Imbalanced Researcher-Practitioner Relationships: Biasing the Data

By Cheryl Burk, Lori Dees, and Laura Kalbaugh

Many of the CCRC's research protocol and assessment method choices were not mutually agreed upon selections within the boundaries of the researcher-practitioner partnership.

Cheryl Burk
Professor of Developmental Reading and English/
    Student Success
caburk@waketech.edu

Lori Dees
Associate Professor of English

Laura Kalbaugh
Dean of Academic Success and Transition Resources

Wake Tech Community College
6600 Louisburg Road
Raleigh, NC 27616

ABSTRACT: As part of the Community College Research Center's (CCRC) Analysis of Statewide Developmental Education Reform Learning Assessment Study, Wake Technical Community College's partner team believes the recommendation and implications in the CCRC's Developmental Reading and English Assessment in a Researcher-Practitioner Partnership. Working Paper No. 85 (Perin, Raufman, & Kalamkarian, 2015) are biased as a direct result of an imbalanced researcher-practitioner partnership. In this first-hand account, we provide insight into how valuable a mutually respectful researcher-practitioner partnership is to the design of research methodologies and subsequently the authenticity of a study's data.

The purpose of this paper is to share three practitioners' first-hand experience as members of Wake Technical Community College's (WTCC) partner team in the researcher-practitioner partnership formed with the Community College Research Center (CCRC). Research end-users who review the CCRC's *Developmental Reading and English Assessment in a Researcher-Practitioner Partnership Working Paper No. 85* (Perin, Raufman, & Kalamkarian, 2015) should be aware of research methodologies used to generate the data before unilaterally accepting the CCRC's final conclusion that Developmental Reading and English (DRE) courses taught in the community college setting are not adequately and effectively preparing developmental students for college-level reading and writing demands. The sharing of our experience offers a unique vantage from which to view the CCRC's findings as published in Working Paper No. 85: This article will focus on the necessity of a mutually respectful relationship between a researcher and a practitioner in order to attain the best available empirical evidence from a study.

In a formal report published December 2015, the authors (Perin et al., 2015) of the CCRC's Working Paper No. 85 asserted that the DRE student volunteers studied at two community colleges in North Carolina (NC) "still had quite a way to go" (p. 53) in order to be prepared for college-level reading and writing demands. As two DRE instructors and a Dean at WTCC representing the practitioners at one of the community colleges in the aforementioned study, we assert the study we participated in with the CCRC was not designed nor implemented in such a manner that accurately measured a DRE student's level of readiness for the rigors of college-level reading and writing coursework. We contend that the CCRC's study results—generated, in part, from two project-designed text-based student reading/writing assessments administered during the study—are potentially invalid for informing any future decisions relating to DRE instruction in the community college setting. Many of the CCRC's research protocol and assessment method choices were not mutually agreed upon selections within the boundaries of the researcher-practitioner partnership.

Commencing Fall semester 2013 and continuing through Fall semester 2014, WTCC was in partnership with the CCRC to develop and sponsor an outcome evaluation study aimed at exploring "the nature of student learning in developmental education courses." The study represented an attempt to ascertain not just the level of student learning but also the process of student learning (as stated in the information for lead faculty partners document, Analysis of Statewide Developmental Education Reform Learning Assessment Study Overview). In 2015, the CCRC formally published Working Paper No. 85, listing eight key findings (pp. 53-54) resulting in three summative implications (p. 62): the DRE students, and thus the DRE program, were in "need."

The sharing of our experience as the practitioners in this study provides empirical insight into how an imbalance in the ownership over final research protocol and assessment method decisions in a researcher-practitioner relationship can impact finding. Such an imbalance can render a study's results unsatisfactory for informing other practitioners, educational practices, and policy-makers at local, state, and national levels. All stakeholders depend on valid research studies to make decisions for students.

To be clear, as we discuss the shortcomings and flaws of the CCRC's research protocol and assessment method choices and our role as practitioners in partnership with the CCRC, our intent is not to imply that the CCRC team intentionally sought to manipulate student outcomes or numerical assessment values in their study to produce certain results. We offer this commentary to inspire and

challenge future practitioners in research partnership studies to be more cognizant of their important role. Practitioner partners should intentionally consider each decision made concerning research methodologies in order to generate authentic research evidence for informing educational practices and educational policy-makers working in the best interest of students.

Evidence-Based Practices in Education

An Evidence-Based Practice (EBP) is the goal of many public educational institutions in the United States. EBP, a term first originating in the medical field, is a professional practice that utilizes the "the integration of best research evidence with clinical experience and client values" (Yates, 2013, p. 42). Since its conception, EBP has "gained prominence in [other] helping professions" (Nevo & Slonim-Nevo, 2011, p. 1177). Spencer, Detrich, and Slocum (2012) traced the increase in the role of research evidence in the educational profession to the mandates listed in the No Child Left Behind (NCLB) legislation in 2001 and the Individuals with Disabilities Education Improvement Act (IDEIA) in 2004 (p. 128). Grover J. Whitehurst (2002), in a white paper presented at the Student Achievement and School Accountability Conference, is credited with adapting the original definition of EBP from the medical profession for use in the educational profession. According to Whitehurst, an evidence-based educational practice is "the integration of professional wisdom with the best available empirical evidence in making decisions about how to deliver instruction" (slide 3). Community colleges, like WTCC, use empirical evidence to make curriculum decisions in the best interest of their students.

Research evidence to inform educational practice is extremely desirable and valuable in the education profession because research provides professional educators with the knowledge to create the highest quality instructional practices possible for their students, a responsibility for which all professional educators should be accountable. Social researchers believe the following: "Practitioners can, and do, value research that helps them gain new insights into problems and [gaining new insights] facilitates the search for new kinds of solutions to persistent problems" (Weiss & Bucuvalas as cited in Penuel, Allen, Coburn, & Farrell, 2015, p. 186). Additionally, research-derived evidence can inform policy-makers. Educational researcher Palmer (1999) argues the value of empirical evidence "equips policy makers with insights that will encourage them to adopt responsible approaches" (pp. 379-380). In fact, since 1996, the CCRC's main research objective, as stated on their website, has been to "contribute to the development of practice and *policy* [empasis added] . . ." (CCRC, 2017, para. 2). However, as a result of the increased demand for research evidence to inform educational practices, particularly since NCLB and IDEIA, Spencer et al. (2012) warn that "there is an ethical responsibility to base [educational] decisions on *the best available evidence* [empasis added]" (p. 132). According to EBP analysts Nevo and Slonim-Nevo (2011), "the idea is that the activities of practitioners should, as a matter of both professional and ethical responsibility, be closely guided by *the best empirical findings* [emphasis added] in their fields" (p.1177). We concur and add that intentionally producing the most authentic research evidence possible is not only a researcher-practitioner partnership's professional and ethical responsibility but a professional and ethical directive considering the amount of resources – time, money, and expertise – provided and expended executing a research study.

Best Available Empirical Evidence

To produce the best available empirical evidence within a researcher-practitioner partnership study,

> *An evidence-based educational practice is "the integration of professional wisdom with the best available empirical evidence in making decisions about how to deliver instruction."*

a mutually respectful relationship must exist. Educational researchers Coburn, Penuel, and Geil (2013) point out, "Mutualism is important because it helps ensure that different perspectives–practitioners' and researchers'–contribute to defining the focus of the work that research-practice partnerships do" (p. 3). Social researcher Epstein (2011) asserts, "the route to successful practice-research integration [is] via an appreciative and respectful collaboration" (p. 285). In addition, a mutually respectful researcher-practitioner relationship requires trust: "the most critical element of a successful collaboration" (Sullivan, Khondkaryan, & Fisher, 2013, p. 3). Researchers Crooke and Oslwang (2015) also report, "the team [i.e., researcher-practitioner partnership] needs to constantly strive for a rigorous process that results in trustworthy evidence" (p. S1874). Ultimately, Sullivan et al. (2013) conclude, partnerships without trust and without shared expectations and goals diminish the "potential [for a study] to yield profound benefits" (p. 7). Clearly, researcher-practitioner partnerships without mutualism and respect, trust, and rigor significantly increase the opportunity for research bias. Based on this premise, we contend that the inauthentic research protocol and assessment method choices made both before and during the study we participated in with the CCRC—either against our recommendations or without our knowledge—has weakened the strength of the research design and created bias. Further, it thereby, has decreased the potential for the study to yield the best empirical evidence for informing the future of college DRE practices.

Inauthentic Research Protocol Choices

Setting purpose. In our preliminary team discussions regarding the purpose of the study, partners focused on the intent to evaluate DRE student learning processes for reading and writing (i.e., how students comprehend printed texts and how students express that comprehension in writing) plus students' skill levels (i.e., what students comprehend from printed texts and what writing skills students use to articulate what they comprehend). We did not understand the primary purpose was to ascertain the students' college readiness and, by extension, the quality of the DRE curriculum/instruction. However, once we read the CCRC's published working paper, we realized that the CCRC's primary purpose for assessing DRE students' skill level on the academic reading and writing tasks developed for the study was to judge the students' "college readiness." If indeed the CCRC's primary purpose was to meaningfully quantify the students' college-readiness, the CCRC team never engaged us in a conversation that led directly to a mutually acceptable characterization of student college readiness. Granted, one collaborative discussion led to a discovery of the educational learning tasks that accurately measures reading comprehension and writing skill, but the discussion was not about specific expectations from those learning tasks that would reveal an agreed upon measure of college readiness. Even though the term "college readiness" is often used by educational professionals in a manner that conveys a universal acknowledgment of its definition, a cursory look at the extant literature exploring the concept of college-readiness reveals a large variance in its accepted meaning and understanding, both contextually and measurably. In fact, in their final working paper (Perin et al., 2015, p. 6), the CCRC authors even acknowledge that currently there is no commonly agreed upon quantitative measure of college readiness.

The fact that this research was funded by the Bill & Melinda Gates Foundation, under the project entitled "Analysis of Statewide Developmental Education Reform: Learning Assessment Study," could possibly explain why the study's data was primarily used to evaluate student college-readiness and, by extension, DRE program quality. Coburn et al. (2013) contend, "Those who receive the grant funding often have

a greater voice in decisions related to the focus of the work" (p. 16). Penuel et al. (2015), citing their own past work from 2007, further posit that funding can "undermine efforts to maintain mutualism": when funding in the form of grants goes to the researcher, the funding stream "gives the researcher both the authority and ultimate responsibility for guiding projects, even when projects aim to involve practitioners as collaborators in the work" (p. 193). Also, Coburn et al. (2013) reported that differences in perceived partnership status can negatively affect the foundation of trust that is so essential to a mutually respectful relationship. In fact, Coburn et al. point out: "Practitioners, especially teachers, can fall silent in rooms filled with researchers" (p. 6). In retrospect, our reasoning for not engaging the CCRC in even more discussions about research protocol and assessment method choices was our belief that the study was designed to explore DRE student learning processes and levels, not specifically college readiness. Believing that this study would provide DRE instructors with greater insight about their students' thought processes during reading and writing tasks and how those thoughts correlated with their performance on those tasks certainly did not match the CCRC's "major goal" (Perin et al., 2015, p. 56): to measure DRE student college readiness.

Measuring college readiness. To measure the DRE students' college readiness, the CCRC team created a multi-pronged assessment tool, including two standardized tests and two project-designed reading/writing assessments. The standardized tests were used to establish general reading and writing levels; the project-designed assessments were used to determine reading/writing competency. A multipronged approach to establishing a student's reading/writing college-readiness skills seems more prudent than the reliance on one high-stakes assessment. However, researchers Hathcoat and Penn (2012), after conducting a study designed to determine the validity of using project-designed writing assessments instead of standardized testing to measure student writing achievement, caution that since "evidence implies that a single observation of writing performance fails to generalize to performance across other [writing] tasks, a similar issue may exist with standardized test scores since these scores may also fail to generalize to scores observed on similar tasks outside a controlled testing environment" (p. 24). This research suggests that a student's performance can be highly variable across writing tasks and even change across similar writing tasks. This variability must be noted as a potential source of measurement error and the resulting data viewed considering this bias, particularly when attempting to correlate data between standardized and project-designed assessments.

Authentic research methodology is heavily dependent on an understanding of how and where measurement error can occur in the design of the research. Researchers Schmidt and Hunter (1996) contend that any "progress in theory development and cumulative knowledge is impossible without careful attention to problems created by measurement error" in research (p. 222). Further, educational researcher Slavin (2008) is adamant that, if educators and policy-makers are going to make wise decisions about programs for students, they must not ignore any form of study bias simply because other valuable study design features exist. Ignoring bias, Slavin argues, "does a disservice to the readers and to the cause of evidence-based reform" (p. 48). Research end-users should be aware that no significant measurement limitations related to the standardized or project-designed study assessments were noted in Working Paper #85.

Also, contributing to potential measurement error is the idea that the reading and writing

> *Though academic competency may deem a student "college-eligible," academic competency alone may not deem a student "college ready."*

assessment tools chosen and created by the CCRC for the study were based on the premise that academic skills "are a central indicator" of college-readiness (Perin et al., 2015, p.1). Educational researcher Conley (2008) believes academic skills and knowledge (e.g., engaging in textual material critically and creating organized and supported written work) is but *one part* [emphasis added] of a more comprehensive conception of college readiness; and, even though academic competency may deem a student "college-eligible," academic competency alone may not deem a student "college ready" (pp. 25-26). Moreover, university writing instructors Hassell and Giordano (2015) caution that "academic and rhetorical readiness does not always translate into [future] academic success because lack of success in writing courses can sometimes be tied to other non-academic factors" (p. 67). A collaborative discussion in the planning stages of this research study focusing on authentic methods of measuring college readiness might have highlighted the CCRC's limited premise and possibly resulted in different project-designed assessments.

The most troubling source of measurement error not noted in the CCRC's working paper is the

fact that DRE097-level student data were included in the data set when reporting DRE students not being adequately prepared for college-level reading and writing demands. Students in their first-semester, DRE097-level classes are not expected to demonstrate college-ready reading/writing skills and academic success behaviors until the end of their subsequent DRE098 course. DRE097-level students were solicited for the study at the CCRC's request to help raise the total number of study participants (ultimately 211 students participated in this study) because larger numbers, according to researchers, produce more valid research results. The 2003 U.S. Department of Education, Institute for Education Services, National Center for Education Evaluation and Regional Assistance's user-friendly guide—*Identifying and Implementing Educational Practices Supported by Rigorous Evidence*—suggests at least 300 students are needed for valid [educational] data results (p. 8). We willingly agreed to solicit DRE097-level students as an acceptable way to bolster student numbers because, again, we believed the purpose of the study was to ascertain student learning processes and levels, not specifically college readiness. We now argue that distortions to the overall meaningfulness of the CCRC's study data were created by including DRE097-level students who are not expected to demonstrate college-readiness until the end of the DRE098 course.

We contend that because of the current debate associated with qualifying or quantifying a student's readiness for college, reporting college readiness as a numerical value based on several study assessments does not definitively determine whether any student, much less a developmental student, can negotiate college-level reading and writing tasks successfully. The risk of publically stating implications and recommendations from student assessment data derived from two standardized tests and two project-designed assessments, without a discussion of the bias introduced by measurement error, has the potential to negatively inform a research end-user's beliefs about developmental students and the programs that support these students. Clearly, establishing a mutually respectful partnership for the purpose of designing a research study to determine the quality of a program with as little measurement error as is possible is critical for effectuating and reporting the best empirical evidence

Premature timing of study. Certainly noteworthy is the fact that the CCRC conducted their study during the first semester that WTCC started offering DRE courses. We questioned the CCRC about this fact and were told that, as long as students were enrolled in DRE courses at the time of the assessment, the length of time WTCC had been offering DRE courses was inconsequential to the study. Based on how the CCRC chose to evaluate the DRE program, the limited time the department had been managing a brand new integrated

Reading/English paradigm—complete with new curriculum, new instructor expectations, and newly compressed course timelines—is, in our opinion, very consequential to the understanding of many of the students' seemingly subpar statistical data. In fact, Martin (2015), a research associate in the National Institute of Justice's Office of Research and Evaluation, affirms that the timing of evaluations has consequences: "Conducting certain evaluations, like outcome evaluations, is difficult when a program is *too new* [emphasis added] because program elements, strategies or procedures are often still being adjusted and finalized" (p. 2). However, research and evaluation consultants Russ-Eft and Preskill (2009) postulate *both* [emphasis added] formative and summative outcome evaluations of a program's design, are valuable for informing program design changes (p. 30). We contend, a more rigorous and beneficial use of the CCRC's time and effort when our program was brand new would have been to design a study and use findings formatively; then plans to conduct a summative study once our DRE program became established could have been included as part of the research project. In using such a design, many former, possibly "low skills" (Perin et al., 2015, p. 62) DRE students identified in the formative phase of the study currently persisting in their college-level courses could be studied to ascertain potential reasons for their success despite earlier evidence of academic skill deficiency.

Inauthentic Assessment Design Choices

Time constraints. The CCRC administered six assessment tasks (two being the text-based writing tasks that ultimately were scored to determine college-readiness) in 2.5 hours including breaks, then assessed the two writing tasks using "final draft" scoring expectations with no consideration for the lack of time students had to read two articles and execute writing behaviors that reveal writing mastery. To create a comprehensive reading and writing assessment and then use scoring tools that had no quantitative adjustment for the time necessary to bring one's writing to full fruition shows a misunderstanding of what should be expected from a student's writing obtained within a relatively short amount of time. In addition, time constraints can work against students whose reading behaviors are still developing. Slower reading rates are typical of developmental students (Conlon & Sanders, 2011, pp. 194-197) who have not routinely spent a large amount of their previous school years reading for class assignments or even for pleasure (Gallik, 1999, p. 482). We believe, and even pointed out several times prior to administration of the assessments, that being asked to read two articles and to write both a text-based summarization and a text-based persuasive essay for one of the articles, all under time constraints, is an aggressive agenda even for

college-level students. Our concerns as practitioner partners were not addressed in the final assessment protocol.

In order to minimize a potential researcher-practitioner partnership imbalance and, thus, flawed research methodology, communication is key. According to Crooke and Oslwang (2015), researchers and practitioners must engage in a "two-way feedback loop" (p. S1847) in which both researchers and practitioners are "willing to educate one another about the limitations and needs of the research" (p. S1874). For the relationship to stay balanced, practitioners input, Crooke and Oslwang contend, is important (p. S1881). For the CCRC team to dismiss our concerns about time constraints, we believe, showed a lack of trust in our professional opinions as practitioners.

Word processing. Students were expected to compose their essays without the use of any word processing technology, an expectation that is inconsistent with WTCC's current DRE writing

> *For the CCRC team to dismiss our concerns about time constraints … showed a lack of trust in our professional opinions as practitioners.*

practices. A conversation was never initiated by the CCRC team to discuss the consideration of using word processing technology to mimic the students' DRE classroom experience. In fact, not using technology for word processing appears to contradict one of the CCRC researcher's own previous published findings. In a literature review, Perin and a coauthor (Graham & Perin, 2007) cited that, compared with composing by hand, the use of word processing had a consistently positive impact on writing quality because "word processing produces a neat legible script [and] allows the writer to add, delete, and move text easily, [and word processors] include spell checkers as well" (p. 17). Additionally, although book-style dictionaries were provided at the assessment site for students to explore vocabulary choices, students use, and are largely more comfortable using, electronic dictionaries/thesauruses found on their computers or cell phones, which were not allowed during the assessment session.

Student assessment readings. The two readings that were selected for the project-designed assessments were both "social scientific," one from the psychology domain and one from the sociology domain; however, this classification of article is not

representative of the broad spectrum of readings used in our DRE courses. This point was registered with the CCRC team during the assessment drafting stage of the study, but we could not come to a mutual agreement. Therefore our request for an alternate second reading, one more closely related to the students' topic knowledge, was not granted. We now have become aware that the articles chosen by the CCRC team were based on topics discussed in foundational college-level Psychology and Sociology courses required of most students working toward a college credential. This means the CCRC reading selection team was assessing our developmental students' college readiness by asking them to relate to college-level content area topics not yet formally introduced to them.

Writing assessment prompts. We also worked in collaboration with the CCRC team to design writing tasks that would require critical thinking. The two types of assessment writing tasks mutually agreed upon were (a) to write a text-based summary and (b) to write a text-based essay. However, the prompts for these writing tasks were not mutually agreed upon and our practitioner team found them problematic.

First, the text-based summary prompt not only requested that the students summarize the article but also requested that students reflect on and postulate an evidence-based response to the issues discussed in the article. We question the inclusion of the second "argumentative" directive. We contend that an academic summary writing task should not contain any personal opinions.

Second, the text-based essay prompt was supposed to elicit written responses that would demonstrate whether a student could "persuade" through essay-style writing; however, we question whether the students understood they were expected to write an essay from the prompt. The text-based essay prompt was actually a request—"Please answer this question"—followed by not one, but two requests and an imperative. One request posed a text-based comprehension question, the other request queried the student's opinion about the answer to the comprehension question, and the imperative instructed students to "persuade a friend to agree." Nowhere on the student's assessment page was the word "essay" printed; in fact, the term "answer" was used three times: "In your **answer** [emphasis added]," "with your written **answers** [emphasis added]," and "Please **answer** [emphasis added]." We question if any college-level student would be inspired to write an academic essay with this "prompt." The choice to ask students to answer instead of prompting them to write an essay shifts the entire dynamic of the expectation. "Please answer" potentially signals and elicits a markedly less formal style of written response than does, "write an essay." Rank and

Pool (2014) contend that the guidelines for writing assignments (e.g., a prompt) affect the quality of a student's written response: – "using terms precisely in a writing assignment *does matter* [emphasis added]" (p. 676). In Rank and Pool's professional opinion, "It is the instructor's responsibility to use terms that clearly state what we [instructors] want, make obvious what those terms mean, and *to not penalize students if they do what is asked rather than what we wanted them to intuit* [emphasis added]" (p. 677).

Also, the students were not provided with a template, rubric, or example for the "argumentative" persuasive-style essay desired so that the students could conceivably understand and write to the assessment's expectations. The assessment prompt lacked these features despite the fact that Graham and Perin's (2007) previously mentioned literature review recommends setting a product goal that "involves assigning students specific, reachable goals for the writing they are to complete" to include "the purpose of the assignment (e.g., to persuade) as well as the characteristics of the final product" (p. 17) as an effective writing strategy that promotes student writing success. We argue that without knowledge of what is expected to be included in a requested writing sample, a student's chances of performing "well" are greatly diminished.

Unrealistic Assessment Scoring Tools

Main ideas. Of the eight assessment scoring tools (four for the essay and four for the summarization), one scoring tool for the summarization task was counting the number of acceptable main ideas from the reading included in each student's summary. According to Working Paper No. 85, to determine the acceptable number of main ideas in the reading for the summary writing task, three English language arts professionals had to come to a consensus "on which ideas to retain, eliminate, or add" (Perin et al., 2015, p. 27). This team of professionals eventually agreed on nine. Each student's summary score reflected the number of acceptable main ideas out of the nine the professionals had identified. We acknowledge that one could argue a team of content-specific professionals discussing and sharing their ideas contributes to assessment scoring consistency. However, we raise issue with the validity of using an article that contains as many as nine main ideas that are seemingly so contestable that even reading and writing experts needed to debate each idea's function in the article. If understanding and identifying the assessment article's main ideas required a discussion between three language arts professionals, first, we question why we were not also consulted to determine main ideas in the reading as the practitioners in the partnership, and second, we question if any first-year college-level student would be able to identify all nine main ideas.

Essay guidelines. The scoring tool used to assess the text-based essay was extremely problematic. Although the essay "prompt" was intended to elicit students' opinions, the fact that each student's essay was going to be graded using a "holistic persuasive [argumentative] quality rubric" (Perin et al., 2015, p. 25) was never mutually discussed prior to or after administration. Our DRE essay expectations—aligned with NC Student Learning Outcomes (SLOs)—encourage critical thinking and writing guidelines that explore and combine a variety of patterns of development to offer one's opinion in response to text-based reading material, including (but not limited to) exemplification and cause/effect. Also, as previously stated, the students were not provided with a template, rubric, or term-driven prompt for the argumentative style essay so that they could understand and address the assessment's expectations. In addition, we maintain that since several prominent NC universities list instruction of argumentative skills as a writing objective taught

> *The assumption made by researchers who believe the skills taught in remedial courses [DRE] are equal to their respective gatekeeper courses [ENG 111] is erroneous.*

as part of their introductory college-level English courses, DRE students should not be expected to show mastery of this competency prior to entering their first college-level English and/or freshman-level content-specific courses. Developmental education specialists Goudas and Boylan (2013) point out the assumption made by researchers who believe the skills taught in remedial courses [DRE] are equal to their respective gatekeeper courses [ENG 111] is erroneous (p. 29). All text-based writing assessments designed by the CCRC team to measure our student's academic writing level were scored based on this erroneous assumption. Furthermore, according to the U.S. Department of Educational Institute for Education Services National Center for Education Evaluation and Regional Assistance (2003), "Whenever possible, a study should use objective, 'real-world' measures of the outcomes that the intervention is designed to affect" (p. 6). Moreover, research evaluation consultant Massey (1995) asserts that outcome evaluations, like this partnership study, "should be designed based on the [studied] program as implemented" (p. 115). Therefore, we contend that the CCRC's choice to measure college readiness by

expecting a sample of DRE students to demonstrate mastery of argumentative essay writing skills (not a specifically stated NC DRE SLO) without a writing template or rubric, directions containing precise prompt terminology, or prior classroom instruction was unrealistic.

Academic words. An additional scoring component of both the essay and the summarization tasks that may have contributed to inaccurate findings was the examination of the ratio of "academic words" (words that appear frequently in college-level academic texts) to the number of total written words in a student's essay. The CCRC was proceeding under the assumption that less mature writing contains a larger number of high frequency words and fewer low frequency words (Perin et al., 2015, p. 27). Noting that this scoring qualifier is an assumption is important. Using an automated essay scoring program, referred to as an e-rater®, Powers, Burstein, Chodorow, Fowles, and Kukich (2002) decided to test the notion that pretentious language (often a characteristic of low frequency words) does not necessarily create a more academically mature essay. Powers et al.'s results "suggest that gratuitous interjections of such phrases and cue words [like 'unwarranted assumption' and 'fallacy of equivocation'] can affect e-rater® evaluations" by inflating scores (p. 113). As practitioners, we believe clarity of ideas is a truer indicator for whether a student understands a writing topic than the use of ostentatious words; therefore, we instruct students accordingly. Because we were not consulted about this scoring criterion, we never shared our expertise. Plus, to help students understand that a quality essay is first about its structure and organization, students are taught to focus on word choice as a late-stage editing activity. Because editing one's written work for word choice and usage is a late-stage writing behavior, students' limited time to finish the writing tasks may be a contributor to low "academic word" usage.

Word count. Yet another disconcerting scoring component of both the essay and the summarization tasks was word count. According to Working Paper No. 85 (Perin et al., 2015), educational researchers believe lower-achieving students produce very short writings containing limited information (p. 27). Even though we agree short writings tend to represent a developing student's writing behavior, we also believe the word count of an essay or summary is often conditional rather than always indicative of low performance. One such condition is the student's familiarity with the reading's topic and the length of time a student has had to internalize the points in a reading; students tend to write more about what is familiar to them. Olinghouse, Gillespie, and Graham (2015) provide support that "discourse and topic knowledge are important elements in the writing of … developing writers" (p. 404). DRE course SLOs require that students develop knowledge of successful

reading and writing skills only, not knowledge of a specific content-area topics. Specific topic knowledge necessary to support the students' understanding of an article's topic focus takes place in the DRE classroom prior to a required reading. "Low" word counts on the students' writing samples, used for assessing reading comprehension, could be due to the students' underdeveloped familiarity with the topics addressed in the two social-scientific themed articles. A second condition of word count is an instructor's writing directives to students. Instructors will set the word count of an essay according to the intention of the writing. No word count guidance was included in either of the assessment writing-task instructions.

Furthermore, total word count is a meaningless measure of writing sophistication if the words included in the count are words forming sentences devoted to repetitive information or words forming sentences devoted to information that is unrelated or indirectly related to the topic of the writing. Perelman (2012) confirms that "common sense tells us that development and organization are much more complex features than mere verbiage. A horde of rambling, unconnected sentences does not develop an idea" (p. 126).

Powers et al. (2002) found the e-rater® positively favored "essays that 'rambled,' 'missed the point,' used faulty logic, or were 'haphazard' in their progression, but used relevant content words, complex sentence structure, or other features valued by e-rater®" (p. 112). Additionally, Gansle, Noell, VerDerHeyden, Naquin, and Slider (2002) found that although there is valid evidence to support word count as an index of a student's skill level in writing, "total words written [is] not the *best choice* [emphasis added] for predicting skill in written language…other variables have been found that far surpass its predictive power" (p. 494). Again, clarity of ideas over word count is emphasized in the DRE classroom.

Moreover, the possibility that the text-based essay task instructions implied a limited word count must be noted. Each student's instruction page included the directive to hand in the article with his or her "written answers." The pluralized form of the word "answer" potentially implies the students were expected to respond to multiple prompts, thus potentially causing a student to reduce the number of words per response in order to respond well or at least completely to each prompt. Had practitioner partners been consulted about using word count as a scoring tool, we would have suggested that a low word count is not necessarily analogous with substandard writing skills. Since the CCRC team determined that word count would be a scoring tool and no word count expectation would be printed in the assessment writing instructions, a statement included in the CCRC's final working paper alerting research reviewers to the fact that the writing directives for both the text-based essay and the text-based summary tasks lacked a word count range should have been included to acknowledge the potential bias. In addition, the possible word count measurement limitations revealed in the extant literature were never noted in the CCRC's final working paper.

Conclusion

As developmental educators, we fully support the imperative to develop the highest quality, most effective, and most efficient programs of study for students who enter community colleges underprepared for the rigorous expectations of college-level coursework. Few educators and policy-makers would disagree that evidence-based research is a valuable tool for informing the practices and the practitioners in community college developmental programs that support underprepared students. However, we surmise, few educators (including ourselves) and/or policy-makers often question the degree of research measurement error, particularly when potential measurement errors are not duly

---

Never should any college educational program, particularly college-offered developmental educational programs whose existence is often questioned, be at risk of public misrepresentation by biased research data.

---

noted in the published findings, by recognizing limitations and their potential consequences on a study's implications. Our personal experience as the practitioners in the partnership formed with the CCRC has greatly contributed to our understanding of why critics of EBPs contend that when research end-users are reviewing research studies or program evaluations for validity in a specific context, educators and policy-makers must examine each study "on the strength of the research design" (Penuel et al., 2015, p. 186). Before our partnership experience, we would have had very little reason to question the validity of the data in researcher-practitioner studies, much less the validity of the partnerships established within those studies because the very essence of the word "partnership" implies a mutualism and trust between professionals.

Going forward, we want our experience as the practitioners in the CCRC's Developmental Reading and English Assessment in a Researcher-Practitioner Partnership study, results of which have been formally published, to inspire a heightened awareness of the potential for an imbalance in a researcher-practitioner partnership study as well as a heightened understanding of its potential negative impact (i.e., bias) on the resulting data, data that ultimately informs the practitioners and policy-makers at local, state, and national levels. We also want our experience to encourage educational practitioners who are asked to be in a research partnership, with any research team, to intentionally query all facets of the impending study – including those facets not seemingly within one's professional purview – to ask many questions both before and during the study process, and to require that a statement of limitations or measurement error be added to the resulting study if a professional-based request is deemed non-negotiable by the research team. Never should any college educational program, particularly college-offered developmental educational programs whose existence is often questioned, be at risk of public misrepresentation by biased research data because the research protocols and methodologies were poorly designed, due in large part, to an imbalanced researcher-practitioner relationship.

References

CCRC (Community College Research Center). (2017). *About us.* Retrieved from http://:ccrc.tc.columbia.edu/About-Us. Html

Coburn, C. E., Penuel, W. R., & Geil, K. E. (2013). *Research-practice partnerships: A strategy for leveraging research for educational improvement in school districts* (White paper). New York, NY: William T. Grant Foundation. doi:10.1002/sce.2128

Conley, D. T. (2008). Rethinking college readiness. *The New England Journal of Higher Education, 22*(5), 24-26. doi:10.1002/he.321

Conlon, E., & Sanders, M. (2011). The reading rate and comprehension of adults with impaired reading and visual discomfort. *Journal of Research in Reading, 34*(2), 193-214. Retrieved from doi:10.1111/j.1467-9817.2009.01421.x

Crooke, P. J., & Olswang, L. B. (2015). Practice-based research: another pathway for closing the research-practice gap. *Journal of Speech, Language, and Hearing Research, 58,* S1871-S1881. doi:10.1044/2015_JSLHR-L-15-0243

Epstein, I. (2011). Reconciling evidence-based practice, evidence-informed practice, and practice-based research: The role of clinical data-mining. *Social Work, 56*(3), 284-288. Retrieved from https://search-proquest-com./stable/922421950

Gallik, J.D. (1999). Do they read for pleasure? Recreational reading habits of college students. *Journal of Adolescent & Adult Literacy, 42*(6), 480-488. Retrieved from http://www.jstor.org/stable/40014062

Gansle, K.A., Noell, G.H., VanDerHeyden, A.M., Naquin, G,M., & Slider, N. J. (2002). Moving beyond total words written: The reliability criterion validity, and time cost of alternate measures for curriculum-based measurement in writing. *School Psychology Review, 31*(4), 477-497. Retrieved from https://search-proquest-com./stable/219652835

Goudas, A. M., & Boylan, H. R. (2013). A brief response to Bailey, Jaggers, and Scott Clayton. *Journal of Developmental Education, 36*(3), 28-32. Retrieved from http://www.jstor.org./stable/42775424

for developmental writing, but that simply doesn't seem possible. In an environment in which many state legislatures are exploring ways to further limit developmental writing, it does not seem possible that we will ever return developmental education to the universities that have removed it, but there are steps we can take to mitigate the negative effects of the current situation.

First, graduate programs in English need to recognize that many of their graduates will be teaching developmental writing at a community college. At the least they should offer a course, preferably more than one, in the teaching of developmental writing. They also could establish internships for graduate students at nearby community colleges. Perhaps they could consider hiring, at least on a part-time basis, faculty who have extensive records of accomplishment in community colleges. And, even though little developmental writing is going on at universities today, scholars there could still choose to do research in this important area. In recent years, many studies of developmental education have been conducted by the Community College Research Center at Columbia. We are lucky to have those, but many more universities could still be partnering with community colleges to produce developmental writing scholarship.

Community colleges could decide to create positions with reduced teaching loads to allow time for research, as Piedmont Central Community College recently did. We could also, regardless of what accrediting agencies say is the minimum credential, hire more qualified faculty to teach developmental writing, and we could encourage our most experienced faculty to teach developmental writing on a regular basis. Recognizing that most newly graduated faculty will arrive with little or no preparation to teach developmental writing,

Lack of socioeconomic mobility is a problem.

we should establish robust programs of faculty development. We could, for example, give newly hired English faculty a one-course reduced load and then require that they participate in a rigorous program to prepare them to teach developmental courses.

Many in this country have, in recent years, warned about the danger that the American Dream—the ideal that everyone in this country should have a chance to succeed, the idea that anyone willing to work hard can improve their state in life, the idea that we want to close the increasingly large gap between the rich and everyone else—is endangered. This is one of the few issues today on which politicians of both parties agree. The two parties may have different approaches to the problem, but they seem to agree that the lack of socioeconomic mobility is a problem.

Unfortunately, just when the political climate appears ready to address socioeconomic equity, the focus of higher education has moved on. Few in higher education would express opposition to developmental education, but few seem to recognize that developmental education is the place in higher education where we most directly address this problem. For the sake of our developmental programs, our developmental students, and social equity in our society, we must regain that important focus.

Reference

Traub, J. (1995). *City on a Hill: Testing the American Dream at City College*. New York, NY: Perseus Books.

Peter Adams (padams2@ccbcmd.edu) is Professor Emeritus at the Community College of Baltimore County, 800 South Rolling Road, Baltimore, MD 21228-5317. You can follow him on Twitter at @ALPAdams.

Graham, S., & Perin, D. (2007). *Writing Next: Effective strategies to improve writing of adolescents in middle and high schools – A report to Carnegie Corporation of New York*. Washington, DC: Alliance for Excellent Education. Retrieved from https://www.carnegie.org/media/filer_public/3c/f5/3cf58727-34f4-4140-a014-723a00ac56f7/ccny_report_2007_writing.pdf

Hassell, H., & Giordano, J.B. (2015). The blurry borders of college writing: Remediation and the assessment of student readiness. *College English, 78*(1), 56-80. Retrieved from https://search-proquest-com./stable/1705118501

Hathcoat, J.D., & Penn., J.D. (2012, Winter). Generalizability of student writing across multiple tasks: A challenge for authentic assessments. *Research and Practice in Education, 7*, 16-28. Retrieved from https://search-proquest-com./stable/1505321347

Martin, A.B. (2015, September). *Plan for program evaluation from the start NIJ Journal, 275*. Washington, DC: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice. Retrieved from https://nij.gov/journals/275/pages/plan-for-program-evaluation.aspx

Massey, O. T. (1995). *Evaluating human resource development programs: A practical guide for public agencies*. Boston, MA: Allyn & Bacon.

Nevo, I., & Slonim-Nevo, V. (2011). The myth of evidence-based practice: towards evidence-informed practice. *British Journal of Social Work, 41*, 1176-1197. doi:10.1093/bjsw/bcq149

Olinghouse, N., Gillespie, A., & Graham, S. (2015). The relationship of discourse and topic knowledge to fifth graders' writing performance. *Journal of Educational Psychology, 107*(2), 381-406. Retrieved from http://dx.doi.org/10.1037/a0037549.

Palmer, J. (1999). Research matters: A call for the application of empirical evidence to the task of improving the quality and impact of environmental education. *Cambridge Journal of Education, 29*(3), 379-395. doi: 0305-764/99/030379-17

Penuel, W. R., Allen, A. R., Coburn, C. E., & Farrell, C. (2015). Conceptualizing research-practice partnerships as joint work at boundaries. *Journal of Education for Students Placed at Risk, 20*(1-2), 182-197. doi:10.1080/10824669.2014.988334

Perelman, L. (2012). Construct validity, length, score, and time in holistically graded writing assessments: The case against automated essay scoring (AES). In C. Bazerman, C. Dean, J. Early, K. Lunsford, S. Null, Pl Rogers, & A. Stansell (Eds.), *International Advances in Writing Research: Cultures, Places, Measure* (pp. 121-150). Fort Collins, CO: WAC Clearinghouse & Parlor Press. Retrieved from wac.colostate.edu/books/wrab2011/

Perin, D., Raufman, J., & Kalamkarian, H. S. (2015). *Developmental reading and English assessment in a researcher-practitioner partnership* (CCRC Working Paper No. 85). New York, NY: Community College Research Center, Teachers College, Columbia University.

Powers, D. E., Burstein, J. C., Chodorow, M., Fowles, M., & Kukich, K. (2002). Stumping e-rater: Challenging the validity of automated essay scoring. *Computers in Human Behavior, 18*(2), 103-134. Retrieved from http://doi:10.1016/S0747-5632(01)00052-8

Rank, A., & Pool, H. (2014). Writing better writing Assignments. *American Political Science Association, 47*(3), 675-681. doi:10.107/S1049096514000821

Russ-Eft, D., & Preskill, H. (2009). *Evaluations in organizations: A systematic approach to enhancing learning, performance, and change* (2nd ed.). New York, NY: Basic Books.

Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods, 1*(2), 199-223. doi:10.1037/1082-989X.1.2.199

Slavin, R. E. (2008). Evidence-based reform in education: Which evidence counts? *Educational Researcher, 37*(1), 47-50. doi:10.3102/0013189X08315082

Spencer, T.D., Detrich, R., & Slocum, T. (2012). Evidence-based practice: A framework for making effective decisions. *Education and Treatment of Children, 35*(2), 127-151. Retrieved from http://doi.org/10.1353/etc.2012.0013

Sullivan, T.P., Khondkaryan E., & Fisher, B.S. (2013). *Recommendations for collaborating successfully with academic researchers, findings from the researcher-practitioner partnership study (RRPS)*. Washington, DC: U.S. Department of Justice. Retrieved from https://www.ncjrs.gov/pdffiles1/nij/grants/243911.pdf

U.S. Department of Education, Institute of Education Services, National Center for Education Evaluation and Regional Assistance. (2003). *Identifying and Implementing Educational Practices Supported by Rigorous Evidence: A User Friendly Guide*. Washington, DC: Author. Retrieved from https://www2.ed.gov/rschstat/research/pubs/rigorousevid/rigorousevid.pdf

Whitehurst, G.J. (2002, October). *Evidence-based education* [PowerPoint slides]. Retrieved from http://www2.ed.gov/nclb/methods/whatworks/eb/edlite-index.html

Yates, C. (2013). Evidence-based practice: the components, history, and process, *Counseling Outcome Research and Evaluation, 4*(1), 41-54. doi:10.117712150137812472193