include the following information alongside scores (or other evaluative output) on student writing:

- **Evidence from the student-authored text** (e.g., specific words, phrases, or sections) used by the GAI instrument to generate the score or other evaluation.
- **Chain of thought (CoT) reasoning**, which requires GAI to provide a list of reasoning steps that led it to an outcome (i.e., a

particular score or evaluation). CoT reasoning can be executed in one prompt; across several prompts that, in essence, are a CR; or even across different GAI platforms (see Renze & Guven, 2024, for recent best practices).

- **Comparable dataset examples**, which give further evidence of a reasoning process. Requiring the GAI instrument to identify which sections of a dataset were used to produce the output can provide literacy program coordinators and multidisciplinary team members with useful information to understand and evaluate the logic used. Team members can apply that information to adjust individual scores, the dataset used for training the GAI instrument, or the contextual information provided in a scoring session.

5.3.4 Human-in-the-Loop

Including at least one human-in-the-loop is necessary for responsible, human-centered AI-assisted writing assessment (von Davier & Burstein, 2024; see also Mollick, 2024). As von Davier and Burstein (2024) have written, human-in-the-loop procedures

> involve integrating human intelligence with AI systems to create a symbiotic relationship that enhances the performance and capabilities of both parties . . . This collaboration between human and machine learning algorithms allows for a continuous improvement loop, as human experts can correct the AI system's mistakes, while the AI system can learn from these corrections and update its recommendations. The human weighs in on the final decision, through direct grading or through proctoring and evaluation. (p. 100)

Literacy educators who engage GAI in writing placement must calibrate and affirm that the trained GAI instrument is calculating correctly and consistently. Because LLMs work from probability and generate differing output, there is no guarantee that even an extensively trained GAI instrument can accurately detect or evaluate writing traits and discourse features valued at local sites or do so consistently. It is necessary for humans to evaluate those accountability elements (§5.3.3) to ensure fairness by identifying issues that need overriding, future finetuning, or revised prompting—or to determine whether the unintended negative consequences of using a particular approach to GAI-assisted writing placement are so significant as to require a different method.

We cannot overstate the importance of disciplinary expertise in decisions made about student writing placement. Postsecondary educational institutions in the United States must always have human experts in writing pedagogy, writing assessment, and machine learning involved when

incorporating GAI into writing placement, and human experts in writing pedagogy must always be involved in making decisions based on the results of such placement processes.

5.3.5 Comparison with Historical Data

Whether prompting GAI platforms, integrating a RAG system, finetuning existing LLMs, or building SLMs, humans must supervise all writing placement efforts to attend to evidence related to fairness, validity, and reliability. Literacy program coordinators must additionally regularly compare the evaluations generated by GAI instruments with the results of their institution's historical placement methods. As discussed throughout this white paper, in particular in Principle 2 and in §5.2, postsecondary educational institutions have a responsibility to ensure all students have an opportunity to demonstrate their learning in any writing placement process, including one that incorporates GAI.

Finally, it is important to recognize that most GAI platforms are hosted and revised externally, outside the control of a local literacy program. Because their outputs vary (e.g., due to the degree of randomness in model output, updates, and deprecations), we cannot assume a model's behavior will remain stable over time. This instability means that the GAI platform on which a literacy program's placement system is built could change at any time and in any way without notice. So far, larger open-source companies (e.g., OpenAI, Google, Anthropic) have been fairly upfront about changes in their platforms by giving them new names (e.g., GPT3, GPT4.), but they are not obligated to do so and might not continue to; other companies might not be as transparent, now or in the future. Ongoing validation of model output against historical benchmarks is therefore essential.

5.3.6 Scoring Theory and Evidence

As discussed in §4, we encourage literacy program coordinators and multidisciplinary teams designing GAI-assisted writing placement processes to incorporate assessment practices that resonate with a given theory and collect forms of evidence related to fairness, validity, and reliability. For placement processes that involve scoring, we believe the scoring theory—from typology to collection to pedagogy—should be amenable to contemporary sociocultural assessment methods (Bennett et al., 2025). That is, the assessment should strive toward inclusion of connected identities; expression of substantive knowledge; design for deeper learning; adaptation to personal characteristics; and characterization of results as interaction among what a student brings to the assessment, the tasks engaged, and the conditions and context of that engagement. As Montenegro and Jankowski (2024) articulated, "By being mindful of how culture affects students' meaning-making processes, cognition, and demonstrations of learning, we can better understand and appreciate the learning gains that students make" (p. 13), which we see as an essential step in transforming writing placement (§5.4). In such an

environment, it is possible to collect evidence related to fairness, validity, and reliability as specified in the *Standards for Educational and Psychological Testing* (AERA et al., 2014), as well as in social justice expansions of such standards (Poe et al., 2023) and as articulated in §2.2, §4.0–§4.4, and §5.2.

5.4 Beyond Scoring: Transformative Uses for GAI in Writing Placement

While scoring has a long history in writing assessment, including decades of innovation and research on AWE, GAI presents opportunities to move beyond a focus on scoring as the sole, primary, or default form of evaluation in writing placement processes. New possibilities emerge from GAI's potential to provide near-instantaneous formative feedback. Imagine, for example, a placement process in which a student submits an essay, not hastily scrawled in a tense 40 minutes, but thoughtfully written and revised with ample time and space to do their best possible work. Upon submission, GAI-powered analytics immediately illuminate patterns in language, pinpoint strengths in sentence structure, highlight sophisticated understanding of rhetorical concepts, or suggest areas that could benefit from elaboration or other revision to meet an identified discourse community's needs as explicitly presented in the CR task. The student might then make use of these insights to aid self-reflection and revision, as other students have done with AI applications such as DocuScope (Kaufer et al., 2006, 2018; Wetzel et al., 2021) and Writing Mentor (Burstein et al., 2018). They could, as Sperber and colleagues have suggested (2025), use responses from the GAI instrument to review the CR prompt, consider and even discuss with others how their writing fulfills *their own goals* as a writer, and consider how the response might be revised to better achieve those goals.

Precedent exists for such dialogic GAI-assisted writing assessment: Duolingo's Interactive Writing task. First, test takers are asked to respond to a prompt for five minutes. Their first response is analyzed in real-time by the company's GAI model for topic-relevant themes, and students are asked a related follow-up question based on their initial response. They then have three additional minutes to respond. However constrained, this interaction is intended to more authentically reflect real-world writing scenarios (see Contexts in Figure 3 above) and to elicit greater evidence of test takers' writing proficiency by prompting elaboration (Runge et al., 2025).

Current research consistently points to specific linguistic features—such as sentence complexity, nuanced phrasing, and cohesive structure—as markers of proficient academic writing (Aull, 2020; Brown & Aull, 2017; Crossley, 2020). GAI algorithms, carefully trained on extensive student data, might be able to identify these patterns as such—or patterns valued in a variety of other social and cultural contexts beyond academia—in ways traditional assessment methods might not be able to (Wanderley & Epp, 2021; Katinskaia & Yangarber, 2021). This functionality allows for

targeted, near-instantaneous *formative* feedback (Hazelton et al., 2021). We can ask whether automated feedback should be viewed as a collaboration between writer and technology and, in turn, recognized as a legitimate approach for students to demonstrate their knowledge, skills, and abilities. If our local writing construct allows for this recognition, we might consider how a GAI-assisted writing placement process can facilitate such human-machine collaboration. GAI-driven analysis of student writing might reveal that a student favors descriptive language, a trait that signals readiness for more advanced work (Aull & Lancaster, 2014; Biber & Conrad, 2019). Alternatively, the analysis could pinpoint specific metatextual structures that encourage students to reconsider their drafting approach (Schleppegrell, 2004, 2013). An interactive approach to GAI writing placement, drawing on formative rather than (exclusively) summative feedback, might offer students something transformative: an opportunity for introspection. Instead of merely receiving a score or placement verdict, students could see their writing analyzed in ways that highlight their choices and habits, and they would get this feedback much more quickly than, for example, the lone writing placement coordinator at a large, urban two-year college might ever be able to provide. Students could consider questions designed to support metacognition and self-efficacy and thus get a glimpse into the kind of reflection they will continue to experience in their institution's literacy program. In this way, AI—generative or otherwise—could facilitate a learning experience in its own right, one that hospitably and interactively communicates the local construct of writing with students entering this new literacy context.

5.4.1 Curricular Connections

Because writing placement is embedded within assessment ecologies (see Inoue, 2015; Poe et al., 2018), it is important to postulate the pedagogical sequences that follow this initial "threshold" interaction with a literacy program in ways that will enable and encourage writers to thrive. Considering curricular connections can help ensure that placement is not merely a sorting activity, separated from pedagogical contexts and replicating decades of social stratification based on judgments about language.

For example, in the first weeks of the semester, writing instructors could devote class time to helping students understand the role that the GAI-assisted instrument, driven by human processes and decisions, played in their placement experiences. Students could reflect on—and critique—the feedback provided, as they have in a section of Composition I Jessica taught in Spring 2024 (Deshantae et al., 2024) and as is emerging in the scholarship (e.g., see Li, 2025; Messeri, 2023). Instructors could involve the class in identifying confidence levels for CR-task evaluations across various GAI platforms. De-identified samples could be discussed in class through score analysis, and lexical, grammatical, and syntactic language patterns (features other

than the trait scores) could be analyzed, using the same GAI instrument supporting the placement process. Such analyses could then be compared to GAI-generated feedback or evaluations. Instructors could look to the Peer and AI Review and Reflection (PAIRR) model, developed by Sperber and colleagues (2025), as an example of this type of instructional activity. With PAIRR, students elicit feedback from peers and from GAI, compare and reflect on that feedback, make revision plans, revise, and reflect on their processes.

Such processes mirror the continuous improvement cycle discussed in §5.0 and can be a starting point for instruction in ethical GAI use as part of students' ongoing development as college writers, or as part of an introductory GAI literacy course, which some postsecondary U.S. institutions are currently developing. In this vision of placement, students use an institution's GAI-assisted writing placement process to evaluate their writing and then critically consider their relationship with GAI through the improvement cycle. Students submit their writing to the GAI-assisted placement instrument, review the ways in which GAI has evaluated their work, engage in critique or validation of the process, and suggest improvements to prompts or training or decision trees. Such a practice affords students the opportunity to engage with critical GAI literacy practices and the rhetorical work of prompt engineering, which could mitigate some of the concerns regarding the learning opportunities students lose through cognitive off-loading with extensive, uncritical use of GAI while writing academically (Gerlich, 2025; Kosmyna et al., 2025).

5.4.2 Placement as Conversation

Traditional placement tests end abruptly, often with no further dialogue, no closing the loop of the improvement cycle for individual students as writers. The student receives a score and corresponding placement verdict; an institution opens or closes doors as a result. GAI-assisted writing placement could offer an opportunity to turn these moments of assessment into more dialogic processes, introducing descriptive lexical, grammatical, and syntactic insights; allowing for new evaluations of discourse coherence, topic, and rhetorical awareness to inform placement conversations; and inviting more meaningful interconnections between students and literacy programs. In this approach, GAI does not replace human relationships but rather becomes a means of connecting students with meaningful, contextualized human learning interactions. An instructor reviewing a GAI instrument's output might sit down with a student to discuss the particular patterns that instrument identified in the student's writing. These insights are not judgments but invitations: "Let's talk about these writing choices. What do you think about this analysis? What were you aiming for here?"

If integrated as a support for metacognitive introspection, GAI-assisted writing placement could act as a starting point for conversation and transformation, rather than an isolated testing

experience. The focus shifts from assessing student writing in a vacuum to interacting with students as writers, fostering conversations among students and instructors about why certain writing choices matter and how students might want to grow as writers in the years to come (Inoue, 2015). Recently, Huang, Palermo, and Wilson (2025) found students who *meaningfully engaged* with a computer-mediated instrument while revising their drafts demonstrated deeper uptake of revision strategies and improvement in their writing. GAI is relatively new, and this is an early study situated within a middle school district; we remain curious about its applicability to writing placement in postsecondary contexts. The point, though, is that in a GAI-assisted writing placement process, students do not have to be passive recipients of judgment decisions or scores. Students can ask questions about and even disagree with GAI-generated feedback related to their writing.

Taking inspiration from colleagues in corpus analysis writing research (e.g., see Lang et al., 2023), we can also imagine how GAI could offer literacy programs opportunities to conduct sentiment analysis across large swaths of student writing. Such analyses could, for example, provide opportunities for instructors, advisors, writing center consultants, and librarians to discuss how students' writing confidence manifests when they write as part of the initial placement process compared to when they submit an end-of-term portfolio. These conversations could be illuminated by conversations with students about how they felt and what they were thinking about as they completed the placement process, as well as how they implemented various strategies throughout the semester to achieve their goals. We see possibilities for expanding existing research on designing prompts for reflective writing (e.g., Messina et al., 2023) or strategies for revision (e.g., Holcomb & Buell, 2018), as well as using GAI to aid instructors in analyzing writing at program and institutional scales.

Indeed, we can imagine an approach to GAI-assisted student self-placement (SSP) that dispenses with scoring writing entirely to embrace the process as an explicitly non-violent, nonjudgmental conversation between students, instructors, literacy program coordinators, and the wider student support ecology in and beyond the institution. This GAI-assisted SSP process could provide a rich, interactive introduction to the localized literacy context that attends to all four of Aull's (2023) domains for a robust writing construct: cognitive and discursive (reasoning, memory, and knowledge, including discourse and genre knowledge), interpersonal (collaborative), intrapersonal (reflective and self-monitoring), and health or well-being (including material security). The placement instrument could provide an engaging multimodal welcome to the institution and literacy program, and then invite students to write responses to a series of reflective CR tasks relating to areas such as the following, all of which encompass several if not all of Aull's (2023) domains:

- Students' prior reading and writing experiences
- What students already know about their strengths and preferences as learners
- Students' interests and literacy practices in and beyond school
- Students' goals and motivations for being in college
- Students' worries or anxieties about college, both relating to literacy and more broadly
- Students' out-of-school responsibilities, community commitments, and support systems
- Potential challenges or barriers students might face in the coming term and throughout their time in college
- Examples of reading and/or writing assignments that represent the kinds of literacy tasks students are likely to encounter in their college courses

CR tasks could also include opportunities for students to upload writing they have done in other contexts for immediate GAI-generated commentary about how that writing relates to the local writing construct, and tailored reflective questions could invite students to consider that relationship more deeply (for more on SSP question design, see Toth, 2019; Toth & Aull, 2014). A locally-trained GAI instrument, or one that has been provided rich contextual framing, could, in real-time, execute the following: affirm the importance of the student's experiences, interests, concerns, and out-of-school communities; request more information or detail; prompt additional reflection; offer relevant insights about the local construct of writing in terms that align with the language the student is using; and ultimately present the student with an accessible overview of multiple possible combinations of support tailored to their responses. In this scenario, feedback from the GAI instrument would include a direct explanation of how its recommendations relate to the information and reflections students have provided in their writing.

In this hypothetical GAI-assisted SSP process, the interactive assessment goes well beyond matching students with courses. Support options presented by the GAI instrument could include combinations of particular course options with possibilities like concurrent studio hours, writing or learning center consultations, specific student services focusing on access and accommodations, extra- and co-curricular opportunities like student organizations, basic needs support from the college and/or community organizations, and/or career planning, undergraduate research, and community-engaged learning opportunities. Students could ask questions about the different options presented in response to their writing and receive immediate GAI-generated answers, rather than having to navigate the often-murky institutional processes or the not-clearly-marked hallways on campus to try to find the right person to help address their questions. If needed, the GAI instrument could also provide direct referrals and

embedded on-the-spot scheduling of meetings with human advisors or other people working in relevant student support units or community-based social services (e.g., disability services, identity-based student centers, TRiO and other programs for first-generation college students, financial aid, counseling and mental health support, free laptop and wifi hotspot loan programs, childcare services, campus or community food pantries, housing assistance, or medical care). Ultimately, students would be able to choose the combinations of support and resources they decide will best meet their needs and preferences, potentially a far more informed and individualized decision-making process than non-GAI SSP experiences have so far been able to support.

GAI could also help ensure that the interactive writing students undertake through this process directly informs local literacy instruction and program development. We find promise in this application of the technology, particularly as we recognize the material and labor conditions of postsecondary literacy educators at access-oriented institutions like community colleges. For instance, in the *TYCA Workload Survey* (N = 1,062), distributed to two-year college English faculty, Suh and colleagues (2020) found 42% of respondents teaching full-time were contractually obligated to teach 28–35 credit hours annually (p. 4) and 36% always or frequently taught beyond their required credit hours as an overload (p. 5). Hassel (2020) found first-year composition classes averaged around 22 students for all responding institutions (N = 435); the average was 25 for two-year colleges, with course caps ranging from 18 students to 39 students. Furthermore, instructors are typically off-contract until just shortly before a term begins, and students are able to enroll up through the second week of class at many institutions. Simply put, an instructor may not have enough time to carefully review student writing across five sections of twenty-some students, synthesize patterns of learning, and revise their lesson plans or curriculum in the first weeks of a semester. Contingent faculty and graduate student instructors at most four-year institutions —who, combined with two-year college faculty make up what Hassel and Giordano (2013) and Hassel and Phillips (2022) have called the field's *teaching majority*—face similar material constraints on their capacities to respond to individual students' needs and preferences.

A GAI-assisted SSP instrument could support responsive teaching by offering fast and usable feedback to literacy instructors. For instance, the instrument could compile a synthesized report for each course section that offers instructors an accessible and actionable overview of responses this particular group of students wrote in the SSP. That report could identify key recurring themes in students' writing as well as differentiated experiences, preferences, interests, and needs. It could also provide an account of the feedback generated—for instance, a count of how many students received information about basic needs resources or service-learning experiences—which could help instructors follow up with students on

how they are navigating the institution and whether they have received the support they need. Given the right training and contextualization, the GAI instrument could offer the instructor concrete pedagogical and curricular suggestions for that specific section grounded in the local writing construct and available programmatic and institutional resources. A literacy program coordinator could offer guidance to instructors on how best to use these reports within their local context, and the multidisciplinary GAI-assisted writing placement team could identify professional development opportunities for faculty seeking deeper insights into the team's process or the analytic capabilities and limits of the instrument.

Likewise, the GAI-assisted SSP instrument could produce program-level course reports for coordinators at the start of each term—reports programs would have near-immediate access to, and which can offer valuable insight into patterns of teaching, learning, and support. These reports could include suggestions for creating tailored professional development opportunities, recommendations about individual instructor needs based on section-level responses, and possibilities for curricular revisions in the short-term. Reports could also provide suggestions about what that term's responses indicate about the kinds of resources and support the program might need from other units on campus and in the community. Annual GAI-generated reports on SSP responses could inform iterative curriculum design and faculty professional development in and beyond the literacy program. This information could be used to inform annual assessment reports, cyclical program reviews, and regular accreditation reports, as well as institutional efforts related to general education learning outcomes assessment and writing in the disciplines or writing and reading across the curriculum efforts.

Ultimately, this kind of GAI-assisted SSP process would help move us out of the business of evaluating incoming students' writing for course-based sorting purposes. That enterprise remains fraught with the potential inequities that often result from consequential value judgments (human or GAI) about language, which is always entangled with differences across communities, identities, ideologies, and power. Instead, the process could provide a welcoming enactment of a locally responsive construct of writing that fully attends not just to cognitive and discursive domains, but to the interpersonal, intrapersonal, and wellbeing domains that experienced literacy instructors know are often *at least* as important as the cognitive and discursive in shaping students' writing, as well as their feelings of belonging, and ultimately their success in postsecondary education. This kind of SSP process could be an interactive, responsive learning experience that helps students understand and navigate the institution they are entering—and facilitates their connections and emerging relationships with actual humans. Rather than expecting students to undertake uncompensated literacy labor for our institutional purposes, slotting them into standardized courses with Taylorist efficiency, we can labor to become who they want and need us to be, who our *communities*

want and need us to be. Such a GAI-assisted process could move us toward more dialogic assessment *for* learning (§2.6.1), continuously seeking to learn with and from students to make our literacy programs and our institutions places where all students can thrive (§2.6.2).

5.4.3 Making Places Where All Students Can Thrive

As we think about the opportunities for conversation and reflection that GAI could open up in our writing placement processes, we imagine how these technologies might create new avenues for radical listening. As Williams (2018) wrote, "Listening, after all, is not just a matter of considering differing opinions; listening is an ongoing process that creates space for silence and reflection as well as creates opportunities for others to speak" (p. 412). GAI-assisted writing placement could potentially disrupt the binary between success and failure, knowing and not knowing. It could invite students to enter long-underway conversations about writing while offering faculty and institutions new possibilities for understanding the linguistic, cultural, and intellectual spaces students are already navigating. These insights, in turn, could help us provide students with nuanced and individualized support while centering their self-determined educational goals and sustaining the complex and diverse linguistic, discursive, and rhetorical knowledges they bring into our shared places of inquiry. In other words, we can continuously *remake our local construct of writing* by learning with and from our students and their communities.

At our most optimistic, we wonder if GAI can help dismantle placement as verdict, offering new means of attending to what students are saying, both within large swaths of data and in conversations with instructors. We wonder if it can offer opportunities to "explore how the praxis of listening offers socio-political agency . . . to make heard the voices of marginalized populations and reveal opportunities for growth deriving from difference" (Boehr, 2021). These kinds of radical listening might be possible at scale with GAI support, and such GAI-assisted listening might help us identify and redress institutional injustices in ways that even the most rigorously validated and equity-minded placement assessments have failed to deliver. It could provide additional footing for the ongoing work of unmaking oppressive and inequitable educational spaces and reshaping them based on the needs and desires of the least advantaged students entering them. It could help us work with all students to shape the institution's future on something closer to their terms. Such a shift could transform our initial writing assessment interactions with students from course placement to place-making.

Section 6 • Next Steps

Collaboration sustains us. As teacher-scholars who have found and sought to foster community across various subfields and institutional sites of writing studies, we know that statement is no platitude. Collaboration helps us see and experience writing as a social act taking place in a community (Graham, 2018). It helps us feel connected in an era when "loneliness epidemic" has become a stock phrase. It causes us to learn, to begin to understand perspectives different from our own, to create and nourish relationships with one another that enable us to do intellectual and programmatic work more than the sum of our individual parts. It helps us work through ideas as we hypothesize and modify and float new insights based on our colleagues' observations and feedback. In some cases, it helps us find a way forward when we are confronted with ethically dubious institutional demands that we have neither the expertise nor the political capital to confront alone. Collaboration requires vulnerability and trust. Co-authorship is one form of scholarly collaboration that we are deeply committed to and have undertaken in this document—and so is peer review.

Because of the rapidly developing technological landscape of GAI and research regarding its implications for writing assessment, *The Journal of Writing Analytics* designed a unique review process for this white paper. Editors Norbert Elliot and Alaina Tackitt sought out a group of twelve reviewers, many more than the typical one to three, from across several disciplines. We are grateful for the wisdom and generosity of specialists representing fields across writing studies, writing assessment, computers and writing, applied linguistics, translingual literacies, AI literacies, digital cultures, corpus linguistics, writing program administration, writing analytics, rhetorical theory, feedback literacy, and two-year college literacy studies.

This white paper documents one small constellation of colleagues coming together amid many other teaching, programmatic, and personal challenges in the summer and fall of 2025. While we initially imagined a short set of guidelines, we soon realized that the core concepts we needed to articulate, the amount of explanation required for audiences without GAI or writing assessment expertise, and the practical suggestions we wanted to provide required what amounted to a novella-length document. Some of us needed this document to exist so we could proceed ethically within our own institutional contexts; others, so we could imagine how we might proceed in the future, or how we might better support colleagues. We have worked together to articulate principles and practices we hope can help guide decision making when colleagues at a range of institution types encounter—with varying degrees of volition and enthusiasm—the prospect of incorporating GAI into writing placement. Our positions reflect our individual and collective lived experiences, beliefs, and institutional positions, as well as the many thoughtful resources, suggestions, and critiques our peer reviewers volunteered amid their

own responsibilities in late summer and early fall. We offer the resulting white paper as a way to make sense of our current moment—and as a way to make meaning together going forward.

The next step in our ongoing validation process is turning this white paper over to readers like you. Throughout the 2025–2026 academic year, we have and will be creating opportunities for colleagues across and beyond the United States to come together and discuss the perils and possibilities of GAI-assisted writing placement. These in-person and virtual events will include the following:

- **Listening sessions**, during which participants can share their experiences, ideas, and concerns about GAI and writing placement.
- **Town hall meetings**, for which we will circulate the white paper in advance and members of our co-authorship team can answer questions and invite feedback and revisions to the document.
- **Workshops**, during which participants can draft their own theories of action, game plans, and assessment materials for GAI-assisted writing placement.

Table 6. *Tentative Schedule for Peer Review Events in 2025–2026*

Event	Location	Dates	Modalities
TYCA–Northeast	Baruch College, NY	October 2–3	Face-to-face
TYCA–Midwest	College of DuPage, IL	October 17	Virtual
TYCA–Pacific Northwest	Columbia Basin College, WA	October 17–18	Face-to-face
Developmental Education Reform Act (DERA; H.B. 2170) Symposium	Harper College, IL	October 24	In-house event
International Writing Analytics Conference	Tampa/St. Pete, FL	February 19–20	Face-to-face
TYCA National	Cleveland, OH	March 4	Face-to-face
TYCA Placement Network at CCCC	Cleveland, OH	March 5	Face-to-face
TYCA–West*	TBA	Spring / TBD	Face-to-face
CWPA*	University of Wyoming	July 19–25, 2026	Face-to-face

** denotes possible opportunities*

At each of these in-person and virtual meetings, we have and will share an electronic means of collecting feedback, which will be housed in a cloud storage drive only accessible to this white paper's co-authors. You can also find this feedback form here: https://humutah.co1.qualtrics.com/jfe/form/SV_b2SW3vkxYvl8brM. We are particularly curious about what GAI-writing placement path you are contemplating or already on (§1.4); how the decision-making principles in this document align with your individual or programmatic values (§2); how you might refine the theory of action (§4.1) for your local context; how the possible uses of GAI in writing placement (§3 and §5) resonate with you; and what concerns, objections, or reasons for resisting GAI placement you think we have failed to fully account for.

Our team will compile perspectives that participants share with us throughout the year, as well as insights gleaned from emerging international, U.S., and local research on GAI technologies (§2.5). Christie Toth and Tiffany Barney have taken on leadership of their department's GAI reading group; they will incorporate insights from faculty and graduate students at the University of Utah's Department of Writing and Rhetoric Studies, as well as the corresponding working group in Salt Lake Community College's Department of English, Linguistics, and Writing Studies that occasionally meets jointly. Christie will also convene an undergraduate research team in Spring 2026 that will review this document and be part of the multidisciplinary team planning the University of Utah's pilot GAI-assisted writing placement process. During the spring semester, Tiffany will undertake her dissertation study of reliability in GAI-assisted writing assessment. The findings of both these studies will inform our revisions to this document. Meanwhile, Jessica Nastal will be integrating aspects of this white paper into first-year composition classes at the College of DuPage in Spring 2026 and seek out students' perceptions, reactions, and ideas. In Summer 2026, we will reconvene, review information gathered over the academic year, and revise the white paper for a 2.0 update to be published in *The Journal of Writing Analytics* in Fall 2026.

Ultimately, we encourage you to expand the scholarship on GAI in writing placement with research from your own contexts. *JWA* accepts submissions for a range of genres (https://wacclearinghouse.org/jwa/submisssions/), and we imagine members of this community offering valuable contributions on GAI-assisted writing placement over the next year. For instance, a multidisciplinary team developing open-source code for SLMs that prioritize fairness, validity, and reliability could publish their research as an Innovation in Analytics report. A team focused on developing a local dataset and finetuning a GAI model might offer a Research Note. A literacy program coordinator might develop a GAI literacy curriculum that connects placement assessment and first-year writing for a Writing Analytics in the Classroom article. With enough relevant submissions, we have the opportunity to publish a special section of *JWA* on GAI and writing

placement to accompany the revised white paper in late 2026. Likewise, Jessica is serving as incoming co-editor of *Teaching English in the Two-Year College*, and she welcomes submissions focusing on GAI and writing placement at community colleges and other access-oriented institutions.

As you find ways to use *General Artificial Intelligence, Writing Placement, and Principled Decision Making in U.S. Postsecondary Contexts* this year, please connect with us. We are eager to learn more about your and your students' experiences.

References

Adisa, K., Byrd, A., Flores, L., Green, D., Hassel, H., Johnson, S. J., Kirschenbaum, M., Lockett, A., Losh, E. M., & Mills, A. (2024a). *Generative AI and policy development: Guidance from the MLA-CCCC Task Force* [Working paper 2]. MLA Knowledge Commons. https://aiandwriting.hcommons.org/working-paper-2/

Adisa, K., Byrd, A., Ene, E., Flores, L., Giordano, J., Green, D., Hassel, H., Hendrickson, J., Johnson, S. J., Kirschenbaum, M., Losh, E., Mckoy, T., Mills, A., Mina, L., Wynn Perdue, S., Ruttenberg, J., Wang, Z., Ward, J., & William, J. (2024b). *Building a culture for generative AI literacy in college language, literature, and writing* [Working paper 3]. MLA Knowledge Commons. https://aiandwriting.hcommons.org/working-paper-3/

Adelman, C. (1993). Kurt Lewin and the origins of action research. *Educational Action Research, 1*(1), 7–24. https://doi.org/10.1080/0965079930010102

Aguilar, G. L. (2024). Rhetorically training students to generate with AI: Social justice applications for AI as audience. *Computers and Composition, 71*(102828), 1–10.

Ali, A.M., Ali, A. M., Adan, A. H., Adan, A. O., Guyo, A. M., Labatt, A., Kairu, Al., Abdirahim, A., Joda, B. S., Tesfaye, B., Benard, B., Mulinya, B. K., Ranta, B., Pieterse, C., Arulogun, D. S., Kamau, D., Nkurunziza, D., Dawit, Ezekiel, E., . . . Anonymous signatory. (2024, May 24). Open letter to President Biden from tech workers in Kenya. *Foxglove*. https://www.foxglove.org.uk/open-letter-to-president-biden-from-tech-workers-in-kenya/?tpcc=NL_Marketing

Alim, H. S., Paris, D., & Wong, C. P. (2020). Culturally sustaining pedagogy: A critical framework for centering communities. In N. S. Nasir, C. D. Lee, R. Pea, & M. M. de Royston (Eds.), *Handbook of the cultural foundations of learning* (pp. 261–276). Routledge.

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

Andreotta, A. J., Kirkham, N., & Rizzi, M. (2022). AI, big data, and the future of consent. *AI & Society, 37*, 1715–1728. https://doi.org/10.1007/s00146-021-01262-5

Argyris, C. (1997). Learning and teaching: A theory of action perspective. *Journal of Management Education, 21*(1), 9–26. https://doi.org/10.1177%2F105256299702100102

Arieli-Attali, M., Ward, S., Thomas, J., Deonovic, B., & Von Davier, A. A. (2019). The expanded evidence-centered design (e-ECD) for learning and assessment systems: A framework for incorporating learning goals and processes within assessment design. *Frontiers in Psychology, 10*, 1–17.

ASA News (2016). American Statistical Association releases statement on statistical significance and p-values. https://www.amstat.org/asa/files/pdfs/p-valuestatement.pdf

Aull, L. L. (2020). *How students write: A linguistic analysis*. Modern Language Association.

Aull, L. L. (2023). *You can't write that: 8 myths about correct English*. Cambridge.

Aull, L. L., & Lancaster, Z. (2014). Linguistic markers of stance in early and advanced academic writing: A corpus-based comparison. *Written Communication, 31*(2), 151–183.

Badenhorst, C., Moloney, C., Rosales, J., Dyer, J., & Ru, L. (2015). Beyond deficit: Graduate student research-writing pedagogies. *Teaching in Higher Education, 20*(1), 1–11. https://doi.org/10.1080/13562517.2014.945160

Bahr, P. R., Fagioli, L. P., Hetts, J., Hayward, C., Willett, T., Lamoree, D., Newell, M. A., Sorey, K., & Baker, R. B. (2019). Improving placement accuracy in California's community colleges using multiple measures of high school achievement. *Community College Review*, *47*(2), 178-211.

Baker-Bell, A. (2020). *Linguistic justice: Black language, literacy, identity, and pedagogy*. Routledge.

Baker-Bell, A., Williams-Farrier, B., Jackson, D., Johnson, L., Kynard, C., & McMurtry, T. (2020). *This ain't another statement! This is a DEMAND for Black linguistic justice!* Conference on College Composition and Communication. https://cccc.ncte.org/cccc/demand-for-black-linguistic-justice

Bansal, P. (2024). Prompt engineering importance and applicability with generative AI. *Journal of Computer and Communications, 12*(10), 14-23.

Barney, T. B. (2023, September 29–30). *The large language model writing assistant* [Conference presentation]. TYCA West 2023 Conference, Salt Lake City, Utah, United States.

Basgier, C. (2025). "Never let a good [literacy] crisis go to waste": Writing across the curriculum administration amid artificial intelligence anxiety. In C. Wang & Z. Tian (Eds.), *Rethinking writing education in the age of generative AI* (pp. 97–109). https://doi.org/10.4324/9781003426936

Bedekovics, G., & Ragland, W. (2025, July 25). Mapping federal funding cuts to US colleges and universities. *Center for American Progress*. https://www.americanprogress.org/article/mapping-federal-funding-cuts-to-us-colleges-and-universities/

Beigman Klebanov, B., Ramineni, C., Kaufer, D., Yeoh, P., & Ishizaki, S. (2019). Advancing the validity argument for standardized writing tests using quantitative rhetorical analysis. *Language Testing, 36*(1), 125–144. https://doi.org/10.1177/0265532217740752

Belfield, C., & Crosta, P. (2012). *Predicting success in college: The importance of placement tests and high school transcripts.* Community College Research Center, Columbia University, Teachers College. http://ccrc.tc.columbia.edu/publications/predicting-success-placement-tests-transcripts.html

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big?🦜. *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610–623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Bennett, R. E., Darling-Hammond, L., & Badrinarayan, A. (2025). *Socioculturally responsive assessment: Present and future*. Taylor & Francis.

Ben-Shalom, R. (2011). Me' ir Nativ: The First Hebrew Concordance of the Bible and Jewish Bible Study in the Fifteenth Century, in the Context of Jewish-Christian Polemics. *Aleph: Historical Studies in Science and Judaism, 11*(2), 289-364.

Berthelot, A., Caron, E., Jay, M., & Lefevre, L. (2025). Understanding the environmental impact of generative AI services. *Communications of the ACM, 68*(7), 46–53.

Biber, D., & Conrad, S. (2019*). Register, genre, and style*. Cambridge University Press.

Boehr, C. (2021). The praxis of listening in feminist-relational research. *Peitho, 23*(3). https://wac.colostate.edu/docs/peitho/article/recoveries-and-reconsiderations-the-praxis-of-listening-in-feminist-relational-research/

Brown, D. W., & Aull, L. L. (2017). Elaborated specificity versus emphatic generality: A corpus-based comparison of higher- and lower-scoring advanced placement exams in English. *Research in the Teaching of English, 51*(4), 394–417.

Burstein, J. (2003). The e-rater(R) scoring engine: Automated essay scoring with natural language processing. In M. D. Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 113–122). Lawrence Erlbaum Associates.

Burstein, J., Elliot, N., Klebanov, B. B., Madnani, N., Napolitano, D., Schwartz, M., Houghton, P., Molloy, H. (2018). Writing Mentor™: Writing progress using self-regulated writing support. *Journal of Writing Analytics, 2*. https://wacclearinghouse.org/docs/jwa/vol2/bursteinetal.pdf

Burstein, J., & LaFlair, G. T. (2024). *Where assessment validation and responsible AI meet.* arXiv. https://doi.org/10.48550/arXiv.2411.02577

Byrd, A. (2023). Truth-telling: Critical inquiries on LLMs and the corpus texts that train them. *Composition Studies, 51*(1), 135–142.

Byrd A., Flores, L., Green, D., Hassel, H., Johnson, S.Z., Kirschenbaum, M., Lockett, A., Losh, E. M., & Mills, A. (2023). *MLA-CCCC joint task force on writing and AI working paper: Overview of the issues, statement of principles, and recommendations*. MLA Knowledge Commons. https://hcommons.org/app/uploads/sites/1003160/2023/07/MLA-CCCC-Joint-Task-Force-on-Writing-and-AI-Working-Paper-1.pdf

Carter-Tod, S. (2021, March 29). Rhetoric(s): A broader definition. *FEN Blog*. https://compstudiesjournal.com/2021/03/29/rhetorics-a-broader-definition/

Cassano, F., Gouwar, J., Lucchetti, F., Schlesinger, C., Freeman, A., Anderson, C. J., Feldman, M. Q., Greenberg, M., Jangda, A., & Guha, A. (2024). *Knowledge transfer from high-resource to low-resource programming languages for code LLMs*. arXiv. https://doi.org/10.48550/arXiv.2308.09895

Cauoette, B. L. (2019). Directed self-placement, corequisite models, and curricular choice. *Journal of Basic Writing, 38*(1), 56–77.

Center for Digital Thriving. (2024). *Teen and young adult perspectives on generative AI*. Graduate School of Education, Harvard University. https://digitalthriving.gse.harvard.edu/wp-content/uploads/2024/06/Teen-and-Young-Adult-Perspectives-on-Generative-AI.pdf

Center for Urban Education. (2020). *Equity-minded inquiry series: Data Tools*. Rossier School of Education, University of Southern California.

Chang, T., & Wiens, J. (2024). From biased selective labels to pseudo-labels: An expectation-maximization framework for learning from biased decisions. *Proceedings of Machine Learning Research, 235*, 6286–6324.

Chen, W.-L., Wei, Z., Zhu, X., Feng, S., & Meng, Y. (2025). *Do LLM evaluators prefer themselves for a reason?* arXiv. https://doi.org/10.48550/arXiv.2504.03846

Cohen, I. G. (2019). Informed consent and medical artificial intelligence: What to tell the patient? *Georgetown Law Journal, 108*, 1426–1469.

Cole, K., Hassel, H., & Schell, E. (2017). Remodeling shared governance: Feminist decision making and resistance to academic neoliberalism. In K. Cole & H. Hassel (Eds.), *Surviving sexism in academia: Strategies for feminist leadership* (pp. 13–28). Routledge.

Combahee River Collective (1981). A Black feminist statement. In C. Moreaga & G. Anzaldúa (Eds.), *This bridge called my back: Writings by radical women of color* (pp. 234–244). Kitchen Table Press.

Committee on CCCC Language. (1974). Students' Right to Their Own Language. *College Composition and Communication, 25*(3), 1–18. https://doi.org/10.58680/ccc197417210

Community College Research Center. (2026). *Community college FAQs: Community college enrollment and completion.* Columbia University. https://ccrc.tc.columbia.edu/community-college-faqs.html

Cottier, B., Rahman, R., Fattorini, L., Maslej, N., Besiroglu, T., & Owen, D. (2024). *The rising costs of training frontier AI models*. arXiv. https://doi.org/10.48550/arXiv.2405.21015

Cox, A., & Reidner, R. (2023). Persistence, coalition and power: Institutional citizenship and the feminist WPA. *Peitho, 25*(2). https://cfshrc.org/article/persistence-coalition-and-power-institutional-citizenship-and-the-feminist-wpa/

Crawford, K. (2021). *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Cronbach, L. J. (1988). Internal consistency of tests: Analyses old and new. *Psychometrika, 53*(1), 63–70.

Crossley, S. A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research, 11*(3), 415–443.

Cruz, T. M., & Webb, S. (2025). Social interventions in times of crisis: Community perceptions of integrating social determinants and equity into the COVID-19 response. *Social Science & Medicine, 118456*. https://doi.org/10.1016/j.socscimed.2025.118456

Cushman, E. (2016). Decolonizing validity. *Journal of Writing Assessment, 9*(1), https://escholarship.org/uc/item/0xh7v6fb

Dai, S., Xu, C., Xu, S., Pang, L., Dong, Z., & Xu, J. (2024). Bias and unfairness in information retrieval systems: New challenges in the LLM era. *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining*. Association for Computing Machinery. https://doi.org/10.1145/3637528.3671458

de Roock, R. S. (2024). To become an object among objects: Generative artificial "intelligence," writing, and linguistic White supremacy. *Reading Research Quarterly, 59*(4), 590–608.

Deshantae, S., Murphy, S., Arias, A., Bolega, E., Burley, E., Cannone, F., Contreras, M., Ferrer, A., Fonseca, R., Halpin, P., Hernandez, E.B., Liskey, T., Molinari, A., Verde, M., Zhang, S., & Nastal, J. (2024, May 9). *Artificial intelligence, academic integrity, and faculty*. COD|COMP [Composition Program at College of DuPage]. https://sites.google.com/view/codcomp/best-practices-resources/generative-ai-principles-and-recommendations?authuser=0#h.a4njgjxa9hq3

Deutsch, D., Dror, R., & Roth, D. (2022). On the limitations of reference-free evaluations of generated text. *Proceedings of the 2022 Conference on empirical methods in natural language processing*. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.emnlp-main.753

Diederich, P. (1974). *Measuring growth in English*. National Council of Teachers of English.

Dikli, S. (2006). An overview of automated scoring of essays. *The Journal of Technology, Learning and Assessment, 5*(1).

Diller, H-J., De Smet, H., & Tyrkkö, J. (2010). A European database of descriptors of English electronic texts. *The European English Messenger, 19*(2), 29–35.

Dong, Q., Li, L., Dai, D., Zheng, C., Ma, J., Li, R., Xia, H., Xu, J., Wu, Z., Chang, B., Sun, X., Li, L., & Sui, Z. (2024, November). A survey on in-context learning. *Proceedings of the 2024 conference on empirical methods in natural language processing* (pp. 1107–1128). Association for Computational Linguistics. https://aclanthology.org/2024.emnlp-main.64/

Edwards, D. W. (2020). Digital rhetoric on a damaged planet: Storying digital damage as inventive response to the Anthropocene. *Rhetoric Review, 39*(1), 59–72.

Elliot, N. (2016). A theory of ethics for writing assessment. *Journal of Writing Assessment, 9*(1). https://escholarship.org/content/qt36t565mm/qt36t565mm.pdf

Elliot, N., & Klobucar, A. (2013). Automated essay evaluation and the teaching of writing. In M. Shermis & J. Burstein (Eds.), *Handbook of automated essay evaluation* (pp. 13–35). Taylor and Francis.

Ericsson, P. F., & Haswell, R. H. (2006). *Machine scoring of student essays: Truth and consequences.* Utah State University Press.

Faison, W. (2022). Full disclosure: Black rhetoric, writing assessment, and Afrocentric rubrics. *Rhetoric Society Quarterly, 52*(3), 270–281. https://doi.org/10.1080/02773945.2022.2077627

Fernandes, M., & McIntyre, M. (2025). Recoveries and reconsiderations: Linguistic justice and storying resistance to generative AI. *Peitho, 27*(2), 66–76.

Flodén, J. (2025). Grading exams using large language models: A comparison between human and AI grading of exams in higher education using ChatGPT. *British Educational Research Journal, 51*(1), 201–224.

Flores, L., Gibson, A., Green, D., Johnson, S. Z., Kirschenbaum, M., & Losh, L. (2025, January 9–12). *Actionable AI in the humanities classroom: A workshop with the MLA-CCCC Task Force on AI* [Conference presentation]. MLA 2025 Convention, New Orleans, LA, United States. https://docs.google.com/presentation/d/1vSBk38rjpwQsMNhuketh8iq_ThYKKk19fhcwJxbopqQ/edit?slide=id.g2a4e596f38d_0_4#slide=id.g2a4e596f38d_0_4

Flores, N., & Rosa, J. (2015). Undoing appropriateness: Raciolinguistic ideologies and language diversity in education. *Harvard Educational Review, 85*(2), 149–171.

Foltz, P. W., Streeter, L. A., Lochbaum, K., & Landauer, T. (2013). Implementation and applications of the intelligent essay assessor. In M. Shermis & J. Burstein (Eds.), *Handbook of automated essay evaluation* (pp. 68–88). Taylor and Francis.

Forzani, E., Corrigan, J., Slomp, D., & Randall, J. (2024). Prioritizing equitable social outcomes with and for diverse readers: A conceptual framework for the development and use of justice-based reading assessment. *Educational Psychologist, 59*(4), 291–314.

Gadiraju, V., Kane, S., Dev, S., Taylor, A., Wang, D., Denton, R., & Brewer, R. (2023, June). "I wouldn't say offensive but . . .": Disability-centered perspectives on large language models. *Proceedings of the 2023 ACM conference on fairness, accountability, and transparency* (pp. 205–216). Association for Computing Machinery. https://dl.acm.org/doi/10.1145/3593013.3593989

Gao, L., Biderman, S., Black, S., Golding, L., Hoppe, T., Foster, C., Phang, J., He, H., Thite, A., Nabeshima, N., Presser, S., & Leahy, C. (2020). *The Pile: An 800GB dataset of diverse text for language modeling*. arXiv. https://doi.org/10.48550/arXiv.2101.00027

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., & Wang, H. (2024). *Retrieval-augmented generation for large language models: A survey*. arXiv. https://doi.org/10.48550/arXiv.2312.10997

Gegg-Harrison, W., & Shapiro, S. (2025). From policing to empowerment. In C. Wang & Z. Tian (Eds.), *Rethinking writing education in the age of Generative AI* (pp. 26–41). Routledge.

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies, 15*(1). https://doi.org/10.3390/soc15010006

Gilman, H., Giordano, J. B., Hancock, N., Hassel, H., Henson, L., Hern, K., Nastal, J., & Toth, C. (2019). Two-year college writing placement as fairness. *Journal of Writing Assessment, 12*(1). https://escholarship.org/uc/item/4zv0r9b2

Giordano, J. B., Horton, A. E., Lehman, E., & Phillips, C. (2025, July 10–11). *Redefining literacy coordination work: Mapping diversity in program labor* [Conference presentation]. Council of Writing Program Administrators 2025 Conference, online.

Godfrey, J. M. (2024). *Prompt and circumstance: Investigating the relationship between college writing and postsecondary policy* (Publication no. 31631201) [Doctoral dissertation, University of Michigan]. ProQuest Dissertations & Theses Global. https://www.proquest.com/pqdtglobal/docview/3101493300/abstract/8D9989DA72EF4E87PQ/1

Gong, W, Li, L., Dai, D., Zheng, C., Ma, J., Li, R., Xia, H., Xu, J., Wu, Z., Chang, B., Sun, X., Li, L., & Sui, Z. (2024). A survey on in-context learning. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 1107–1128. https://aclanthology.org/2024.emnlp-main.64.pdf

Graham, S. (2018). A revised writer(s)-within-community model of writing. *Educational Psychologist, 53*(3), 258–279. https://doi.org/10.1080/00461520.2018.1481406

Graham, S. S. (2023). Post-process but not post-writing: Large language models and a future for composition pedagogy. *Composition Studies, 51*(1), 162-168.

Gross, D. M. (2024). Construct validity and the demise of the Analytical Writing Placement Examination (AWPE) at the University of California: A tale of social mobility. *Journal of Writing Assessment, 17*(2). https://escholarship.org/uc/item/3gm126sq

Guo, K. (2024). EvaluMate: Using AI to support students' feedback provision in peer assessment for writing. *Assessing Writing, 61*. https://doi.org/10.1016/j.asw.2024.100864

Gupta, A. (2024). Translinguo: Critically making chatbot prototypes by learning how to write generative AI prompts. In C. Schnitzler, A. Vee, & T. Laquintano (Eds.), *TextGenEd: Continuing experiments*. The WAC Clearinghouse. https://doi.org/10.37514/TWR-J.2024.2.1.01

Gupta, A,, & Shivers-McNair, A. (2024). "Wayfinding" through the AI wilderness: Mapping rhetorics of ChatGPT prompt writing on X (formerly Twitter) to promote critical AI literacies. *Computers and Composition, 74*, 102882. https://doi.org/10.1016/j.compcom.2024.102882

Hammond, J. W. (2019). Making our invisible racial agendas visible: Race talk in *Assessing Writing*, 1994–2018. *Assessing Writing, 42*, 100425. https://doi.org/10.1016/j.asw.2019.100425

Hammond, Z. (2014). *Culturally responsive teaching and the brain: Promoting authentic engagement and rigor among culturally and linguistically diverse students.* Corwin.

Harrington, S. (2005). Learning to ride the waves: Making decisions about placement testing. *WPA: Writing Program Administration, 28*(3), 9-29.

Hartwell, K., & Aull, L. (2024). Navigating innovation and equity in writing assessment. *Assessing Writing, 61*, 100873. https://doi.org/10.1016/j.asw.2024.100873

Hassel, H. (2020). *Postsecondary writing course caps* [Google sheets spreadsheet]. https://docs.google.com/spreadsheets/d/16anb0UR

Hassel, H., & Giordano, J. B. (2013). Occupy writing studies: Rethinking college composition for the needs of the teaching majority. *College Composition & Communication, 65*(1), 117–139.

Hassel, H., & Phillips, C. (2022). *Materiality and writing studies: Aligning labor, scholarship, and teaching*. National Council of Teachers of English.

Hassel, H., Klausman, J., Giordano, J. B., O'Rourke, M., Roberts, L., Sullivan, P., & Toth, C.(2015). TYCA white paper on developmental education reforms. *Teaching English in the Two-Year College, 42*(3), 227.

Haswell, R., & Elliot, N. (2019). *Early holistic scoring of writing: A theory, a history, a reflection*. University Press of Colorado.

Haswell, R., & Wyche-Smith, S. (1994). Adventuring into writing assessment. *College Composition & Communication, 45*(2), 220–236.

Hazelton, L., Nastal, J., Elliot, N., Burstein, J., & McCaffrey, D. F. (2021). Formative automated writing evaluation: A standpoint theory of action. *Journal of Response to Writing, 7*(1), 37–91.

Heritage, M., & Wylie, C. (2018). Reaping benefits of assessment for learning: Achievement, identity, and equity. *ZDM: Mathematics Education, 50*(2), 729–741. https://doi.org/10.1007/s11858-018-0943-3

Herrington, A., & Moran, C. (2001). What happens when machines read our students' writing? *College English, 63*(4), 480–499.

Higgins, K., & Warnke, A. (2020). Reform as access, reform as exclusion: Making space for critical approaches to the neoliberal moment. *Basic Writing E-Journal, 16*(1), 1–27.

Hodges, R., Payne, E. M., McConnell, M. C., Lollar, J., Guckert, D. A., Owens, S., Gonzales, C., Hoff, M. A., O'Donnell Lussier, K., Wu, N., & Shinn, H. B. (2020). Developmental education policy and reforms: A 50-state snapshot. *Journal of Developmental Education, 44*(1), 2–17.

Holcomb, C., & Buell, D.A. (2018). First-year composition as "Big Data": Towards examining student revisions at scale. *Computers and Composition, 48*, 49–66.

Hood, S., & Hopson, R. K. (2008). Evaluation roots reconsidered: Asa Hilliard, a fallen hero in the "Nobody Knows My Name" project, and African educational excellence. *Review of Educational Research, 78*(3), 410–426.

Hosseini, M., Gao, P., & Vivas-Valencia, C. (2025). A social-environmental impact perspective of generative artificial intelligence. *Environmental Science and Ecotechnology, 23*, 100520.

Hu, S., Tu, Y., Han, X., He, C., Cui, G., Long, X., Zheng, Z., Fang, Y., Huang, Y., Zhao, W., Zhang, X., Thai, Z. L., Zhang, K., Wang, C., Yao, Y., Zhao, C., Zhou, J., Cai, J., Zhai, Z., … Sun, M. (2024). *MiniCPM: Unveiling the potential of small language models with scalable training strategies*. arXiv. https://doi.org/10.48550/arXiv.2404.06395

Huang, Y., Palermo, C., Liu, R., & He, Y. (2025a). An early review of generative language models in automated writing evaluation: Advancements, challenges, and future directions for automated essay scoring and feedback generation. *Chinese/English Journal of Educational Measurement and Evaluation, 6*(2). https://doi.org/10.59863/FAMJ7696

Huang, Y., Palermo, C., & Wilson, J. (2025b). Identifying active ingredients and uptake patterns in the implementation of an AI-based writing support tool: Insights from a randomized controlled trial. *Computers and Education: Artificial Intelligence, 9*, 100479. https://doi.org/10.1016/j.caeai.2025.100479

Hubrig, A. A., Olivas, B., Van Dyke, T., Lovett, S., Rousculp, T., Bernstein, S. N. (2022). Symposium: Cultivating anti-ableist action across two-year college contexts. *Teaching English in the Two-Year College, 49*(3), 257-272.

Huddleston, E. M. (1954). Measurement of writing ability at the college-entrance level: Objective vs. subjective testing techniques. *The Journal of Experimental Education, 22*(3), 165–213.

Huot, B. (1994). A survey of college and university placement practices. *WPA: Writing Program Administration, 17*(3), 49–65.

Huot, B. (1996). Toward a new theory of writing assessment. *College Composition and Communication, 47*(4), 549–566.

Huot, B. (2003). *Rearticulating writing assessment for teaching and learning*. University Press of Colorado.

Inoue, A. B. (2009). The technology of writing assessment and racial validity. In C. Schreiner (Ed.), *Handbook of research on assessment technologies, methods, and applications in higher education* (pp. 97–120). IGI Global.

Inoue, A. B. (2012). Racial methodologies for composition studies: Reflecting on theories of race in writing assessment research. In G. Kirsch (Ed.), *Writing studies research in practice: Methods and methodologies* (pp. 125–139). Southern Illinois University Press.

Inoue, A. B. (2015). *Antiracist writing assessment ecologies: Teaching and assessing writing for a socially just future*. Parlor Press.

Inoue, A. B. (2019). 2019 CCCC chair's address: How do we language so people stop killing each other, or what do we do about White language supremacy? *College Composition and Communication, 71*(2), 352–369.

Inoue, A., & Poe, M. (2012). Racial formations in two writing assessments: Revisiting White and Thomas' findings of the English placement test after 30 years. In N. Elliot & L. Perelman (Eds.), *Writing assessment in the 21st century: Essays in honor of Edward M. White*. (pp. 343–362). Hampton Press.

Isaacs, E. J. (2018). *Writing at the State U: Instruction and administration at 106 comprehensive universities*. Utah State University Press.

Israel, M., & Hay, I. (2006). *Research ethics for social scientists*. Sage.

Jiang, P., Sonne, C., Li, W., You, F., & You, S. (2024). Preventing the immense increase in the life-cycle energy and carbon footprints of LLM-powered intelligent chatbots. *Engineering, 40*, 202–210. https://doi.org/10.1016/j.eng.2024.04.002

Johnson, G. P. (2023). Don't act like you forgot: Approaching another literacy "crisis" by (re)considering what we know about teaching writing with and through technologies. *Composition Studies, 51*(1), 169–175.

Kane, M. T. (2006). Validation. *Educational Measurement, 4*(2), 17–64.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1–73.

Katinskaia, A. & Yangarber, R. (2021). Assessing grammatical correctness in language learning. *Proceedings of the 16th Workshop on Innovative Use of NLP for Building Educational Applications,* 135–146. https://aclanthology.org/2021.bea-1.15.pdf

Kaufer, D., Geisler, C., Vlachos, P., & Ishizaki, S. (2006). Mining textual knowledge for writing education and research: The DocuScope project. In L. van Waes, M. Leijten, & C. Neuwirth (Eds), *Writing and Digital Media* (pp. 115–129). Brill.

Kaufer, D., Ishizaki, S., Brown, D., Werner, N., & Wetzel, D. (2018). DocuScope. *Department of English, Dietrich College of Humanities and Social Sciences, Carnegie Mellon University*. Carnegie Mellon University. http://www.cmu.edu/dietrich/english/research-and-publications/docuscope.html

Kelly-Riley, D., & Elliot, N. (2014). The WPA Outcomes Statement, validation, and the pursuit of localism. *Assessing Writing, 21*, 89–103.

Kelly-Riley, D., & Elliot, N. (2020). *Improving outcomes: Disciplinary writing, local assessment, and the aim of fairness*. Modern Language Association.

Kim, M., & Chon, Y. V. (2025). The impact of self-revision, machine translation, and ChatGPT on L2 writing: Raters' assessments, linguistic complexity, and error correction. *Assessing Writing, 65*. https://doi.org/10.1016/j.asw.2025.100950

Klausman, J., Toth, C., Swyt, W., Griffiths, B., Sullivan, P., Warnke, A., Williams, A. L., Giordano, J. B., & Roberts, L. (2016). TYCA white paper on placement reform. *Teaching English in the Two-Year College, 44*(2), 135–157.

Klimt, B., & Yang, Y. (2004). The Enron corpus: A new dataset for email classification research. In J.-F. Boulicaut, F. Esposito, F. Giannotti, & D. Pedreschi (Eds.), *Machine learning: ECML 2004* (pp. 217–226). Springer. https://doi.org/10.1007/978-3-540-30115-8_22

Klopfer, E., Reich, J., Abelson, H., & Breazeal, C. (2024). Generative AI and K-12 education: An MIT perspective. *An MIT Exploration of Generative AI*. https://doi.org/10.21428/e4baedd9.81164b06

Koretz, D., Yu, C., Langi, M., & Braslow, D. (2014, August 26). Predicting freshman grade-point average from high-school test scores: Are there indications of score inflation? [Working paper]. *Education Accountability Project*. Harvard Graduate School of Education.

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X. H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). *Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing tasks*. arXiv. https://doi.org/10.48550/arXiv.2506.08872

Kryger, K., Mitchum, C., & Higgins, A. (2024). Localizing directed self-placement: UX stories and methods. *Journal of Writing Assessment, 17*(1). https://doi.org/10.5070/W4jwa.1582

Kryger, K., & Zimmerman, G. X. (2020). Neurodivergence and intersectionality in labor-based grading contracts. *Journal of Writing Assessment, 13*(2). https://escholarship.org/content/qt0934x4rm/qt0934x4rm.pdf

Kučera, H, & Francis, W. N. (1967). *Computational analysis of present-day American English*. Brown University Press.

Kuru, T. (2024). Lawfulness of the mass processing of publicly accessible online data to train large language models. *International Data Privacy Law, 14*(4), 326–351. https://doi.org/10.1093/idpl/ipae013

Lan, G., Li, Y., Yang, J., & He, X. (2025). Investigating a customized generative AI chatbot for automated essay scoring in a disciplinary writing task. *Assessing Writing, 66*, 100959.

Lang, S., Buell, D. A., & Elliot, N. (2023). Computer-assisted corpus analysis: An introduction to concepts, processes, and decisions. *IEEE Transactions on Professional Communication, 66*(1), 94-113. https://doi.org/10.1109/TPC.2022.3228026

Latour, B. (1994). On technical mediation—Philosophy, sociology, genealogy. *Common Knowledge, 3*(2), 29–64.

Latour, B. (1999). *Pandora's hope: Essays on the reality of science study*. Harvard University Press.

Lederman, J., & Warwick, N. (2018). The violence of assessment: Writing assessment, social (in)justice, and the role of validation. In M. Poe, A. B. Inoue, & N. Elliot (Eds.), *Writing assessment, social justice, and the advancement of opportunity* (pp. 229–255). The WAC Clearinghouse; University Press of Colorado. https://doi.org/10.37514/PER-B.2018.0155.2.07

Lee, D., & Palmer, E. (2025). Prompt engineering in higher education: A systematic review to help inform curricula. *International Journal of Educational Technology in Higher Education, 22*. https://doi.org/10.1186/s41239-025-00503-7

Lee, J. W., & Rüdiger, S. (2025). *Entangled Englishes* (1st ed.). Taylor & Francis.

Leech, G., & Smith, N. (2005). Extending the possibilities of corpus-based research on English in the twentieth century: A prequel to LOB and FLOB. *ICAME Journal, 29*, 83–98.

Leggett, J. M. (2023). Toward the co-construction of assessment: Equity, language ideology and culturally sustaining pedagogy at the community college. *Currents in Teaching & Learning, 15*(1), 83–97.

Lepagnol, P., Gerald, T., Ghannay, S., Servan, C., & Rosset, S. (2024). Small language models are good too: An empirical study of zero-shot classification. arXiv. https://doi.org/10.48550/arXiv.2404.11122

Lewin, K. (1946). Action research and minority problems. *Journal of Social Issues, 2*(4), 34–46. https://doi.org/10.1111/j.1540-4560.1946.tb02295.x

Lewis, J. E., Arista, N., Pechawis, A., & Kite, S. (2018). Making kin with the machines. *Journal of Design and Science.* https://doi.org/10.21428/bfafd97b

Li, H., Chen, C. H., Fan, K., Young-Johnson, C., Lim, S., & Feng, Y. (2025). *Agreement between large language models and human raters in essay scoring: a research synthesis.* arXiv. https://doi.org/10.48550/arXiv.2512.14561

Li, R. (2025). Critiquing ChatGPT compositions: Collaborative annotation as an approach to enhancing students' metalinguistic awareness of AI-generated writing. *Thresholds in Education, 48*(1), 7–24.

Li, Z., Shi, Y., Liu, Z., Yang, F., Payani, A., Liu, N., & Du, M. (2024). *Language Ranker: A metric for quantifying LLM performance across high and low-resource languages.* arXiv. https://doi.org/10.48550/arXiv.2404.11553

Lillis, T., & Scott, M. (2007). Defining academic literacies research: Issues of epistemology, ideology and strategy. *Journal of Applied Linguistics, 4*(1), 5–32.

Liu, Y., Lu, X., & Qi, H. (2025). Comparing GPT-based approaches in automated writing evaluation. *Assessing Writing, 66*, 100961. https://doi.org/10.1016/j.asw.2025.100961

Lloyd-Jones, R. (1977). Primary trait scoring. In C. Cooper & L. Odell (Eds.), *Evaluating writing: Describing, measuring, judging* (pp. 33–66). National Council of Teachers of English.

Maloy, J., Comeau-Kirschner, C., & Anderst, L. (2024). Pandemic realities in the midst of developmental education reform: Documenting the labors of basic writing faculty. *Journal of Basic Writing, 43*(1), 1–8.

Mansour, W., Albatarni, S., Eltanbouly, S., & Elsayed, T. (2024). *Can large language models automatically score proficiency of written essays?* arXiv. https://doi.org/10.48550/arXiv.2403.06149

Mark, M. M., Hopson, R. K., Caracelli, V. J., & Miller, R. L. (2023).The oral history of evaluation: The influence of Edmund Wyatt Gordon on evaluation. *American Evaluation Association, 23*(2), 175–189.

Martin, M. (Host). (2023, November 21). People disagree about the risks and benefits of artificial intelligence. In *Morning Edition*. National Public Radio. https://www.npr.org/2023/11/21/1214341015/people-disagree-about-the-risks-and-benefits-of-artificial-intelligence

McIntyre, M. (2024). Equitable writing classrooms and programs in the shadow of AI. *Computers and Composition, 75*, 102908. https://doi.org/10.1016/j.compcom.2024.102908

McIntyre, M., Fernandes, M., & Sano-Franchini, J. (2025). Practicing GAI refusal in writing studies. *Refusing Generative AI in Writing Studies*. https://refusal.blog/practicing-refusal/

Messer, K. (2024). "Easy and understanding": Everyone has power in this space. *Teaching English in the Two-Year College, 52*(1), 44–69.

Messer, K., Gallagher, J., & Hart, L. (2022). A path to equity, agency, and access: self-directed placement at the Community College of Baltimore County. In J. Nastal, M. Poe, & C. Toth (Eds.), *Writing Placement in Two-Year Colleges: The Pursuit of Equity in Postsecondary Education* (pp. 85–105). The WAC Clearinghouse; University Press of Colorado. https://doi.org/10.37514/PRA-B.2022.1565

Messeri, L. (2023). Teaching with ChatGPT: Critiquing generative artificial intelligence from the classroom. *Anthropology Now, 15*(1), 84–92. https://doi.org/10.1080/19428200.2023.2230098

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). Macmillan; American Council on Education.

Messina, C. M., Jones, C. E., & Poe, M. (2023). Prompting reflection: Using corpus linguistic methods in the local assessment of reflective writing. *Written Communication, 40*(2), 620–650. https://doi.org/10.1177/07410883221149425

Metz, C. (2025, September 5). Anthropic agrees to pay $1.5 billion to settle lawsuit with book authors. *The New York Times*. https://www.nytimes.com/2025/09/05/technology/anthropic-settlement-copyright-ai

Micus, C., Dekova, A., Böttcher, T., & Krcmar, H. (2024). Chat your data: Prompt engineering for standardized GenAI results. *AMCIS 2024 Proceedings, USA, 7*. https://aisel.aisnet.org/amcis2024/fow/fow/7

Miller, K. L., Wender, E., & Finer, B. S. (2017). Legislating first-year writing placement: Implications for Pennsylvania and across the country. *Journal of Writing Assessment, 10*(1). https://escholarship.org/uc/item/2vs4k4gg

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

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). *A brief introduction to evidence-centered design* [Research report]. Educational Testing Services.

Mislevy, R. J., Oliveri, M. E., Slomp, D., Crop Eared Wolf, A., & Elliot, N. (2025). An evidentiary-reasoning lens for socioculturally responsive assessment. In R. E. Bennett, L. Darling-Hammond, & A. Badrinarayan (Eds.), *Socioculturally responsible assessment: Theory, research, and practice* (pp. 199–241). Routledge. https://doi.org/10.4324/9781003435105-13

MIT Sloan Teaching & Learning Technologies. (2025, May 30). *Effective prompts for AI: The essentials.* https://mitsloanedtech.mit.edu/ai/basics/effective-prompts/

Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics, 2*(2), 100050. https://doi.org/10.1016/j.rmal.2023.100050

Mollick, E. (2024). *Co-intelligence: Living and working with AI*. Penguin.

Mansour, W. A., Albatarni, S., Eltanbouly, S., & Elsayed, T. (2024). Can large language models automatically score proficiency of written essays? In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)* (pp. 2777–2786).

Mokher, C., McCoy, K., Henning, H., Fluker, C., & Park-Gaghan, T. (2024). Placing students for success: Florida's statewide reform of alternative methods for college course placement. *Community College Journal of Research and Practice, 48*(3), 146–163.

Montenegro, E., & Jankowski, N.A. (2024). Equity and assessment: Moving towards culturally responsive assessment [Occasional paper #29]. National Institute of Learning Outcomes Assessment.

Moss, P. A. (1998). Testing the test of the test. A response to "Multiple inquiry in the validation of writing tests." *Assessing Writing, 5*(1), 111–122.

Moxley, J., Elliot, N., Eubanks, D., Vezzu, M., Elliot, S., & Allen, W. (2017). Writing analytics: Conceptualization of a multidisciplinary field. *Journal of Writing Analytics, 1*. https://wacclearinghouse.org/docs/jwa/vol1/intro.pdf

Naismith, B., Mulcaire, P., & Burstein, J. (2023). Automated evaluation of written discourse coherence using GPT-4. In E. Kochmar, J. Burstein, A. Horbach, R. Laarmann-Quante, A. Madnani, V. Tack, Z. Yaneva, Z. Yuan, & T. Zesch (Eds.), *Proceedings of the 18th workshop on innovative use of NLP for building educational applications (BEA 2023)* (pp. 394–403). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.bea-1.32

Nastal, J. (2019). Beyond tradition: Writing placement, fairness, and success at a two-year college. *Journal of Writing Assessment, 12*(1). https://escholarship.org/uc/item/4wg8w0ng

Nastal, J. (2024). A sociocognitive grading model for first-year writing classes. *Pedagogy, 24*(3), 405–426.

Nastal, J., Evans, J., & Gravely, J. (2022a). No reform is an island: Tracing the influences and consequences of evidence-based placement reform at a two-year Predominantly Black Institution. In J. Nastal, M. Poe, & C. Toth (Eds.), *Writing placement in two-year colleges: The pursuit of equity in postsecondary education* (pp. 35–58). The WAC Clearinghouse; University Press of Colorado. https://doi.org/10.37514/PRA-B.2022.1565.2.01

Nastal, J., & Messer, K. (2025). Afterword: Finding the right note in writing placement. *Journal of Writing Assessment, 18*(1). https://doi.org/10.5070/W4.jwa.47034

Nastal, J., Poe, M., & Toth, C. (2022b). *Writing placement in two-year colleges: The pursuit of equity in postsecondary education*. The WAC Clearinghouse; University Press of Colorado. https://doi.org/10.37514/PRA-B.2022.1565

National Center for Education Statistics. (2022). Digest of education statistics: 2022 Tables and figures. U.S. Department of Education. https://nces.ed.gov/programs/digest/2022menu_tables.asp

National Center for Education Statistics. (2024). Digest of education statistics: 2024 Tables and figures. U.S. Department of Education. https://nces.ed.gov/programs/digest/2024menu_tables.asp

National Center for Education Statistics. (2026). Student financial aid survey: Percentage of first-time, full-time undergraduate students awarded financial aid at Title IV degree-granting institutions in the United States, by type of financial aid and by level and control of institution: Academic year 2021–22. U.S. Department of Education. https://nces.ed.gov/ipeds/survey-components/12

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research*. U.S. Department of Health and Human Services. https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html

Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., & Mian, A. (2024). *A comprehensive overview of large language models*. arXiv. https://doi.org/10.48550/arXiv.2307.06435

Occhipinti, J. A., Hynes, W., Prodan, A., Eyre, H., Green, R., Burrow, S., ... & Heffernan, M. (2025). Generative AI may create a socioeconomic tipping point through labour displacement. *Scientific Reports, 15*(1), 26050.

Oliveri, M. E., Mislvey, R. J., & Elliot, N. (2020). After admissions: What comes next in higher education? In M. E. Oliveri & C. Wendler (Eds.), *Higher education admissions practices: An international perspective* (pp. 347–375). Cambridge.

Ostman, H. (2013). *Writing program administration and the community college*. Parlor Press.

Owusu-Ansah, A. L. (2023). Defining moments, definitive programs, and the continued erasure of missing people. *Composition Studies, 51*(1), 143–148.

Pack, A., Barrett, A., & Escalante, J. (2024). Large language models and automated essay scoring of English language learner writing: Insights into validity and reliability. *Computers and Education: Artificial Intelligence, 6*, 100234. https://doi.org/10.1016/j.caeai.2024.100234

Page, E. B. (1966). The imminence of ... grading essays by computer. *The Phi Delta Kappan, 47*(5), 238–243.

Palmer, M., & Xue, N. (2010). Linguistic annotation. In A. Clark, C. Fox, & S. Lappin (Eds.), *The handbook of computational linguistics and natural language processing* (pp. 238–270). Wiley-Blackwell. https://doi.org/10.1002/9781444324044.ch10

Palmquist, M. (2020). Learning analytics in writing instruction: Implications for writing across the curriculum. In L. E. Bartlett, S. L. Tarabochia, A. R. Olinger, & M. J. Marshall (Eds.), *Diverse approaches to teaching, learning, and writing across the curriculum: IWAC at 25* (pp. 55–72). The WAC Clearinghouse. https://doi.org/10.37514/PER-B.2020.0360.2.04

Pantelides, K. L., & Whittig, E. (2024). Placement is everyone's business: A love letter to our SSP coalition. *Journal of Writing Assessment, 17*(1). https://escholarship.org/uc/item/7v9830xh

Pantelides, K. L., & Whittig, E. (2025). Collaboratively building our SSP scholarship (because placement is still everyone's business). *Journal of Writing Assessment, 18*(1). https://escholarship.org/uc/item/1nm1x2zp

Papadimitriou, I., Lopez, K., & Jurafsky, D. (2023). Multilingual BERT has an accent: Evaluating English influences on fluency in multilingual models. In A. Vlachos & I. Augenstein (Eds.), *Findings of the Association for Computational Linguistics: EACL 2023* (pp. 1194–1200). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-eacl.89

Paris, B. S., Conti-Cook, C., Greene, D., Jones, K. M., Justie, B., Kirschenbaum, M., Kresge, L., May, E., Nguyen, A., Reynolds, R., Sargent, S., Weinberg, L., West, S. M., & Widder, D. G. (2025). Artificial Intelligence and Academic Professions. *Academe, 111*(3), 49–59.

Paris, D. (2012). Culturally sustaining pedagogy: A needed change in stance, terminology, and practice. *Educational Researcher, 41*(3), 93–97.

Paris, D., & Alim, H. S. (2014). What are we seeking to sustain through culturally sustaining pedagogy? A loving critique forward. *Harvard Educational Review, 84*(1), 85–100.

Paris, D., & Alim, H. S. (Eds.). (2017). *Culturally sustaining pedagogies: Teaching and learning for justice in a changing world*. Teachers College Press.

Parthasarathy, V. B., Zafar, A., Khan, A., & Shahid, A. (2024). *The ultimate guide to fine-tuning LLMs from basics to breakthroughs: An exhaustive review of technologies, research, best practices, applied research challenges and opportunities*. arXiv. https://doi.org/10.48550/arXiv.2408.13296

Perelman, L. (2012). Mass-market writing assessments as bullshit. In N. Elliot & L. Perelman (Eds.), *Writing assessment in the 21st century: Essays in honor of Edward M. White* (pp. 425–438). Hampton Press.

Perelman, L. (2014). When "the state of the art" is counting words. *Assessing Writing, 21*, 104–111.

Pete, J. S. (2025, April 22). New data center expected to use as much power as half of Indiana's households. *The Munster Times*. https://nwitimes.com/news/local/article_21fb8493-846f-488e-a5b8-c5e8198d9482.html

Pierce, J., Martin, D. G., & Murphy, J. T. (2011). Relational place-making: The networked policies of place. *Transactions of the Institute of British Geographers, 36*(1), 54–70.

Pletz, J. (2025). Privacy law drives AI data centers away: Amazon, Microsoft, others stay away from Chicago, citing liability under BIPA. *Crain's Chicago Business, 48*(38).

Poe, M., & Elliot, N. (2019). Evidence of fairness: Twenty-five years of research in *Assessing Writing*. *Assessing Writing, 42*, 100418. https://doi.org/10.1016/j.asw.2019.100418

Poe, M., Nastal, J., & Elliot, N. (2019). Reflection. An admitted student is a qualified student: A roadmap for writing placement in the two-year college. *Journal of Writing Assessment, 12*(1). https://escholarship.org/uc/item/31c793kf

Poe, M., Oliveri, M. E., & Elliot, N. (2023). The *Standards* will never be enough: A racial justice extension. *Applied Measurement in Education, 36*(3), 1–23.

Price, C., & Chao, S. (2023). Multispecies, more-than-human, nonhuman, other-than-human: Reimagining idioms of animacy in an age of planetary unmaking. *Exchanges: The Interdisciplinary Research Journal, 10*(2), 177–193.

Project Gutenberg. (2026, January 12). Project Gutenberg. https://www.gutenberg.org/

Ran, F. X., & Lee, H. (2024). Does corequisite remediation work for everyone? An exploration of heterogeneous effects and mechanisms [EdWorkingPaper No. 24-928]. Annenberg Institute at Brown University. https://edworkingpapers.com/sites/default/files/ai24-928.pdf

Ranade, N., Saravia, M., & Johri, A. (2024). Using rhetorical strategies to design prompts: A human-in-the-loop approach to make AI useful. *AI & Society, 40*, 711–732. https://doi.org/10.1007/s00146-024-01905-3

Randall, J. (2021). Color-neutral is not a thing: Redefining construct definition and representation through a justice-oriented critical antiracist lens. *Educational Measurement: Issues & Practice, 40*(4), 82–90.

Randall, J., Poe, M., Oliveri, M. E., & Slomp, D. (2023). Justice-oriented, antiracist validation: Continuing to disrupt White supremacy in assessment practices. *Educational Assessment, 29*(1), 1–20. https://doi.org/10.1080/10627197.2023.2285047

Randall, J., Poe, M., Slomp, D., & Oliveri, M. E. (2024). Our validity looks like justice. Does yours? *Language Testing, 41*(1), 203–219.

Reifman, J. B., Wittstock, S., Serviss, T., Pearsall, B., & Melzer, D. (2025). Constructivist

writing placement: Repositioning Agency for more equitable placement through collaborative writing placement practices. *College Composition & Communication, 76*(3), 423-451.

Reppen, R, Ide, N., & Suderman, K. (2005). *American national corpus (ANC) second release* (LDC2005T35) [Dataset]. Linguistic Data Consortium. https://doi.org/10.35111/251H-G440

Rose, M. (1983). Remedial writing courses: A critique and a proposal. *College English, 45*(2), 109–128.

Rowberry, S. (2025). The value of books in the age of generative AI training data. *Convergence: The International Journal of Research into New Media Technologies*, 31(6), 1935–1950. https://doi.org/10.1177/13548565251358020

Rudniy, A. (2018). De-identification of laboratory reports in STEM. *Journal of Writing Analytics, 2*, 176–202. https://wacclearinghouse.org/docs/jwa/vol2/rudniy.pdf

Rudniy, A. (2024). Artificial Intelligence for automated scoring and feedback in chemistry courses. *Journal of Writing Analytics, 7*. https://doi.org/10.37514/JWA-J.2024.7.1.02

Runge, A., Goodwin, S., Attali, Y., Poe, M., Mulcaire, P., Lo, K.-L., & LaFlair, G. T. (2025). A multi-stage interactive writing task for the assessment of English language writing proficiency. *Language Testing, 42*(4), 423–446. https://doi.org/10.1177/02655322251349908

Saini, A. K., Cope, B., Kalantzis, M., & Zapata, G. C. (2024, May 15). *The future of feedback: Integrating peer and generative AI reviews to support student work*. EdArXiv. https://doi.org/10.35542/osf.io/x3dct

Sano-Franchini, J., McIntyre, M., & Fernandes, M. (2025). *Refusing GAI in writing studies: A quickstart guide*. Refusing Generative AI in Writing Studies. https://refusal.blog/

Sanyal, S., Shwartz-Ziv, R., Dimakis, A. G., & Sanghavi, S. (2024). *When attention collapses: How degenerate layers in LLMs enable smaller, stronger models*. arXiv. https://doi.org/10.48550/arXiv.2404.08634

Sassi, K. (2018). Bending the arc of writing assessment toward social justice: Enacting culturally responsive professional development at Standing Rock. In M. Poe, A. B. Inoue, & N. Elliot (Eds.), *Writing assessment, social justice, and the advancement of opportunity* (pp. 317–352). The WAC Clearinghouse; University Press of Colorado.

Sauer, G. (2025). How technical communicators may work safely within the "fog" of emerging GenAI ethics [Keynote lecture, Tech Comm AI Symposium]. Techne Forge. https://techneforge.com/proceedings/keynote-how-technical-communicators-may-work-safely-within-the-fog-of-emerging-genai-ethics/

Schick, T., & Schütze, H. (2021). It's not just size that matters: Small language models are also few-shot learners. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2339–2352. https://doi.org/10.18653/v1/2021.naacl-main.185

Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistics perspective*. Routledge.

Schleppegrell, M. J. (2013). The role of metalanguage in supporting academic language development. *Language Learning, 63*(s1), 153–170. https://doi.org/10.1111/j.1467-9922.2012.00742.x

Scott-Clayton, J. (2012). *Do high-stakes placement exams predict college success?* [Working paper no. 41]. Community College Research Center.

Seamon, S. (2025, July 24). Texas is still in drought, and AI data centers are quietly guzzling up water. *The Austin Chronicle*. https://www.austinchronicle.com/news/2025-07-25/texas-is-still-in-drought-and-ai-data-centers-are-quietly-guzzling-up-water/

Shafik, W. (2024). Generative AI for social good and sustainable development. In K. Raza, N. Ahmad, & D. Singh (Eds.), *Generative AI: Current trends and applications* (Vol. 1177) (pp. 185–217). Springer.

Shermis, M. D., Burstein, J., & Bursky, S. (2013). Introduction to automated essay evaluation. In M. D. Shermis & J. Burstein (Eds.), *Handbook of automated essay evaluation: Current applications and new directions* (pp. 1–15). Routledge.

Shermis, M. D., & Wilson, J. (Eds.). (2024). *The Routledge international handbook of automated essay evaluation*. Taylor & Francis.

Sloane, M., Danks, D., & Moss, E. (2024). Tackling AI hyping. *AI and Ethics, 4*(3), 669–677. https://doi.org/10.1007/s43681-024-00481-y

Slomp, D. H. (2012). Challenges in assessing the development of writing ability: Theories, constructs and methods. *Assessing Writing, 17*(2), 81–91.

Slomp, D. (2016). An integrated design and appraisal framework for ethical writing assessment. *Journal of Writing Assessment, 9*(1). https://escholarship.org/uc/item/4bg9003k

Slomp, D., Oliveri, M. E., & Poe, M. (2025). Working toward culturally sustaining workplace assessment programs using an integrated design and appraisal framework. In C. M. Evans & C. S. Taylor (Eds.), *Culturally responsive assessment in classrooms and large-scale contexts* (pp. 177–201). Routledge.

Solano-Flores, G. (2008). Who is given tests in what language by whom, when, and where? The need for probabilistic views of language in the testing of English Language Learners. *Educational Researcher, 37*(4), 189–199.

Sperber, L., MacArthur, M., Minnillo, S., Stillman, N., & Whithaus, C. (2025). Peer and AI Review + Reflection (PAIRR): A human-centered approach to formative assessment. *Computers and Composition, 76*, 102921. https://doi.org/10.1016/j.compcom.2025.102921

Srinivasan, K. P. V., Gumpena, P., Yattapu, M., & Brahmbhatt, V. H. (2024). *Comparative analysis of different efficient fine tuning methods of large language models (LLMs) in low-resource setting*. arXiv. https://doi.org/10.48550/arXiv.2405.13181

Stahl, M., Biermann, L., Nehring, A., & Wachsmuth, H. (2024). Exploring LLM prompting strategies for joint essay scoring and feedback generation. In E. Kochmar, M. Bexte, J. Burstein, A. Horbach, R. Laarmann-Quante, A. Tack, V. Yaneva, & Z. Yuan (Eds.), *Proceedings of the 19th workshop on innovative use of NLP for building educational applications (BEA 2024)* (pp. 283–298). Association for Computational Linguistics. https://aclanthology.org/2024.bea-1.23/

Starch, D. (1913). Reliability and the distribution of grades. *Science, 38*(983), 630–636.

Sternglass, M. S. (1997). *Time to Know Them: A Longitudinal Study of Writing and Learning at the College Level.* Lawrence Erlbaum Associates.

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. In A. Korhonen, D. Traum, & L. Màrquez (Eds.), *Proceedings of the 57th annual meeting of the Association for Computational Linguistics* (pp. 3645–3650). Association for Computational Linguistics. https://doi.org/10.18653/v1/P19-1355

Suh, E. K., Giordano, J. B., Griffiths, B., Hassel, H., & Klausman, J. (2020). *The profession of teaching English in the two-year college: Findings from the 2019 TYCA workload survey.* https://ncte.org/wp-content/uploads/2020/12/TYCA_Final_Report_December_2020.pdf

Suh, E. K., Williams, J. L., & Owens, S. (2021). *Raciolinguistic justice in college literacy and learning: A call for reflexive practice* [White paper]. College Reading and Learning Association. https://cdn.ymaws.com/crla.net/resource/resmgr/white_papers/crla_2021_whitepaper_racioli.pdf

Sullivan, P. (2015). The Two-Year College Teacher-Scholar-Activist. *Teaching English in the Two-Year College, 42*(4), 327–350.

Sullivan, P., & Adams, P. (2023). National report on developmental education: Corequisite reform is working. *Teaching English in the Two-Year College, 50*(3), 224–254.

Sun, Y., Sheng, D., Zhou, Z., & Wu, Y. (2024). AI hallucination: Towards a comprehensive classification of distorted information in artificial intelligence-generated content. *Humanities and Social Sciences Communications, 11*(1), 1–14.

Sweeney, M., & Colombini, C. (2024). (Re)placing personalis: A study of placement reform and self-construction in mission-driven contexts. *Journal of Writing Assessment, 17*(1). https://doi.org/10.5070/W4jwa.1578

Tayles, M. (2021). Trauma-informed writing pedagogy: Ways to support student writers affected by trauma and traumatic stress. *Teaching English in the Two-Year College, 48*(3), 295–313.

TED. (2023, November 6). *AI is dangerous, but not for the reasons you think. | Sasha Luccioni* [Video]. YouTube. https://www.youtube.com/watch?v=eXdVDhOGqoE

Tinkle, T., Godfrey, J., Hammond, J. W., Moos, A., & College, C. (2024). Self-characterization in the self-placement assessment ecology. *Journal of Writing Assessment, 17*(1). https://doi.org/10.5070/W4jwa.1625

Tinkle, T., Godfrey, J., Menon, A. R., Moos, A., Romaine, L., & Sprouse, M. (2022). (In)Equities in directed self-placement. *Assessing Writing, 54*, 100671. https://doi.org/10.1016/j.asw.2022.100671

Tomlinson, B., Black, R. W., Patterson, D. J., & Torrance, A. W. (2024). The carbon emissions of writing and illustrating are lower for AI than for humans. *Scientific Reports, 14*(1), 3732.

Toth, C. (2018). Directed self-placement at "democracy's open door": Writing placement and social justice in community colleges. In M. Poe, A. B. Inoue, & N. Elliot (Eds.), *Writing assessment, social justice, and the advancement of opportunity* (pp. 139–172). The WAC Clearinghouse; University Press of Colorado.

Toth, C. (2019). Directed self-placement at two-year colleges: A kairotic moment. *Journal of Writing Assessment, 12*(1).

Toth, C., Andrus, J., Clawson, N., Fochs, A., Fraser, P., Onwuzuroha, N., & Aguilar, S. R. (2024). Informing self-placement: A polyvocal narrative case study. *Journal of Writing Assessment, 17*(1). https://doi.org/10.5070/W4jwa.1637

Toth, C., & Aull, L. (2014). Directed self-placement questionnaire design: Practices, problems, possibilities. *Assessing Writing*, *20*, 1–18. https://doi.org/10.1016/j.asw.2013.11.006

Toth, C., Nastal, J., Hassel, H., & Giordano, J. B. (2019). Introduction: Writing assessment, placement, and the two-year college. *Journal of Writing Assessment, 12*(1). https://escholarship.org/uc/item/8393560s

Toulmin, S. E. (2003). *The uses of argument* (2nd ed.). Cambridge University Press.

Tran, K.-T., O'Sullivan, B., & Nguyen, H. D. (2025). *IRLBench: A multi-modal, culturally grounded, parallel Irish–English benchmark for open-ended LLM reasoning evaluation*. arXiv. https://doi.org/10.48550/arXiv.2505.13498

Tremain, L., Anderson, J., Eschenbach, B., Amann, N., & Marsden, K. (2023). What counts as literacy in the polytechnic Hispanic Serving Institution? Culturally sustaining frameworks for writing assignments, assessment, and language use. *Humboldt Journal of Social Relations, 45*, 119–135.

Valdivia, A. (2024). The supply chain capitalism of AI: A call to (re)think algorithmic harms and resistance through environmental lens. *Information, Communication & Society, 28*(12), 2118–2134.

Vee, A., Laquintano, T., & Schnitzler, C. (2023). *TextGenEd: Teaching with text generation technologies*. The WAC Clearinghouse. https://doi.org/10.37514/TWR-J.2023.1.1.02

Vieira, K., Heap, L., Descourtis, S., Isaac, J., Senanayake, S., Swift, B., & West, G. B. (2019). Literacy is a sociohistoric phenomenon with the potential to liberate and oppress. In L. Adler-Kassner & E. Wardle (Eds.), *(Re)Considering what we know: Learning thresholds in writing, composition, rhetoric, and literacy* (pp. 36–55). Utah State University Press.

Villanueva, V. (1997). Maybe a colony: And still another critique of the comp community. *Journal of Advanced Composition, 17*(2), 183–190.

von Davier, A. A., & Burstein, J. (2024). AI in the assessment ecosystem: A human–centered AI perspective. In P. Ilic, I. Casebourne, & R. Wegerif (Eds.), *Artificial intelligence in education: The intersection of technology and pedagogy* (pp. 93–109). Springer.

Wanderley, L. F. & Epps, C. D. (2021). Identifying negative language transfer in learner errors using POS information. *Proceedings of the 16th Workshop on Innovative Use of NLP for Building Educational Applications,* 64–74. https://aclanthology.org/2021.bea-1.7.pdf

Wang, Y., Han, Y., Han, K., & Shen, J. (2025). Does DeepSeek curb the surge of energy consumption in data centers? *The Innovation, 6*(9), 100944.

Warnke, A., & Higgins, K. (2018). A critical time for reform: Empowering interventions in a precarious landscape. *Teaching English in the Two-Year College, 45*(4), 361–384.

Wataoka, K., Takahashi, T., & Ri, R. (2025). *Self-preference bias in LLM-as-a-judge*. arXiv. https://doi.org/10.48550/arXiv.2410.21819

Wei, J. W., Wei, J., Tay, Y., Tran, D., Webson, A., Lu, Y., Chen, X., Liu, H., Huang, D. Zhou, D., & Ma, T. (2023). Larger language models do in-context learning differently. arXiv. https://doi.org/10.48550/arXiv.2303.03846

West-Puckett, S., Caswell, N. I., & Banks, W. P. (2023). *Failing sideways: Queer possibilities for writing assessment*. University Press of Colorado.

Wetzel, D., Brown, D., Werner, N., Ishizaki, S., & Kaufer, D. (2021). Computer-assisted rhetorical analysis: Instructional design and formative assessment using DocuScope. *The Journal of Writing Analytics, 5*.

White, E. M., Elliot, N., & Peckham, I. (2015). *Very like a whale: The assessment of writing programs*. Utah State University Press.

Willard-Traub, M., Decker, E., Reed, R., & Johnston, J. (1999). The development of large-scale portfolio placement assessment at the University of Michigan: 1992–1998. *Assessing Writing, 6*(1), 41–84.

Williams, D. M. (2018). The unheard voices of dissatisfied clients: Listening to community partners as feminist praxis. In K. L. Blair & L. Nickoson (Eds.), *Composing feminist interventions: Activism, engagement, praxis* (pp. 409–426). https://doi.org/10.37514/PER-B.2018.0056.2.21

Williamson, M. (1994). The worship of efficiency: Untangling theoretical and practical considerations in writing assessment. *Assessing Writing, 1*(2), 147–173.

Willingham, W. W. (1974). *College placement and exemption*. New York: College Entrance Examination Board.

Won Chung, H., Hou, L., Longpre, S., Zoph, B., Tay, Y., Fedus, W., Li, Y., Wang, X., Dehghani, M., Brahma, S., Webson, A., Gu, S. S., Dai, Z., Suzgun, M., Chen, X., Chowdhery, A., Castro-Ros, A., Pellat, M., Robinson, K., … Wei, J. (2022). *Scaling instruction-finetuned language models*. arXiv. https://doi.org/10.48550/arXiv.2210.11416

Wood, S. (2024, May 13). Some colleges are requiring test scores again: What it means for applicants. *U.S. News & World Report*. https://www.usnews.com/education/best-colleges/applying/articles/some-colleges-are-requiring-test-scores-again-what-it-means-for-applicants

Wood, T. (2017). Cripping time in the college composition classroom. *College Composition & Communication, 69*(2), 260–286.

World Medical Association. (2024). *WMA declaration of Helsinki: Ethical principles for medical research involving human participants*. https://www.wma.net/policies-post/wma-declaration-of-helsinki/

Yamashita, T. (2025). Exploring potential biases in GPT 4o s ratings of English language learners. *Language Testing, 42*(3), 1–15. https://doi.org/10.1177/02655322251329435

Yancey, K. B. (1999). Looking back as we look forward: Historicizing writing assessment. *College Composition & Communication, 50*(3), 483–503.

Yancey, K. P., Laflair, G., Verardi, A., & Burstein, J. (2023). Rating short L2 essays on the CEFR scale with GPT-4. *Proceedings of the 18th workshop on innovative use of NLP for building educational applications (BEA 2023)* (pp. 576–584). https://doi.org/10.18653/v1/2023.bea-1.49

Yoshida, L. (2024, July). The impact of example selection in few-shot prompting on automated essay scoring using GPT models. In *International Conference on Artificial Intelligence in Education* (pp. 61–73). Cham: Springer Nature Switzerland.

Young, I. M. (2011). *Responsibility for justice*. Oxford University Press.

Young, V. A., & Robinson, M. B. (2018). *The Routledge reader of African American rhetoric: The longue duree of Black voices*. Routledge.

Zanders, C. J., & Wilson, E. (2019). Holistic, local, and process-oriented: What makes the University Utah's writing placement exam work. *Assessing Writing, 41*, 84–87.

Zanettin, F. (2014). *Translation-driven corpora: Corpus resources for descriptive and applied translation studies*. Routledge. https://doi.org/10.4324/9781315759661

Zheng, L., Chiang, W.-L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., Lin, Z., Li, Z., Li, D., Xing, E. P., Zhang, H., Gonzalez, J. E., & Stoica, I. (2023). *Judging LLM-as-a-judge with MT-Bench and Chatbot Arena*. arXiv. https://doi.org/10.48550/arXiv.2306.05685

Zhou, Z., Li, L., Chen, X., & Li, A. (2024). *Mini-giants: "Small" language models and open source win-win*. arXiv. https://doi.org/10.48550/arXiv.2307.08189

Zhu, Y., Kiros, R., Zemel, R., Salakhutdinov, R., Urtasun, R., Torralba, A., & Fidler, S. (2015). Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. arXiv. https://doi.org/10.48550/arXiv.1506.06724

Appendix A: Standard Writing Assessment Evaluation Methods for Reliability

1. Measure agreement and reliability of humans and GAI instrument by:

    a) Establishing a Gaussian distribution, or a bell curve. To do so would require both the humans and the GAI instrument to read and evaluate a sufficiently large sample of student texts (e.g., at least 30) and then plot their scores or ratings on an x-axis and the frequency of those scores on a y-axis. Plotting would visually represent how both sets of scores are distributed. It is likely scores of a random sample of writing would cluster around the middle.

    OR

    b) Reading enough student texts to establish targeted confidence levels of 80%, 90%, or 95%, with a 5% margin of error for both human and GAI scoring. To identify the number of samples required to assure the greatest confidence that results are representative of the entire set, practitioners can use a simple sample size calculator (e.g., from calculator.net). While confidence levels may not be as high in other scoring scenarios, the nature of writing placement, particularly if mediated by a GAI instrument, has such significant consequences for students that systems must be held to high degrees of accuracy.

2. Select sampling to represent the entire population and decide if there is any overrepresentation of gender, race and ethnicity, and/or socio-economic-status groups for any trait score by

    - Using random sampling for the entire population (a subset of the whole group that is selected by chance)

    AND

    - Using purposive sampling for specific groups (a subset of the whole group that represents a specific trait to be examined for fairness) (NOTE: If these can be identified in advance (§5.2.3), then random sampling can be used independently.)

3. Use counts and percentages to report scoring differences of GAI instrument-to-human and/or GAI platform-to-GAI platform and determine discrepancy range for:

    - Exact scores (scores that match exactly)

    AND

    - Adjacent scores (scores have a 1-point difference up or down)

    AND

    - Beyond-adjacent scores (scores that have a 2-or-more-point difference up or down)

4. Report adjudication scoring with:

    - Quadratic Weighted Kappa for adjudicated scores (scores that fall out of the determined discrepancy range and were resolved by a third rater)

    AND

    - Pearson's Product Moment Correlation for nonadjudicated scores (scores that fell outside of the determined discrepancy range and were not resolved by a third rater) (NOTE: If the datasets are too small, Pearson's will not work correctly; instead, Spearman's rho should be used.)

5. Determine acceptability level for agreement and reliability, noting that:

    - The standard acceptability level for reliability is 0.7 or higher

    AND

    - There may be reason for local placement processes to determine a different acceptable level for reliability

6. Compare gathered evidence for types of validity including:

    - Concurrent validity by comparing results of human-to-human scores to results of human-to-GAI instrument scores

    AND

    - Predictive validity by comparing GAI instrument scores-to-GAI instrument scores

    AND

    - All of these for the selected subgroup (§5.2.3)

Appendix B: Evaluating GAI-Assisted Scoring for Fairness

Sampling Plan

For both descriptive and inferential methods, beginning with these two sampling selection steps will help set up the work:

- Start with random sampling for the entire population.
- Use purposive sampling, dividing the population into subcategories based on IPEDS and institution-specific groups.

Descriptive Statistics

To use descriptive statistics for fairness studies of GAI-writing placement processes that involve GAI-assisted scoring, researchers can choose to use counts and percentages to identify:

- **Exact scoring differences** (How many students did not get the same score from the GAI instrument that they got from the human scorer(s)?)
- **Adjacent scoring differences** (How many students got a GAI-generated score that was one above or one below the human scorer(s)?)
- **Beyond-adjacent scoring differences** (How many students got a GAI-generated score that was two or more points above or below the human scorer(s)?)

These types of measurements can give valuable information to use to interpret scoring, the effects of cut-scores, and the ways in which different populations are influenced by reliability issues. To evaluate for reliability issues regarding fairness, apply the methods described above to the purposive (selective) sampling group(s).

Inferential Statistics

There are several types of inferential statistical methods that can be used for analyzing correlation between individual demographics and test scores (e.g., ANOVA, structural equation modeling, t-tests). Regression modeling may be particularly useful.

The type of score indicates the type of regression model used. **Ordinal scores** are scores that go in a specific ranked order, but the amount between each score is not necessarily quantifiably equal: for example, a Likert Scale, where writing might be scored as good or very good, with a rubric describing qualitative differences between those two score points. **Continuous scores** are scores that have equal, measurable amounts between them, like 1–100, where scores may be between integers like 88.25 or 93.5.

For both multiple ordinal regression (used for ordinal scores) and multiple linear regression (used for continuous scores) models, using student demographic traits as independent variables to

identify their relationship to the dependent variable (the student text) can give insights into meaningful relationships between groups of students and scoring practices.

These general explanations can serve as a starting point for regression analysis:

Prepare the data. Be sure the data is student-level (each essay and score is directly correlated to the demographics of the student). Create a spreadsheet with each student's information along the same line. Format the data to the program you'll be using to run the regression.

Check the assumptions associated with the regression model. (i.e., Is my independent variable completely independent from the dependent variable (no correlation)? Does the size of error remain fairly constant (homoscedasticity)? Is the best fit line straight (linearity)? Is the distribution normal throughout (mean, median, mode)?)

Run the regression. This analysis can be done by mathematical formula by hand, but a number of programs are capable of running these types of regressions (e.g., Excel, STATA, SPSS, R). When using a program, the external calculations that give further evidence for interpreting the results are also done automatically (r, r^2, p-value).

Interpret the results. Look for p-values greater than 0.05 to indicate statistical significance, but keep in mind that statistical significance alone does not indicate the importance of a result (ASA, 2016). Keep in mind that a low r^2 may still be important when considering fairness. Return to the assumptions for more guidance considerations for evaluating the results.