

# Rhetorical Judgments: Using Holistic Assessment to Improve the Quality of Administrative Decisions

Roger J. Klurfeld & Steven Placek\*

Federal, state, and local governments issue hundreds of thousands of administrative decisions annually. Considering the number of encounters the public has with administrative appeal agencies, administrative decisions may be the largest category of legal writing and reading interaction the public has with the legal system. Usually, appellants come to appeal organizations after they have been denied some form of benefit. Federal agencies, for example, have hearing procedures established to conduct hearings in person, by telephone, by video, or by a case review of the documentary record.<sup>1</sup> These interactions often generate a two- to five-page legal decision that identifies issues in dispute and applies legal reasoning to the factual pattern of the case to reach a conclusion. Many of these agencies have identified writing quality — however they define it — as a priority in their strategic plans,<sup>2</sup> but the overwhelming number of hearings and decisions, coupled with regulatory guidelines for timeliness, may subordinate this goal to other management priorities.

In ad hoc attempts to improve writing, many government agencies send writers to a slew of training programs in writing and legal rhetoric at community colleges, law schools, the National Judicial College, and other private contractor or government-run training facilities. This training is cumulatively expensive,

---

\* © Roger J. Klurfeld & Steven Placek 2008. Roger J. Klurfeld is the Director of the National Appeals Division (NAD) at the United States Department of Agriculture (USDA). Steven Placek is a Training Specialist of the USDA-NAD, and a former Assistant Professor of English at the United States Military Academy. The views expressed in this article do not necessarily represent the views of the United States Department of Agriculture.

We would like to acknowledge our colleague Jennifer Nicholson for her comments and insights on this paper.

<sup>1</sup> Our example in the Department of Agriculture is typical. The National Appeals Division is an independent office within USDA that conducts hearings and reviews of USDA sub-agency adverse decisions, mainly the denial of benefits to farmers and firms engaged in rural development. To accomplish this function, NAD employs over 65 hearing officers located around the country, most of whom function as one-person offices and conduct hearings with authority to issue written decisions upholding or reversing the initial action. After the receipt of a hearing officer decision, appellants may also request a review from the NAD director. Typically, one-third of appellants seek review after receiving hearing officer determinations. The director's decisions are drafted by a group of nine review officers. In all, NAD issues over 2,500 written decisions annually.

<sup>2</sup> For examples, see USDA-NAD, *Strategic Plan*, [http://www.nad.usda.gov/about\\_stratplan.html](http://www.nad.usda.gov/about_stratplan.html) (last updated Jan. 18, 2007) and James P. Terry, *Fiscal Year 2006 Report of the Chairman, Board of Veteran's Appeal* 5, <http://www.va.gov/Vetapp/ChairRpt/BVA2006AR.pdf> (Jan. 10, 2007).

fragmented, and often ineffective when the agency does not integrate the course writing perspective into an overall program.

Moreover, these programs rarely integrate the training with the day-to-day activities of the organization's writers. To control the quality of written products, many government legal agencies dictate that hearing officers write decisions in boilerplate format, often contorting factual patterns and legal analysis into templates at the paragraph and sentence syntax level. Remedial instructional efforts may bludgeon writers by pointing out how previously written administrative decisions contain faulty grammatical traits, syntax errors, improper use of legal terminology, and deviation from the boilerplate. Indeed, at our agency, the National Appeals Division of the U.S. Department of Agriculture (NAD-USDA), we discovered a disturbing trend: the fragmented training approach and boilerplating parts of decisions actually caused stronger, more experienced and educated writers to write poorer decisions. These writers needed to be freed to perform under a different set of writing measures.

Improving the quality of administrative decisions at these agencies, therefore, presents a practical legal business challenge as well as the theoretical challenges embedded in the nuances of legal writing pedagogy. Amidst the pressure of issuing a large volume of decisions, agencies must contend with improving the varying skills of writers; delivering well-reasoned, clear, and reader-friendly decisions to the public; and measuring organizational performance based upon the quality of its written products.

This article proposes that a formal holistic assessment program can be an effective tool for confronting these challenges. In Part I, we describe holistic assessment and argue that adopting and emphasizing an evaluation strategy can be a powerful component of a legal writing program that results in improvements in writing across the organization. In Part II, we show the advantages of applying holistic assessment to administrative decisions. In Part III, we propose some guidelines for establishing a rubric and discuss how we adapted it to conform to the traits of legal readers and to judge the quality of legal discourse. In Part IV, we describe how to integrate holistic assessment for administrative decisions into a writing program. Part IV includes a discussion about metrics, reader protocols and training, and the use of evaluation session results by a writing program manager. Finally, in Part V, we briefly offer some suggestions about other potential applications of holistic assessment to legal writing.

## I. Description of Holistic Assessment

Formal holistic evaluation has been well embedded in the writing industry for many years,<sup>3</sup> but less so in law schools and legal writing organizations. Legal

---

<sup>3</sup> By the early 1980s, over 90 percent of English departments responding to a survey stated they used holistic scoring. See Edward M. White, *Holistic Scoring: Past Triumphs, Future Challenges* 83

writing professors and program managers may have adopted some elements also found in holistic evaluation, such as a rubric-based evaluation, employing multiple readers for high-stakes writing tasks, or portfolio grading. To distinguish these elements from the holistic assessment method, as it is used as a systematic theoretical and practical evaluation strategy for a writing program, it is helpful to begin our discussion by describing holistic assessment.

Holistic assessment comprises a scoring method based upon a rubric of identified writing criteria applicable to the subject area.<sup>4</sup> Raters, or readers, are encouraged to view the writing sample as more than the mere sum of its elementary parts; readers do not judge separately singular factors — such as treatment of topic, selection of rhetorical method, word choice, grammar and mechanics — that constitute a piece of writing. Rather, evaluators are asked to consider these and other factors as elements that work together; they score the writing sample on the “total impression” it makes upon the reader.

The scoring scale is usually a six-point scale, divided into two halves (a four-point scale is also common). Decisions that fall into the upper half — those scored four, five, or six — are satisfactory or labeled “mastery.” Lower-half decisions are unsatisfactory, or labeled “non-mastery.” Each score is described in terms important to readers. For example, a “two” might be described as “flawed writing,” while a submission that earns a “five” might demonstrate “clearly proficient writing.” After an informed reading, a rater first decides whether the writing sample is above or below the line. Based upon the pre-established holistic rubric of agreed-upon conventions, the rater then scores the essay.

To ensure statistical reliability, writing samples usually receive two or three “reads,” each by different readers. Adjacent scores (ratings that are within one point) and discrepant scores (ratings that vary by more than one point) receive thorough statistical scrutiny and inform test managers whether further reads are necessary. Training or “calibration” exercises precede evaluation sessions, giving readers a chance to apply the holistic rubric to previously scored essays and thus fostering consistency. Session leaders integrate “monitor papers,” essays with previously agreed-upon scores, into the sample pool to track reader reliability. In the end, scores can be analyzed across all writing samples by traditional data dispersion measurements. Overall judgments about good and poor writing, therefore, reflect the systemized results of reader and text interactions for specific documents assessed under controlled statistical conditions.

Holistic assessment emerged at the same time that teachers began applying new methods of teaching composition and writing to accord with modern

---

in *Validating Holistic Scoring for Writing Assessment: Theoretical and Empirical Foundations* 83 (Michael M. Williamson & Brian A. Huot eds., Hampton Press, Inc. 1993) [hereinafter *Validating Holistic Scoring*]. The prevalence of holistic scoring continues. The Scholastic Aptitude Test (SAT), for example, began holistic scoring of essays in 2005.

<sup>4</sup> There are many texts that provide an explanation of the holistic evaluation process. See Erika Lindeman, *A Rhetoric for Writing Teachers* 245-58 (4th ed., Oxford U. Press 2001).

language theory.<sup>5</sup> Many writing disciplines have embraced it as the primary standardized formal assessment tool.<sup>6</sup> Legal writing programs, however, can also take advantage of the attributes of holistic assessment. It privileges the reader's role in determining writing quality, adopts an inherently judgmental disposition, and — through a rubric — weighs writing standards as they affect rhetorical and content-driven aspects of the written product. After all, legal rhetoric is intensely judgmental and self-reflective, often calling explicitly upon the audience to evaluate and weigh both the content and form of an argument. Indeed, the main purpose of much legal rhetoric is to engender a particular response in a reader or group of readers.

Some of our earliest examples of legal texts show the same kind of judgmental and rhetorical self-consciousness. For example, in introducing his defense against the charge of corrupting Athenian youth, Socrates implores the judges, in evaluating his case, to subordinate his rhetorical style to the truth of his words:

[Am I making] an unfair request [of you?] Never mind the manner, which may or may not be good; but think only of the justice of my cause, and give heed to that: let the judge decide justly and the speaker speak truly.<sup>7</sup>

The setting for the above passage is a forum, similar to all legal proceedings, that reflects a method of pleading similar to the holistic evaluation method: the speaker submits the argument before a panel; the panel represents an interpretive community; and individual members of the panel make quantifiable assessments that eventually result in a single overall judgment.<sup>8</sup> The passage reflects a protocol very similar to the procedures we see in holistic assessment sessions.

In analyzing the passage further, we see also that Socrates addresses a cognitive element of judgment that compares to holistic assessment. As he cleaves the “manner of speech” from “truth,” Socrates exposes the millennia-old polemic in legal discourse: he asks the audience to decide what is good and true, inviting members to develop a scheme that evaluates both rhetorical

---

<sup>5</sup> We use the term “modern language theory” as shorthand for the description of several linguistic philosophies named by Gerald Wetlaufer: Saussurian linguistics, structuralism, post-structuralism, and semiotics. See Gerald Wetlaufer, *Rhetoric and Its Denial in Legal Discourse*, 76 Va. L. Rev. 1545, 1546 (1990).

<sup>6</sup> For a history of how holistic assessment emerged in twentieth century writing assessment, see Norbert Elliot, Maximino Plata & Paul Zelhart, *A Program Development Handbook for the Holistic Assessment of Writing* 27-43 (U. Press of Am., Inc. 1990). For a more personal view of the history of holistic assessment, see White, *supra* note 3.

<sup>7</sup> Plato, *Apology*, in *Euthyphro, Apology, Crito, Phaedo* 28 (Benjamin Jowett trans., Prometheus Books 1988).

<sup>8</sup> In this evaluation, the final judgment may only be a two-point scale (guilty or not guilty; one or zero), but it reflects the myriad of smaller judgments made by the citizen-judges of Athens. It is always important to remember that Socrates, in losing his case, failed to impress the citizen-judges.

effectiveness and content.<sup>9</sup>

While Socrates' explicit invitation is one that writers and texts extend implicitly to all readers, legal readers seem to bring to the task a heightened sense of judgment. The rhetoric of law often adopts modes of discourse that emphasize objectivity and syllogistic logic, commitments that drive readers to closure.<sup>10</sup> The emphasis on logic and closure is perhaps one reason why holistic assessment, as typically used in other disciplines, does not immediately appear suitable for legal writing evaluation. The fear may be that traditional holistic assessment may favor a form of rhetorical effectiveness over the logical and syllogistic content privileged in legal discourse. And it is true that we have found that in order to complement the judgmental disposition of legal writers and readers, holistic assessment needs some tweaking. It has to be adapted to readers confronted by legal rhetoric — whether they are judges, lawyers, professors, jurors, or students — who find themselves resolving both the effectiveness of the rhetoric and the outcome of the dispute between the parties. Holistic assessment must align the evaluator's judgment with the disposition of traditional legal readers as they judge the rhetorical effectiveness of an argument in relation to, or as opposed to, the truth of a matter at issue.

### *Emphasizing an Evaluation Strategy*

Some of the benefits of emphasizing an evaluation strategy for a writing program may seem obvious: formal writing assessment and measurement can become linchpins for evaluating writing quality and transferring legal writing theory and pedagogy to a plan for action and improvement. The benefit is that assessments can provide a way to measure the efficiency of the legal writing program that has been validated and statistically shown to be sound. The results of these assessments can therefore become the basis for a continuous cycle of organizational change and improvement in writing.

A formal assessment strategy promulgated throughout the writing program, however, can benefit the individual skills and cognitive processes of writers too. If implemented properly, the evaluation standards and protocols become part of the writing program interaction between reader, writer, and text. Over time, writers in the organization can serve as readers and evaluators. They read and judge the quality of their peer's texts and base their assessments on the overall

---

<sup>9</sup> This opposition is well established in the foundations of forensic rhetoric. Both Aristotle and Cicero also address this classical dilemma: Aristotle dissects the elements of truth, facts, and persuasive art as elements that play a part in forensic judgment. See Aristotle, *The Rhetoric and The Poetics of Aristotle* 20-23 (W. Rhys Roberts & Ingram Bywater, trans., Random House 1954). And he recommends that orators introduce forensic speech with an appeal to the audience, stating that the content of the speech may seem paradoxical. *Id.* In creating a fictitious dialogue between a lawyer and a rhetorician, Cicero devotes considerable space to arguing the relationship between legal knowledge, facts, and rhetorical effectiveness. See 3 Marcus Tullius Cicero, 3 *De Oratore* 169-95 (Loeb Classical Library) (E.W. Sutton & H. Rackham trans., rev. ed., Harv. U. Press 1945).

<sup>10</sup> Wetlaufer, *supra* n. 5, at 1550-52.

quality goals of the evaluation standards. When assessment is used properly in a writing program, writers know how their products will be evaluated, who will be the audience, and what will be the contextual conditions of the assessment. This information fosters a recursive writing process between reader, writer, and text that promotes change and growth throughout the life of a text.<sup>11</sup> Thus, by teaching the evaluation standards and practicing assessment protocols, the holistic assessment program also supports individual writing skill development. This approach is perfectly in accord with the theoretical view that writing development is a process-oriented activity. Further, writers can bring to bear on the assessment mechanism some of the information and writing tools they have acquired from other writing courses, which adds value to previously questionable training activities.

## II. Holistic Assessment of Administrative Decisions

Applied in a legal context, holistic assessment has four advantages for evaluating administrative decisions. The first advantage is that administrative decisions look very similar to analytic essays, a rhetorical form with which holistic assessment has an established history. Usually three to five pages in length, administrative decisions include an introduction, case narrative, background information, and a factual pattern. They point out the main arguments that need resolving, and consider and evaluate all sides of an issue. In coming to a reasoned conclusion, they apply regulatory criteria to a factual pattern through traditional logical and organizational patterns.

These are precisely the same elements found in analytic essays for other disciplines. For example, the analytic writing exercise for the Graduate Management Admissions Test (GMAT) requires two essays: one essay must analyze an issue question; the second essay analyzes an argument. The criteria for assessing both essays include the writer's use of organization, logic, and analysis as well as the writer's ability to understand and identify the complexities of an argument. Interestingly, the "issue" in legal writing compares with the analytic "issue" only in narrowing the field of topics that might be considered. Analytical essays are also common in academic areas, such as English and History, where writers summon facts and textual evidence to support or refute a thesis or proposition that is often based upon a theoretical model.

The second advantage is that holistic assessment also provides a practical business advantage for organizations that issue administrative decisions. Holistic assessment sessions can evaluate large samples of writing and produce measurable results that can be used and analyzed at higher management levels.

---

<sup>11</sup> See e.g. Linda L. Berger, *Applying New Rhetoric to Legal Discourse: The Ebb and Flow of Reader and Writer, Text and Context*, 49 J. Leg. Educ. 155, 155-56 (1999). Berger suggests that writers can self-impose an internal dialectic in the composition process, alternating positions as writer and reader.

One reason holistic assessment was implemented for the standardized Scholastic Aptitude Test (SAT), for example, was its speed and low cost in assessing large numbers of writing samples.<sup>12</sup> In evaluating large numbers of samples, holistic assessment can produce a higher level of statistical confidence in measuring the quality of organizational writing, and at a cheaper cost than other assessment techniques. The management of holistic assessments is relatively straightforward, which makes the task less daunting for managers. Managers can put systems into place for selecting and training assessment teams and coordinating high- and low-stakes evaluation sessions. They can integrate rubric elements into performance management systems, such as appraisals and awards or bonuses.<sup>13</sup> And since the output of these sessions is quantitative, they can integrate quality improvements into future organizational initiatives, such as training.

The holistic rubric and calibration sessions conducted before assessments provide a way to manage organizational writing quality because they provide a common language for discussing the quality of written work. Developing a common language is one of the most important aspects in increasing organizational writing quality. Every writer and reader in an organization possesses a common definition of writing quality; no longer is quality defined by what each senior partner or manager likes. Since holistic assessment provides a basis for writers and readers to judge the relationship between conventions, based on the whole document, using commonly defined and understood terms, organizational discussion about quality is more coherent, interactive, and dynamic. These discussions are very useful at our agency since we choose holistic assessment readers from the pool of writers in the organization.

The third advantage in using holistic assessment for administrative decisions is the similarity of specific disputed issues. Traditional holistic assessment evaluation exercises rely upon a prompt to which writers respond. The prompt is an important aspect for ensuring validity and reliability in holistic assessment because it controls the testing condition. Unlike a holistic testing environment, there is, of course, no universal prompt that generates a real life written work product for administrative decisions. For many administrative agencies, however, disputes center on similar patterns of disagreement about groups of regulatory criteria and language. Even when the case disputes vary, the culture of administrative decision writing creates a pattern of writing and analysis. This

---

<sup>12</sup> For administrative decisions, our experience is that readers need to take two to three times longer for reading and evaluating administrative decisions than analytical essays for other disciplines. In the reader protocols, we state that readers should conduct an “informed” uninterrupted read of the whole decision first. They may put small marks on the decision to keep track of areas they may later wish to return, but the goal is to give readers a chance to form an impressionistic assessment first. But even taking into account the longer time required to read decisions holistically, it is still more time efficient than other assessment methods.

<sup>13</sup> At the National Appeals Division, a portion of the annual performance bonus has been based for several years on a holistic assessment of samples submitted by hearing and review officers.



pattern serves a function similar to the prompt and lends itself to reliable holistic assessment results.

The final advantage for holistic assessment pertains to the wider and more diverse reading audience that administrative decisions often garner. They are written for appellants with differing educational levels who may or may not have legal representation. Poor writing signals to appellants that a given appeals system is unfair or designed to be intentionally ambiguous. Avoiding these impressions demands that authors make style (word choice, sentence and paragraph length, etc.) and clarity choices that meet the needs of both legal and non-legal readers. Also, many government appeal organizations post administrative decisions on the Internet to foster an impression of consistency, public fairness, and accountability.<sup>14</sup> Internet publication forces writers and managers to consider how decisions will survive wider public scrutiny. Holistic evaluation easily supports an intense and varied reader awareness approach to writing quality by simulating how a community of readers might react to an administrative decision.

### III. The Administrative Decision Rubric

In a holistic assessment, we can assume that legal readers will bring their heightened sense of judgment to assessing the quality of legal writing. The legal reading judgment is often reacting to a particular mode of discourse with a prescriptive voice. A reader reacting to the prescriptive voice, often being forced to agree or disagree with the narrator, will find that the assessment of rhetorical effectiveness becomes intertwined with evaluating the truth of the argument or validating the content of the case. The challenge for developing or adapting the rubric for holistic assessment in a legal context, therefore, is to integrate the components of truth and rhetorical effectiveness into that judgment.

In order to be successful in a legal context, holistic assessment must harmonize the impulse to judge legal writing as either a function of truth or content or both or simply on the basis of its rhetorical effectiveness — the dilemma that Socrates first identified. In emphasizing syllogistic logic and other modes of discourse that drive readers to closure, the commitment to these logical elements implies that lawyers are advocating objective truths. Thus, there may be a tendency to judge legal rhetoric based on one kind of language only. The norm for that kind of discourse, therefore, would erroneously become the standard for the rest of the language.<sup>15</sup> Thus, a writing assessment tool that neglects to

---

<sup>14</sup> The Administrative Procedures Act, 5 U.S.C. § 552(a)(2)(A) (2006), requires federal agencies to make public final opinions made in the adjudication of cases.

<sup>15</sup> See Emily Grant, *Toward a Deeper Understanding of Legal Research and Writing as a Developing Profession*, 27 Vt. L. Rev. 371, 382-84 (2003). Grant argues that historically the structure of law schools prioritizes speech over writing. In the oral Socratic tradition, writing is one step removed from speech, which is one step closer to thought or objective reality. In order to form valid and sound arguments and conclusions, classical syllogistic logic, for example, depends upon propositions and conclusions that are “true.” And of course it is important to remember that the



evaluate the truth, which for this discussion, we may now call the *apparent truth* component of legal writing, would be considered inadequate.

Yet, history and practice show that rhetorical effectiveness often wins the day in legal rhetoric too. The history of law practice confirms that rhetoric is an art that requires skills and techniques that have persuasive impacts on the readers. The various legal subject areas, such as contract, administrative, and criminal law, present different rhetorical challenges that require varying discursive forms. So an assessment instrument must have the flexibility to evaluate the effectiveness of these forms and conventions too.<sup>16</sup>

It is important, therefore, that a holistic assessment in a legal context, and especially the rubric used, permit prescriptive and persuasive elements to play out on the reader's battleground for textual meaning and value. They must compete as elements of the discourse with other conventions, such as fact vs. fiction, organization vs. style, or logic vs. passion. If the rubric is constructed carefully, it can integrate elements of legal rhetoric that would normally encompass apparent objective truths or content judgments. Some elements of such a rubric that are candidates for integrating truth and content are *case issue*, *logic*, *analysis*, and *organization*. These elements in a rubric can include descriptions that legal readers value as conventions when judging the quality of legal writing. These elements are parts of rubrics for other kinds of writing too; their emphasis in reader protocols and pre-assessment training sessions for legal writing can produce the same kind of statistical results as in other writing disciplines.

The proposed holistic rubric for administrative decisions comes from three main sources: 1) shared traits of rubrics from various other analytic essay rubrics in the writing industry; 2) results of surveys about writing problems in the legal field; and 3) our experience with writing and managing the writing of administrative decisions in a large organization.

The rubric is a six-point scale. Mastery scores are four (Competent), five (Strong), and six (Superior). Unsatisfactory scores are three (Marginal), two (Weak), and one (Incompetent). Since at our agency we focus on quality Internet publication of administrative decisions, we often state that we would be "pleased" to see upper-half decisions posted on the web; we would be "embarrassed" to see lower-half decisions posted on the web. These guidelines are simply another way to guide readers to forming an overall impression that

---

root of *logic* is *logos*, which means *speech* or *truth*.

<sup>16</sup> See Edward L. Rubin, *The Practice and Discourse of Legal Scholarship*, 86 Mich. L. Rev. 1835, 1838 (1988). This approach to judgment of discourse is similar to Rubin's approach to legal scholarship. Rubin argues that "normative discourse," a system of socially constituted modes of argument shared by a community of readers, may harmonize the opposition between objective truth and rhetorical effectiveness. *Id.* at 1891-1905. From this viewpoint, both aspects are embedded in the writer's judgment and the reader's evaluation of that judgment in the practice of the discourse. Similarly, a suitable holistic evaluation instrument would capture the value of normative discourse.

resonates with our real world application of holistic assessment. Administrative decisions that fall in the upper half (four, five, six) may have some errors that “distract” readers, but they do not “obscure” for readers the meaning, issue, or basis of a conclusion. Lower-half decisions exhibit characteristics that obscure for the reader the meaning, issue, or basis for a conclusion.

We propose the following five elements be included in an administrative decision evaluation rubric:

**Issue:** The decision clearly and correctly identifies all matters in dispute upfront and responds specifically to all aspects of the issues throughout the decision.

**Organization:** The case narrative has a clear beginning, middle, and end, and it connects parts with clear transitions. It has strong ideas to introduce and organize paragraphs.

**Analysis & Logic:** The decision effectively analyzes all sides of the issues with thoughtfulness and depth. It uses valid logical reasoning that integrates well-chosen facts and regulations to support sound conclusions. It effectively responds to a faithful representation of the parties’ point of view and refutes those views when appropriate.

**Style:** The decision employs a readable style that is clear and concise for the level of evidence. It demonstrates control of language, including appropriate word choice and sentence variety.

**Mechanics:** The decision is generally free from errors in mechanics, usage, and sentence structure. It is free from grammar or spelling issues that would be highlighted in Microsoft Word.

For these criteria, the rubric identifies strengths and weaknesses that characterize the gradations of the scores for each element. For example, a decision that scores six tends to fully satisfy the basic definitions of the elements. Using the *organization* element as an example, a “six” decision has a clear beginning, middle, and end, with clear transitions and strong ideas that introduce paragraphs. A “four” decision, however, may have some unnecessary repetition or “breaks” in the story that may distract, but not confuse, the reader. Or, main ideas in some paragraphs in a “four” decision may not always be evident. In the lower-half range of *organization*, the decision may be poorly organized or have gaps that confuse the reader. It may also have poor paragraph organization. At the lower end of the *organization* spectrum, the “story” of the case may lose the reader. There are other gradations for other holistic categories.<sup>17</sup>

---

<sup>17</sup> For a full delineation of the NAD-USDA rubric for evaluating administrative decisions, consult the NAD Style Manual, <http://www.nad.usda.gov/Forms/NAD%20Style%20Guide>

It is important to remember that the criteria guide an overall evaluation of the writing. Holistic assessment does not establish a catalogue of precise individual errors in each category that might appear; instead, the criteria help the reader decide what impact any errors have on the overall quality of the writing sample. With continuous close reading, even a “six” decision can contain minor errors or distractions. Readers learn that upper-half decisions may have errors in one or more elements that distract readers, but still retain their overall coherence and persuasiveness. For lower-half decisions, the writing errors may include such obvious flaws that they obscure the reading experience. The relationship of the elements gives holistic assessment its dynamic life.

Two rubric categories, *issue* and *analysis and logic*, address elements of discourse that legal readers particularly value when judging administrative decisions. Both these categories ask legal readers to weigh elements of objectivity and content, aspects of the prescriptive voice that must be integrated into a legal writing holistic rubric. Issues and issue statements are analogous to thesis statements and proposals in other types of analytic writing. Like other rubrics for those essays, the legal writing rubric asks readers to evaluate some rhetorical traits of issue statements. For example, good administrative decisions state issues upfront; the rubric favors specific issue statements over general issue statements; and in exploring both sides of the argument, the rubric calls for the discussion to respond specifically to the issues throughout the decision.<sup>18</sup>

But for the *issue* category, the administrative decision rubric also asks the reader to judge whether the writer has “correctly” identified all the matters in dispute. This judgment is often possible, even though the reader has neither access to the case record nor independent knowledge of a case. Decision writers choose to select specific factual patterns, regulatory citations, and organizational patterns for arguments that provide insight into whether the dispute was correctly identified. For example, a writer who states that the issue is about Appellant’s income threshold, but selects facts and argues the rest of the decision around an Appellant’s medical condition, has clearly missed the issue.<sup>19</sup> This contradiction will likely confuse the reading experience, thus driving the assessment to an unsatisfactory level. As a component of issue effectiveness, a

---

%20Manual.pdf (June 2005).

<sup>18</sup> At the NAD, we have also used the holistic rubric to evaluate writing for over 150 writing samples submitted by job applicants. Most applicants submit a brief, memorandum, or prior administrative decision. The rubric approach of analysis is quite telling: for example, approximately 20 percent of candidates thus far have submitted samples without explicit issue statements. Of those samples, about half seem to display a “sense of the issue,” either through some other convention or by the reader inferring the issue from the arguments; the other half simply have unfocused legal rhetoric. Yet, virtually all candidates are law school graduates with significant experience practicing in the legal field.

<sup>19</sup> We have found that disagreement among the readers about the issue is often the main reason why scoring may be discrepant, especially in instances where one or two readers score the administrative decision as mastery, while a third reader may score it as unsatisfactory. It is important to investigate these discrepancies as part of continuous reliability monitoring.

judgment about issue correctness supports the basic principles of holistic assessment: the elements of the rubric, properly executed, work with each other to impress the reader.

Most legal readers who assess administrative decisions find that the element of *issues* in the rubric resonates strongly with the element of *logic and analysis*. The *logic and analysis* element puts the apparent objective truth at play in writing assessment. While readers, legal or otherwise, want texts to make sense on formal and intuitive levels, legal writing must employ stricter standards of formal logic to achieve this goal. As a result, the administrative decision rubric for legal writing contains language asking a reader to evaluate whether the text demonstrates valid and sound legal syllogisms to support a conclusion or whether facts and analysis of factual patterns aptly apply a rule or law to reach a sound and valid conclusion. These criteria compel readers to determine if writers have adopted appropriate rhetorical modes in stating an objective factual pattern. They also evaluate whether legal citations properly support the premises of legal reasoning displayed in analytical paragraphs.

The rubric refrains from encouraging a specific format or rhetorical pattern for displaying logic or legal reasoning. One such pattern, for example, might be the IRAC method (Issue, Rule, Analysis, Conclusion), or something similar, for displaying logic and demonstrating a valid conclusion by reasoning from a rule through analysis to conclusion on an issue. Formats like IRAC privilege rhetorical form, but it is the logical content of a passage or paragraph that produces a persuasive impression upon the legal reader. Paragraphs written in prescribed forms may very well be unsound, invalid, or confusing.<sup>20</sup> Conversely, a text may demonstrate sound and valid logic even if it varies from the form (i.e., all the premises and conclusion are not contained in one paragraph or a sequence of paragraphs). In instances where premises and conclusions may appear to be scattered or fragmented, the distraction to the reader may be traced to the element of *style*, which encourages writers to use topic sentences with appropriate detail and conclusions, a criterion that supports the rhetorical delivery of logical syllogisms.<sup>21</sup>

As readers judge the relationship between the elements of *issues* and *logic*, we see the connections within elements of the rubric merge with assessing rhetorical effectiveness. Legal reasoning produces conclusions that answer definitively the

---

<sup>20</sup> See Kristen K. Robbins, *Paradigm Lost: Recapturing Classical Rhetoric to Validate Legal Reasoning*, 27 Vt. L. Rev. 483, 483-87 (2003). Robbins makes the excellent point that instruction in the precise mechanics of syllogisms is necessary to understand fully the form of legal reasoning and the basis for short-cut formulaic conventions such as IRAC. At the NAD, we were able to demonstrate that some initial training in classical reasoning and logic contributed to an increase in holistic scores in less than a year's time.

<sup>21</sup> See *id.* at 487-517. Robbins shows and categorizes some examples of faulty reasoning typical of legal writing. We have gone over these examples, judging them through the lens of the holistic rubric, and have determined that the rubric would take these kinds of reasoning errors into account.

questions that issues ask. As readers evaluate whether the issues were correct, they see that the writer adopted a mode of logical discourse in an administrative decision that ensured objective reasoning to arrive at conclusions. These elements spill over to more ordinary, but no less important, rhetorical traits too. For example, logical conclusions — the end result of a proper syllogism — become the topic sentences for paragraphs that display deductive reasoning. This analysis supports elements of *style* and *organization* in the rubric. The issue questions and the legal syllogisms become the focus of paragraph and decision organization. In the end, the holistic rubric permits both content and rhetorical effectiveness to work together in a reader-writer-text medium to increase perceived quality.

#### IV. Integrating Holistic Assessment into a Writing Program

Holistic assessment works best as a linchpin for improving legal writing when an organization is oriented toward its basic principles in other parts of the legal writing process. As a starting point, all writers should know the content of the rubric and have opportunities to assess and analyze decisions based on its components. This exercise places writers in the position of readers. Since most administrative law organizations publish administrative decisions on the Internet, these discussions can extend to how the rubric supports readability for a wider audience.

The holistic rubric can also promote the process and discussion that supports organizational reviews of drafts that provide feedback to the writer before a decision becomes published. These reviews can be conducted by peers and supervisors. Even though it is helpful for reviewers to use the language of the rubric when providing feedback, that feedback should not include a holistic score or some final quality judgment about the decision. These scores would not be reliable and consistent as they are not the product of a calibration exercise and more than one reader. They would also tend to advance the writer too hastily through the draft-to-publication process.<sup>22</sup> Holistic assessment is most effective during high-stakes sessions that are defined by the organization and preceded by

---

<sup>22</sup> See Peter Elbow, *Everyone Can Write: Essays Toward a Hopeful Theory of Writing and Teaching Writing* (Oxford U. Press 2000). At the NAD, we borrow heavily from Peter Elbow's guidelines for integrating the writing process and feedback into our writing program. We allocate time in the writing process so that writers can touch all the important markers between "private writing" and publication. We do not mix a scoring assessment with peer review feedback. Supervisors conduct supervisory peer reviews with the mindset of facilitators who are helping writers prepare final drafts before Internet publication; thus, their feedback comes in the form of a cordial letter — with complete sentences and well-formulated paragraphs — that emphasize their "reader response" to the final draft of the administrative decision. And their supervisory peer reviews are reviewed quarterly with the same kind of feedback. These are strategies to avert the "war between readers and writers" about which Elbow is concerned.

training.<sup>23</sup> High-stakes sessions in our organization, for example, are assessment exercises that affect awards and performance evaluations or sessions that are designed to identify model decisions for future assessments.

One advantage an administrative agency may have over law schools or individual legal writing courses is that its employees stay with the organization and the writing environment over a significant amount of time. Longevity provides an opportunity for the organization to conduct periodic writing quality checks each year to assess whether writers in the organization consistently recognize strong and weak writing. The Internet or e-mail provide convenient and low-cost methods to distribute a decision, collect scores, and inform the organization about the results. At our agency, most people find these calibration sessions conducted during the year more informative and interesting than actual high-stakes assessments. Because of the efficiency of these periodic calibration sessions, we have been able to assess over 300 decisions per high-stakes assessment.

These periodic quality checks help speed up calibration when the organization intends to conduct a holistic assessment for high-stakes writing. The pool of evaluators can come from the same pool as the writers in the organization. With proper statistical control and management, we have seen the same levels of reliability as with holistic assessment results in other disciplines. For example, at one of our agency national training conferences, we conducted an analytic writing exercise with members of a nationally recognized testing service. This session validated our rubric and the ability of our own employees to score essays and administrative decisions with the same statistical reliability as the consultants.

### *A. Metrics for Administrative Decisions*

As stated previously, one of the advantages that holistic assessment provides for large legal organizations is that systematic implementation can report valid and reliable data on the writing strengths and weaknesses of an organization. Any discussion about holistic assessment metrics, however, must begin with a brief description of holistic assessment as a psychometric evaluation instrument.

Classified as a form of direct assessment, holistic assessment gains support from linguistic perspectives such as reader response, semiotics, and other views that connect meaning to a process view of language. From most of these perspectives, the act of reading and writing both construct meaning. Further, authorial intention dissipates as the reader becomes more prominent. Proponents of direct assessment focus on the cognitive processes of writing, including the social and linguistic contexts in which it occurs. Reader and writer become more

---

<sup>23</sup> We do formal holistic assessments at national training conferences, for end of the year awards, and for periodic contests. Much of this work is facilitated by electronic transfer, storage, and validation of assessments.

like information-processing mechanisms, processing a complex set of semiotic cues. The assessment of a text, therefore, is an assessment of the interplay between the reader, the writer, and the linguistic and social context of the discourse.<sup>24</sup>

Direct assessments can be highly contextualized. For writing assessments, the context may come from a prompt or some other impetus that causes the writing. The prompt means that a direct assessment occurs under controlled conditions. Since there is not one right answer to a test question or prompt, direct assessments measure divergent knowledge. In assessing texts, readers evaluate and compare samples produced under the same conditions. Through a rubric, and after training or instruction, readers consider the complete systematic conditions that produce a text and how various elements work together to affect quality. The strength of holistic assessment is that it calls upon the full array of writing skills that reflect real world writing conditions.

As do other direct assessments, holistic assessment relies heavily upon a modern linguistic understanding of the reader. From this theoretical perspective, the reader is no longer a biographical person, but the name of a place where semiotic codes are located and processed. The reader becomes a function that processes signs and enables them to have meaning. Holistic assessment, therefore, applies the perspective of semiotic inquiry. Semiotic inquiry describes how achieving a system of conventions is responsible for meaning;<sup>25</sup> holistic assessment describes how well those systems of conventions achieve meaning for the reader. If it is true that the reader becomes the repository for the processing of codes that account for the intelligibility of a text, then holistic assessment is taking the pulse of quality at a key place in the process.

In using the rubric, holistic assessment also adopts some principles of cognitive processes from the early twentieth century Gestalt school of psychology. Holistic assessment applies these psychological principles to the readers of texts. According to these principles, cognitive processes are not additive or elemental. Instead, cognition perceives a phenomenon as greater than the sum of its parts.<sup>26</sup> Based upon these cognitive principles, proponents argue that it is possible for readers in a holistic assessment to rank writing samples if the samples are produced under controlled conditions. Further, readers can identify similar characteristic of papers and agree upon the value of these characteristics for any given particular assessment. After training exercises, readers are able to agree on adjacent scores for individual essays. And finally, readers accomplish all of the above while weighing the relationships among

---

<sup>24</sup> Michael M. Williamson, *An Introduction to Holistic Scoring: The Social, Historical, and Theoretical Context for Writing Assessment*, in *Validating Holistic Scoring*, *supra* n. 3, at 9-25.

<sup>25</sup> For more information on the reader's role in the reading act, see Jonathan Culler, *The Pursuit of Signs: Semiotics, Literature, Deconstruction* 38-43 (Cornell U. Press, 1981).

<sup>26</sup> Elliot, Plata & Zelhart, *supra* n. 6, at 15-17. The authors compare the basic principles of the holistic rubric to the cognitive processes of Gestalt psychology.



hypothetical standards, the total effect and impression of the writing sample, and the varying social-linguistic conditions of the testing environment. Moreover, writers know how their products will be evaluated, who will be the audience, and what will be the contextual conditions of the assessment.

Holistic assessment and direct assessments bring a specific approach to the issue of context in writing assessment to achieve valid and reliable test results. The direct assessment approach varies slightly from the approach used in indirect assessments. Generally, objective empiricism is a presumed strength of indirect assessment, because proponents claim to eliminate context, therefore ensuring that the results of assessments are objective evidence of writing ability. On the other hand, direct assessments, like holistic assessment, acknowledge and manipulate the context, through a prompt, to trigger the semiotic mechanisms that produce a writing sample.<sup>27</sup> It is not surprising, therefore, that holistic assessment continuously strives to address concerns about objectivity and reliability and validity of its results. Proponents are trying to show that the semiotic mechanisms employed by readers and writers in holistic assessment behave consistently and predictably.

The main statistical challenges for holistic assessment are predictive validity and instrument and inter-rater reliability. How well the test predicts future writing success, analysis of the holistic prompt, and agreement among readers are all areas of inquiry that address these challenges. Validity and reliability in holistic assessment provide a rich canvas for academics to discuss statistics, testing conditions, and the reporting of results. This discussion can become quite complex even though complex statistical analysis is usually not a trait associated with writing practitioners. As a practical matter, however, a good writing program that uses holistic assessment must collect data that reports information in the following areas: 1) the test must predictably assess writing results over time; 2) the writing program must analyze whether different sets of readers consistently score essays produced under similar conditions; and 3) the program must analyze how consistently readers agree upon individual scores for essays. These statistics take into account all sources of measurement error — the writer, the test, and the scoring protocols.<sup>28</sup>

---

<sup>27</sup> Williamson, *supra* n. 24, at 29. Williamson provides an excellent discussion that analyzes how distinctions between indirect and direct assessments suffer under the weight of their own defining characteristics. In linguistic theory, it is now common to point out that these disputes themselves are dependent upon the signs that convey them. At an elemental level, they depend upon constructs that are shaken at the outset by the indeterminacy of the sign itself. The reasoning that supports this perspective can be found in Jacques Derrida, *Of Grammatology* 36 (Gayatri Chakravorty Spivak trans., The Johns Hopkins U. Press 1976). It may be helpful to apply Derrida's notion of the *graph*, to this discussion. Derrida argues that the graphic form of words is unstable and with an ungraspable point of origin.

<sup>28</sup> Roger D. Cherry & Paul R. Meyer, *Reliability Issues in Holistic Assessment in Validating Holistic Scoring*, *supra* n. 3, at 109-38. As part of this thorough discussion of reliability and holistic assessment, Cherry & Meyer go through the advantages and disadvantages of several options for

### ***B. Reader Protocols and Training Sessions***

In line with results for the writing industry, we have found that an administrative law organization can achieve statistically valid and reliable results with holistic assessment. These results depend, however, upon the proper implementation of protocols and training sessions over time.

The common practice for training readers prior to evaluation sessions is first to familiarize them with the scoring rubric. Writing program managers then submit previously scored decisions to the reader-judges for practice scoring. After the judges score the decisions, session leaders provide justification for the accepted score. A give-and-take session usually follows to allow judges and session leaders to discuss their variances.

One challenge that confronts new assessment programs is finding model decisions that reflect representative scores. Most likely, legal writing programs that use holistic assessment systematically will fall into this category.<sup>29</sup> Meeting the challenge of finding model decisions is partly art and partly science. These model decisions should come from previous assessments that have had similar prompts; or, as is often the case, they come directly from the current pool of samples during the assessment. A small team of “experts” determines the scores and provides in-depth justification of the strengths and weaknesses of the samples. This information is passed on to prospective judges in training sessions.

Administrative legal organizations have other sources for finding representative decision writing that reflects various scores. Government appeal organizations often provide one or more levels of review of initial appeal decisions. As these decisions filter through the review process, either being upheld or reversed by higher authority, writing program managers can pay attention to both the positive and negative responses reviewers have to the writing. These decisions are often good candidates for training sessions. More complex or controversial decisions that stand the test of time and the administrative review process are usually the sign of mastery level decisions.

As the writing program matures, legal writing professionals will be able to assemble a database of representative writing. Astute writing program managers will ensure that model decisions receive many “looks” and feedback in various forums before they are submitted to a team that will determine their final representative score. Only then will the team provide feedback for the decision by applying the rubric to the content of the decision.

### ***C. Evaluation Session Results***

We have found that when inter-rater reliability goes below 85 percent in a holistic assessment, the prior scores of decisions identified by the mismatched

---

resolving discrepant scores by readers.

<sup>29</sup> Initially, in the NAD, it was difficult finding decisions at both ends of the scoring continuum: decisions rated one or six.

pair of readers need to be analyzed.<sup>30</sup> For administrative decisions that provoke discrepant scores, introducing additional reads may actually increase variance. The best solution is to refer those decisions to a previously identified small team of calibration experts to resolve differences. Most likely this team will be the readers who identified representative decisions at the beginning of the session. In these instances, the team should attempt to articulate why the decision may have received discrepant scores. Our experience has often been that administrative decisions in this category have an element of complexity or a particular aspect of the issue and analysis has compelled some readers to react more harshly than others.

Some relatively straightforward analysis of central tendencies can demonstrate whether the decision writers in an organization are calibrated. For example, assume that the writing program manager in an appeal agency distributes an administrative decision (Decision X) to all the writers in the organization as part of a quarterly quality writing check. From past holistic assessments, training classes, or an expert team analysis, the known score of the decision is five. Sixty-four writers in the organization — writers who are now acting as readers — might typically submit the following assessment scores:

Score	Number of Times Scored
Six	12
Five	25
Four	19
Three	8
Two	0
One	0

---

<sup>30</sup> In high-stakes assessments conducted over the past three years, we have achieved inter-rater reliability exceeding 90 percent.

A quick spreadsheet analysis and histogram can show some useful information:

Results for Decision X

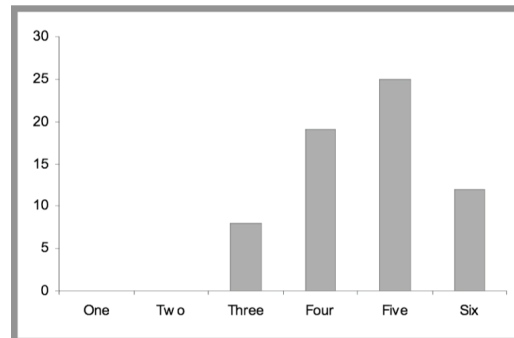
Mean: 4.6

Mode: 5

Median: 5

Standard deviation: .93

Skew: -.78



Our experience with these exercises demonstrates the following guideline: when the mean score of the assessment approaches the known score of the decision, and the standard deviation is less than 1.0, the organization approaches appropriate consensus about the quality of the writing. For a large pool of readers, the 1.0 guideline would mean that over 67 percent of the organization has submitted adjacent scores. If the standard deviation is less than 1.0 with other representative samples, combined with other favorable statistics, the program manager can have reasonable assurance that a holistic assessment conducted for hundreds of decisions will produce reliable results. Moreover, ensuring statistical reliability is important so writers can be confident of quality when they see strong decisions and make adjustments to their own writing.

As the histogram and the statistics above show, there are several signs from Decision X that the agency is almost calibrated. First, both the mean and median scores, important measures of central tendency, are equal to the known score of the decision. As the standard deviation reflects, fifty-six readers have scored the decision within one point of the known score. In this instance, the writing program manager can take heart that much of the organization has a consensus about a strongly written decision.

There are also indications, however, that another round of calibration should continue. First, approximately 15 percent of the organization scored Decision X as unsatisfactory. These “three” scores brought the mean score below the known score, but they also show some confusion about what constitutes satisfactory writing. Even though the standard deviation (.93) is within the suggested 1.0 guideline, the measure of skewness suggests that the scores for this decision are skewing negatively. (Generally, measures of skewness between .5 and 1.0 or -.5 and -1.0 show moderate skewness.) Since Decision X was clearly a mastery decision, enough so to warrant twelve readers to submit scores of six, the writing program manager must investigate what elements in the decision affected a group of readers negatively. To investigate these elements, the writing program manager should discuss the scoring with readers or have further training sessions. With repeated calibration sessions that display similar data, organizations can attain satisfactory calibration after four or five decisions.

An “ideally” calibrated organization will submit scores with central tendency statistics that approximate and support the known score. This ideal is never achieved, of course, so there are some additional issues to consider when analyzing scores in calibration exercises. First, the program must take into account the number of readers in a calibration session. For example, in the above session, the writing program manager was calibrating an entire organization of sixty-four writers. For calibrating fewer judges, statistics will show more sensitivity, as the number of calibrated readers becomes smaller. Conclusions about calibration, therefore, must adjust accordingly. For example, if the exercise was calibrating only ten judges, and two of those judges were continuously submitting discrepant scores, the writing program manager should investigate and resolve those discrepancies more quickly.

The known score of previously scored decisions will also affect data interpretation. Decisions at the “six” level, for example, will certainly skew negatively. (There is no more room on the right side of the scale for the data distribution.) Often in calibration sessions, readers initially have difficulty submitting scores at the extreme ends of the spectrum. In the case of a “six” decision, however, the mean score should be above 5.5 and the writing program manager should certainly analyze all scores that fall below a four.

Finally, once reliability has been achieved, an administrative appeals agency over time may monitor the quality of its written products, and, hopefully, demonstrate increased quality in its decision-making. As noted previously, at first our agency had difficulty finding decisions that rated as a six, the highest score. Since holistic assessment has been implemented over the last three years, the mean score of decisions has risen over one point, the number of “six” decisions has increased dramatically, and the number of writers who have received high performance awards based on decisions rated as a six has also markedly increased.

## V. Future Issues for Holistic Assessment of Legal Writing

Future research in the use of holistic assessment for legal writing can follow the areas of inquiry already established in writing assessment for other disciplines. Some basic questions emerge: 1) in addition to administrative decisions, can holistic assessment produce the same standards of validity and reliability with other rhetorical forms of legal writing? 2) what writing prompts call upon the appropriate writing skills to substantiate claims about the effects of pedagogy and training for law students and lawyers? 3) in law schools, can holistic assessment be used to allow students to graduate from legal writing programs, to place them in advanced programs, or to judge briefs submitted in moot court exercises? 4) can holistic assessment be used with other forms of writing assessment and with other parts of the law school curriculum to inculcate a consensus about writing

quality throughout the curriculum?<sup>31</sup>

One of the more interesting initiatives in holistic assessment is the automated scoring of essays. Many academic assessments have an essay portion that receives two scores — one from a human reader and one from a computer. Using natural language software programming, computer scoring is able to predict with a very high rate of reliability the scores that human readers would have given an essay. Educational testing companies now offer computer-graded scoring and feedback to student writers who submit essays on line.

We have investigated computer-graded scoring for administrative decisions. Our research shows that while it is theoretically possible, some customization and additional modeling must occur for this form of assessment to become effective and reliable. One component of the computer-graded scoring initiative compares to the judgmental dilemma legal readers face when evaluating texts: generally, computer-grading programs evaluate rhetorical markers in texts that can reliably predict a holistic reader's response; but as we have argued, legal readers, while judging writing, also evaluate content and apparent truth, especially for issues and logic. Computer assessment, therefore, must become more content-driven. Although content-based analysis is a present component of natural language programming, it is not as fully developed as assessment based on rhetorical markers.

It may be that the future of legal writing assessment will be implemented through some form of artificial intelligence. If that is true, however, then artificial intelligence applications will have to don the same components of the judgmental mindset that legal readers have imposed upon texts for thousands of years. If computer assessment software is to find acceptance by the legal community, writers will have to gain the same level of comfort and faith that Socrates did when he asked the judges to decide “justly” about his rhetoric. Socrates submitted his case, his rhetoric, and inevitably his life to an informed community of readers. And in doing so, he acknowledged the validity of the evaluation protocols placed upon his rhetoric and the evaluators' right to impose judgment upon him.

---

<sup>31</sup> We do not intend for holistic assessment to preclude other forms of indirect assessment or portfolio grading. In fact, some of these other evaluation tools, especially portfolio grading, can complement holistic assessment.