Title
A Theory of Ethics for Writing Assessment

Permalink

Journal

ISSN

Author
Elliot, Norbert

Publication Date
2016

Copyright Information

Peer reviewed

Volume 9, Issue 1: 2016

A Theory of Ethics for Writing Assessment

by Norbert Elliot, New Jersey Institute of Technology

This paper proposes a theory of ethics for writing assessment. Based on a definition of fairness as the identification of opportunity structures created through maximum construct representation under conditions of constraint--and the toleration of constraint only to the extent to which benefits are realized for the least advantaged--the theory is expressed in terms of its tradition, boundary, order, and foundation. To examine the force of the theory, a thought experiment demonstrating action based on the theory is offered so that its weaknesses and strengths are identified. Intended for the research specialization of writing assessment, the theory has generalization implications for the field of writing studies.

1.0 Introduction

Fairness is the first virtue of writing assessment. Conceived as the structuring of opportunity, the aim of fairness unifies foundational measurement concepts of validity and reliability into a framework of principled inquiry organized to achieve an ethical outcome. To advance that outcome, a proposed theory of ethics is the subject of this paper.

Because of the complexities involved in providing the core of the theory, it is helpful to begin with its location, scope, statement, and aim. Some readers may wish to approach the theory deductively by turning directly to the sections referenced in this introduction; others may prefer a more inductive approach in which the theory is considered along a trajectory from its origins to its challenges. Still others may want to begin with the thought experiment and examine the potential of the theory in application. Whatever the method of reading, cross-reference by sections allows discrete ideas to be examined in detail while allowing connections among them to be established.

1.1. Location

As part of the profession of English Language and Literature/Letters, the proposed theory is located within the field of Rhetoric and Composition/Writing Studies, hereafter referred to as writing studies.[1] As part of the field, writing assessment is defined under the theory of ethics as a research specialization dedicated to the investigation of fairness, validity, and reliability as these foundational concepts are used to structure opportunity and thereby advance opportunity to learn (§3.1). Following the United Nations Educational, Scientific and Cultural Organization (2015), universal education and the achievement of literacy are understood as human rights and are therefore non-negotiable under the proposed theory.

1.2 Scope

As the introduction to this *JWA* special issue argues, the proposed ethical theory of writing assessment must meet seven obligations: it must explore validity and reliability through ontological, epistemological, and axiological perspectives of fairness; it must provide an overarching referential frame; it must have a systems orientation; it must provide a unifying function; it must account for stakeholder perspectives; it must have value for a range of assessment contexts; and it must hold assessment users to actionable ethical principles. These obligations align with the ongoing discussion in the field regarding the purposes and aims of writing assessment and its connection with student learning (Condon, 2013; Gallagher, 2014; Haswell, 2005; Huot, 2002; Inoue & Poe, 2012; Inoue, 2015; Lynne, 2004; White, Elliot, & Peckham, 2015; Yancey, 2012). While §5 demonstrates how these obligations have been met, it is best to begin with the idea of theory itself to demonstrate how abstraction anneals practice.

Any theory, Morton (1980) has written in his theory of theories, "is a body of assertions whose terms refer to individuals and properties, and which is transmitted and evolves in accordance with the intention that it asserts truth about them" (p. 5). However, he adds,

> a theory need not be stated or statable by any single person. It need not be precise. Nor need everyone who subscribes to it know quite what his subscription commits him to. But one must use the theory as if those one had obtained it from intended its terms to refer to objective realities, and the changes one makes in the theory, by removing existing beliefs or adding new ones, must be made with the intention of increasing the proportion of true to false assertions about those realities. (p. 5)

In the case of theory-building as part of humanistic inquiry, Cole (2015) identified three features of theory: its ability to capture the difficulties of thinking; its realization that we conceptualize in language and thus conjure materiality; and its ability to historicize that very materialization.

While the proposed theory for writing assessment is deeply constructivist, as discussed in §3.2.3, it operates as if the present proportion of true to false assertions is well balanced. In this way, Morton's concept of theory aligns with the present effort. In reviewing the theory, readers will recognize the difficulties of expressed thought: While ideas can be made clearer, they often cannot be made simpler. As to the conceptual power of categories, readers will notice that some corral nicely and other roam freely; regardless of domestication, the very existence of an idea is fascinating to watch. In terms of materialization, theory becomes a map across difficult terrain. Following the map in practice determines its worth and occasions the need for new directions. Here at the beginning, it is best to view the present theory as an invitation for others to do participate in the process of work undertaken in the service of difference.

1.3. Statement

While the theory and its benefits are given in §3.1, statement of the theory here will be of structural help. The proposed theory rests on this definition of fairness: Fairness in writing assessment is defined as the identification of opportunity structures created through maximum construct representation. Constraint of the writing construct is to be tolerated only to the extent to which benefits are realized for the least advantaged. In his review of this manuscript, psychometrician Robert J. Mislevy expressed the core of the theory thus: $Fairness_1 \grave{1} Validity \grave{1} Fairness_2$, where $Fairness_1$ is understood as the received measurement view explained in §3.1 and Validity understood as the traditional aim of assessment explained in §3.3.2 (personal communication, August 14, 2015). The present theory further holds that $Reliability \grave{1}$ $Validity \grave{1} Fairness$, where Fairness integrates the foundational principles of Reliability and Validity. Rhetorically based, this aim yields focus on the language of materiality and sites of that materiality in which fairness is the first virtue and consequence must be the first concern.

1.4. Aim

This definition has four aims. Each occurs at different points as associated with the theory's articulation, but a straightforward statement here will be useful.

First, the theory requires statement and subscription and so differs from the tacit understanding identified by Morton. The proposed theory holds that stated principles for action are best identified and achieved through philosophical analysis associated with the concept of fairness. The definition (§3.1), boundary (§3.2), order (§3.3), and foundation (§3.4) of the theory are, taken together, intended to provide a principled, integrated system of research. This theory is therefore not intended to be a code, nor is it intended to be a method of investigation used in the service of validity or reliability inferences. The theory is intended to be a complete statement to be used by those who subscribe to it as a distinct framework for writing assessment. While the rich measurement traditions of validity and reliability remain in play, the theory seeks to integrate these traditions under the definition of fairness and delineate the ethical dimensions that follow through a principled approach. As argued in §3.2.4, both standard reporting guidelines needed for comparative analysis and common empirical techniques needed to analyze evidence associated with fairness are often absent from writing assessment. The theory is therefore intended to advance research both systematically and substantially so a defined referential frame may be

maintained and modified as needed from one generation of researchers to the next.

With the aim of statement and subscription, the theory is also intended to yield a unified research approach leading to programs of research across all writing assessment genres.[2] This unified approach is the second aim of the theory. Cautious of the legacies of imperialism noted by Cushman in this *JWA* special issue (§1) and the hegemonic views that accompany them (§2), the theory advocates an approach to research that is based on the definition of fairness established above. Specifically, the theory acknowledges that method alone will not achieve this unification and will, more likely than not, result in technological determinism. The theory is therefore intended to advance an approach to research that structures opportunity. A brief example demonstrates the distinction between technique and aim. Imagine need is voiced for a locally-developed placement test for admitted students at a post-secondary United States institution. Using traditional techniques, scores would be examined using the general linear modeling approaches noted in §3.2.3 and score use would be justified according to reliability and validity evidence. Beyond this emphasis on technique, the theory calls for justification of the placement test based on construct representation, reasonable assurance that the students have had the opportunity to learn the construct specified in the test, and advanced resource allocation to help those who are disadvantaged by the test secure appropriate instruction. While general linear modeling techniques would surely be used, regression techniques (§3.2.4) would be required to examine the potential for disparate impact as defined by Poe and Cogan in this *JWA* special issue. If these conditions could not be met, than the need for the test would be reexamined and attention would turn, specifically, to co-requisite options (Complete College American, 2015) and, more broadly, to the reasons that so little faith was placed in the high school teachers and their grades through which students had already been admitted.

The example gives rise to the third aim of the theory: reduction of over-testing (Hart, Casserly, Uzzell, Palacios, Corcoran, & Spurgeon, 2015; Lazarin, 2014). By integrating construct representation with fairness, the theory aims to place ethical demands on assessments so that, when they are used, faith is justifiably placed in score use. If demands for fairness cannot be met, then the assessment cannot be used and other criterion measures must be explored.[3] Before characterizing this consequence as utopian, it is important to recall that, while federal law requires state-mandated testing in schools, many writing assessments are in complete control of state, district, and local leaders. This return to the local is especially reinforced by the recent reauthorization of the Elementary and Secondary Education Act of 1965 by the Every Student Succeeds Act (2015). The same is true for regional accreditation in which post-secondary assessments remain locally controlled. In such contexts, the theory might play a major role in restoring the centrality of teacher knowledge and the role of that knowledge in assessing student learning. In the example above, a desirable outcome would be to examine the specific reasons why an admitted student is not considered a qualified student.

The fourth aim of the theory is based on perspective gained from the Civil Rights Movement (Lewis, 1999): The theory is intended to get in the way of that which does harm. In the case of assessment, it is imperative to insist on explicit connections to the identification of opportunity structures leading to the advancement of opportunity to learn. When learning becomes the stated focus of assessment, opportunities arise to lessen educational inequalities. Examined in §3.1, the view of opportunity to learn used in the theory welcomes collaboration across sociocultural and psychometric perspectives and celebrates the profoundly situated nature of learning (Moss, Pullin, Gee, Haertel, & Young, 2008).

More précis than introduction, this summary of the location, scope, statement, and aim of the theory allows us turn to its origins.

2.0 Traditions

Two traditions associated with the proposed theory—educational measurement and philosophy—are historically concerned with ethics, and attention to each provides origin and basis. In standards and codes, measurement deals almost exclusively with safety and risk to participants through attention to respect, beneficence, and justice associated with clinical research (Levine, 1986). In theoretical propositions, philosophy deals with abstract systems that provide perspectives necessary to grapple with dominant moral cultures (MacIntyre, 2007). Close attention reveals connections between both, especially in terms of fairness. With unique attention to the concept of fairness, especially as it is examined in the work of John Rawls (1971/1999; 2001) connections may be identified that establish the basis for an integrated, unified ethical theory for writing assessment. Significant to the theory, a third tradition, the law, is covered by Poe and Cogan in this *JWA* special issue and is therefore not treated here.

2.1 Measurement Traditions

Focusing on designing, implementing, and evaluating assessments of human learning, three research traditions have historically been concerned with fairness. With the 1969 American Psychological Association report of Baxter and his colleagues taken as a hallmark event—the report concluded that job testing of minority groups failed construct validity requirements of test content, lacked content relevance regarding employment appraisal, and thus frustrated accurate interpretation of scores—the development of attention to fairness in measurement traditions may be traced from that point.

2.1.1 Educational measurement. As a field of study, educational measurement originated at the end of the nineteenth century. By the time of the publication of Wood's *Measurement in Higher Education* (1923), the field had identified investigative principles that would stand until the time of the Baxter report, when it became clear that the goal of objectivity resulting from validity and reliability assurances was in question in terms of diverse groups. In the American Psychological Association's 1974 *Standards for Educational and Psychological Tests* (the third edition), attention was given to the social impact of test scores. Such attention continued until the present 2014 edition in which fairness in testing has been given its own section along with the foundational concepts of validity and reliability. In the *Standards for Educational and Psychological Testing,* fairness is defined as

> the validity of test score interpretations for intended use(s) for individuals from all relevant subgroups. A test is fair that minimizes the construct-irrelevant variance associated with individual characteristics and testing contexts that otherwise would compromise the validity of scores for some individuals. (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014, p. 219)

As the definition reveals, fairness is understood in terms of validity and does not emerge as its own developed foundational concept (Elliot, 2015).

Following the *Standards*, key educational measurement stakeholders such as the Educational Testing Service (ETS) have provided statements of quality and fairness. In the case of ETS, the non-profit organization's statements of fairness originate in October of 1981. In its present form, the definition of fairness is founded on both the presence of subgroup difference and elimination of construct irrelevant variance. For ETS, fairness is defined as "the extent to which a product or service is appropriate for members of different groups" (p. 57). Distinct from the AERA, APA, and NCME *Standards*, fairness and validity are not understood solely in terms of each other in the *ETS Standards for Quality and Fairness* (2014). Indeed, specific attention is drawn to diversity of populations. To assure fairness for all examinees, technical attention is given to the importance of differential item functioning to investigate score differences for African-American, Asian-American, Hispanic-American, and Native-American (as compared to White) product users, and female (as compared to male) users. Similar analyses are called for in terms of product users with disabilities and English-language learners. Issued just after the publication of the 2014 *Standards*, the ETS document is lucid in its analysis of critical issues surrounding score use and meaningfully advances fairness as its own foundational principle. Intended solely to guide employees of the non-profit organization, the *ETS Standards* is an example of an institutional guide regarding responsibility, transparency, and accountability. Further analysis and critique of the *ETS Standards* is provided by Broad in this *JWA* special issue.

Internationally, the Research Council of Norway has produced *Assessment and Learning* (Baird, Hopfenbeck, Newton, Stobart, and Steen-Utheim, 2014), a comprehensive review of assessment and learning. In their proposal for ethical evaluation of assessment policy, the authors offer a neo-Messickian framework. Their framework is based on an influential analysis of validity and the ethics of assessment by Messick (1980), who stressed the importance of construct validity as a rational foundation for prediction and the significance of accounting for value implications in score interpretations. In a well-known matrix illustrating the intersection of the

sources of justification to the interpretation and use of scores (Figure 1, p. 1023), Messick distinguished between the evidential and consequential basis of test use. Extending this evaluative framework, Baird and colleagues examined foci for the evaluation of assessment policy with aspects of technical and social evaluation in the pursuit of an overall judgment: "Is it acceptable to implement (or continue implementing) the assessment policy?" (Figure 10, p. 90). As the authors recognize, evaluative judgment involves ethical judgment. Mislevy and his colleagues (2014) have extended this work to offer a framework of fairness based in conditional inferences. In §4, an extended thought experiment similar to that of the neo-Messickian framework is offered as an examination of the proposed ethical model.

2.1.2 Language assessment. The International Language Testing Association (ILTA, 2000) has adopted a *Code of Ethics* that "draws upon moral philosophy" to guide professional conduct (p. 1). Emphasizing fundamental principles of beneficence and justice, the authors recommend that the code be used to inform the *ILTA Guidelines for Practice* (2007) and its identification of basic considerations for good testing practice and the rights and responsibilities of test takers. The Association of Language Testers in Europe (2010) similarly created a Code of Practice that identifies the responsibilities of association members (developers who construct and administer examinations) and examination users (those who select examinations and make decisions based on scores). Reviewing such codes, McNamara and Roever (2010) found that they highlight the social context of assessment, offer a systematic process of identifying biased items, increase the transparency of the profession of language testing, and reassure stakeholders of the moral framework of those belonging to that profession. Each of these benefits of codes is also gained by the ethical theory proposed here. Recent work has focused on the broad influence of philosophical ethics (Spolsky, 2014) and the influence of fairness and justice as specific ethical concepts (Kunnan, 2014b).

2.1.3 Writing assessment. Each year, the National Council of Teachers of English (NCTE) develops a policy platform to guide literacy education advocacy efforts. The 2015 platform emphasized capacity building, educational equity, evidence-based literacy education, and assessments for learning and accountability. In terms of assessments, attention is drawn to using standardized tests only for purposes for which they have been proven valid. Tests designed to measure school performance, the platform cautions, should not be used to evaluate individual teachers—a finding in accordance with that of Haertel (2013) regarding value added models.

An affiliate of NCTE, the Conference on College Composition and Communication (CCCC) has produced position statements on assessment and ethical issues. Among the position statements on assessment, the authors of *Writing Assessment: A Position Statement* (2009) emphasized evidence-based decisions to insure that scores are "valid, fair, and appropriate to the context and purposes for which they are designed." Guiding principles stress assessments' primary usefulness as a means of improving teaching and learning, the social nature of writing, the fluid nature of writing ability in context, the relationship of this fluidity to evaluative methods, and the need for use of best practices. Among the position statements on ethical issues, the authors of the *CCCC Guidelines for the Ethical Conduct of Research in Composition Studies* (2015a) articulated a common commitment to protecting "the rights, privacy, dignity, and well-being" of those involved in studies conducted by CCCC members. With emphasis on compliance and competency, attention is given to a diverse range of topics, from obtaining informed consent to indicating possible conflicts of interest. Principled action yielding respect for examinees is a benefit of the proposed ethical theory.

In the research specialization of writing assessment, Schendel and O'Neill (1999) were the first to call attention to the problematic separation of ethical inquiry from the technical discussions of reliability and validity—a separation also observed by many in the educational measurement community at the present time (Borsboom, Cramer, Kievit, Scholten, & Francic, 2009; Cizek, 2012). The divide is artificial, Schendel and O'Neill wrote, "because ethical inquiry is needed to evaluate the consequences and implications of an assessment" (p. 210). Emphasizing the uses and consequences of self-assessment procedures such as directed self-placement (Royer & Gilles, 1998), Schendel and O'Neill extended the premise that empirical research is, in itself, an ethical obligation (Beason, 2000). As they observed, we must "tease out through ethical and validity inquiry how methods of gathering data and the way writing professionals use such data impact students, teachers, writing programs, and even the larger public discussions about literacy and education" (p. 221). The obligation to perform empirical research, informed by principles of ethical philosophy, is examined in §3.3.4.

In her comprehensive general theory of writing assessment, Lynne (2004) called for an ethics of assessment that would focus on relationships among those affected by the assessment. Based on that focus, ethical principles would then provide a method of organizing and understanding the conduct of the participants. Concentrating on the definition of validity itself, Inoue (2009) has called for a new concept of racial validity as "an argument that explains the degree to which empirical evidence of racial formations around our assessments and the theoretical frameworks that account for racial formations support the adequacy and appropriateness of inferences and actions made from the assessments" (p. 110). Such principles of organizational inquiry are further developed by Inoue and Poe in a landmark edited collection (2012), in a special issue of *Research in the Teaching of English* on the consequences of writing assessment (Poe, 2014), and in a forthcoming special issue of *College English* on writing assessment and social justice (Poe & Inoue, in press).

In his comprehensive theory of classroom-based writing assessment, Inoue (2015) focused on the investigation and elimination of racism in the classroom. To eliminate racist pedagogies, he proposed a new assessment ecology:

> a complex political system of people, environments, actions, and relations of power that produce consciously understood relationships between and among people and their environments that help students problematize their existential writing assessment situations, which in turn changes or (re)creates the ecology so that it is fairer, more livable, and sustainable for everyone. (p. 82)

Focusing on counter-hegemonic structures, Inoue offered a heuristic centered on purpose (connections between the classroom assessment and learning opportunities), process (student labor), places (status created through judgment), parts (codes, texts, and artifacts that are developed and exchanged during assessment), power (relationships that advance student learning), people (roles within the ecology), and products (material consequences of judgment). Emplacing micro and macro connections, the heuristic is to be used to guide instructional design.

2.2 Philosophical Traditions

Traditions of educational measurement, language testing, and writing assessment reveal important benefits for a theory of writing assessment. As well, philosophical inquiry is important if we are to establish a theory of ethics grounded in the analytic tradition of moral philosophy—what Rawls (1971/1999) termed "the study of the conception and outcome of a suitably defined rational decision" (§40, p. 221). The theory relies on the tradition of philosophy as a source of critical reflection. Archetypally, philosophy—the investigation of basic principles that underlie all branches of knowledge—is divided into five fields: logic (methods of thought); epistemology (methods of knowledge); metaphysics (the nature of reality); aesthetics (the nature of beauty); and ethics (the nature of the virtuous). Although fruitful inquiry may be made within each of these fields regarding a moral theory for writing assessment, the proposed theory concentrates exclusively on ethics.

Beyond codes, philosophical analysis allows integrated, extended, and principled investigation through unified referential frames that have been debated and used since their origins in the ancient world. Whether we do or do not agree that Aristotle (2014) and Aquinas (2009) allow us to understand the predicaments of moral modernity, the analysis presented by MacIntyre in *After Virtue* (2007) remains a created ethical worldview that demonstrates the power of theory described by Cole (2015). While a review of the origins, nature, justification, perspectives, and challenges to ethics as the philosophical analysis of morality is beyond the scope of the present effort, three approaches to rational decisions involving ethical conduct lend important elements to the theory (Hursthouse, 1999).

2.2.1. Kantianism. In his *Groundwork of the Metaphysics of Morals* (1785/1997), Kant advanced his categorical imperative: "Act only in accordance with that maxim through which you can at the same time will that it become a universal law" (p. 31). This universal imperative of duty, an absolute rule of rationality derived from Enlightenment adherence to the power of rationality, has exerted enormous influence on conceptions of our obligations to each other. Despite its shortcomings of conflict with other moral imperatives and embrace of absolute rules, Kant's categorical imperative endures.

Kant is important to the proposed theory for three reasons. First, as Rachels (1986) observed, the categorical imperative highlights the fact that moral reasons are both rational in origin and binding across time and circumstance. From a moral point of view, exceptionalism and contingency are not desired ends; communal decision-making and principled action are desired ends. Second, as Rawls (1971/1999) noted, communal decisions are best viewed within the perspectives of contractarian traditions of both Rousseau and Kant (§40). For Rawls, the categorical imperative was closely related to the principles of justice, a relationship examined in §2.3.3. Emphasizing the choice of exerting the categorical imperative as indicative of "a free and equal rational being," Rawls identified the non-teleological basis for the categorical imperative and the principles of justice (p. 222). While both presume to have some conception of moral pursuit, nothing is known about the final ends of rationality pursued under such frameworks. Indeed, both Kant and Rawls incorporate an assumption of mutual disinterest that allows freedom of choice in a system emphasizing principles of conduct, a premise of the theory offered in §3.1.

2.2.2 Utilitarianism. In *Utilitarianism* (1861/1879), Mill advanced his greatest happiness principle:

> the ultimate end, with reference to and for the sake of which all other things are desirable (whether we are considering our own good or that of other people), is an existence exempt as far as possible from pain, and as rich as possible in enjoyments, both in point of quantity and quality. (p. 17)

This imperative of utility, derived from the Industrial Revolution and its adherence to the importance of social good, remains the dominant philosophy of property-owning capitalism. As Sidgwick (1884) observed, utilitarianism does not require an action as right or wrong based on an impersonal law; rather, a utilitarian acts for the welfare of those for whom some degree of fellow feeling is felt. As Rawls (1971/1999) concluded, the utilitarian view is psychologically understandable (§72, p. 417).

As is the case with both the theory of justice of Rawls and the proposed theory of writing assessment, utilitarianism stands as a contrast for two reasons. First, the aim of achieving the greatest net balance of satisfaction summed over individuals, as Rawls put it (1971/1999, §5, p. 22), ignores the fate of the individual. As such, utilitarianism is incompatible with the idea of justice. Second, utilitarianism is a teleological theory, defining the good locally as a homogeneous quality maximized to achieve the totality of greatest happiness. In contrast, the contractarian theory of Rawls and of the proposed theory for writing assessment,"moves in the opposite fashion by identifying a sequence of increasingly specific structural forms of right conduct each set within the preceding one, and in this manner working from a general framework for the whole to a sharper and sharper determination of its parts" (Rawls, 1971/1999, §85, p. 496).

2.2.3 Social justice. Abjuring teleological assumptions associated with hedonistic utilitarianism, the theory of social justice stands in counterpoint to dominant capitalistic theories of social good. Because he addresses distributive principles based on social advantage, the work of John Rawls is especially important to the proposed theory.

Rawls' theory of justice may is offered in two parts: *A Theory of Justice* (1971/1999), containing the essence of the theory in its considerable revision following a German translation in 1975; and *Justice as Fairness: A Restatement* (2001), containing emendations, corrections, and improvements to the original theory. While it is impossible to provide even a précis of the theory of justice as proposed by Rawls, three important concepts illustrate its force.

First, the theory of justice is a form of political philosophy that stands as a rebuttal to utilitarianism. The term "justice as fairness" is a contractarian theory brought into service to advance the aims of a democratic, constitutional social compact—a conveyance that the principles of justice are agreed to in an initial situation that is fair (Rawls, 1971/1999, §2, p. 11). Its assumptions, therefore, are to advance maximum liberty under realistically constrained conditions necessary to maintain the social contract as advanced by Rousseau (1762/2006). For Rawls, liberty is understood as a primary good (2001, §17.2, p. 58).

Second, the theory rests on a sequenced, lexicographic process. Justice as fairness operates in a precise sequence in which two principles must be advanced and met before others. Here we find the first and second principles of justice. The first, a constitutional principle, states "each person has the same indefeasible claim to a fully adequate scheme of equal basic liberties, which scheme is compatible with the same scheme of liberties for all" (Rawls, 2001, §13, p. 42). The second, a difference principle, states:

> social and economic inequalities are to satisfy two conditions: first, they are to be attached to offices and positions open to all under conditions of fair equality of opportunity; and, second, they are to be to the greatest benefit of the least advantaged members of society. (Rawls, 2001, §13, pp. 42-43).

Hence, equal rights must be satisfied before benefits to the least advantaged can be distributed.

Third, the theory of justice is communal. As opposed to the singular action of the categorical imperative, the enactment of justice thus is to take place under a veil of ignorance in which "no one knows his place in society, his class position or social status, nor does anyone know his fortune in the distribution of natural assets and abilities, his intelligence, strength, and the like"(Rawls, 1971/1999, §3, p.11). In this "original position," decisions are to be rationally made that yield a "well-ordered society" (Rawls, 1971/1999, §3, p. 13). This communal interaction is the very embodiment of the idea of "justice as fairness" and explained more fully in §3.4.3.

The theory of ethics for writing assessment is greatly indebted to the theory of justice. It is impossible to think of the former without acknowledging its debt to the latter. A triumph of compassion and reason, the theory of justice makes one proud to be human. What Rawls observed of his own theory may be also observed of the theory proposed here: The ideas are classical and familiar, and the intention has been to organize these ideas into a general framework, using certain devices, so their force can be demonstrated for the specialization of writing assessment.

2.3 Departure from Rawls
With debt acknowledged, the theory of ethics for writing assessment differs from the theory of justice in four ways. The differences arise not from flaws in the logic of Rawls but, rather, in the demands of particular instances that the proposed theory must address.

2.3.1 Insistence on moral basis. Rawls did not intend his theory of justice as fairness to be a comprehensive religious, philosophical, or moral doctrine. Rather, his is a political conception of justice for the special case of contemporary democratic society (Rawls, 2001, §5.2, p. 14). For Rawls, the fundamental questions of political philosophy, such as specification of the fair terms of social cooperation, are abstracted from features of the social world and idealized "to gain a clear and uncluttered view of a question seen as fundamental by focusing on the more significant elements that we think are most relevant in determining its most appropriate answer" (Rawls, 2001, §2.3, p. 8).

In contrast, the proposed theory of ethics aligns its definition, boundary, order, and foundation with the systematic tradition of moral philosophy. Deriving the theory from philosophical traditions affords a wellspring of authoritative sources and principled thought. The theory can best be examined, disputed, extended, or rejected through rational inquiry derived from the ancient directions of Plato (2009) on how we ought to live to the modern empathies of Rorty (1989) to think of traditional differences as unimportant when compared with similarities of pain and suffering. Required, then, is the conception of a moral basis—of suitably defined, rational decisions attuned to what must be done and alert to the consequences of those decisions on each of those affected.

2.3.2 Rejection of Platonism. On one hand, Rawls is explicit in his rejection on the metaphysics associated with Platonism. "There is no necessity to invoke theological or metaphysical doctrines," he wrote, to support the theory of justice—"nor to imagine another world that compensates for and corrects the inequalities" of this world. "Conceptions of justice," he concluded, "must be justified by the conditions of our life as we know it or not at all" (Rawls, 1971/1999, §69, p. 398). This is

precisely the position of the proposed theory. Methodologically, however, Rawls used language that suggests an implicit reliance on a metaphoric concept of forms as the privileged way to obtain knowledge. The term "justice as fairness" itself is conceived in simile—a device that Rawls consciously employed: "The name does not mean that the concepts of justice and fairness are the same, any more than the phrase 'poetry as metaphor' means the concepts of poetry and metaphor are the same" (Rawls, 1971/1999, §3, p. 11). "Fairness is realistically utopian," he conceded; in pursuit, we probe the limits of that which is realistically practicable to achieve "democratic perfection" (Rawls, 2001, §5, p.12). While the form of metaphysics is rejected, its function remains embedded in language.

As noted elsewhere (Elliot, 2015), this embedding is equally implicit in the language of educational measurement, from its pragmatic presence in the *Standards* (2014) to its theoretical use in the description of latent variable models that dominate psychometric traditions (Spearman, 1904). In the case of the latter, identification of Platonism is of great significance if we are to use other forms of language to describe assessment research. As Baird and colleagues (2014) have observed, postmodern psychometricians might very well use latent trait models while understanding that the traits themselves may not exist but are, instead, socially constructed (p. 58). If so, then there is a willing suspension of disbelief that is unnecessary. Straightforwardly specifying construct domains and examining correlational and predictive models justifies the conditions under which constructs are behaviorally manifested under social and cognitive conditions. There is no need to postulate utopian domains and hold ourselves accountable when we fail to glimpse that which is hidden from us. This topic is revisited in §3.2.3 and §5.4, but it is important to note here that the absence of correlational and predictive variable models associated with defined models of the writing construct, disaggregated by groups, is related to that very Platonism by which the construct is originally postulated within the system of latent variable modeling. While metaphor is important to its structure, as established in §3.4, the theory of ethics rejects the Platonic concept of forms.

2.3.3 Insistence on localism. In identifying the limits of his theory, Rawls is clear: In his focus on political justice, he declared his commitment to "leave aside questions of local justice" (Rawls, 2001, §5, p. 12). In this insistence, his logic is Platonic: "the limits of the possible are not given by the actual" (Rawls, 2001, §1.4, p. 5). The chain of causal logic justifying insistence on the possible is evident by example. Firms, labor unions, churches, universities, and the family are bound by constraints arising from the principles of justice, but these observed constraints "arise indirectly from just background institutions within which associations and groups exist, and by which the conduct of their members exist" (Rawls, 2001, §4.2, p. 10). By example, Rawls specified equality between men and women in sharing the work of society and, as such, special provisions are needed in family law "so that the burden of bearing, raising, and educating children does not fall more heavily on women, thereby undermining their fair equality of opportunity" (Rawls, 2001, §4.2, p. 11). In determining deep structures for justice as fairness, due attention is not necessarily provided on contextualized perception.

Nevertheless, depending on how childcare is framed (as burden or blessing) and how marriage is conceptualized (as the union of opposite or same-sex partners) determines the structure of family law. As Mossman (1994) demonstrated in her review of a report sponsored by the Canadian Bar Association concerning equality of access to justice in family law,

> "Effective strategies for gender equality require a re-assessment of ends as well as means. The paradigmatic liberal prescription—equal opportunity—is a necessary but never sufficient social objective. The ultimate goal is not simply to ensure women's full participation in organizations that wield social, economic, and political power; it is rather to change the nature of those organizations and the way power is distributed and exercised" (p. 368).

It is precisely this view that is provided by attention to local context and theory-building from the ground up. As the thought experiment in §4 demonstrates, structuring of opportunity depends both on reference to general principles of action and on modification of those principles to fit local contexts. Deep contextualization, as Mossman argued, allows referential frames to shift and facilitates important redistribution of power. In its insistence on localism, the theory of ethics seeks, when appropriate, just such particularized redistribution of power in the service of fairness. Situating fairness as the first principle of writing assessment allows this redistribution to occur as an action concurrent with design and score use—not as an action to be undertaken after the assessment episode is completed and the scores are used. Indeed, the proposed theory approaches assessment as a programmatic activity. While the episodic model of Ruth and Murphy (1988) is invaluable in the study of agents and their actions in complex environments, integrating these episodes within a program of assessment allows fairness to be the aim across time and circumstance.

2.3.4 Identification of the least advantaged. Again, Rawls is explicit: Defining expectations as life-prospects, the "least advantaged are those belonging to the income class with the lowest expectations" (Rawls, 2001, §17.2, p. 59). Further clarification is provided by an extensive note in which he specified that the least advantaged are "never identifiable as men or women, say, or as whites or blacks, or Indians and British." Nor are they identified "by natural or other features (race, gender, nationality, or the like)" (Rawls, 2001, §17.2, note 26, p. 59). For Rawls, wealth is the primary cause of opportunity failure.

The theory of ethics for writing assessment departs from Rawls and holds that there must not be fixed categories for the least advantaged lest agency be denied. There are many reasons that opportunity is frustrated, and so the theory demands disaggregation of assessment scores by sex assignment at birth, gender identity, sexual orientation, race, ethnicity, socioeconomic status (SES), special program enrollment (such as English language learning and physiologic difference), and health perceptions.[4] Explication of categories of disaggregation reveals the necessity of such practice.

Because sexual orientation and gender identity constitute ontological, epistemological, and axiological systems of belief, information about group differences is important to both assessment and instruction. As Camilli (2006) demonstrated, large differences in test scores among different races and ethnicities makes it important to understand if these reported differences are artifacts of the test rather than measures of proficiency. In the case of ethnicity, for example, field test scores analyzed from the Smarter Balanced Assessment Consortium (2014) on assessments of English language arts in the Common Core State Standards Initiative (CCSSI) revealed wide performance differences among Asian, White, Hispanic, and Black students. While SES is an indicator associated with writing performance, as Condon (2012) demonstrated in the case of the relationship between SAT scores and family income (Table 13.1, p. 235), there are other indicators associated with writing ability that may be used to identify the least advantaged according to special program enrollment (Inoue & Poe, 2012).

Information about special program enrollment is closely related to the physiologic sphere of the four-domain model advanced by the theory of ethics in §3.1. Traditionally, special program enrollment is understood as addressing test-takers with disabilities or health-related needs. With the force of law enacted by the ADA Amendments Act of 2008, individual are assured protection against discrimination. Extending this assurance conceptually requires the perspective of disabilities studies research, a theoretical and empirical field of study characterized by "the assumption that disability is a social construct offering the opportunity to investigate important questions, rather than a biological fact or medical deficit" (Wood, Price, Johnson, 2012, p. 1). Theoretically, as Solomon (2012) observed, this perceptual shift means that the social model of disability based on legislation creates only the first step of empathy. As he noted,

> Many disabled people say that the social disapprobation they experience is much more burdensome than the disability from which they suffer, maintaining simultaneously that they suffer only because society treats them badly, and that they have unique experiences that set them apart from the world—and that they are eminently special and in no way different. (p. 32)

Developing what Solomon termed "horizontal identities" (p. 33)—perceptions critical to our larger understanding of differences—is a collective category of information essential to understanding group differences in writing assessment. Unfortunately, empirical investigation of such difference, especially those differences related to neurodiversity (Silberman, 2015), traditionally lies outside the boundary of formal representation of the writing construct. Because providing equivalent assessment conditions does not provide evidence about examinees or allow methodological investigation of the interpretative processes of those with horizontal identities associated with cognitive difference, accommodation alone will not change this shortcoming. Recent advances in accessibility research, however, demonstrate the potential of psychometric innovations associated with disability studies. Using a combined universal design (Rose, Meyer, & Hitchcock, 2005) and evidence-centered design framework (Mislevy, Steinberg, & Almond, 2002), Mislevy and his colleagues demonstrated that fairness can be achieved for students with different cognitive

functions through attention to distinct categories of student abilities (perceptual, expressive, language and symbols, cognitive, executive functioning, and affective) required for successful performance on assessment tasks (Mislevy et al., 2013). Adapting task features through technological assessment delivery, the researchers found, facilitates much-needed self-regulation for students who access constructs differentially. When assessments adopt such approaches to assessment—approaches that are beyond mere modification in considering the cognitively different student as an important source of information—opportunity to learn is advanced for all. As knowledge about the construct domain and its access is advanced, the programs of research can develop opportunity structures for all students (Nussbaum, 2002).

Related to horizontal identity, information regarding health is relevant to assessment design, score interpretation, and information use. Sternberg (2010) made the case plain in reflecting on his study on the effects of intestinal parasitic infections and academic ability in Jamaican children:

> The data showed that the cumulative effect of missing much of what happens in school probably cannot be reversed quickly. Indeed, students in all societies who suffer from health problems, including poor nutrition, or who feel unsafe at home or school, do not have equal chances to succeed. (p. 33)

To corroborate this conclusion, Hoetz (2014) estimated that millions of low-income U.S. residents suffer from parasitic infections such as toxocariasis, toxoplasmosis, congenital cytomegalovirus infection, neurocysticercosis, and West Nile virus infection—neglected infections of poverty associated with diminished cognitive functions, mood disorders, hearing and vision loss, epilepsy, and depression (Table 1, p. 1100). In the case of toxocariasis alone, Hoetz has estimated that 2.8 million African-American individuals are infected. As the case with all background information that stands a proxy to the life trajectories, attention to the continuum of health dispels the meritocratic myth of cruelty that everyone has an equal chance to succeed. While all are created with equal agentic ability, social forces destroy opportunity structures. The nature and extent of this destruction must be understood if the structures, especially the opportunity to learn, are to support the powerful intellective force present in all.

Identification of membership along a continuum of groups is not intended to obviate the racialization processes that have accompanied score use since Spearman (1904) posited his theory of general intelligence. Rather, the spectrum of group categories is intended to avoid logical aporia. As an unresolvable internal contradiction, aporia occurs when identification of a structure undermines the structure itself. In the case of the race, Haswell established the aporia as one in which racism cannot be eliminated without constructing the notion of race, yet the construction of race furthers racism. Because multiple memberships are always under scrutiny, the theory of ethics holds that score disaggregation along a continuum allows aporia to be avoided. This is the perspective held by Poe and Cogan in this *JWA* special issue, in their discussion of intersectionality (§4.3).

To return to the departure of the proposed theory from the theory of justice, examination of groups reveals a long tradition of performance differences that cannot be resolved by identification of economic status, race, or any singular factor. To respond to questions regarding how many identity investigations need be made to assess writing, answers must be made in terms of study design and information use. If such studies appear difficult, that is perhaps because metaphors of learning are not yet aligned with ecological models such as that developed by Bronfenbrenner and Morris (2006) in its emphasis on process, person, context, and time as factors influencing individual development. Lest such model appear more metaphoric than actionable, it is important to recognize that they are driving innovative differential item functioning techniques that allow researchers to highlight processes and forms of influence of the student that are absent from the standard view of item response theory (IRT) discussed in §5.3 (Zumbo, Liu, Wu, Shear, Astivia, & Ark, 2015). If such studies appear overly challenging, that may also be because patterns of racialization identified by Inoue (2015) remain. Continued absence of information needed to identify the least advantaged and expressed abandonment when faced with the complexities of obtaining it are, in reality, patterns of neglect associated with racism.

Accompanied by Institutional Board Review, use of institutional data and survey responses allows examination of group differences. In turn, this examination yields identification of actual performance differences as distinct from differences associated with construct underrepresentation or other forms of construct-irrelevant variance associated with bias, as discussed in §3.2.4. In this *JWA* special issue, Poe and Cogan (§ 4.1) demonstrate the complexities of group specification and argue for meaningful categorization. To provide the kinds of evidence needed to pursue fairness and make claims regarding its achievement, the theory agrees and holds that there must indeed be a commitment to ensuring that the categories used to define groups are meaningful across groups.

3.0. Ethical Theory

Reliance on a moral basis, adherence to realistic conditions, embrace of the local, and identification of multiple causes of opportunity failure are departures from the theory of justice envisioned by Rawls. These departures allow identification of distinctions as well as commonalities between justice as fairness and fairness as the singular force unifying the proposed theory. Attention may now be turned to matters of definition, boundary, order, and foundation to establish the theory.

3.1 Definition: Identification of Fairness
Fairness, not justice, is the first virtue of writing assessment.[5] This claim calls attention to a single aim and does not associate concepts through simile as does Rawls with the formulation of "justice as fairness." As well, attention to fairness alone allows continuity with longstanding measurement, philosophical, and legal traditions. This continuity also allows integration of assessment design (actions pursued under the proposed theory, perhaps by a stakeholder of an assessment program) and advocacy (actions pursued as distinct from the assessment, perhaps undertaken in the pursuit of justice on behalf of key stakeholders). Under the proposed theory, the achievement of fairness is integrated within the assessment context and, when achieved, incorporates advocacy processes within the design.[6] The proposed theory rests on this definition of fairness:

> Fairness in writing assessment is defined as the identification of opportunity structures created through maximum construct representation. Constraint of the writing construct is to be tolerated only to the extent to which benefits are realized for the least advantaged.

The definition is derived from the difference principle of Rawls (2001, §13) and the Social Structure and Anomie Theory of Merton (1938, 1996). For Rawls, the difference principle is a form of distributive justice in which inequalities in wealth and income may be allowed to exist only to the extent to which those benefiting from the inequality contribute to the benefit of the least advantaged (2001, §18). Unless these benefits are accrued, "the inequalities are not permissible" (p. 64).

While a treatment of Social Structure and Anomie Theory is beyond the scope of the present theory, it is important to account for a sociological theory first posited in 1938 that remains a rich source of explanation to the present day. Reflecting on a theory that had consumed his career much as the theory of justice had consumed the life of Rawls (Pogge, 2007), Merton realized that a society that held opportunity and then systematically denied it was creating an environment of demeaning frustration. As he wrote, "the paradigm [of social structure and anomie theory] centers on the interactions between aspirations for upward mobility being normatively defined as legitimate for all—'the American Dream'—and on structural differentials in the probability of actually realizing those aspirations" (1996, p. 7). The creation of authentic, realizable opportunity structures—means of success demonstrably open to all (Cloward, 1959)—is thus central to the definition of fairness. In the proposed theory of ethics, the identification of opportunity structures and obligation to take positive action to minimize group differences dually extend from task design to consequence anticipation. The creation of the constructed-response task (Bennett, 1993), for example, should allow all student groups to display their knowledge of the specified construct in writing performance; for those students whose scores reveal that they are at a disadvantage because of the task design, opportunity structures are to be created that include a range of support, from resources allocated for individual tutoring to curricular support for teams of instructors.

As offered, the definition has six interrelated advantages. First, and most significant, the definition follows Messick (1989) and welds construct representation to fairness, the first principle of assessment, as discussed in §3.3.1. As a precise expression, the term construct representation is used to define the nomothetic span of writing in its four domains: cognitive (genre, task, audience, writing process, problem solving, information literacy, conventions, metacognition), interpersonal

(collaboration, social networking, leadership, diversity, ethics), intrapersonal (openness, conscientiousness, extraversion, agreeableness, and stability), and physiologic (nerve, attention, and vision capacity).[7] Because writing is a sociocognitive construct, its manifestation also includes consideration of composing environments (digital, print, and blurred), integrative language arts frameworks (writing, reading, speaking, and visualization), and rhetorical conceptualizations (discursive and nondiscursive practices) (White, Elliot, & Peckham, 2015). Linking construct representation to fairness ensures that the assessment will incorporate a rich view of writing. By reducing inequalities associated with narrow construct representation and eliminating construct irrelevant variance (Haladyna & Downing, 2004), researchers ensure that robust construct representation is understood as moral imperative to promote opportunity (Condon, 2013; Elliot, Deess, Rudniy, & Joshi, 2012; Inoue & Poe, 2012). When resources are directed to the least advantaged, identified in §2.3.4, it must be guaranteed that subsequent instruction is not to be based solely on the assessment, which could be designed according to a narrow view of the writing construct. Rather, instruction should be based on maximum construct representation. This guarantee lessens the possibility of teaching to tests that incorporate narrow views of the writing construct and ensures that those disadvantaged by such construct representation will not be further victimized by constrained pedagogical practices.

Second, the definition further positions construct underrepresentation as the enemy of valid assessment, a position taken by the measurement (AERA, APA, & NCME, 2014; Kane, 2013; Messick, 1980) and writing communities (CCCC, 2009; White, Elliot, & Peckham, 2015). Construct representation is thus foregrounded as an essential component of fairness, and lack of robust representation is subsequently positioned as potentially contributing to the absence of fairness. However, even in cases of robust construct representation, certain groups may still be disadvantaged in relation to the construct. In such cases fairness demands score disaggregation by group to identify those who are least advantaged so that positive action may be taken to minimize those differences.

Third, validity and reliability are unified within a framework of fairness in which construct representation is understood as an obligation to promote equality and lessen inequalities associated with narrow construct representations (Elliot, Deess, Rudniy, & Joshi, 2012). The definition thus has the potential to make a contribution to an integrated, unified theory of validity—the subject of much analysis in educational measurement (Borsboom, 2005; Markus, 1998; Markus & Borsboom, 2013; Messick, 1989)—by extending the focus of assessment toward a specific aim. In other words, for writing assessment purposes, validity and reliability become vehicles for the creation of opportunity structures. In cases where these vehicles are compromised by circumstance, there is assurance that benefits will be provided for those whose scores reveal that they are among the least advantaged.

Fourth, the definition acknowledges that cost-benefit analysis is a part of all assessment. Construct underrepresentation resulting from limited resources will remain part of U.S. education for the foreseeable future. The risks of these assessments are well-known. As Kane (2006) has observed in the case of language arts assessments, "The performances involved in answering questions based on short passages under rigid time limits are legitimate examples of literacy, but they constitute a very narrow slice of the target domain for literacy" (p. 31). As a premise, the theory holds that only an informed instructor, watching a student develop over time, can hope to make a valid claim about writing ability of that student (White, Elliot, & Peckham, 2015). As an aim, the definition of fairness and the theory supporting it is designed to decrease assessments in high-stakes situations and to result in increased use of course grades and other longitudinal measures. Beyond such an aim intended to restore faith in instructor judgment, the theory takes a further radical turn based on economic imperative: Constraint of the writing construct must result in resource allocation to the least advantaged. The formulation has, in turn, two consequences: If benefits cannot be guaranteed in advance of the assessment, it should not be undertaken—a requirement intended to end the assessment-saturated environment discussed in §4.3.2. When such benefits are realized, such as funding for additional instruction for under-performing students, then resources will be encumbered in advance of issued scores to lessen their impact and leverage future equity. These encumbrances are understood as demonstrable measures of stakeholder commitment to opportunity to learn.

Fifth, the direct link between assessment and instruction in terms of resource allocation allows direct and immediate attention to opportunity to learn. As Pullin (2008) indicated, the interrelationships between assessment, learning, and the opportunity to learn are complex, contextual, and evolving. From this perspective, emphasis on the opportunity to learn is both a reflection of the learning environment and a concept demanding articulated connections between the assessment and the instructional environment. For the assessment to proceed, resonance must be demonstrated among the following: the design of the assessment; the opportunity to learn; and the educative intent to improve and continue that learning. This resonance positions score interpretation and use as a vehicle for examining what Gee (2008) identified as the rights of students in terms of opportunity to learn: universal affordances for action, participation, and learning; assurances of experiential ranges; equal access to relevant technologies; emphasis on both information communication and the communities of practice that manage that information; and emphasis on identity, value, content, and characteristic activities associated with language across academic areas. Thus, emphasis on opportunity to learn holds the potential to play an important role in lessening social inequality resulting from writing assessment. Again, in cases where resources are allocated to the least advantaged, subsequent writing pedagogy must be based on a defined, construct-rich model of writing domains, or on a related construct model, that offers the opportunity to learn through maximum construct representation. The default design of a writing assessment program using a single, timed-impromptu task combined with a multiple-choice section would thus be forbidden by the theory on the grounds that robust construct representation is not present and subsequent opportunity to learn is frustrated due to the impoverished initial model. Documented by Sternglass (1997), Hillocks (2002), and the Complete College America Project (2012), long-term impacts associated with such assessments—impacts that no amount of compensation can mitigate—would be understood under the proposed theory as evidence of moral failure.

Finally, the definition acknowledges the importance of evidence-based decision-making, especially as that evidence is used in interpretation and use arguments (Kane, 2013), which are often based on the Toulmin model of persuasion (Mislevy, 2007). In its emphasis on identification of the least advantaged, the definition demands score disaggregation by group membership; therefore, the definition insists that interpretation and use arguments must be made in terms of each group and each proposed score use before interpretations are made and decisions are authorized. Rushed claims and poor decisions resulting from them are thus tempered by sincere pursuit of qualification and falsifiability (Popper, 1962). While beyond the scope of the theory in its present form, further attention might be devoted to modifying the framework for interpretation and use arguments based on the work of evidence scholars (Anderson, Schum, & Twining, 2005). Given the overarching role of narrative in the context of litigation, it might be more accurate to speak of exposition, interpretation, and justification narratives in which argumentation plays but one role among other discourse forms.

3.2 Boundary: Selection of Domain

Following the conclusions of Popper, the proposed theory identifies restrictions that have value because their specification approves certain patterns of reasoning and forbids others. The more a theory specifies and restricts patterns of reasoning, the better for those it may influence.

3.2.1 Theory-building through philosophical traditions.

In §2.2 attention was given to the importance of the key theories of Kantianism, utilitarianism, and social justice. Identification of these theories advances the importance of philosophical theory in writing assessment. While attention to measurement traditions is necessary, philosophically informed development in writing assessment is an important and often under-examined activity, as Huot (2002), Lynne (2004), and Inoue (2015) have shown in their book-length investigations of the role of theory in the specialization.

A preference for philosophically based theory-building in writing assessment has distinct advantages. In that writing assessment is described as a rhetorical act—a deeply situated, language-embedded activity attentive to aim, genre, and audience—attention to the language of assessment becomes very important (Huot, 2002; Yancey, 2012). While beyond the scope of the theory of ethics at the present time, attention to philosophers associated with the linguistic turn in philosophy—Austin (1962), Wittgenstein (1921/2014), and Ryle (1949)—may prove beneficial in analysis of writing assessment documents. As is evident in §2.3.2, philosophical analysis with special attention to rhetorical forms holds the potential to influence assessment within and beyond our field. Philosophical traditions featuring the linguistic turn have deep connections with the rhetorical turn in writing studies, and emerging scholarship featuring extended rhetorical analyses of validity is most promising (Miller, 2016).

3.2.2 Adoption of non-teleological perspective.

As a specification, advocacy of a non-teleological position seems counter-intuitive in a theoretical framework offering definition, boundary, order, and foundations. It may appear as if these structures are provided precisely to guarantee an outcome of fairness. Such is not the

case.

As an account given to justify a purpose, a teleological account is structured, consciously or unconsciously, to justify an outcome. Often associated with determinism, especially in cases of technological manifestations, teleological perspectives identify an end. Conversely, non-teleological accounts provide structures but do not guarantee precise outcomes. Often associated with deontological reasoning, especially in cases of morality, non-teleological accounts identify standards and processes but do not seek to guarantee outcomes. In the present case, the definition, boundary, order, and foundation are offered to advance principled inquiry, as discussed in §3.2.4. As the case study described in §4 illustrates, the extract outcome of the theory is not pre-structured in accordance with adherence to the liberty, freedom, and agency of stakeholders involved in the assessment.

3.2.3 Fairness through empirical research. Although writing studies has a troubled history regarding empiricism, the empirical tradition is a methodological mainstay of writing assessment. The proposed theory follows Beason (2000) in his belief that empirical research and assessments are ethical obligations. These obligations, he wrote, are not undertaken

> to further just our own search for knowledge, nor just to define effective teaching (as important as these are). Empirical research and assessment are required to meet a crucial ethical duty—namely, to help us be informed enough to determine what a campus community considers valuable about composition courses. (p. 113)

As Schendel and O'Neill (1999) correctly observed in quoting this critical observation, it is important to evaluate assessments themselves through "ethical and validity inquiry" to determine the consequences of information gathering and score use for students, teachers, writing programs, and the larger public (p. 221). Such consequences are analyzed by Slomp in this *JWA* special issue. Following such lines of thought, Sánchez (2012) claimed empiricism is both a producer and recipient of theoretical insight. A strategy is needed, he argued, that allows researchers to grapple with "articulating and rearticulating relations between and among components (including oneself) in constantly proliferating and changing systems" (p. 238). The theory of ethics encourages just such engagement in which researchers are obliged to use empirical research techniques for writing assessment so that information is replicable, aggregable, and data supported (Haswell, 2005).

Such a strategy is to be found in the definition, boundary, order, and foundation of the theory of ethics. Understood as the first aim of writing assessment, fairness disallows the value dualism between ethical and validity inquiry. Further, both validity and reliability are unified under the core referential frame of fairness. In its adherence to the non-teleological reasoning identified in §3.2.2, the theory allows a refreshed version of empiricism to emerge. As the thought experiment in §4 illustrates, the task of what Sánchez (2012) termed the new empiricism emerges as the documentation of existential conditions which are inscribed as referential frameworks, not as privileged perspectives on a given context. Following this perspective, the theory adheres to the concept of constructive empiricism of van Fraassen (1980). While scientific realism may be thought of as the pursuit of truth gained through objective reality, constructive empiricism holds that such beliefs need not go, as Rawls put it in his treatment of concepts of justice, "all the way down" (Rawls, 2001, §13. p. 32). That which is empirically adequate—in the case at hand, that which advances the definition, boundary, order, and foundation of fairness—is all that is required. In this way, empiricism is perceived as a shared point of view, a key principle of community identified in §3.4.3. Following Broad and Boyd (2005), this view of empiricism may be described as complementary, "a rhetorical and democratic process for establishing knowledge, value, and meaning" once relegated to logical positivism (p. 10). In the Forum section of this special *JWA* issue, Broad further examines the role of qualitative research in its relationship to the theory of ethics. While special attention is given to quantitative analysis in the present presentation of the theory, it is important to state that no hierarchy is implied in relationships between quantitative and qualitative empirical techniques.

While topics such as empiricism and community may seem yoked by will to the concept of fairness, nothing could be further from the truth. Emphasizing voluntary association between free individuals, Camilli (2006) began his comprehensive review of test fairness with an exposition as philosophical as it is political. With its emphasis on differences in group performance, research on bias deeply informs the empirical requirement of the proposed theory (Cole, 1981). Extending from the qualitative review of item content (Ramsey, 1993) to the qualitative study of differential prediction (Cleary, 1968), examination of test fairness is a mainstay of educational measurement and writing assessment. General linear modeling and statistical tests related to it—from tests of significant difference to multivariate analyses —are based on group comparisons.

Group comparisons are, however, not often linked to the development of variable models. It is here that the theory offers an important direction for writing assessment. For purposes of identification, educational measurement scholars distinguish between two types of models: consensus models and latent variable models (National Research Council, 2012). In a consensus model, experts propose variables that, when combined, provide coverage of a given construct. While the model is expert, it is not necessarily empirical. As a result, the variables may (or may not) be related and they may (or may not) predict an outcome measure at levels of statistical significance. In writing studies, the CCSSI, (National Governor's Association, 2015a), *WPA Outcomes Statement for First Year Composition* (Council of Writing Program Administrators, 2014), and the *Framework for Success in Postsecondary Writing* (Council of Writing Program Administrators, National Council of Teachers of English, & National Writing Project, 2011) are each consensus statements. Conversely, latent variables models are identified through statistically significant correlations among scores from a set of tasks and further studied through factor analysis models, or, more robustly, with multivariate techniques such as structural equation modeling (Abbot & Berninger, 1993) and principal component analysis (Hoffman & Lowitzki, 2005). In the field of personality measurement, intrapersonal factors, for example, have been identified using these methods (McCrae & Costa, 1987.)

The creation of variable models, disaggregated by group, is an essential program of research for writing studies that, at present, has not been addressed. Such preliminary models have been developed for young writers (Abbott & Berninger, 1993), and there are traditions of studies in the cognitive domain (Berninger, 2012), interpersonal domain (Storch, 2005), and intrapersonal domain (Hulleman, Godes, Hendricks, & Harackiewicz, 2010). However, no such models exist that address the writing domains identified in §3.1 and the potential of those domains to predict successful writing performance for the groups identified in §2.3.4. Invaluable as consensus statements are, stakeholders deserve empirically-derived models to justify the inclusion of variables in constructed-response tasks and writing rubrics. The absence of such models is evident in the English Language Arts instructional sequence of the CCSSI, as noted by Applebee (2013) and discussed in §5.4. Their absence is also evident in the disjuncture between high quality field studies documenting the importance of narrative and reflective writing in non-White cultures—such as *Writing from These Roots: Literacy in a Hmong-American Community* (Duffy, 2007) and *Del Otro Lado: Literacy and Migration across the U.S.-Mexico Border* (Meyers, 2014)—and the emphasis in the CCSSI on argumentative writing (National Governor's Association, 2014b). When diverse populations are considered in terms of curricular initiatives and assessment programs, there is little alternative but to agree with Cushman, Juzwik, Macaluso, and Milu (2015) that the aim of decolonization must extend from task design to consequence anticipation. The theory of ethics holds that domain-based models of writing, disaggregated by group, are necessary to draw meaningful inferences about student ability. The absence of these models reveals important directions for research gained from adopting the perspective of *la frontera* (§3), as Cushman explains in this *JWA* special issue.

A program of research leading to such models is easily imagined. The first step in the definition of such models would be to design a project based on constructive empiricism in which variable models are not manifestations of the Platonic assumptions discussed in §2.3.2— a rejection in accordance with the belief of Borsboom that metaphysics should be abandoned (2006, p. 84). The term "latent" would therefore be discarded, although the statistical modeling techniques themselves might well remain largely unchanged (Finch & French, 2015). As a critical part of this first step, findings from the qualitative empirical tradition in writing studies (e.g., Berninger, 2012; Flower, 1994; Meyers, 2014; Rose, 1984/2009; Sternglass, 1997) would be used to establish the span of the initial construct model. Following the observation of Poe and Cogan in this *JWA* special issue that homogeneity and stability are not to be assumed (§ 1), the second step, perhaps the most crucial one, would be to plan to disaggregate the variables in terms of meaningful categorization of group performance. This step would assure that the variables are not latent or universal but, rather, drawn from the constructed-response tasks and therefore an expression of variability according to specific tasks, genres, and groups. The third step would be to ensure that the elicited samples are robust enough so that researchers will have faith that the variables are manifested in constructed- response tasks that provide more than a narrow slice of the target domain. Following the fourth step of empirical analysis through structural equation techniques and principal component analysis, these models could then be used to sequence curricula and design tasks that truly serve students by structuring opportunities that allow them to

demonstrate literate acts. Here is an important program of research for writing studies, and writing assessment can play a significant role in the design and implementation of such studies.

The thought experiment described in §4 could, with some planning, result in the creation of such a variable model that would be extremely useful in assessment and curricular design. Nevertheless, while the circumstance of an assessment such as that used in the case study provides a research path, overreliance on exigence as the impetus for inquiry produces unnecessary limiting temporality. With fairness as an extended aim, research is given the enduring scope identified in §1.2 and thus becomes a programmatic pursuit not bound by demand.

3.2.4 Research through principled action. As a specification, the definition, boundary, order, and foundations of fairness are taken as evidence of principled action. That is, outcomes resulting from application of the proposed theory may be traced back to articulated assumptions, limits, processes, and beliefs. While educational measurement has focused on actions centered around interpretation and use arguments (Kane, 2013) and their instantiation in evidence-centered design (ECD) (Mislevy, Steinberg, & Almond, 2002), writing assessment has not historically proposed a similar series of referential frames that can be used to design assessment and to provide exposition, interpretation, justification, and use of information from them. Only recently has the writing community offered Design for Assessment (DFA) and its categories of evidence (White, Elliot, & Peckham, 2015) and the Integrated Design and Appraisal Framework (IDAF) (Slomp, in this *JWA* special issue) with its emphasis on sociocultural perspectives, access, opportunity to learn, maximum construct representation, data disaggregation, and justice.

Requirements for principled action extend beyond embrace of constructive empiricism and include research methods themselves. The need for such methods is illustrated by O'Neill's response to threshold concepts in writing studies (Adler-Kassner & Wardle, 2015). In concretizing concepts necessary for participation in the field, colleagues developed thirty-seven definitions that may be said to constitute the first statement of a body of knowledge for writing studies. The publication of these concepts may be understood as fulfillment of the prescient observation of Condon (2011) that a tipping point is at hand. In her analysis of the usefulness of these concepts to the specialization of writing assessment, however, O'Neill (2015) reached a sobering conclusion: "While writing studies' threshold concepts are central to understanding writing assessment, they are not sufficient to such understanding because writing assessment lies at the intersection of threshold concepts specific to writing studies and those specific to educational assessment" (p. 158). The solution she proposed is multidisciplinary awareness: Writing studies professionals must understand critical concepts such as validity and reliability; conversely, educational measurement specialists must understand threshold concepts in writing studies.

The theory of ethics supports and extends this position. Foundational principles of educational measurement such as those found in the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 2014) are necessary for research in writing studies. Of equal importance are the threshold concepts identified in *Naming What We Know* (Adler-Kassner & Wardle, 2015). Between these two sides of the triptych—standards and concepts—the missing panel contains guidelines and methods.

The reporting guidelines for writing assessment identified in White, Elliot, & Peckham (2015, Figure 4.1, p. 135) are, in fact, proposed standards for reporting research. Here is found part of the missing integrative frame. Information regarding study aim and design, comparative studies, sampling plan design, null hypothesis statements and rejection principles, and criterion variable identification are basic reporting requirements prerequisite to any consideration of information used to make inferences about student writing ability. Further, these guidelines are intended to bind regardless of assessment classification as large-scale or local. Examples of such reporting guidelines in writing studies, it may credibly be argued, find their origin in the large-scale state assessments in California (e.g., White, 1973) and New Jersey (Hecht, 1980) in the 1970s. Notably, these early state-based assessments were sensitive to racial disparities associated with construct representation (White & Thomas, 1981). Contemporary models exist on a continuum from digital writing assessments designed for large-scale use (Deane et al., 2015) to directed self-placement designed for a specific institutional site (Gere, Aull, Green, & Porter, 2010). The proposed theory holds that unless these guidelines have been addressed and documented in reports similar to these exemplar models, all subsequent inferences and uses of information are invalid. As is the case with construct representation, adherence to these essential requirements is a moral imperative. Because such reporting requires justification of the writing construct under examination in the assessment, the existence of the structure ensures that challenges to robust construct representation are met and that opportunity to learn is advanced.

These basic requirements are, in fact, an artifact of research methods known but not widely adopted in field of writing studies and, regrettably, not always present in the specialization of writing assessment. An example drawn from the work of Poe and Cogan in this *JWA* special issue supplies more of the missing integrative frame and illustrates the necessity of method (§3). In their example of writing placement, assessment results were examined according to the four-fifth rule and the rule of practical significance. Results demonstrated the presence of disparate impact for African-American students. Their study identifies another equally compelling analysis developed by Cleary (1968), briefly noted above in §3.2.3. In essence, this analysis requires that an assessment such as the writing placement examination must have equal predictive validity for all groups. "If the intercepts are not equal," Cleary wrote, "then consistent nonzero errors of prediction are being made within the groups and the test must be considered biased by the definitions of [her] study" (p. 118). In other words, in a regression analysis between the examination (the independent variable) and the course grade (the dependent variable), all groups must have the same regression line. If the intercepts differ, as Camilli (2006) noted, "the phrase differential predication" accurately conveys the unfairness of the selection procedure (p. 231). While it is important to realize that the Cleary model cannot reveal what is wrong with the examination (Meade & Tonidanel, 2010), evidence of differential prediction is a critical juncture demanding that further analysis will be needed before scores can be used.

The Poe and Cogan illustration leads to a conclusion regarding the need for principled action advanced by the theory of ethics: If differential prediction is identified through regression analysis, scores from the writing assessment genre under examination should not be used to draw any inferences whatsoever about student ability. [8] Whether the examination is purchased or local, whether the construct representation is narrow or robust, the same consideration of fairness embodied in the Cleary regression model—as well as other models related to (Darlington, 1976) and distinct from (Novick & Peterson, 1976; Sawyer, Cole, & Cole, 1976) it—holds across all writing assessment settings and therefore must be applied before score use can be considered.

The theory of ethics is designed to advance precisely these techniques of analysis as the first order of analysis. Far from a mere statistical technique, attention to bias in selection—and the family of techniques used to identify occasions for differential impact—are core to the theory of ethics. Without these methods of systematic analysis—and guidelines for action attendant to them—specialists in writing studies and researchers in writing assessment will be unable to advance the theory quantitatively. Diminishing in force, its concepts are rendered more theoretical then actionable. With these methods in hand, however, writing researchers will be able to conduct inquiry along the lines of the specific and restricted patterns of reasoning embodied in the theory.

3.2.5 Rejection of value dualism. As a restriction, the theory of ethics holds that disjunctive pairs in which terms are seen as oppositional rather than complementary are to be rejected. With a firm conceptual history rooted in political philosophy (Hirschmann, 1992), environmental ethics (Warren, 1990), writing studies (Berthoff, 1990), and writing assessment (Leydens & Olds, 2012), the identification and censure of value dualism allows attention to the rhetorical formulation of the assessment, as noted in §3.2.1, with special attention to divisive formulations. As such, validity and reliability are not seen as potentially disjunctive pairs but, rather, as empirical contributions serving the pursuit of fairness within an integrative, principled framework. Inferences used to support validity and reliability claims, therefore, cannot be judged as adequate until the fairness of the assessment is judged as adequate. In terms of the dramatistic theory of Burke (1945), all actions, agents, agencies, scenes, and purposes are unified under a single aim of fairness in which the logic remains but is limited regarding tensions produced by value dualism. Accordingly, all ratios are viewed in terms of their relationship to fairness.

3.3 Order: Structure of Process

As is the case with the theory of justice (Rawls, 1971/1999, §8), intuition is very limited in the theory of ethics. It cannot be assumed, even under the conditions of community described in §3.4.3, that stakeholders of an assessment will have intuitive assumptions and judgments similar to each other; it is best to assume, in fact,

that they do not because they will hold different positions in society. This limiting factor can be somewhat compensated for by the principled actions following the specified definition, boundary, order, and foundation of the proposed theory. Yet these are categories—of knowledge and of evidence—and they do not necessarily include processes. The proposed theory includes but one process.

3.3.1 *Primus inter pares*. To address the issue of process in a theory of justice, Rawls (1971/1999, §8) used the term lexicographical order, a procedural qualification requiring satisfaction of one principle before a second is satisfied, satisfaction of the second before the third, and so forth. This serial ordering, he correctly claimed, avoids the issue of balance because those earlier in the ordering have absolute rank. The history of moral philosophy has many examples, as he illustrated, of moral worth as lexically prior to non-moral values. The selection of the first principle as a moral one is justified by tradition.

In practice, the selection of fairness as *primus inter pares* in considerations of validity and reliability is justified for three reasons. First, fairness as the first among equally important foundational principles identified in §3.4 provides a focused aim to the rhetorical formulation of the assessment process. Conceptualizing any event as a rhetorical situation formed by exigence, audience, and constraint, as Bitzer (1968) proposed, is essential to understand any writing assessment program and the episodes within it. Advancing aim as the reason the rhetorical situation exists in the first place directs the occasion, participants, and boundaries of the assessment. Second, advancing fairness provides a much-needed integration to the Trinitarian foundational principles of validity, reliability, and fairness advanced by the educational measurement community (AERA, APA, & NCME, 2014). While the search for a unified theory of validity has yet to be realized (Markus; 1998; Messick, 1989; Newton, 2012), unifying assessment under a single aim of fairness contributes to such efforts. Barriers to this unifying function nevertheless remain, as identified in §5.3. Third, advancing fairness as the first principle contributes to both principled inquiry (identified in §3.2.4) and non-teleological structure (identified in §3.2.2), thus lending specificity to the boundaries of the theory.

It must, however, be noted that the theory adheres to the rejection of value dualism while it does not concurrently demand absence of value hierarchy. This departure is in contradiction to the scholars identified in §3.2.5, as well to scholars of standpoint feminism (Intemann, 2010). While value dualism and value hierarchy are concurrently advanced as strategies to prevent hegemonic standpoints, adherence to displacing value hierarchy is less required when fairness is understood as the first virtue of writing assessment. Empirical evidence of the promise of fairness as *primus inter pares* is evident in Slomp, Corrigan, & Sugimoto (2014) in their model of collecting and evaluating consequential validity evidence. By inserting sources of consequential evidence in the assessment design process itself, these scholars demonstrated the range of information that can be collected within a given assessment program. An extension of that model is offered by Slomp in this *JWA* special issue.

3.3.2 Implications of fairness first. Since fairness is truly and deeply intended to be the first principle of writing assessment—one that is overarching and integrative but nevertheless primary—then it is possible that concepts of validity and reliability will shift. While such changes are beyond the scope of the theory in its present form, two observations may be made at this time regarding the integrative impact of fairness on validity and reliability.

In terms of validity, definitions remain contested. Kane (2015) derived his definition from Cronbach (1971) and Messick (1989) and conceptualized validity as "the extent to which the proposed interpretations and uses of test scores are justified. The justification requires conceptual analysis of the coherence and completeness of the claims and empirical analyses of the inferences and assumptions inherent in the claims" (p. 1). As such, it is not the assessment that is validated; rather, it is the interpretations and uses that are validated. To achieve clear statements of these uses, Kane (2013) has advanced the idea of interpretation and use arguments discussed in §3.1. Finding this system overly complicated and excessively dependent on argument (Borsboom & Markus, 2013), Borsboom (2005) claimed that validity is a property of tests themselves: "A valid test can convey the effect of variation in the attribute we intend to measure" (p. 162). Since causal relationships, not correlational ones, are evidence of validity, "a test is valid for measuring an attribute if variation in the attribute causes variation in the test scores" (p. 163). Invoking a technical sense of validity, Newton (2012) has provided yet another definition: "[A]n assessment-based decision-making procedure is valid if the argument for interpreting assessment outcomes (under stated conditions and in terms of stated conclusions) as measures of the attribute entailed by the decision is sufficiently strong" (p. 1). Whether we take our definition of validity from Kane, Borsboom, or Newton, it is clear that fairness as defined in §3.1 is absent from consideration. The process of evaluating the plausibility of proposed interpretations and uses of test scores, or the process of identifying variation, seems without aim aside from reifying the presence of validity. However, were either theory re-defined—as evaluating the plausibility of proposed interpretations and uses of test scores to ensure fair treatment of examinees, or as identifying variation to identify performance gaps and leverage resources to ensure opportunity to learn—then there would be an aim beyond solipsism. Fairness would thus be defined as validity for all (Newton, personal communication, July 15, 2015), with interpretation and use arguments holding for each individual (Mislevy, personal communication, January 26, 2016).

In terms of reliability, Haertel (2006) defined this foundational concept as concerned "with how the scores resulting from a measurement procedure would be expected to vary across replications of that procedure" (p. 65). At the present writing, Generalizability Theory (G theory, Brennan, 2001) has provided the best, most nuanced framework for reliability, complete with conceptual and statistical tools for analysis. In terms of standardization, G theory encourages consequential analysis and encourages investigators to consider different study designs and assumptions about observations (p. 14). Accepting that invitation, Kane (2011) examined the implications of standardization and warned that it will not alone control sources of random error associated with reliability. As is the case with validity, however, aim is lacking in reliability theory. If controlling sources of error is the aim of reliability—assurances, for instance, that tasks are stable across administrations and that writing samples have been read with consensus and consistency—then this control could be directed toward fairness. If fairness were the aim, then score use and disaggregation by group would be a primary consideration. Standards of reliability would then change, with fewer standard gauge reliability demands placed on use of scores for formative or program assessments (Elliot, Rupp, Williamson, 2015). In cases where summative assessments demanded higher levels of reliability, tradeoffs with validity would be balanced in terms of fairness as the assessment aim.

Under the aim of fairness, reliability would thus be defined as attention paid to variation associated with assessment aim and score disaggregation by group. Disaggregation by group advances the role of background information of individuals as a proxy for differences in their experience that would, in turn, broaden the horizon of score interpretation. Conversely, failure to disaggregate scores and interpret information related to such disaggregation would be understood to be a source of unfairness (Mislevy, 2015). Beyond the identification of error presently associated with reliability, such variations would be analyzed as occasions to enhance the opportunity to learn. These and other related proposals to reconsider present psychometric models are revisited in §5.3 in terms of challenges to the theory.

3.4 Foundation: Perspective of Inquiry
Metaphors provide both structure and disposition (Gentner & Grudin, 1985; Kahneman, 2011; Lakoff & Johnson, 1980; Leary, 1990). Following the constructivist boundary identified in §3.2.3, the theory holds that metaphors structure cognition, occasion invention, and create knowledge stabilization. Because writing assessment is a rhetorical act of occasion, audience, and constraint (Huot, 2002; Yancey, 1999), the language of assessment—and the theory that informs it—must be subject to careful scrutiny.

Rather than define sources of evidence, as have the authors of the *Standards for Educational and Psychological Measurement* (AREA, APA, & NCME, 2014, pp. 13-22), the proposed theory turns to the metaphor of vision derived through four interrelated perspectives that, taken together, improve the attainment of fairness. Achieved by adherence to the definition offered in §3.1, the boundaries set in §3.2, and the process established in §3.3, these four perspectives are intended to serve as foundation of principled action.

3.4.1 Consequence. The anticipation of consequence and the achievement of consequence are inextricably related. As White, Elliot, and Peckham (2015) have proposed, anticipation of consequence is important in the design of writing programs and assessment processes in terms of sources of validity evidence (Figure 3.3, p. 85) and assessment design (Figure 5.1, p. 155). Alone, consequence takes four forms: unintended negative, unintended positive, intended negative, and intended

positive. The obligation to identify opportunity structures created through maximum construct representation provides an aim to the four consequences. That is, there are no longer simply outcomes to be identified; rather, they are moral obligations to be pursued; unintended and intended negative consequences to be eliminated; and unintended positive and intended positive consequences to be achieved. Proactively constructing a theory of change and theory of implementation, as proposed in the thought experiment of §4, thus becomes obligatory.

It has been compellingly argued, for example, that disparate impact accompanies some standardized assessments (Kidder & Rosner, 2002). Although not universally accepted, this consequence is often recognized, and compensatory effort is made to justify the assessment impact (Bridgeman, Pollack, & Burton, 2008). Under the proposed theory, such an assessment would not be allowed in a specific institution if it could not be shown that it did not, in itself, allow the identification of opportunity structure; nor would the assessment be allowed if compensation, not advanced opportunity, was the tacit environment of the assessment. Attention to consequence is not to say, of course, that standardized assessments necessarily have adverse consequences. As Bridgeman (2014) showed, systematic identification of threats to the validity of standards-based entrance assessments can result in sound admission decisions. Following §3.4.3, local communities must make these decisions based on principled analysis.

It is also important to establish that attention to consequence does not necessarily mean that failure can be eliminated. As Inoue (2014) has shown, theory-building focused on quality-failure (associated with judging the quality of drafts) and labor-failure (associated with estimating work and effort) yields better consequences for learners. As he proposed, productive failure—signaling opportunities to learn discussed in §3.1— is an important ethical standpoint. The proposed theory attends only to the identification of opportunity structure; there are many reasons that such structures cannot be achieved even when identified. Stakeholders need to be alert to the fact that that failure is a complex phenomenon that cannot necessarily be eliminated. Multiple means of support, from instructor consultation to advisement recommendations, will be needed to reconceptualize failure as a form of the shared outcome of opportunity to learn.

As Rawls (2001) observed, a well-ordered social structure is enhanced by the identification of a shared outcome:

> For whenever there is a shared final end, the achievement of which calls on the cooperation of many, the good realized is social: it is realized through citizens' joint activity in mutual dependence on the appropriate actions being taken by others. (§60.3, p. 201)

As contributing to the realized good of opportunity structure, identification of consequence, framed in terms of fairness, thus becomes integral to contractual, communal, and economic foundations.

3.4.2 Contractarianism. Broadly conceived, a constitutional regime exists when the authority to determine social policy resides in a body of individuals representative of the larger electorate, limited by terms, and accountable to that same electorate (Rawls, 1971/1999, §36). In his restatement of the theory of justice, Rawls (2001) claimed that the theory cannot be used in certain kinds of social systems, such as laissez-faire capitalism, because the aim of economic efficiency and growth rejects principles of fairness and equal political liberties (§41.2). As is the case with the theory of justice, the proposed theory relies on established social contracts resulting from constitutional democracy. As noted in §5.1, the theory will not work in the absence of constitutional and legal restrictions in which individuals are at liberty to take action (or not to take action) under conditions intended to protect those actions from interference (Rawls, 1971/1999, §32).

In local settings, these constitutional rights of liberty are protected by key stakeholder groups. In post-secondary institutions, academic governance is protected by faculty and student senates, as well as by human resource policies and codes of student conduct. Under the proposed theory, these governance forms establish a constitutional system. These systems, under the theory, are then obliged to undertake review of all present and proposed forms of writing assessment under the identified definition, boundary, order, and foundation. When necessary, the governance documents should be altered to reflect changes in policy resulting from the review.

For instance, if a writing assessment program is in place, stakeholders would review the program for its unintended negative, unintended positive, intended negative, and intended positive consequences, including the presence of disparate impact on key groups. While that review process is traditionally undertaken by members of the faculty working with institutional researchers, the process would be opened to include all internal stakeholders (students, instructors, writing program administrator, departmental and senior administration) and external stakeholders (state and regional high schools, state and regional community colleges, workplace advisory boards, and professional organizations) identified in Table 1 (§4). This review would be informed by research documenting that 1.7 million beginning students are placed into remediation annually, most of whom will not graduate (Complete College America, 2012). In light of such information and its applicability to the local setting, stakeholders might then begin discussion of alternatives to remediation in which the admitted student is viewed as a qualified student and resources are reallocated to corequisite courses, smaller class size, and additional tutorial support (Complete College America, 2015). As non-credit classes are eliminated under the principle of fairness, new instructional contexts may emerge as contributing to fairness under models of writing attentive to robust construct representation.

Under such conditions, governance documents are not merely indications of constitutional premises. They become subject to review to incorporate enlivened perspectives resulting from the inclusion of stakeholders alert to the adverse consequences of writing assessment.

3.4.3 Community. Inoue (2005) has articulated the value of community-based assessment pedagogy in which criterion-based outcomes are collaboratively developed with students and grades are the result of extended conferencing. Because student writers are integrally involved in the assessment process, the resulting outcome is associated with collaborative activity. Theoretically, such an assessment is informed by the concept of critical consciousness (Freire, 1970). In applying this concept to writing studies, Villanueva (1993) explained,

> critical consciousness is the recognition that society contains social, political, and economic conditions which are at odds with the individual will to freedom. When that recognition is given voice, and a decision is made to do something about the contradiction between the individual and society's workings against individual freedom, even if the action is no more than critical reflection, there is *praxis*. (p. 54)

Related to critical consciousness, the idea of an original position of reasoning and the pursuit of reasonable overlapping consensus both contribute to an understanding of community. For Rawls, postulation of an original position was the "initial status quo" adopted to insure that fundamental agreements were fair (Rawls, 1971/1999, §4, p. 15). In the original position, parties assembled to reach a decision are

> not allowed to know the social positions or the particular comprehensive doctrines of the person they represent. They also do not know persons' race and ethnic group, sex, or various native endowments such as strength and intelligence, all within the normal range. We express these limits on information figuratively by saying that the parties are behind a veil of ignorance. (Rawls, 2001, §6, p. 15)

As an elimination of bargaining advantages, the original position and the descending veil of ignorance are prerequisite to producing fair agreement.

The theory of ethics extends the imagined context for deliberation into actuality by requiring that the design, execution, interpretation, and use of scores be conducted by a range of internal and external stakeholders. While it may be difficult, if not impossible, to create an original position by bracketing out the positions and doctrines of stakeholders of writing assessment, it is quite possible to gather stakeholders together to engage in shared, networked decision-making (Gallagher, 2011). Specifying stakeholders relieves the burden of the original position as a theoretical construct. While a preliminary model is found in Elliot, Briller, and Joshi (2007) that includes broad-based instructor support, under the proposed theory the stakeholder group is substantially expanded.

Following the conclusion of Tinder (1980) that community is unattainable when conceptualized as an ideal, the present theory holds that community is formed under conditions of directed action aimed at the identification of truth. "Truth is that which links human beings when they rise above confusion and dishonesty," Tinder wrote, "and in that way it is the substance of community" (p. 35). Inquiry then becomes the unifying element in community formation. As discussed in §3.2.3, empirical adequacy, not absolute truth, is the aim of fairness. Because we inquire "haltingly and ineffectively," as Tinder acknowledged, it is focused inquiry, not the achievement of truth, that is the essence of community (p. 36). For the English profession, Haswell and Haswell (2010, 2015) have offered a comprehensive theory of community featuring the potentiality and singularity of writers and the hospitable environment that nurture them. This theory stands as one in which the scores from assessment could be trusted because the context from which they issued was intellectual, transformative, and compassionate.

3.4.4 Economics. Attention to market forces and use of economic analyses are both important to the fourth foundational principle. Inclusion of economic arrangements in the theory will, no doubt, be seen as controversial. In his call for a reconceptualization of writing studies, Horner (2015) identified the occlusion of the field by economic forces: "Commodifications of knowledge and learning are substituted for the ongoing work of knowing and learning, and dispositions of flexibility in keeping with fast capitalism dictates are pursued as ideals" (p. 452). As noted in §3.4.4, the impact of laissez-faire capitalism is incongruent with both the theory of justice and the theory proposed here.

Nevertheless, programs of writing assessment exist within conditions of market arrangements, and ignoring such arrangements—or condemning them outright—isolates the proposed ethical theory from economic theory. Such isolation is detrimental to the community formation necessary for the theory, especially in cases where external stakeholders may be drawn from workplace advisory boards in which economic forces play an important role in decision-making. Acknowledgement of economic realities augments the importance of thought experiments of the kind described in §4. A thought experiment in which designers imagine what might be accomplished in a given assessment with unlimited time, money, and expertise is a useful exercise because it helps stakeholders identify categories of inevitable design compromises, such as constrained construct representation, that are made in a finite world (Mislevy, 2015). Identification of such compromises informs future score interpretation and facilitates resource allocation, under the principle of fairness, for the opportunity to learn.

Characterizing the U.S. economic enterprise as a static instance of resource allocation under conditions of scarcity fails to capture current critique and future conceptual possibilities. Harvey (2010) has been clear in identifying the current crisis of American capitalism due to its assumptions of 3 percent annual normal growth. His language is compelling. Driven by an irrational belief in market forces that will result in the achieved annual growth, remorseless maintenance of present conditions in which surplus labor and surplus capital exist side by side results in conditions in which "human lives are disrupted and even physically destroyed, whole careers and lifetime achievements are put in jeopardy, deeply held beliefs are challenged, psyches wounded and respect for human dignity is cast aside" (p. 215). The fiction of annual compound growth has also been demonstrated historically by Piketty (2014). The central contradiction of capitalism, he notes, can be described in the inequality $r > g$—that the private rate of return on capital (r) can be significantly higher for long periods of time than the rate of growth of income and output (g). "The consequences for the long-term dynamics of the wealth distribution are potentially terrifying," he concluded (p. 571). Because there are no simple solutions, a keen understanding of the nature of social change is necessary whether the views of Piketty are held or challenged.

One possibility rests in financial innovation designed to provide incentives of social good. Achieved by vehicles such as social impact bonds, direction of resources to social needs rehabilitates the image of capitalism as a financial arrangement benefitting the few at the expense of the many. From novel policies that help biotechnology firms raise funds for risky clinical trials to human capital contracts that help college students raise tuition dollars without borrowing from banks, Palmer (2015) identified practices of financial innovation that are beneficial because they address needs that are not narrowly perceived as individualistic. Sarasvathy (2008) has examined such practices in relationship to creative and entrepreneurial practices, identifying strategies that allow decisions to be made under conditions of uncertainty. Operating beyond causal models, effectuation sees environments as emerging, constructed, agentic, and venturesome.

As a means to identify social needs and entrepreneurial solutions for the least advantaged, economic engagement becomes an ethical obligation under the proposed theory. Especially suited to meet this obligation is cost-effectiveness analysis (Levin & McEwan, 2001). Defined as "the evaluation of alternatives according to both their costs and their effects with regard to producing some outcome" (p. 10), cost-effectiveness analysis is well suited to writing assessment. In place of the traditional system of cost-benefit analysis with its emphasis on value in monetary terms, cost-effectiveness analysis emphasizes alternatives for meeting effectiveness criteria. In its emphasis on ingredient identification, cost-benefit analysis holds that any statement of the costs of assessment that ignores its value for the improvement of teaching and the advancement of student learning is unacceptable. Further, these benefits must be realized at the level of the individual students. Ecological in conceptualization (§2.3.4), the ingredients method focuses on the identification of costs as well as units of effectiveness, benefit, and utility. Harmonious with the emphasis on economic awareness as a foundational principle of the proposed theory, cost-effectiveness analysis is integral to the assessment process.

<center>**4.0 Thought Experiment**</center>

Demonstration of the theory's power resulting from its definition (§3.1), boundary (§3.2), order (§3.3), and foundation (§3.4) may be illustrated through an examination of action. At its best, theory exposition is accompanied by a demonstration of action. The thought experiment presented in §4.1 is designed as example of what Weiss (1998) has termed a theory of change. As an articulation of the set of beliefs that underlie action, a theory of change should establish the key components of a specific intervention as a way to manifest a desired outcome. By defining processes targeted in an intervention, the theory holds, the possibility of achieving desired outcomes is increased. In order to stress agency, Weiss advised combining implementation theory (McGraw et al., 1996)—emphasizing service and delivery in action—with the theory of change to emphasize how the intervention can be implemented in practice.

In the following thought experiment, elements of change and implementation are included by transforming the four foundational elements of the domain model (§3.4) into interventions. These interventions are then taken as functional—that is, behavioral—obligations of the proposed theory by internal and external stakeholders. The theory departs from Weiss in her belief that it is possible to identify "beliefs that underlie action" (1998, p. 55). Adhering to the logic of the intentional fallacy (Wimsatt & Beardsley, 1946) and the theory of reasoned action (Fishbein & Ajzen, 2010), the theory of ethics rejects the assumptions of Platonism examined in §2.3.2 in favor of identifying action as observed in individual and group behaviors.

4.1 Thought Experiment
As an imaginative way to investigate agency in a specific scene, thought experiments were used by the Presocratic philosophers to conjecture explanations, demonstrate negative reasoning, identify reductionism, introduce skepticism, illustrate axiological reasoning, and analyze value dominance (Rescher, 1991). From the time of Thales of Miletus to the present, thought experiments have been used by philosophers and scientists alike to extend accepted empirical concepts to new conceptualizations and new knowledge. In making his case for the priority of shared practice, Kuhn (1964/1977) used thought experiments to establish the presence of paradigm shifts themselves—innovative changes in the use of analytic procedures that occur when received views are insufficient to explain a given phenomenon. In writing assessment, Poe, Elliot, Cogan, and Nurudeen (2014) used a thought experiment to demonstrate the usefulness of disparate impact analysis in locally-developed writing placement tests; Poe and Cogan continue their use of vehicle in the present *JWA* special issue. Anson (2010) has also used thought experiments based on metaphors arising from reflection on the Möbius strip, a non-orientable surface. As we understand that stakeholders must envision themselves in terms of each other—a premise similar to the original position of Rawls (1999, §15)—we quickly come to the conclusion, Anson wrote, that "no program can improve from the

<center>12</center>

unarticulated and disparate efforts of a confederation of teachers, no matter how strong their individual contributions" (p. 4). Needed are perspectives gained from clarity of purpose and fundamental agreements.

Following the tradition of thought experiments, imagine the assessment design described in Table 1.

Table 1. A Theory of Change

Assessment Design		
Design Principles ↓	Specification ↓	
Aim	To meet regional accreditation requirements by assessing writing program effectiveness. To use this assessment as an occasion to structure opportunity to learn for all students.	
Design	To elicit construct-rich writing samples at end of first year (from students with fewer than 30 semester hours) and end of junior year (between 61-90 semester hours) through a descriptive, base-line design. To score the samples based on a trait model mapped to a specified construct domain. To use information obtained from student logs and corpus linguistics within a digital course management system to study construct-related performance.	
Comparative Studies	To undertake a literature review based on a four-domain construct model of writing: cognitive, interpersonal, intrapersonal, and physiologic. To identify post-secondary research related to longitudinal patterns of writing growth.	
Sampling Plan	To use a stratified random sample to capture and disaggregate meaningful group profiles. To use power analysis yielding a 95 percent confidence interval.	
Null Hypothesis Statements	To propose a series of fairness, validity, and reliability studies assuming that no statistically significant relationships will be found among the writing domains and the criterion measures. In addition, no statistically significant longitudinal improvement will be assumed.	
Rejection Principles	To reject the null hypotheses, confidence levels must reach a 95 percent for the relational and improvement studies. Effect sizes will also be calculated as further evidence of result significance.	
Criterion Variables	To use course grades and scores from purchased tests in comparison to writing sample scores, system logs, and corpus linguistics.	

Ethical Implementation							
	Internal Stakeholders				**External Stakeholders**		
Interventions ↓	Students	Instructors	Writing Program Administrators	Departmental and Senior Administration	State and Regional High Schools	State and Regional Community Colleges	Workplace Advisory Boards	Professional Organizations
1. Consequential	Experiment with simulation before assessment.	Ensure opportunity to learn within the curriculum.	Align curriculum and implement writing construct model.	Ensure appropriate score use.	Coordinate Common Core State Standards Initiative.	Publicize writing model and curricular design.	Advise on significance of writing and career success.	Sponsor development of programmatic research.
2. Contractarian	Gather information about writing domain experiences.	Embed assessment in the curriculum.	Analyze information from digital course management system.	Specify stakeholder rewards.	Investigate dual credit for first-year writing.	Investigate dual credit for upper-division writing.	Emphasize writing continuity.	Support for effective practice.
3. Communal	Communicate significance of regional accreditation.	Strengthen WAC efforts.	Employ diverse instructor knowledge in assessment design.	Coordinate WID efforts.	Share instruction and assessment resources.	Share instructional and assessment duties.	Coordinate workplace genres with curricular design.	Identify cross-institutional collaboration.
4. Economic	Identify incentives.	Reward student success.	Reward instructor success.	Allocate stakeholder resources.	Prepare social good and effectuation studies.	Prepare social good and effectuation studies.	Incentivize by internships and fellowships.	Fund sustained collaborative research.

The setting for the case study is classified as MFT4/S/HTI (Medium full-time four-year, selective, higher transfer-in) under the Carnegie Classification of Institutions of Higher Education™. Comparative fall enrollment information demonstrates that 60–79 percent of undergraduates enroll full-time at these bachelor's degree granting institutions; test scores indicate that these institutions are selective in admissions for first-year students (most of these institutions in roughly the middle two-fifths of baccalaureate institutions); and at least 20 percent of entering undergraduates are transfer students. With an average enrollment of 11,510 students, the institution is one of 114 other such similar institutions. In the aggregate, these institutions enroll 1,312,151 students, or 2.5 percent of the 4,634 institutions of higher education surveyed by the National Center for Education Statistics. As well, the institution is very diverse, with men and women distributed equally, and the following race/ethnicity distributions: White students = 30 percent; Hispanic students = 22 percent; Black students = 20 percent; Asian students = 13 percent; Pacific Islander = 10 percent; and American Indian = 5 percent. With students from over 120 countries, 6 percent of the students are international. Total undergraduate enrollment is made up of 25 percent Pell Grant recipients, a designation that the institution is serving low-income students (Horn, 2006). The campus is known for its LGBTQ community. Six hundred equally diverse tenured and tenure-track faculty members serve the students at this urban, publically controlled, primarily non-residential university. Let's call this institution Garden State University.

Because it is subject to decennial review by its regional accreditation organization, Garden State has decided to investigate the impact of its writing assessment programs at the end of the first year and the end of the third year of undergraduate study—curricular locations of general university requirements for first-year and upper-division writing courses. Although fully accredited, the accreditation visiting team had encouraged the college to continue to develop its outcome assessment efforts following its Spring 2015 visit. Beginning in AY 2016, Garden State administrators determined that a new emphasis on outcomes had to be in place by AY 2018, with the first new assessment ready by AY 2019 and the refined system ready by AY 2020. That way, the administrators reasoned, their five-year interim report to the accreditation agency would demonstrate substantial attention to the assessment of mission-related outcomes, including the claim that students would be effective communicators in a variety of forms for a variety of audiences. As part of the effort to improve the assessment of outcomes, administrators had joined the Voluntary System of Accountability (VSA, 2008), an organization founded to leverage collective evidence and commitment to institutional transparency. As part of the VSA membership, administrators were obliged to use a standardized test purchased from ACT, the Council for Aid to Education, or the Educational Testing Service—three organizations providing tests for the VSA—as part of the comprehensive system of comparative assessment.

Reports of such studies (Arum & Roska, 2011) and critique by members of the writing community (Haswell, 2012) led the Department of English to begin plans to develop a test of its own so that an institutionally appropriate, robust measurement of the writing construct might be available to members of the Garden State University community. Realizing they had a scant three years until the first assessment series would begin—using both the purchased tests and the locally-developed assessment—colleagues began work in the fall of 2016 to have the completed assessment system ready by AY 2020. Their work was informed by the theory of ethics for writing assessment as delineated in this paper.

4.2 Design
Working with the Office of Institutional Research at Garden State, writing specialists in the Department of English knew that the effort to design, implement, and use scores from a locally-developed assessment would take considerable effort. To justify allocation of resources in order to achieve the aim of fairness, the assessment

would need to capture the writing construct in ways demonstrably more robust than the two-hour purchased assessment.

To design an assessment with construct representation sufficient to achieve the aim of fairness, writing specialists agreed that the effort had to incorporate a sociocognitive view of writing to be used throughout the undergraduate curriculum. Adhering to the program of research originating with the social cognitive theory of Flower (1994) meant that both the curriculum and its assessment had to consider writing as a process dependent upon robust construct representation (Duckworth & Yeager, 2015; National Research Council, 2012). As Mislevy and Durán (2014) emphasized, this perspective holds that learning is built around "recurring linguistic, cultural, and substantive patterns" in which individuals, as part of communities, become attuned to recurrent structure—and thus, over time, prove capable of creating meaning through participative structures (p.564). Task design for the assessment therefore had to elicit construct-rich writing that allowed students to deal with complex cognitive patterns, to demonstrate their tenacity with these tasks, and to work collaboratively with others in preparation of submitting their final response to the task. As well, because Garden State students were accustomed to working with *My Reviewers*—a web-based software tool designed to facilitate peer review (Moxley, 2013)—the constructed-response task would be delivered within that platform to allow students extended simulation time for drafting, collaboration, and submission. Designed according to a mode of knowledge construction termed datagogies—the use of digital platforms to develop pedagogical communities (Moxley, 2008)—*My Reviewers* opened an information threshold that included trait-based rubrics and log patterns that could be used to investigate multiple domains of the writing construct. While, for example, trait scores could be used to study the cognitive domain across multiple sections of courses (Moxley, 2013), log patterns could be used to study patterns of swift trust associated with facets of the intrapersonal domain such as motivation (Coppola, Hiltz, & Rotter, 2004). Similarly, analysis of comments could be used to study interpersonal patterns of collaboration of peer review teams (Dixon & Moxley, 2013). When such multiple sources of information were considered, writing specialists decided to expand the interpretation and score use model of Kane (2013) to include many types of information that would require exposition, interpretation, and justification narratives.

Following an evidence-based model for the assessment design (Mislevy, Steinberg, & Almond, 2002), writing specialists planned ways to use the digital platform to score the writing samples through a trait model matched to the writing construct as instantiated in the constructed- response task. Because intrapersonal and interpersonal domains would be challenging to capture without the use of a digital platform designed to produce information, reports from *My Reviewers* would be analyzed for individual and group activity information. Similarly, corpus linguistics would be used to identify patterns of certainty and expertise in the submitted texts. Identified by Aull (2015) as patterns adding to our knowledge of how various groups of writers "construct a balance of certainty and caution," such analysis can lead to better designed writing tasks and linguistically informed pedagogical applications for all student groups (p. 111). Regarding the sampling plan, measurement specialists in the Office of Institutional Research designed a stratified random sample based on group membership of sufficient power to enable comparisons to be made from the scores in the sample to the broader undergraduate population (Maxwell, 2000). Working with writing specialists, measurement specialists also designed an Institutional Review Board approved survey to find more about perceptions of demographics, academic success, and experiences with writing. These questions would allow further information disaggregation of different groups.

Statistically, information from the assessment would enable the writing specialists to make inferences at a 95 percent confidence interval about Garden State students with fewer than 30 semester hours (approximately 1,500 students) and students who had earned between 61 and 90 semester hours (approximately 2,000 students, including approximately 400 transfer students from area community colleges). Because only a limited number of student performances could be scored and care had to be taken with that phase of the assessment, the sampling plan would become very important if trait scores were to yield inter-reader agreement and inter-reader reliability indices congruent with other assessments in which complex responses were evaluated (Collins, Elliot, Klobucar, & Deek, 2013; Kelly-Riley & Elliot, 2014).

4.3 Ethical Implementation Resulting from Process of Change

Traditionally, assessment design is the consuming consideration in thought experiments such as the one described above. From aim to sample size estimates, from construct validity to inter-reader reliability, planning is designed so appropriate use can be made of scores and key inferences can be drawn from them. Appropriately, these interpretation and use arguments are the central focus of the assessment (Kane, 2013). Yet they need not be the sole or terminal focus.

To continue the thought experiment, now imagine that the assessment design is further extended and substantively modified by the focus on fairness presented in §3 above. Using the theory of change discussed in §4, imagine that the design process did not end with the writing specialists and institutional researchers. Instead, imagine that the design were then shared with the key internal and external stakeholder groups identified in Table 1 in the pursuit of fairness. Further, imagine that representatives from each of these stakeholder groups were present in a series of campus meetings in which the validity and reliability aspects of the design were reviewed in terms of fairness. As a result of those meetings, a network structure emerged that allowed the stakeholders to recognize the dangers of hierarchy and thus "provide a larger, legible logic" in which assessment claims can function (Gallagher, 2011, p. 465). How would the conversation change with this focus on fairness?

    4.3.1 Consequential interventions. In terms of consequences related to the planned assessment—recall that, by lexicographic procedure, consequence is always the first consideration—one possible direction would be a renewed focus on opportunity to learn. As Pullin (2008) observed, the most pressing issue facing U.S. education is providing all students with learning opportunities. Distributed unequally as demonstrated by score patterns for groups, equity remains an elusive goal of education. In the case of the assessment at hand for Garden State and its diverse student population, emphasis on opportunity to learn—the extent to which students have been exposed to the writing construct in the planned assessment, as well as their exposure to language experiences required to succeed on the assessment (AERA, APA, & NCME, 2014, p. 221)—would quickly take two forms: curricular alignment and genre determination.

While the assessment design is innovative, the curriculum in the first and junior years would not necessarily be so. Although, for example, writing from sources may be the focus of both cohorts of courses, workplace leaders might question if those sources are to be solely text, or if those sources are to include figures and tables including quantitative information. If the essay were to be proposed as the exclusive reporting structure, other stakeholders—representatives from professional organizations such as the Accreditation Board for Engineering and Technology (ABET)—might wonder aloud if the curriculum supported the design of constructed-response tasks such as proposals. Because ABET's *Criteria for Accrediting Engineering Programs* (2014) provides guidelines for establishing and assessing outcomes—including the ability to communicate effectively—engineering colleagues might wish to sponsor the development of program goals and curricular standards similar to their own. Objections might also be voiced from colleagues in architecture in which painting, photography, and film are important sources. Because their National Architectural Accrediting Board (2014) also provides guidelines—including an emphasis on the importance of studio culture that may resonate well with writing center culture—colleagues in the visual arts may become invaluable participants in sponsoring new genres to be used in the assessment. Colleagues who must respond to their own program accreditation demands would also be interested in extrapolation inferences to be made across disciplinary areas (Kane, 2012). In the alignment of curriculum and genre, an immediate impact of the planned assessment would be the need, creation, and distribution of a nomothetic map of the writing construct encompassing the span of its domain.

In terms of stakeholder needs, curricular alignment would ultimately turn to genre—not merely as a form of writing but, rather, as traditions of artifact production, use, and interpretation, ideologies that shape the very contexts in which they emerge (Gee, 2012; Spinuzzi, 2003, 2015). If the curriculum is based on traditional essays in which students write from print-based sources, a foundational element of the CCSSI that state and regional high school representatives might well address, additional effort would be needed to work with instructors to transform the curriculum to a place in which many kinds of information, including design and interpretation of quantitative material, would be included in a wide range of genres, from including print-based proposals and digitally-designed wiki sites. As well, writing program administrators would want to ensure that the assessment was not so innovative that new genres were created that were not covered in the curriculum. To avoid the creation of an assessment in search of a curriculum, the new assessment could be field-tested with students as an assurance that they had experiences with the digital system and the assessment genre therein (a guarantee of their opportunity to learn incumbent on writing program administrators) and to pre-test that system and genre before the first AY 2018 administration (a guarantee incumbent on departmental and senior administrators that inferences made from scores would be justified). To assure the opportunity to learn for all students, specialists in English language learning and representatives familiar with assessment of students with disabilities would also review the field test to assure equity.

As part of score exposition and use, administrators would want to assure that the same students taking the local assessment would also take one of the standardized tests approved by the VSA so that generalization inferences could be made between assessments and across institutions (Kane, 2012). Because the standardized tests capture a limited version of the construct, departmental and senior administrators would want to investigate score relationships between the local assessment and standardized test in terms of correlations between measures and predictive ability to the criterion measure of course grades. Regarding transfer of writing knowledge across the curriculum and professions incorporated within it, state and regional community colleges would make substantial contributions to curriculum articulation, especially as the assessment impacts the Garden State high transfer population and the ability of those students to succeed in junior level writing courses. Of special importance to all stakeholders, teaching for transfer would thus become an important consequence of the assessment in its relationship to opportunity to learn (National Research Council, 2012; Yancey, Robertson, & Taczak, 2014).

4.3.2 Contractarian interventions. In terms of contractarian issues related to the assessment, consideration of the concerns and cares of students would allow the assessment to be legitimately grounded. For first-year students, concerns would certainly be voiced on the need for the assessment, especially when recent studies demonstrate massive over-testing in the schools (Hart, Casserly, Uzzell, Palacios, Corcoran, & Spurgeon, 2015; Lazarin, 2014). Attention to the cares of first-year students would lead to discussions of the ways that the assessment would frame the writing construct and how that framing would be distinct from other assessments.

Here would arise an opportunity for the study of response processes—examining the students from various groups about their concepts of their previous writing domain experiences, prior conceptualization of the writing construct, and approaches to the planned assessment under simulation conditions discussed above as a vehicle of opportunity to learn. For instructors, establishing a social compact with students based on their responses would lead to meaningful curricular embedding. For example, if the instructors value a domain mode of the writing construct including reflection and metacognition, they would design a response process study to ensure that students were asked about their experiences with these variables of writing. To leverage student motivation, a key variable in writing assessment, instructors would want to consider embedding the assessment in the courses through the process of enculturation discussed in §4.3.4 (Graham & Williams, 2009). Writing program administrators would want to work with the designers of *My Reviewers* to explore what big data patterns can tell us about the interpersonal and intrapersonal domains through analysis of course logs (Junco & Clem, 2013); similarly, corpus analytics techniques could be used to reveal relationships between cognitive and linguistic patterns (Aull & Lancaster, 2014; Dixon & Moxley, 2014). Physiologic functioning that constitutes a barrier to writing performance could also be identified, compliance with disability requirements could be observed (Lewiecki-Wilson & Brueggemann, 2007), and universal design could be adopted to strengthen the assessment for all student groups (Mislevy et al., 2014). Combined with instructor scores on the assessment, analysis of patterns within the digital environment would address the concerns and cares of students by providing new information about their performance that was impossible to gather in print-based testing environments (Shea et al., 2010). Creation of a variable model of writing, disaggregated by student group, would also be an important research aim that would support both the assessment and occasions for structuring opportunity to learn. In terms of rewards, departmental and senior administrators would want to identify ways to reward both internal and external stakeholders for their participation, including in such considerations research support for conference participation and publication of research.

In terms of external stakeholders, both high school and community college representatives would be interested in the role of dual credit for courses that adopt both the curriculum and its embedded assessment. While Advanced Placement is a long-standing way to achieve college credit for high school courses, innovative local assessments such as the one planned for Garden State might be of interest to teachers who wish to explore instruction and assessment in digital environments (DeVoss, Eidman-Aadahl, & Hicks, 2010). Similarly, community college instructors might be interested in helping prepare their students for the junior level course and thus infuse their own curriculum with the writing construct model in the Garden State assessment (Ignash & Townsend, 2000). Workplace stakeholders might encourage students to do their best on the assessment by assuring them that the vision of the writing construct, with its attention to innovative genres and audiences beyond the classroom teacher, is aligned with workplace needs (Pimentel, 2013). Based on the study design, professional organizations such as the National Writing Project might sponsor studies to identify effective instructional and assessment practices occurring as a result of the Garden State project.

4.3.3 Communal interventions. In terms of communal issues related to the assessment, engaging students as key communicators regarding the significance of regional accreditation in general, with specific attention to the writing assessment effort, would be invaluable to all involved. Instead of passive recipients of the assessment, involving students in the earliest planning stages affords enormous benefits, especially in terms of motivation of other students, as discussed in §4.3.2. To strengthen community, instructors could use the assessment as part of Writing Across the Curriculum (WAC) efforts. Since source-based writing for the assessment includes quantitative information, colleagues from Science, Technology, Engineering, and Mathematics (STEM) could be consulted for representative tasks that could possibly be used in classes other than designated writing courses. Building on that momentum, writing program administrators could seek diverse instructor knowledge—in STEM, management, and the health sciences—to build support for writing and its assessment throughout various Garden State undergraduate curricula. In similar fashion, departmental and senior administrators would want to concretize this effort in support of Writing in the Disciplines (WID) as part of mission integration arising around the college's mission-related outcome of effective communicators, in which all students would prove capable of using a variety of forms designed for a variety of audiences. Based on such networked information, assessment designers might very well decide on two writing tasks: one that is traditional and print-based in its use of academic traditions using argument and exposition; and a second that is innovative and digitally-based in its use of documentary traditions employing description and narration. With two distinctly different writing tasks, both important for student success, scores could be disaggregated by group membership and by genre to advance the opportunity to learn.

In their emphasis on community formation, high school stakeholders would be interested in sharing instruction and assessment resources with their Garden State colleagues, especially in the case of digital assessments of writing in the CCSSI (National Governors Association, 2015) and NAEP writing (National Center for Education Statistics, 2012). In both instances, classroom teachers would have valuable experiences with assessment tasks designed to capture blended constructs associated with language arts, thus providing substantial challenges to accompany integrated literacy models (Cumming, 2013). Taking a leadership role, secondary school teachers would provide valuable information on the barriers and benefits of digital assessment.

In turn, community college instructors would provide valuable curriculum and assessment design and might, along with their high school colleagues, support the scoring duties to come. Through cross-institutional modeling of the writing construct, such inclusion would add additional validity to exposition, interpretation, justification, and use of information related to the assessment; as well, such networks would facilitate expansion of the reader pool as a demonstration of inter-reader consensus and consistency associated with the construct model. Workplace stakeholders would enhance community formation through providing guidance on what genres are most used in non-academic settings and where these genres might best be addressed in the undergraduate curriculum—as well as in high school and community college curricula to ensure the school to workplace connections. Professional organizations known for their community-building capacities—the Association of Teachers of Technical Writing, the Council for Programs in Technical and Scientific Communication, the Society for Technical Communication, and the IEEE Professional Communication Society—could help to expand the capacity of the Garden State assessment through cross-institutional collaboration focusing on curriculum design and innovative assessment.

Perhaps an ultimate measure of the success of community interventions in terms of the theory of ethics is the publication and distribution of a document such as *ETS Standards for Quality and Fairness* (2015), analyzed in §2.1.1. Were Garden State University internal and external stakeholders to write, peer-review, and circulate such a detailed statement of beliefs and practices, the transparency required for the theory to succeed would surely be advanced.

4.3.4 Economic interventions. In terms of broad economic issues related to the assessment, a general determination of the resources needed to conduct the assessment could be made through cost-effectiveness analysis of the ingredients of the assessment and their relationship to the desired aim of program assessment (Levin & McEwan, 2001). These ingredients would include allocation of resources to the least advantaged. That is, following the concept of fairness, stakeholders would realize that even the best-designed assessments result in some form of construct underrepresentation. Because the locally-developed assessment and the VSA assessment were both a necessity of accountability, plans would therefore be made in advance for score disaggregation and meaningful group analysis—especially

according to academic major. It is here that the period between AY2015 and AY2020 would be filled with exploratory studies, preliminary identification of the least advantaged, and resource encumbrance for those students. Because the assessments required no cut scores, established processes of standard setting would allow internal stakeholders to establish levels of performance (Cizek & Bunch, 2007) and establish resource allocation for those at specified score levels. The five-year period would thus be used to create a financial model for the assessment.

In terms of individual students, incentives might be identified in terms of reward structures associated with student motivation. As Graham and Williams (2009) recalled, one of the most well-documented findings in attribution research is that expectance is related to the perceived stability of causes (Weiner, 1986). If the cause of the performance is perceived to be malleable, then expectations can improve. Thus, poorly performing students who believe that they did not try hard enough on an assessment can be encouraged to succeed by trying harder because effort is malleable. As students realize that the assessment they helped to design at Garden State offers maximum construct representation, they may become motivated to do their best. Perhaps, some students may reason, they are not poor writers—a stable condition—as other tests may have suggested.

Recognition of malleability might unfortunately lead to barriers when students are then asked to take the standardized test in fulfillment of the VSA—tests that may be associated with previous poor performance. In such an environment, instructors would have to achieve effectuation in novel ways. Because the assessments are embedded in courses, allocating a percent of the final course grade may increase motivation, as would the award of a semester's worth of books or remission of student fees. Degree of motivation might be determined by the log patterns of the locally-developed assessment; determination would be more difficult to assess on the VSA-related assessment where self-reporting might be used. Such reward structures would then be balanced by resource allocation to students who do not perform at expected levels. From additional individual tutoring to writing center support, efforts would be made to help students improve their performance. Were the students to be assessed after such support, similar incentives would again be provided to increase motivation.

Based on the economic incentives identified in §4.3.2, high school and community college stakeholders would want to design their own social good and effectuation studies to identify processes, principles, and logic that would motivate a shared, collaborative relationship with Garden State (Sarasvathy, 2008). Workplace stakeholders might offer internships to students, as well as their instructors, to enculturate connections between writing in academic and non-academic settings. Working within a framework of fairness, student internships would be tailored to individual student groups so that the least advantaged could benefit by these internships along with those who scored at performance levels indicating proficiency. Through such internships, both motivation and performance could be increased equally among all student groups. Professional organizations such as the Conference on College Composition and Communication (2015b) might support such research through seed grants from its research initiative designed to support sustained, substantial, and equity-based investigation among the internal and external stakeholders.

4.4 Thought Experiment Conclusions
Recalling Kant, Rawls (1999) identified the force of moral philosophy. As Rawls wrote, Kant

> begins with the idea that moral principles are the object of rational choice. They define the moral law that men can rationally will to govern their conduct in an ethical commonwealth. Moral philosophy becomes the study of the conception and outcome of a suitably defined rational decision. This idea has immediate consequences. (§40, p. 221)

Under traditional compartmentalized frameworks of validity, reliability, and fairness (AERA, APA, NCME, 2015) little of the case study presented above would fall within the parameters of that which is allowed. Under an ethical framework, however, these imagined actions take on a moral force by structuring opportunity under the guiding principle of fairness. Viewed as interventions, application of the principles of consequence, contract, community, and economy restructures , validity and reliability. No longer foundational categories, they become pursuits of multiple stakeholders in a moral commonwealth in which rational decisions have real and immediate consequences.

Envisioning the conceptual domain model as a series of interventions yields striking differences in the assessment design. Dealing with consequences as the first form of fairness leads to discussion of the opportunity to learn. Such a perspective leads, concurrently, to analysis of both the curriculum and its assessment through the lens of multiple stakeholders. Within this context, the writing task is thus connected to the curriculum; that is, the constructed-response task is responsive both to domain models of the writing construct and to the curriculum within the domain model is embedded. Constructive alignment—an integrated instructional and assessment framework used to map learning activities to outcomes (Biggs & Tang, 2011)—is thus achieved by considering consequence first. Contractarian emphasis leads to investigation of student responses to the assessment. Again, concepts associated with measurement—identification of sources of validity evidence and explicit determination of contract domain—arise in meaningful ways as a result of operationalizing the implicit social compact we have with students in all instructional and assessment encounters. Envisioned as a continuum of care, community formation becomes a process of support that extends from school through college and on to the workplace. Extending beyond metaphor, an ecological view is thus created that rejects barriers and embraces continuity (Bronfenbrenner & Morris, 2006; Inoue, 2015; Slomp, 2012). In place of the narrow vision of economics as the manipulation of demand, processes of effectuation can be used to determine key aspects of the writing construct that are often ignored, such as motivation, and use those aspects to add value to the assessment; in cases where external funding can be used to support research, experimental assessment projects such as those at Garden State University become competitive because of their unified focus in which validity and reliability are framed in terms of fairness.

5.0 Challenges

The theory of ethics for writing assessment has met the seven obligations identified by Slomp in the introduction to this *JWA* special issue: Foundational measurement concepts of validity and reliability have been integrated through a unified, principled, and integrated framework of fairness (§3, especially §3.3.2); an overarching referential frame has been established for the theory through exposition of its origin (§2.1, §2.2, and §2.3), boundary (§3.2), order (§3.3), and foundation (§3.4); a systems orientation has been provided, including a lexicographic stand on investigative processes (§3.3.1); a unifying function has been established that includes a research framework (§3.2.4); stakeholder perspectives have been identified in a non-hierarchical order through attention to opportunity to learn and community formation (§3.4.3); a case study has established at least one genre to suggest the range of the theory across assessment contexts (§4), with further genres examined in this *JWA* special issue regarding school-based assessment for individual students by Slomp and in placement of admitted postsecondary students by Poe and Cogan; and accountability has been established within non-teleological perspective that allows principled action without predetermining exact outcomes (§3.2.2).

Reservations nevertheless remain about the usefulness of the theory. Failure of the basic social compact, overlap with existing systems emphasizing social consequence, unclear relationship with psychometric frameworks, and disjuncture of principled inquiry among professional stakeholders each pose substantial challenges. If there is a future for the theory, that potential may arise from adhering to philosophical values.

5.1 Absence of Constitutional Processes
As a thought experiment, the case study described in §4 is one of optimality in which agents are grounded in good will. If the fundamental social compact is broken, it is difficult to see how the theory would be of use. A thought experiment could easily be imagined in which students are hostile due to previous unwarranted testing and remediation, instructors are uninterested due to unequal pay, writing program administrators are denied pathways to tenure and are powerless, and departmental and senior administrators see no value in the project and prefer the economy of standardized testing as a way to satisfy VSA requirements. Equally imaginable are high school teachers exhausted from district and CCSSI accreditation demands, community college colleagues who have witnessed failed articulation agreements and seen their students denied transfer credits to Garden State University, workplace stakeholders who complain of employee failures on grammar, and professional organizations who fail to see the scalability of the project in the first place. In such a scenario, the theory would perhaps fail when considered as an intervention. What is certain is that scores from the standardized tests associated with the VSA would be the only evidence used to determine student proficiency during the accreditation

process. Without an alternative measure, inferences from those scores would prevail.

5.2 Redundancy with Existing Systems

Because the theory is specifically designed for U.S. writing assessment, extending it to other settings may be unwarranted. In the United Kingdom, where the market economy is regulated to achieve fairness, the history of writing assessment is far different from that in the United States. As Weir, Vidaković, and Galaczi (2013) have noted, the psychometric tradition with its emphasis on multiple-choice tests was absent for the first half of the 20th century in the U.K. (pp. 56-59). Consequently, sociocognitive frameworks featuring consequential validity are already a part of the Cambridge English examinations (Weir, 2005). More broadly, the Council of Europe (2001) has adopted the Common European Framework of Reference for Languages (CEFR-L) featuring assessments developed according to a comprehensive view of competency and criteria for the attainment of objectives. In such well-articulated, cohesive, and transparent environments, it is unclear if the theory would be of use; if usefulness were determined, the theory would probably be used to address, in pragmatic fashion, the levels of doubt in language testing identified by Fulcher (2015).

More broadly, the efficacy of the theory in countries lacking a constitutional government is unknown. As established in §3.4.3, a prerequisite of the theory is its reliance on principles of participation congruent with a representative body, limited in term, and accountable to an electorate (Rawls, 1999, §36). Among the countries contributing to Volume 4 of *The Companion of Language Assessment* (Kunnan, 2014a), it is uncertain that governments in Latin America, the Caribbean, Africa, the Middle East, Asia, and Eastern Europe lacking constitutional governments would be able to advance the theory in meaningful ways. Equally, it is unknown whether communism or theocracy—to take extremes—may advance principles more appropriate to their cultures for the structure of opportunity through assessment. Bruce and Hamp-Lyons (2015) have noted the problems that arise when international assessments such as the CEFR-L are applied across cultural contexts in China, and the resulting political tensions are associated with imperialist tendencies noted by Cushman in this special issue.

5.3 Uncertain Relationship with Psychometric Models

As noted in §3.3.2, advancing fairness as the first principle of assessment does not demonstrate how the concept itself might be aligned with theories of validity and reliability or integrated into techniques such as item response theory (IRT) (van der Maas, Molenaar, Maris, Kievit, & Borsboom, 2011; Yen & Fitzpatrick, 2006) and G theory (Brennan, 2001). These two latter mainstays of contemporary measurement theory are important techniques in assessments such as NAEP Writing (National Center for Education Statistics, 2012). As a mathematical model of the behavioral relationship between performance on a test item, that item's characteristics, and the subject's understanding of the construct under examination (AERA, APA. & NCME, 2014, p. 220), IRT theory is based on the assumption of signs indicating the presence of latent traits (Snow & Lohman, 1989). As Mislevy (2008) has suggested in his analysis of the challenges of cognitive science to educational measurement traditions, the very idea of latency, as well as the stochastic models of probabilistic reasoning that underlie that IRT, are based on a metaphor in which traits are hidden and their presence inferred through mathematical models. In the case of social cognitive modeling, as is the case of fairness as the first principle of assessment, abandonment of the key structural metaphor of latency results in uncertainty. Defined as the framework for evaluating reliability in which sources of error variance are estimated on items and their generalizability beyond the sample at hand are estimated (AERA, APA, & NCME, 2014, p. 219), G theory also relies on the concept of latent variables discussed in §2.3.2. Again, as is the case with the integration of social cognitive models, the presence of fairness in such models is unknown—although Kane (2011) has analyzed the relationship of G theory and standardization in terms of fairness and argued that standardization alone will not control random error. As Mislevy (2007) noted, required is a shift in interpretative frameworks. At the present, it is unknown what this metaphorical and interpretive shift will entail regarding IRT and G theory as used in writing assessment.

5.4 Absence of Evidence-based, Principled Inquiry

As a basic theory, fairness may not prove useful because the field of writing studies itself has not yet consistently adopted either a long-term use of evidence-based information or a principled system of empirically-oriented research methods. A wide range of scholars, from Fulkerson (2005) to Haswell (2005), has noted the absence of a firm empirical basis for writing studies. In response, as noted in §3.2.4, Sánchez (2012) has called for a reimagining of empiricism as a vehicle for theoretical insight, and Bazerman (2013) has turned to the methods of the social sciences in what may be understood as the first comprehensive theory of writing studies to create an empirical account of the material world as it is instantiated in literate action. In the absence of empirically-based information obtained through widely adopted research methods, the theory of ethics will have limited value beyond its face validity.[9]

As discussed in §2.3.2 and §3.2.4, the absence of an empirical tradition in writing studies poses a systematic and substantial challenge. Perhaps the greatest indication of a challenge resulting from absence of a principled theoretical stance is that our field has not developed correlational and predictive variable models associated with defined writing constructs, let alone models that are meaningfully disaggregated by groups. Consequences related to the absence of such methodological programs of research are serious. Aside from the National Assessment of Adult Literacy (Kutner, Greenberg, & Baer, 2005), we know little about how diverse citizens handle the very literacies demanded by a democratic society. We know less about the connections between the academic curriculum and those skills. Consequences of this absence are clearest in the CCSSI and its dependence on a consensus model (based on expert opinion) rather than a correlational and predictive variable model (based on empirical relationships among variables and disaggregated by groups) as the basis for the English language arts objectives. As Applebee (2013) correctly observed, without an empirically derived sequential model, the curriculum devolves into "trivial grade-to-grade progressions" (p. 29).

Our field, however, cannot be held as solely responsible for creation of the variable model disaggregated by group discussed in §3.2.3. Among many, two reasons for the absence of this model are apparent: Large-scale assessments capable of obtaining sufficient sampling plans inclusive of all groups are often seen as invalid in their representation of the writing construct; and such standardized assessments fail to provide sufficient motivation for students so that the scores can be trusted before they are used. It is unclear how the present theory would address such shortcomings in the absence of sustained federal, regional, and state economic investment.

5.5 A Future for the Theory

If there is potential for the ethical theory presented here, it rests in the observation of Kuhn (1962): "Lack of a standard interpretation or of an agreed reduction to rules will not prevent a paradigm from guiding research" (p. 44). Here, Morton (1980) and Kuhn are as one: Theory does not need to be articulated to be useful. As a young, perhaps even emerging field, writing studies in general—and the research involved in writing assessment—may yet find a center through the pursuit of broad ethical dispositions, as Duffy has argued (2015). Perhaps statement and subscription are not needed after all. For Duffy (2014), our challenge is beyond a disposition toward empiricism and rests with the ontology of writing studies itself. As he believes, there is "a greater disciplinary problem: our failure to explain to the general public, to colleagues in other disciplines, to our students, and perhaps even to ourselves what we do, why our work matters, and what is at stake in the teaching of writing" (p. 212). To right our listing ship, Duffy (in press) believes, the teaching of writing should be conceptualized as the teaching of ethical dispositions—honesty, accountability, compassion, and intellectual courage. Such engagement would facilitate a rediscovery of the ethical tradition. To reach such an end, I would gladly welcome generalization revisions to the theory with the intention of increasing the proportion of true to false assertions about a future for writing studies informed by philosophical traditions of ethics and the pursuit of virtue. That direction would certainly hasten the end of colonialism and racialization broadly associated with empirical educational measurement research, and specifically related to score use, examined by Cushman in this *JWA* special issue.

In terms of principled inquiry, the ethical theory of writing assessment has much in common with ethical dispositions related to virtue. Pursuit of common values, mutual aspirations, and moral standpoints resonate harmoniously with attention to consequences, exposition of constitutional premises, formation of community, and economic effectuation for social good. Philosophical traditions remain a rich resource of reasoning. In terms of generalization inferences, much may be attained by investigating the relationship between ethical theory and the direction of writing studies.

References

Abbot, R. D., & Berninger, V. (1993). Structural equation modeling of relationships among developmental skills and writing in primary- and intermediate-grade writers. *Journal of Educational Psychology*, *85*, 478–504.

Accreditation Board for Engineering and Technology. (2014). *Criteria for accrediting engineering programs*. Retrieved from http://www.abet.org/wp-content/uploads/2015/05/E001-15-16-EAC-Criteria-03-10-15.pdf

ADA Amendments Act of 2008, Public L. No. 110-325 (S 3406). (2008).

Adler-Kassner, L. (2008). *The activist WPA: Changing stories about writing and writers*. Logan, UT: Utah State University Press.

Adler-Kassner, L. & Wardle, E. (Eds.). (2015). *Naming what we know: Threshold concepts of writing studies*. Logan, UT: Utah State University Press.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1974). *Standards for educational and psychological tests* (3rd ed.). Washington. DC: American Psychological Association.

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

Anderson, T., Scum, D., & Twining, W. (2005). *Analysis of evidence* (2nd ed.). Cambridge, UK: Cambridge University Press.

Anson, C. M. (2010). Assessment in action: A Möbius tale. In M. Hundleby & J. Allen (Eds.), *Assessment in technical and professional communication* (pp. 3–15). Amityville, NY: Baywood Publishing.

Applebee, A. N. (2013). Common Core State Standards: The promise and the peril in a national palimpsest. *English Journal*, *103*, 25–33.

Aquinas, T. (trans. 2009) *Summa theologica* (A. J. Freddoso, Trans.). South Bend, IN: St. Augustine's Press.

Aristotle. (trans. 2014). *The Nicomachean ethics* (J. Crisp, Trans.). Cambridge, UK: Cambridge UP.

Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. Chicago, IL: University of Chicago Press.

Association of Language Testing in Europe (2010). *Code of practice*. Retrieved from http://www.alte.org/setting_standards/code_of_practice

Aull, L. (2015). *First-year university writing: A corpus-based study with implications for pedagogy*. New York, NY: Palgrave Macmillan

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

Austin, L. J. (1962). *How to do things with words*. Oxford, UK: Oxford University Press.

Baird, J.-A., Hopfenbeck, T.N., Newton, P., Stobart, G., Steen-Utheim, A.T. (2014). *Assessment and learning: State of the field review*. Oslo, NO: Knowledge Center for Education. Retrieved from http:// www.forskningsradet.no/servlet/Satellite? c=Rapport&cid=1253996755700&lang=en&pagename=kunnskapssenter%2FHovedsidemal

Baxter, B., et al. (1969). Job testing and the disadvantaged. *American Psychologist, 24*, 637–50.

Bazerman, C. (2013). *A theory of literate action*. Fort Collins, CO: The WAC Clearinghouse and Anderson, SC: Parlor Press.

Beason, L. (2000). Composition as service: Implications of utilitarian, duties, and care ethics. In M. A. Pemberton (Ed.), *The ethics of writing instruction: Issues in theory and practice* (pp. 105–138). Stamford, CT: Ablex.

Bennett, R. E. (1993). On the meanings of constructed response. In R. E. Bennett & W. C. Ward (Eds.), *Construction vs. choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment* (pp. 1–27). Hillsdale, NJ: Erlbaum.

Berthoff, A. E. (1990). Killer dichotomies: Reading in/reading out. In K. Ronald & H. Roskelly (Eds.), *Farther along: Transforming dichotomies in rhetoric and composition* (pp. 15–24). Portsmouth, NH: Boynton/Cook.

Berninger, V. W. (Ed.). (2012). *Past, present, and future contributions of cognitive writing research to cognitive psychology*. New York, NY: Taylor and Frances.

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university* (4th ed.). New York, NY: McGraw-Hill.

Bitzer, L. (1968). The rhetorical situation. *Philosophy & Rhetoric, 25*, 1–14.

Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics*. Cambridge, UK: Cambridge University Press.

Borsboom, D., Cramer, A. O. J., Kievit, R. A., Zand Scholten, A., & Francic, S. (2009). The end of construct validity. In Lissitz, R. W. (Ed.), *The concept of validity* (pp. 135–170). Charlotte, NC: Information Age Publishing.

Borsboom, D., & Markus, K. A. (2013). Truth and evidence in validity theory. *Journal of Educational Measurement, 50*, 110–114.

Brennan, R. (2001). *Generalizability theory*. New York, NY: Springer-Verlag.

Bridgeman, B., Pollack, N., & Burton, N. (2008). Predicting grades in college courses: A comparison of multiple regression and percent succeeding approaches. *Journal of College Admissions, 199*, 19–25.

Broad, B., & Boyd, M. (2005). Rhetorical writing assessment: The practice and theory of complementarity. *Journal of Writing Assessment, 2*, 7–20. Retrieved from http://www.journalofwritingassessment.org/archives/2-1.2.pdf

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In R. M. Lerner & W. Damon (Eds.), *Handbook of child psychology: Theoretical models of human development* (Vol. 1, 6th ed.) (pp. 793–828). New York, NY: Wiley.

Bruce, E., & Hamp-Lyons, L. (2015). Opposing tensions of local and international standards for EAP writing programs: Who are we assessing for? *Journal of English for Academic Purposes, 18*, 64–77.

Burke, K. (1945). *A grammar of motives*. New York, NY: Prentice-Hall.

Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 221–256). Westport, CT: American Council on Education/Praeger.

Cizek, G. J. (2012). Defining and distinguishing validity: Interpretations of score meaning and justification of test use, *Psychological Methods*, *17*, 31–43.

Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage Press.

Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and White Students in integrated colleges. *Journal of Educational Measurement, 5*, 115–124.

Cloward, R. A. (1959). Illegitimate means, anomie, and deviant behavior. *American Sociological Review, 24*, 164–176.

Cole, A. (2015). The function of theory at the present time. *Publications of the Modern Language Association, 103*, 809–818.

Cole, N. (1981). Bias in testing. *American Psychologist, 36*, 1067–1077.

Collins, R., Elliot, N., Klobucar, A., & and Deek, F. (2013). Web-based portfolio assessment: Validation of an open source platform. *Journal of Interactive Learning Research,* 24, 5–32.

Complete College America. (2012). *Remediation: Higher education's bridge to nowhere*. Retrieved from http://completecollege.org/docs/CCA-Remediation-final.pdf

Complete College America. (2015). *Corequisite remediation: Spanning the completion divide*. Washington, DC: Complete College America. Retrieved from http://completecollege.org/spanningthedivide/#home

Condon, W. (2011). Reinventing writing assessment: How the conversation is shifting. [Review of the books *Reframing writing assessment to improve teaching and learning*, by L. Adler-Kassner & P. O'Neill; *Organic writing assessment: Dynamic Criteria Mapping in action*, by B. Broad et al.; *On a scale: A social history of writing assessment in America*, by N. Elliot; *Machine scoring of student essays: Truth and consequences*, by P. F. Ericsson & R. Haswell (Eds.); *Assessing writing: A critical sourcebook*, by B. Huot & P. O'Neill (Eds.); *(Re)Articulating writing assessment for teaching and learning*, by B. Huot; *Coming to terms: A theory of writing assessment*, by P. Lynne; *Writing assessment and the revolution in digital texts and technologies*, by M. R. Neal; *A guide to college writing*, by P. O'Neill, C. Moore, & B. Huot; *Assessment of writing*, by M. C. Paretti & K. M. Powell (Eds.); *Assessing writing*, by S. C. Weigle; *Rethinking rubrics in writing assessment*, by M. Wilson; and *An overview of writing assessment: Theory, research, and practice*, by W. Wolcott, with S. M. Legg. *WPA: Writing Program Administration, 34*, 162–182.

Condon, W. (2012). The future of portfolio-based writing assessment: A cautionary tale. In N. Elliot & L. Perelman (Eds.), *Writing assessment in the 21st century: Essays in honor of Edward M. White* (pp. 233–245). New York, NY: Hampton Press.

Conference on College Composition and Communication. (1987). *Scholarship in composition: Guidelines for faculty, deans, and department chairs*. Retrieved from http://www.ncte.org/cccc/resources/positions/scholarshipincomp

Conference on College Composition and Communication (2009). *Writing assessment: A position statement*. Retrieved from http://www.ncte.org/cccc/resources/positions/writingassessment

Conference on College Composition and Communication (2015a). *CCCC guidelines for the ethical conduct of research in composition studies*. Retrieved from http://www.ncte.org/cccc/resources/positions/ethicalconduct

Conference on College Composition and Communication (2015b). *CCCC research initiative*. Retrieved from http://www.ncte.org/cccc/awards/researchinitiative

Coppola, N. W., Hiltz, S. R., & Rotter, N. G. (2004). Building trust in virtual teams. *IEEE Transactions on Professional Communication, 47*, 95–104.

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge, UK: Cambridge University Press.

Council of Writing Program Administrators (2014). *WPA Outcomes Statement for First-Year Composition* (Revisions adopted 17 July 2014). *WPA: Writing Program Administration, 38*, 142–146.

Council of Writing Program Administrators, National Council of Teachers of English, & National Writing Project. (2011). *Framework for Success in Postsecondary Writing*. Retrieved from http://www.nwp.org/img/resources/framework_for_success.pdf

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443–507). Washington, DC: American Council on Education.

Cumming, A. (2013). Assessing integrated writing tasks for academic purposes:

Promises and perils. *Language Assessment Quarterly, 10*, 1–8.

Cushman, E., Juzwik, M., Macaluso, K., & Milu, E. (2015). Decolonizing research in the teaching of English(es). *Research in the Teaching of English, 49*, 333–339.

Darlington, R. B. (1976). A defense of "rational" personnel selection, and two new methods. *Journal of Educational Measurement, 13*, 43–52.

Deane, P., Sabatini, J., Feng, G., Sparks, J., Song, Y., Fowles, M., . . . Foley, C. (2015). *Key practices in the English Language Arts (ELA): Linking learning theory, assessment, and instruction* (ETS RR 15-17). Princeton, NJ: Educational Testing Service.

DeVoss, D. N., Eidman-Aadahl, E., & Hicks, T. (2010). *Because digital writing matters: Improving student writing in online and multimedia environments*. San Francisco: Jossey-Bass.

Dixon, Z., & Moxley, J. (2013). Everything is illuminated: What big data can tell us about teacher commentary. *Assessing Writing, 18*, 241–256.

Duckworth, A. L., & Yaeger, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher, 44*, 237–251.

Duffy, J. (2007). *Writing from these roots: Literacy in a Hmong-American community*. Honolulu, HI: University of Hawai'i Press.

Duffy, J. (2014). Ethical dispositions: A discourse for rhetoric and composition. *Journal of Advanced Composition, 34*, 209–237.

Duffy, J. (2015). Writing involves making ethical choices. In L. Adler-Kassner & E. Wardle (Eds.), *Naming what we know: Threshold concepts of writing studies* (pp. 31–32). Logan, UT: Utah State University Press.

Duffy, J. (in press). The good writer: Virtue ethics and the teaching of writing. *College English*.

Educational Testing Service. (2014). *ETS standards for quality and fairness*. Princeton, NJ: Educational Testing Service. Retrieved from https://www.ets.org/s/about/pdf/standards.pdf

Elliot, N. (2005). *On a scale: A social history of writing assessment in America*. New York, NY: Lang.

Elliot, N. (2015). Validation: The pursuit. [Review of Standards for educational and psychological testing, by American Educational Research Association, American Psychological Association, and National Council on Measurement in Education]. *College Composition and Communication, 66*, 668–685.

Elliot, N., Briller, V., & Joshi, K. (2007). Quantification and community. *Journal of Writing Assessment, 3*, 5–29. Retrieved from http://www.journalofwritingassessment.org/archives/3-1.2.pdf

Elliot, N., Deess, P., Rudniy, A., & and Joshi, K. (2012). Placement of students into first-year writing courses. *Research in the Teaching of English, 46*, 285–313.

Elliot, N., Rupp, A. A., & Williamson, D. A. (2015). Three interpretative frameworks: Assessment of English language arts-writing in the Common Core State Standards Initiative. *Journal of Writing Assessment*, *8*. Retrieved from http://www.journalofwritingassessment.org/article.php?article=84

Embretson, S. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin, 93*, 179–97.

Every Student Succeeds Act of 2015, S. 1177. (2015). Retrieved from https://www.gpo.gov/fdsys/pkg/BILLS-114s1177enr/pdf/BILLS-114s1177enr.pdf

Farrand, W. (1895*). The reform of college entrance requirements: Inaugural address*. Newark, NJ: William A. Baker, Printer.

Finch, W. H., & French, B. F. (2015). *Latent variable modeling with R*. New York, NY: Routledge.

Fishbein, M., & Ajzen, I. (2010). *Predicting and changing behavior: The reasoned action approach*. New York, NY: Taylor and Francis.

Flower, L. (1994) *The construction of negotiated meaning: A social cognitive theory of writing*. Carbondale and Edwardsville, IL: Southern Illinois University Press.

Freire, P. (1970). *Pedagogy of the oppressed*. (M. B. Ramos, Trans.). New York, NY: Herder and Herder.

Fulcher, G. (2015). *Re-examining language testing: A philosophical and social inquiry*. New York, NY: Routledge.

Fulkerson, R. (2005). Composition at the turn of the twenty-first century. *College Composition and Communication, 56*, 654–587.

Gallagher, C. W. (2011). Being there: (Re)making the assessment scene. *College Composition and Communication, 63*, 450–476.

Gallagher, C. W. (2014). Staging encounters: Assessing the performance of context in students' multimodal writing. *Computers and Composition, 31*, 1–12.

Gee, J. P. (2008). A sociocultural perspective on opportunity to learn. In P. A. Moss, D. C. Pullin, J. P. Gee, E. H. Haertel, & L. J. Young (Eds.), *Assessment, equity, and opportunity to learn* (pp. 76–108). Cambridge, UK: Cambridge University Press.

Gee, J. P. (2012). *Social linguistics and literacy: Ideology in discourses* (4th ed.). New York, NY: Routledge.

Gentner, D., & Grudin, J. (1985). The evolution of mental metaphors in psychology: A 90-year retrospective. *American Psychologist, 40*, 181–192.

Gere, A. R., Aull, L., Green, T., & Porter, A. (2010). Assessing the validity of directed self-placement at a large university. *Assessing Writing, 15*, 154–176.

Graham, S., & Williams, C. (2009). An attributional approach to motivation in school. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 11–34). New York, NY: Routledge.

Haertel, E. H. (2006). Reliability. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 65–110). Westport, CT: American Council on Education/Praeger.

Haertel, E. H. (2013, March). *Reliability and validity of inferences about teachers based on student test scores.* William H. Angoff 14th memorial lecture presented at the National Press Club, Washington, DC. Retrieved from https://www.ets.org/Media/Research/pdf/PICANG14.pdf

Haladyna, T. H., & Downing, S. M. (2004). Construct-irrelevance variance in high-stakes testing. *Educational Measurement: Issues and Practice, 23*, 17–27.

Hart, R., Casserly, M., Uzzell, R., Palacios, M., Corcoran, A., Spurgeon, L. (2015). *Student testing in America's Great City Schools: An inventory and preliminary analysis*. Washington, DC: Council of Great City Schools. Retrieved from http://www.cgcs.org/cms/lib/DC00001581/Centricity/Domain/87/Testing%20Report.pdf

Harvey, D. (2010). *The enigma of capital and the crisis of capitalism*. New York, NY: Oxford University Press.

Hassel, H., & Giordano, J. B. (2015). The blurry borders of college writing: Remediation and the assessment of student readiness. *College English, 78*, 56–80.

Haswell, J. & Haswell, R. (2010). *Authoring: An essay for the English profession on potentiality and singularity*. Logan, UT: Utah State University Press.

Haswell, R. (2005). NCTE/CCCC's recent war on scholarship. *Written Communication, 22*, 198–223.

Haswell, R. (2012). Methodologically adrift. [Review of *Academically adrift: Limited learning on college campuses*, by R. Arum & J. Roksa]. *College Composition and Communication, 63*, 487–91.

Haswell, R. (2013). Writing assessment and race studies *sub specie aeternitatis*: A response to *Race and Writing Assessment*. [Review of *Race and Writing Assessment*, ed. by A. Inoue & M. Poe]. *Journal of Writing Assessment Reading List*. Retrieved from http://jwareadinglist.blogspot.com/2013/01/writing-assessment-and-race-studies-sub_4.html

Haswell, R. & Haswell, J. (2015). *Hospitality and authoring: An essay for the English profession*. Logan, UT: Utah State University Press.

Hecht, L. W. (1980). *Validation of the New Jersey College Basic Skills Placement Test* (ED 214 945). Retrieved from http://files.eric.ed.gov/fulltext/ED214945.pdf

Hillocks, G. H., Jr. (2002). *The testing trap: How state writing assessments control learning*. New York, NY: Teachers College Press, 2002.

Hirshmann, N. J. (1992). *Rethinking obligation: A feminist method for political theory*. Ithaca and London: Cornell University Press.

Hoffman, J. L., & Lowitzki, K. E. (2005). Predicting college success with high school grades and test scores: Limitations for minority students. *Review of Higher Education, 28*, 455–474.

Howetz, P. L. (2014). Neglected infections of poverty in the United States and their effects on the brain. *JAMA Psychiatry, 71*, 1099–1100.

Horn, L. (2006). *Placing college graduation rates in context: How 4-year college graduation rates vary with selectivity and the size of low-income enrollment* (NCES 2007-161). Washington, DC: National Center for Education Statistics.

Horner, B. (2015). Rewriting composition: Moving beyond a discourse of need. *College English, 77*, 450–479.

Hulleman, C. S., Godes, O., Hendricks, B., & Harackiewicz, J. M. (2010). Enhancing interest and performance with a utility value intervention. *Journal of Educational Psychology, 102*, 880–895.

Huot, B. (2002). *(Re)Articulating writing assessment for teaching and learning*. Logan, UT: Utah State University Press.

Hursthouse, R. (1999). *On virtue ethics*. Oxford, UK: Oxford University Press.

Ignash, J. M. & Townsend, B. K. (2000). Evaluating state-level articulation agreements according to good practice. *Community College Review, 28*, 1–21.

Inoue, A. B. (2005). Community-based assessment pedagogy. *Assessing Writing, 9*, 208–238.

Inoue, A. B. (2009). The technology of writing assessment and racial validity. In C. Schreiner (Ed.). *Handbook of research on assessment technologies, method, and applications in higher education* (pp. 97–120). Hershey, PA: Information Science Reference.

Inoue, A. B. (2014). Theorizing failure in U.S. writing assessments. *Research in the Teaching of English, 48*, 330–352.

Inoue, A. B. (2015). *Antiracist writing assessment ecologies: Teaching and assessing writing for a socially just future*. Fort Collins, CO: The WAC Clearinghouse and Anderson, SC: Parlor Press.

Inoue, A. B., & Poe, M. (Eds.). (2012). *Race and writing assessment*. New York: Peter Lang.

International Language Testing Association. (2000). *ITLA code of ethics*. Retrieved from http://www.iltaonline.com/index.php/enUS/component/content/article?id=57

International Language Testing Association. (2007). *ILTA guidelines for practice*. Retrieved from https://www.google.com/#q=ILTA+Guidelines+for+Practice

Intemann, K. (2010). 25 years of feminist empiricism and standpoint theory: Where are we now? *Hypatia, 25*, 778–96.

Junco, R., & Clem, C. (2013). *Evaluating how the CourseSmart Engagement Index™ predicts student course outcomes*. San Mateo, CA: CourseSmart. Retrieved from http://blog.reyjunco.com/wp-content/uploads/2010/03/FINAL-CourseSmart_Analytics_White_Paper.pdf

Kahneman, D. (2011). *Thinking fast and slow*. New York, NY: Farrar, Strauss, and Giroux.

Kant, I. (1997). *Groundwork of the metaphysics of morals* (M. Gregor, Ed. and Trans.). Cambridge, UK: Cambridge University Press. (Original work published 1785)

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport, CT: American Council on Education/Praeger.

Kane, M. T. (2011). The error of our ways. *Journal of Educational Measurement, 48*, 12–30.

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

Kane, M. T. (2015). Explicating validity. *Assessment in Education: Principles, Policy, & Practice, 22*, 1–14.

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

Kidder, W. C., & Rosner, J. (2002). How the SAT creates built-in-headwinds: An educational and legal analysis of disparate impact. *Santa Clara Law Review, 43*, 131–211.

Klobucar, A., Elliot, N., Deess, P., Rudniy, O., & Joshi, K. (2013). Automated scoring in context: Rapid assessment for placed students. *Assessing Writing, 18*, 62–84.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Kuhn, T. S. (1977). A function for thought experiments. Reprinted in T. S. Kuhn, *The essential tension* (pp. 240–265). Chicago, IL: University of Chicago Press. (Original work published 1964)

Kunnan, A. J. (2014a). *Assessment around the world*. West Sussex, UK: Wiley.

Kunnan, A. J. (2014b). Fairness and justice in language assessment. In A. J. Kunnan (Ed.), *The companion to language assessment: Evaluation, methodology, and interdisciplinary themes* (Vol. 3) (pp. 1098–1114). West Sussex, UK: Wiley.

Kutner, M., Greenberg, E., & Baer, J. (2005). *A first look at the literacy of America's adults in the 21st century* (NCES 2006-470). Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/NAAL/PDF/2006470.PDF

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Lazarin, M. (2014). *Testing overload in America's schools*. Washington, DC: Center for American Progress. Retrieved from https://cdn.americanprogress.org/wp-content/uploads/2014/10/LazarinOvertestingReport.pdf

Learned, W. S., & Wood, B. D. (1938). *The student and his knowledge: A report to the Carnegie Foundation on the results of the high school and college examinations of 1928, 1930, and 1932.* New York, NY: Carnegie Foundation for the Advancement of Teaching.

Leary, D. (1990). *Metaphors in the history of psychology*. New York, NY: Cambridge University Press.

Levin, H. M., & McEwan, P. J. (2001). *Cost-effectiveness analysis: Methods and applications* (2nd ed.). Thousand Oaks, CA: Sage.

Levine, R. J. (1988). *Ethics and regulation of clinical research* (2nd ed.). New Haven, CT: Yale University Press.

Lewis, J. (1999). *Walking with the wind: A memoir of the movement*. New York, NY: Simon and Schuster.

Leydens, J. A., & Olds, B. M. (2012). Complicating the fail-or-succeed dichotomy in writing assessment outcomes. In N. Elliot & L. Perelman (Eds.), *Writing assessment in the 21st century: Essays in honor of Edward M. White* (pp. 247–258). New York, NY: Hampton Press.

Lewiecki-Wilson, C., & Brueggemann, B. (Eds.). (2007). *Disability and the teaching of writing: A critical sourcebook*. New York: Bedford/St. Martin's Press.

Linn, R. L. (1976). In search of fair selection procedures. *Journal of Educational Measurement, 13*, 53–58.

Lynne, P. (2004). *Coming to terms: A theory of writing assessment*. Logan, UT: Utah State University Press.

MacIntyre, A. (2007). *After virtue: A study in moral theory* (3rd ed.). Notre Dame, IN: University of Notre Dame Press.

Markus, K. A. (1998). Science, measurement, and validity: Is completion of Samuel Messick's synthesis possible? *Social Indicators Research, 45*, 7–34.

Markus, K. A., & Borsboom, D. (2013). *Frontiers of test validity theory: Measurement, causation, and meaning*. New York, NY: Routledge.

McCrae, R. R., & Costa, P. Y. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personal and Social Psychology, 52*, 81–90.

McGraw, S. A., Sellers, D. E., Stone, E. J., Bebchuk, J., Edmundson, E. W., Johnson, C. C., . . . Luepher, R. V. (1966). Using process data to explain outcomes: An illustration from the Child and Adolescent Trial for Cardiovascular Health (CATCH). *Evaluation Review, 20*, 291–312.

McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Oxford, UK: Blackwell Publishing.

Meade, A. W., & Tonidanel, S. (2010). Not seeing clearly with Cleary: What test bias analyses do and do not tell us. *Industrial and Organizational Psychology, 3*, 192–205.

Merton, R. K. (1938). Social structure and anomie. *American Sociological Review, 3*, 672–682.

Merton, R. K. (1996). Opportunity structure: The emergence, diffusion and differentiation of a sociological concept, 1930s–1950. In F. Adler & W. S. Laufer (Eds.), *The legacy of anomie theory: Advances in criminological theory* (pp. 3–78). New Brunswick, NJ: Transaction Publishers.

Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, *35*, 1012–1027.

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

Meyers, S. V. (2014). *Del otro lado: Literacy and migration across the U.S.-Mexico border*. Carbondale, IL: Southern Illinois University Press.

Mill, J. S. (1879). *Utilitarianism* (7th ed.). London, UK: Longmans, Green, and Company. (Original work published 1861)

Miller, C. (1984). Genre as social action. *Quarterly Journal of Speech, 70*, 151–167.

Miller, C. (1994). In A. Freedman & P. Medway (Eds.), *Rhetorical community: The cultural basis of genre* (pp. 67–97). London, UK: Taylor & Francis.

Miller, K. L. (2016). *The rhetoricity and rhetorical histories of writing assessment* (Unpublished doctoral dissertation). University of Nevada, Reno.

Mislevy, R. J. (2007). Validity by design. *Educational Researcher, 36*, 463–69.

Mislevy, R. J. (2008). How cognitive science challenges the educational measurement tradition. Retrieved from http://umdperg.pbworks.com/f/CommentaryHaig_Mislevy.pdf

Mislevy, R. J., & Durán, R. P. (2014). A sociocognitive perspective on assessing EL students in the age of Common Core and next generation science standards. *TESOL Quarterly, 48*, 560–585.

Mislevy, R. J., Haertel, G., Cheng, B., Ructtinger, L., DeBarger, A., Murray, E., …Vendlinski, T. (2013). A "conditional" sense of fairness in assessment. *Educational Research and Evaluation: An International Journal on Theory and Practice, 19*, 121–140.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). Design and analysis in task-based language assessment. *Language Testing, 19*, 477–96.

Morton, A. (1980). *Frames of mind: Constraints of the common-sense conception of the mental*. Oxford, UK: Clarendon Press.

Moss, P. A., Pullin, D. C., Gee, J. P., Haertel, E. H. & Young, L. J. (Eds.). (2008). *Assessment, equity, and opportunity to learn.* Cambridge, UK: Cambridge University Press.

Mossman, M. J. (1994). Gender quality, family law and access to justice. *International Journal of Law and the Family, 8*, 357–373.

Moxley, J. (2008). Datagogies, writing spaces, and the age of peer production. *Computers and Composition, 25*, 182–202.

Moxley, J. (2013). Big data, learning analytics, and social assessment. *Journal of Writing Assessment*, *6*. Retrieved from http://www.journalofwritingassessment.org/article.php?article=68

National Architectural Accrediting Board. (2014). *Guide to the 2014 conditions for accreditation and preparation of an architecture program report*. Washington, DC: National Architectural Accrediting Board.

National Association for College Admission Counseling. (2008). *Report of the commission on the use of standardized tests in undergraduate admissions*. Arlington, VA: National Association for College Admission Counseling. Retrieved from http://www.nacacnet.org/research/PublicationsResources/Marketplace/Documents/TestingComission_FinalReport.pdf

National Center for Education Statistics. (2012). *The Nation's Report Card: Writing 2011* (NCES 2012–470). Washington, DC: U.S. Department of Education. Retrieved from http://nces.ed.gov/nationsreportcard/pdf/main2011/2012470.pdf

National Council of Teachers of English. (2015a). Education policy platform. Retrieved from http://www.ncte.org/positions/statements/2015-policy-platform

National Council of Teachers of English. (2015b). English Language Arts Standards » Anchor Standards » College and Career Readiness Anchor Standards for Writing. Retrieved from http://www.corestandards.org/ELA-Literacy/CCRA/W/

National Governors Association. (2015). Common core state standards initiative. Washington, DC: National Governors Association. Retrieved from http://www.corestandards.org

National Research Council. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. Committee on Defining Deeper Learning and 21st Century Skills, J. W. Pellegrino & M. L. Hilton, Board on Testing and Assessment and Board on Science Education, Division of Behavioral and Social Sciences and Education (Eds.). Washington, DC: The National Academies Press.

Newton, P. (2012). Clarifying the consensus definition of validity. *Measurement: Interdisciplinary Research and Perspectives, 10*, 1–29.

Novick, M. R., & Petersen, N. S. (1976). Towards equalizing educational and employment opportunity. *Journal of Educational Measurement, 13*, 77–88.

Nussbaum, M. (2002). Capabilities and disabilities: Justice for mentally disabled citizens. *Philosophical Topics, 30*, 133–165.

O'Neill, P. (2015). Threshold concepts at the crossroads: Writing instruction and assessment. In L. Adler-Kassner & E. Wardle (Eds.), *Naming what we know: Threshold concepts of writing studies* (pp. 157–170). Logan, UT: Utah State University Press.

Palmer, A. (2015). *Smart money: How high-stakes financial innovation is reshaping our world—for the better*. New York, NY: Basic Books.

Phelps, L. W., & Ackerman, J. W. (2010). Making the case for disciplinarity in Rhetoric, Composition, and Writing Studies: The Visibility Project. *College Composition and Communication, 62*, 180–215.

Piketty, T. (2014). *Capital in the twenty-first century* (A. Goldhammer, Trans.). Cambridge and London: Harvard University Press.

Pimentel, S. (2013). *College and career readiness standards for adult education*. Washington, DC: U.S. Department of Education, Office of Vocational and Adult Education. Retrieved from https://lincs.ed.gov/publications/pdf/CCRStandardsAdultEd.pdf

Plato. (trans. 2010). *The Republic* (T. Griffith, Trans.). Cambridge, UK: Cambridge UP.

Poe, M. (2012). Diversity and international writing assessment [Special issue]. *Research in the Teaching of English, 48*(3).

Poe, M., Elliot, N., Cogan, J. A., & Nurudeen, T. G. (2014). The legal and the local: Using disparate impact analysis to understand the consequences of writing assessment. *College Composition and Communication, 65*, 588–611.

Poe, M., & Inoue, A. B. (in press). Writing assessment as social justice [Special issue]. *College English*.

Pogge, T. (2007). *John Rawls: His life and theory of justice*. New York, NY: Oxford University Press.

Popper, K. R. (1962). *Conjectures and refutations: The growth of scientific knowledge*. New York, NY: Basic Books.

Pullin, D. C. (2008). Assessment, equity, and opportunity to learn. In P. A. Moss, D. C. Pullin, J. P. Gee, E. H. Haertel, & L. J. Young (Eds.), *Assessment, equity, and opportunity to learn* (pp. 333–351). Cambridge, UK: Cambridge University Press.

Rachels, J. *The elements of moral philosophy*. Philadelphia, PA: Temple University Press, 1986.

Rawls, J. (1999). *A theory of justice* (Rev. ed). Cambridge, MA: Cambridge University Press. (Original work published 1971)

Rawls, J. (2001). *Justice as fairness: A restatement* (R. Kelly, Ed.). Cambridge, MA: Harvard University Press.

Ramsey, P. (1993). Sensitivity review: The ETS experience as a case study. In P. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 367–388). Hillsdale, NJ: Lawrence Erlbaum.

Rescher, N. (1991). Thought experiments in Presocratic philosophy. In T. Horowitz & G. Massey (Eds.), *Thought experiments in science and philosophy* (pp. 31–42). Lanham, MD: Rowman & Littlefield.

Rose, D., Meyer, A., & Hitchcock, C. (Eds.). (2005). *The universally designed classroom*. Cambridge, MA: Harvard Education Press.

Rose, M. (2009). *Writer's block: The cognitive dimension*. Carbondale, IL: Southern Illinois University Press. (Original work published 1984)

Rose, S. K., & Weiser, I. (Eds.). *Going public: What writing programs learn from engagement*. Logan, UT: Utah State University Press.

Rousseau, J.-J. (1968). *The social contract* (M. Cranston, Trans.). New York, NY: Penguin. (Original work published 1762)

Royer, D. J., & Gilles, R. (1998). Directed self-placement: An attitude of orientation. *College Composition and Communication, 50*, 54–70.

Rorty, R. (1989). *Contingency, irony, and solidarity*. Cambridge, UK: Cambridge University Press.

Ryle, G. (1949). *The concept of mind*. London, UK: Penguin.

Sánchez, R. (2012). Outside the text: Retheorizing empiricism and identity. *College English, 74*, 234–46.

Sarasvathy, S. D. (2008). *Effectuation: Elements of entrepreneurial expertise*. Cheltenham and Northampton: Edward Elgar.

Sawyer, R. L., Cole, N. S., & Cole, J. W. L. (1976). Utilities and the issue of fairness in a decision theoretic model for selection. *Journal of Educational Measurement, 13*, 59–76.

Shechtman, N., DeBarger, A. H., Dornsife, C., Rosier, S., & Yarnall, L. (2013). *Promoting grit, tenacity, and perseverance: Critical factors for success in the 21st century*. Washington, DC: U.S. Department of Education and Office of Educational Technology. Retrieved from http://pgbovine.net/OET-Draft-Grit-Report-2-17-13.pdf

Schendel E., & O'Neill, P. (1999). Exploring the theories of consequence of self-assessment through ethical inquiry. *Assessing Writing, 62*, 199–227.

Shea, P., Hayes, S. Vickers, J. Gozza-Cohen, M., Uzner, S., Mehta, R. . . . & Rangan, P. (2010). A re-examination of the community of inquiry framework: Social network and content analysis. *Internet and Higher Education, 13*, 10–21.

Sidgwick, H. (1884). *The methods of ethics*. London, UK: Macmillian.

Silberman, S. (2015). *NeuroTribes: The legacy of autism and the future of neurodiversity*. New York, NY: Penguin.

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

Slomp, D. H., Corrigan, J. A. & Sugimoto. T. (2014). A framework for using consequential validity evidence in evaluating large-scale writing assessments: A Canadian study. *Research in the Teaching of English, 48*, 276–302.

Smarter Balanced Assessment Consortium. (2014). Disaggregated data from the Smarter Balanced field test. Retrieved from http://www.smarterbalanced.org/wordpress/wp-content/uploads/2014/12/Disaggregated-FieldTestDataFINAL.pdf

Snow, R. E., & Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 263–331). New York, NY: American Council on Education/Macmillan.

Solomon, A. (2012). *Far from the tree: Parents, children, and the search for identity*. New York, NY: Scribner.

Spearman, C. (1904). General intelligence: Objectively determined and measured. *American Journal of Psychology, 15*, 201–292.

Spinuzzi, C. (2003). *Tracing genres through organizations*. Cambridge, MA: Massachusetts Institute of Technology Press.

Spinuzzi, C. (2015). *All edge: Inside the new workplace networks*. Chicago, IL: University of Chicago Press.

Spolsky, B. (2014). The influence of ethics in language assessment. In A. J. Kunnan (Ed.), *The companion to language assessment: Evaluation, methodology, and interdisciplinary themes* (Vol. 3) (pp. 1571–1585). West Sussex, UK: Wiley.

Sternberg, R. S. (2010). *College admission for the 21st century*. Cambridge, MA: Harvard University Press.

Sternglass, M. S. (1997). *Time to know them: A longitudinal study of writing and learning at the college level*. Mahwah, NJ: Lawrence Erlbaum.

Storch, N. (2005). Collaborative writing: Product, process, and students' reflections. *Journal of Second Language Writing, 14*, 153–173.

Tchudi, S. (Ed.). (1997). *Alternatives to grading student writing*. Urbana, IL: National Council of Teachers of English.

Tinder, G. (1980). *Community: Reflections on a tragic ideal*. Baton Rouge and London: Louisiana State University Press.

United Nations Educational, Scientific and Cultural Organization (2015). *Education for all, 2000-2015: Achievements and challenges*. Paris, France: United Nations Educational, Scientific and Cultural Organization, Retrieved from http://unesdoc.unesco.org/images/0023/002322/232205e.pdf

van der Maas, H. L. J., Molenaar, D., Maris, G., Kievit, R. A., & Borsboom, D. (2011).   Cognitive psychology meets psychometric theory: On the relation between process   models for decision making and latent variable models for individual differences.   *Psychological Review, 118*, 339–356.

van Frassen, B. (1980). *The scientific image*. Oxford, UK: Oxford University Press.

Villanueva, V., Jr. (1993). *Bootstraps: From an academic of color*. Urbana, IL: NCTE.

Voluntary System of Accountability. (2008). Information on learning outcomes measures. Retrieved from https://cp-files.s3.amazonaws.com/21/LearningOutcomesInfo.pdf

Warren, K. J. (1990). The power and the promise of ecological feminism. *Environmental Ethics, 12*, 125–46.

Weiner, B. (1986). *An attributional theory of motivation and emotion*. New York, NY: Springer-Verlag.

Weir, C. (2005). *Language testing and validation: An evidence-based approach*. Basingstoke, UK: Palgrave Macmillan.

Weir, C. J., Vidakovic, I., Galaczi, D. E. (2013). *Measured constructs: A history of Cambridge English Examinations, 1913-2012*. Studies in Language Testing (Vol. 37). Cambridge, UK: UCLES/Cambridge University Press.

Weiss, C. H. (1998). *Evaluation* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

White, E. M. (1973). *Comparison and contrast: The 1973 California State University and Colleges Freshman English Equivalency Examination* (ED 114 825). Retrieved from ERIC database http://files.eric.ed.gov/fulltext/ED114825.pdf

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

White, E. M., & Thomas, L. (1981). Racial minorities and writing skills assessment in the California State University and colleges. *College English, 43*, 276–283.

Wilder, L. (2015). Tangled roots. *College Composition and Communication, 63*, 501–506.

Willingham, W. W., Pollack, J. M., & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. *Journal of Educational Measurement, 39*, 1–37.

Wimsatt, W. K, & Beardsley, M. C. (1946). The intentional fallacy. *Sewanee Review, 54*, 468–488.

Wittgenstein, L. (2014). *Tractatus logico-philosophicus* (F. Ramsey & C. K. Ogden, Trans.). Peterborough, CAN: Broadview Press. (Original work published 1921)

Wood, B. (1923). *Measurement in higher education*. Yonkers-on-Hudson, NY: World Book Company.

Wood, T., Price, M., & Johnson, C. (2012). Disability studies, WPA-CompPile Research Bibliographies, No. 19. *WPA-CompPile Research Bibliographies*. Retrieved from http://comppile.org/wpa/bibliographies/Bib19/DisabilityStudies.pdf

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

Yancey, K. B. (2012). The rhetorical situation of writing assessment. In N. Elliot & L. Perelman (Eds.), *Writing assessment in the 21st century: Essays in honor of Edward M. White* (pp. 475–492). New York. NY: Hampton Press.

Yancey, K. B., Robertson, L., & Taczak, K. (2014). *Writing across contexts: Transfer, composition, and sites of writing*. Logan, UT: Utah State University Press.

Yen, W. M., & Fitzpatrick, A. R. (2006). Item response theory. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 111–153). Westport, CT: American Council on Education/Praeger.

Zumbo, B. D., Liu, Y., Wu, A. D., Shear, B. R., Astivia, O, L. O., & Ark, T. K. (2015). A methodology for Zumbo's third generation DIF analyses and the ecology of item responding. *Language Assessment Quarterly, 12*, 136–151.

Notes

[1]For disciplinary location of Rhetoric and Composition/Writing Studies, see Phelps and Ackerman (2010). The classification associated with the Visibility Project makes no mention of writing assessment, however, and many do not consider writing assessment to be a form of research, as Huot (2002) has observed. This absence of location in the profession has resulted in substantial career challenges for practitioners. An intended consequence of the proposed theory of fairness is the end of this displacement.

[2]Following Miller (1984, 1994) and Spinuzzi (2003, 2015), the term "genre" in the proposed theory is intended to integrate form, response range, and institutional contextualization. Among the traditional writing assessment genres are the following: course alignment (e.g., placement of admitted students); certification of ability for academic progression and graduation (e.g., institutional exit assessment); review of individual student ability to enhance student learning (e.g., formative classroom assessment); evaluation of curricular effectiveness (e.g., program review); and research (e.g., experimental investigation of student performance). Because research is inherent to each genre, the categories should be seen as integrated under a common aim. When this integration fails, that which is not designated as research becomes bureaucratic. As a form of value dualism, this disjuncture is prohibited by the theory in §3.2.5. For the range of responses within each of these genres, see Tchudi (1997). For the contextualization of genres and responses based on institutional ecology, see Inoue (2015).

[3]The choice of criterion measures, or the use of these measures if the assessment cannot demonstrate that it has achieved the aim of fairness under the propose theory, is significant. In his discussion of fairness, Linn (1976) noted the following: "The problem of topic choice for a criterion measure is fundamentally a problem of content validity. It is an issue which has been receiving increased attention in employment settings. It is an area that also requires more attention in educational settings. Easy access is not a sufficient justification for the use of grade point average as the sole criterion variable" (p. 57). Following Linn, Willingham, Pollack, and Lewis (2002) demonstrated that expanding instructor judgment beyond the course grade resulted in substantially higher correlations with the assessment at hand. In the case of writing assessment, Klobucar, Elliot, Dees, Rudniy, & Joshi (2013, Table 4, p. 75) showed that the correlation between the holistic portfolio score and course grade was 0.43 (*p* < .01). While a higher correlation might have been anticipated, it should be recognized that instructors implicitly assign grades based on the four-domain model identified in §3.1. Conversely, portfolios capture largely a single, cognitive domain. It is therefore no wonder that the correlation between portfolio scores and course grades is moderate. Framing the analysis between scores and grades as a matter of construct coverage, as Linn suggests, restores the validity of instructor grades as

rich construct measures. While the rise of large-scale assessment has been accompanied by lack of faith in teacher knowledge, one consequence of the theory of ethics—the reduction of writing assessments if they cannot be demonstrated to achieve fairness—would ideally be accompanied by restoration of faith teachers. There is a rich tradition of research dedicated to using student records for decision-making purposes that began with Farrand (1895), continued with Learned and Wood (1938), and finds its present form in the report of the National Association for College Admission Counseling (2008). In place of drop-from-the-sky assessments, this tradition is in need of examination and restoration. In the case of post-secondary writing placement, Hassel and Giordano (2015) make an excellent case for the use of multiple methods including high school grades.

[4]This list of disaggregation targets should be expanded by necessity of institutional setting and intended score use. To increase the aim of fairness, disaggregation according to writing task may also be warranted to understand responses processes of key groups and individuals.

[5]Referencing fairness as a virtue is a deliberate attempt to call attention to virtue ethics (Duffy, in press; MacIntyre, 2007). As an approach that is neither deontological or consequentialist, virtue ethics is agent centered. As Hursthouse (1999) has observed of modern virtue ethics, while action guidance is not possible under this philosophical orientation, action assessment is indeed within its sphere. In this way, virtue ethics is compatible with the theory of ethics for writing assessment.

[6]Distinct from the theory of ethics, activism is appropriate in cases when stakeholders have not been well served. In such cases, a wide range of strategies are available that are related to, yet distinct from, the proposed theory (Adler-Kassner, 2008; Rose & Weiser, 2010).

[7]The origin of the four-domain model is found in the identification of nomothetic span by Embretson (1983). The three-domain model suggested by the National Research Council (2012) inspired the four domain model proposed by White, Elliot, and Peckham (2014). As the history of writing assessment demonstrates (Elliot, 2005), emphasis on testing led to severe limiting of the writing construct, with weight given almost exclusively to its cognitive domain. Because it lent itself to readily observable facets of the writing construct, cognitive behavior was favored over interpersonal and intrapersonal domains. As to the physiological domain, it was often relegated to the study of disability and thus disenfranchised from the other three domains as notes in §2.3.4. Required are psychometric models of transparent design for special needs students whose abilities are studied as ways to improve the assessment for all students in cognitive, interpersonal, and intrapersonal domains (Mislevy et al., 2013). Key to the theory of ethics, the four-domain model features an integrative, interpretative framework in which the writing construct can be understood in all of its complexity. While print-based assessments made it difficult to capture information about domains other than the cognitive, digital platforms enable rich sources of information to be gathered from each of the domains (through, for instance, performance-related log patterns) and analyzed in relationship to each other (Shechtman, DeBarger, Dornsife, Rosier, & Yarnall, 2013).

[8]This proposed decision rule emphasizes the gravity of all writing assessment. Some disservice has been done to students by classifying score use on a continuum of impact. Anytime an inference is made about a student or a group of students, the impact is serious and the consequences potentially severe.

[9]Failure to recognize the importance of empirical methods, specifically use of descriptive and inferential statistical analysis, continues to be a barrier to the profession of writing studies and the research specialization of writing assessment. In reflecting on the Conference on College Composition and Communication Position Statement on "Scholarship in Composition: Guidelines for Faculty, Deans, and Department Chairs" (1987), Wilder (2015) observed that the document has little impact because it is boldly empirical. Today, as she concluded, "empirical research remains somewhat peripheral to composition" (p. 502).

Author Biography

Norbert Elliot is Professor Emeritus at New Jersey Institute of Technology. He is currently collaborating with Johanna Hillen (Principal Investigator) and Joe Moxley (Project Advisor) on a CCCC Research Initiative for 2015-2016: *Federal Grant Programs and Corollary Institutional Review Board Protocols: An Analysis of Reciprocity in Policy Determination, Implementation, and Impact on Writing Studies Research*.