

# Undergraduate Technical Writing Assessment A Model

**Carol Siri Johnson**

*New Jersey Institute of Technology*

**Norbert Elliot**

*New Jersey Institute of Technology*

**Abstract.** This article describes an assessment process developed for an undergraduate technical writing course at a public research university. To document program outcomes, we used a variety of statistical methods. To describe our process, we present longitudinal performance data collected over five years (fall 2004 through spring 2009) on 636 students. After providing a brief overview of the measurement concepts and statistical tools that we employ, we describe our process in five phases: designing the variable model to ensure construct validation; designing the assessment methodology to ensure content validation; designing the sampling plan to address economic constraint; designing the data analysis to articulate the validation argument; and using the assessment results to ensure consequential validation. Our intention is to provide a model that can be applied to other institutional sites and to encourage others to use it, tailoring the model to their unique needs.

**Keywords.** assessment, constructed response, educational measurement, evidence-centered design, ePortfolios, program assessment, technical communication, writing assessment

**T**echnical writing instruction is increasingly important in the twenty-first century. Often dismissed as a mere skill, technical writing is a vehicle for empowerment in our multinational, multicultural, multilinguistic global culture. As contemporary society has become more dependent on knowledge, Charles Bazerman and Paul Rogers (2008) observed, the economic value of information and the texts reifying that information have both increased. As it became apparent that the digital revolution was to have an impact similar to that of the industrial revolution, writing in the professions began to draw increasing attention. As Anne Beaufort (2007, 2008) has demonstrated, such attention to professional writing has yielded research on the importance of workplace writing, the processes and practices that support it, the impact of

institutional structures, the role of shifting technologies, and the socialization processes that occur as writers gain workplace experience.

As might be expected, teaching and assessing technical writing remains complex.<sup>1</sup> Instruction often requires tasks that use complex print, auditory, and visual processes; assessment requires that such tasks be evaluated. Within a global culture fueled by digital innovation, the traditional demand that students be able to communicate in a variety of ways to a variety of audiences takes a new turn: Students must be able to demonstrate their ability to effectively design digital environments to host their work, must be able to demonstrate clear style and accurate usage in their texts so that no unnecessary burden is given to readers, must be able to demonstrate mastery of complex tasks and bring relevant content to those tasks, and must have a firm control of tone so that audiences are aligned with a document's message. Graphic cohesion must be apparent so that task unification is achieved, and key sources of information must be referenced by the student so that the voyage through the vast digital infrastructure is transparent.

In this article, we describe our assessment-driven instructional model. We have rendered the qualities of technical writing quantifiable to more clearly study them, thereby improving our instructional processes. We separated the features of successful technical writing into a variable model, assessing student performance each year in a system of longitudinal study. The number of students within our five-year research sample is large ( $n=636$ ); due to the long-term nature of the study, we can be fairly assured that our resulting analysis yields important information.

---

<sup>1</sup> We use the term *technical writing* throughout this article. We recognize that technical communication is the overarching construct (Brennan, 2006; Kane, 2006) of *Programmatic Perspectives* and the Council for Professional in Technical and Scientific Communication. In addition, we recognize that the Society of Technical Communication is structuring recent initiatives to build a body of knowledge around concepts of technical communication (Coppola, 2010). A similar emphasis on technical communication is held by the Association of Teachers of Technical Writing. The research reported in this article, however, is based on an undergraduate course that asks students to achieve proficiency with concepts surrounding the construct of technical writing. Technical communication, a related yet distinctly different construct, involves variables different from those described in the model we present in this article. For our work in validating assessment efforts based on the construct of technical communication, see Coppola & Elliot, 2010. Although the design and analytic methods we present in this article hold across both constructs, the practice of assessing, for instance, student ability to demonstrate principles of clear style in a proposal (a task related to technical writing) is quite different from asking a student to demonstrate oral presentation skill in a podcast (a task related to technical communication).

The purpose of this article is to unpack our processes so that others can see them, tailor our model to their specific institutions, and design a process of outcomes assessment that addresses accountability demands from regional and program accreditation agencies (Middaugh, 2010). The method we advocate holds the potential to ensure that technical writing instruction becomes a tool of empowerment for all shareholders—from the students who must be skilled in known and emerging literacies, to the instructors who teach them, to the administrators who must present outcomes assessment for accreditation processes.

Our background assumptions are based on assessment research, especially the work of Brian Huot (2002), Edward M. White (2005), and Robert Broad (2003) in their unified call for contextualism. As Huot (2002) reminded us, writing assessment must be “site-based and locally controlled” (p. 14) because “writing teachers and program administrators must begin to see writing assessment as part of their jobs” (p. 19). This call for localism is true both for the purposes of accountability and for the creation of instructional consensus. Although it is possible to teach without group interaction, we have found that our interactions with instructors have allowed us to expand our pedagogical horizons. The collaborative model upon which our assessment rests has allowed us to theorize and implement our work, thereby rendering third-party intervention unnecessary (see, e.g., Johnson & Elliot, 2004; Coppola & Elliot, 2007). We have shifted from the traditional university culture of isolation to a community-based culture of self-assessment. Instead of repeating the past, self-assessment creates dialogue that enables educators to seek out and incorporate change. This frame of reference is a major shift in educational culture away from the antiquated system of inputs and outputs and has proven to be an approach that can yield important results.

Concurrent with our reliance on research in the field of writing assessment, we have also relied on principles of evidence-centered design (ECD) advanced by the educational measurement community. Fundamental to ECD theory, as advanced by Robert J. Mislevy, Russell G. Almond, and Janice F. Lukas (2003), is that complex assessments must be designed to support course goals *from the beginning*. By designing a performance assessment that will evoke robust student work (i.e., constructed responses requiring precise tasks to be performed rather than general reactions to a prompt) and planning in advance for the kinds of evidence that will be warranted, ECD compelled researchers to think about proof and consequence from the very first stages of program inception (Mislevy, 2007). Thus, the plan for a chain of reasoning (Brennan, 2006, p. 31) to provide evidence should be part of the design itself, not an act that occurs after the design is executed (Messick, 1994). As the *Standards for Educational and Psychological Testing* (American Educational Research Association, American

Psychological Association, and National Council on Measurement in Education [AERA, APA, & NCME], 1999)<sup>2</sup> remind us, validity, or “delineating the knowledge, skills, abilities, processes, or characteristics to be assessed,” is the most fundamental consideration in developing tests (p. 9). Above all, an assessment must be valid—it must capture the subject matter the student is required to learn. Because we have designed our curricular program to yield information, ECD has helped us to design an assessment program that addresses issues of validation.

In addition to focus on localism and evidence-centered design, the program we present is cyclical—the results of the assessment are used in modifying the course, which then modifies the assessment itself. This process, colloquially termed “closing the loop,” embodies a drive towards assessment processes that everywhere connect accountability and instruction. The Accreditation Board for Engineering Technology (ABET) pioneered this idea of using assessment outcomes as input for change: the ABET *Criteria for Accrediting Engineering Programs* (2006) mandates each program under review must publish educational objectives, maintain regular assessment, and—here is what is new—*use evidence to document that the results of the assessment are used to improve the program*. Our regional accreditation agency, the Middle States Commission on Higher Education (MSCHE) (2006), has similar demands for the assessment of educational outcomes. This type of assessment has been explored in engineering, corporate training, and the military, but it can also be applied to the processes involved in technical writing instruction at the upper division level (and in composition instruction at the first-year level). The implementation may be complex, but the process of defining the variables of instruction, creating a curriculum to deliver them, and assessing the outcomes might be likened to community-based group artwork, where all participants have input and the final forms can be seen in the assessment (Chicago Public Art Group, 2009). Because our work is designed to “emulate the context of conditions in which the intended knowledge or skills are actually applied” (AERA, APA, & NCME, p. 137), our program may be categorized as a performance assessment (Lane & Stone, 2006).

After a brief background discussion on measurement concepts and statistical tools, our performance is presented in five acts: designing the variable model as a way to ensure construct validation; designing the assessment methodology as a way to ensure content validation; designing the sampling plan as a way to address economic constraint; designing the data analysis to articulate the validation argument; and using the assessment results to inform the most important consequence of program assessment—improved instruction. By defining our goals (the variables of our assessment), building a system to teach them, and crafting an assessment to test them, we have become more aware of ourselves

---

<sup>2</sup> Hereafter referred to simply as *Standards*.

as teachers. It is in this spirit that we wish to contribute to the continuing Council for Technical and Professional Communication (CPTSC) Research Assessment Project (2008) and the related field of writing assessment (Condon, 2009).

## **Using Measurement Concepts and Statistical Tools in Technical Writing Assessment**

At first glance, statistics and writing seem diametrically opposed: One is analytical, based in math; the other is creative, based in language. Most scholars in the field of technical communication—those who shepherd the construct, or phenomenon, of technical writing—are not well trained in the measurement concepts and statistical terms provided in Appendix 1. The vocabulary is foreign, often appearing to be mere jargon, and learning statistics is simply not intuitive.

Historically, statistical functions were accomplished with complex mathematics because this was the only way large numbers could be processed. Julian Simon and Peter Bruce (1991) found that the origin of the difficulty in teaching statistics was that mathematicians had to develop analytic probability theory and complex formulas to process large combinations. The resultant formulas, which became the foundation of statistics, do not necessarily reflect the purpose of statistics—to provide empirical evidence of phenomenon within complex social systems. Today, with the processing abilities of computers, these complex, hand-calculated formulas are no longer necessary, and statistical analysis is gradually becoming more accessible. In the twenty-first century, it is valuable to gain a basic fluency in statistics because nearly every field uses these measures to drive decisions. Our world is increasingly described in probabilities—in quantum mechanics, in sports, in medicine, in genetics, in the environment, and in the economy.

A basic familiarity with statistics can be achieved in a variety of ways. College textbooks, such as Ralph L. Rosnow and Robert Rosenthal's *Beginning Behavioral Research* (2008), offer the basic understanding we give to students. If researchers are using the most common statistics program in the social sciences, SPSS (now predictive analytics software, PASW) (Norušis, 2011), there is an accessible book written humorously for people who fear statistics (Field, 2005). As well, innovative new ways to understand measurement, such as the fourth edition of *Educational Measurement* (Brennan, 2006), should be required reading for those wishing to become part of a culture of assessment. To remain current in the field, the journal *Educational Researcher* is the best guide.

It is possible, we have found, to break down the elements of a creative act, such as writing, into separate criteria and then collect statistics on those criteria within a community of those trained in English, not in educational measurement.

Although we have used this process with technical writing, it is possible to apply it to music and the visual arts—to any form of assessment where only a performance can allow valid measurement of developed ability (Lane & Stone, 2006).

To begin, a group of professionals must decide on the most important criteria to be assessed and create a rubric with defined variables and an overall holistic score—the variable model of the study. Then the model must be statistically tested in the ways we describe later. After a cohesive model is created, groups of instructors score the creative work, adding another dimension to the critique: a public commitment demonstrating that the curriculum matters.

The primary concept in this campus-based culture of measurement is validation. As noted previously, validation assures that the assessment focuses on capturing the expression of desired student performance: To be valid, a test must be matched to its target behavior. This validity is why program administrators must decide, first, on the most important criteria in any program or course before designing the assessment, which must match those criteria. This process requires consensus on the part of the shareholders in the course or program. Such consensus—in reality, a desire to avoid construct underrepresentation and, instead, ensure that the curriculum will yield optimal student performance—is usually achieved by a series of meetings, online discussions, or both. When completed, this period of planning results in a variable model, an expression of the most important criteria (*X*, or independent) variable and the outcome (*Y*, or dependent) variable, designated as the holistic score of the performance. The criteria of our research, as shown in Figure 1, were labeled ePortfolio design, clear style, accurate usage, task knowledge, relevant content, adapted tone, graphic cohesion, and citation.

After creating the criteria to be assessed, assignments (or tasks, as they are called in the constructed response literature) should be added to the curricula that will allow students to learn—and later to demonstrate—the desired behaviors under performance conditions. The next step is to decide on a method for collecting samples; we chose ePortfolios that could be shown online. Because it is not logistically possible to read and score all student submissions in a single day, we created a method of selecting a number large enough to achieve a 95% confidence level so that the range of scores in the sample would be representative of the larger student population enrolled in our courses. The formula, admittedly complex, uses the number of students enrolled that semester, the mean (or average) score from the previous semester, and the standard error (the researcher's expectation of how much the sample means might vary from the collective mean) to calculate the number of ePortfolios we have to read. We then selected ePortfolios using a list of random numbers easily generated from an internet site such as random.org (Haadr, 2010).

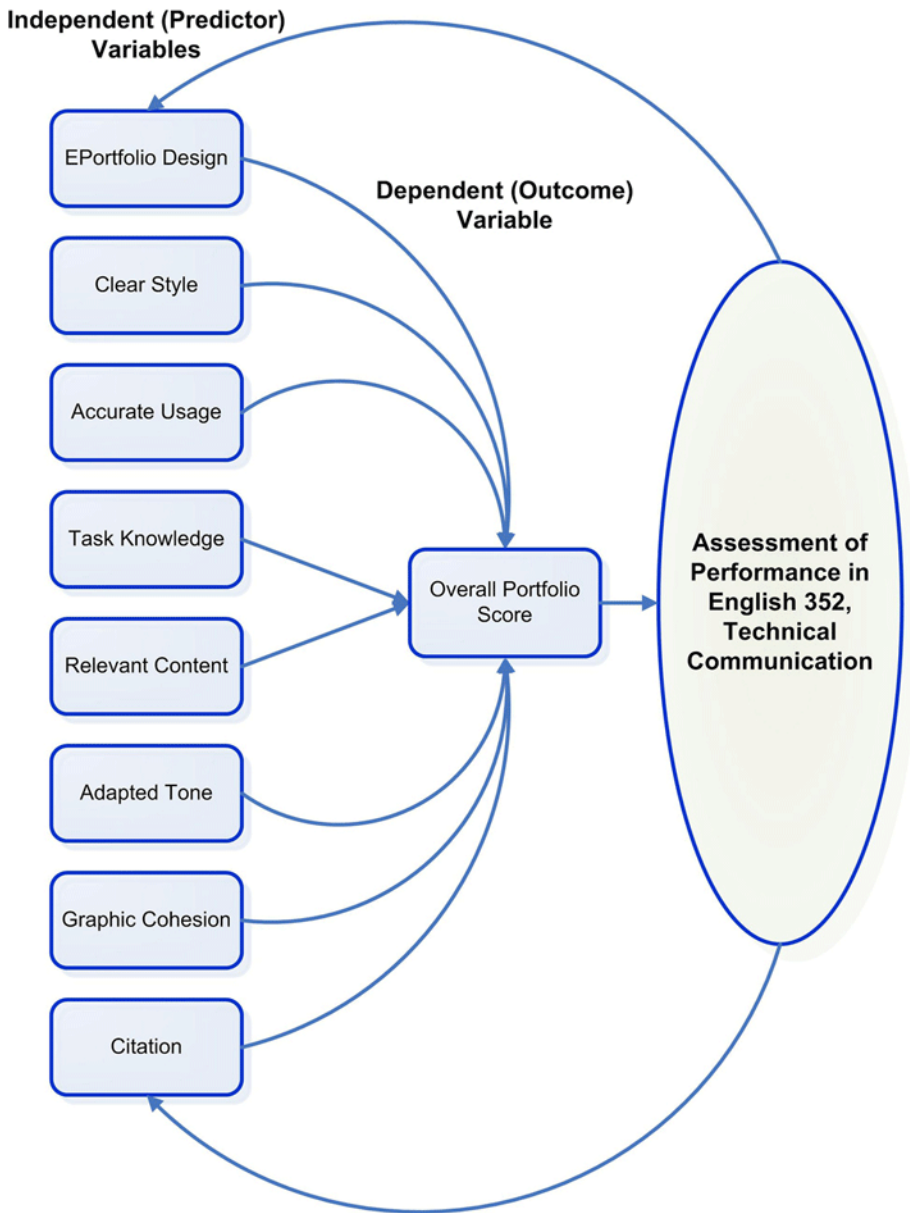


Figure 1. The NJIT model for undergraduate technical writing assessment

As part of our adherence to a unified theory of validation (Messick, 1989), interrater reliability is a key aspect of our work. We have two instructors score every variable separately on a scale from one (the lowest) to six (the highest score); if scores are discrepant (that is, nonadjacent, such as a 6 and a 4) the variable is adjudicated by a third rater. We then examine the raters' scores using a common tool, a consensus estimate that shows the percent of rater agreement documented in Table 1.

**Table 1. Interrater agreement analysis, spring 2009 (n=56)**

<b>ePortfolios Needing No Adjudication, Spring 2009 (n=56)</b>	<b>Interrater agreement</b>	<b>Percent of agreement</b>
No Adjudication on All Variables	19	34%
No Adjudication on ePortfolio (Web Page)	46	82%
No Adjudication on Clear Style	45	80%
No Adjudication on Accurate Usage	45	80%
No Adjudication on Task Knowledge (Understanding Assignments)	49	88%
No Adjudication on Relevant Content	47	84%
No Adjudication on Adapted Tone	47	84%
No Adjudication on Graphic Cohesion	49	88%
No Adjudication on Citation	32	57%
No Adjudication on Overall Score	49	88%

Pearson's *r*, a consistency estimate providing correlations as evidence of reliability, is also used to compare the two columns of nonadjudicated (original) and adjudicated (resolved discrepant) scores. As shown in Table 2, Pearson's *r* is used to show the degree to which the scores of raters are related. The results range from +1 (a perfect relationship) to -1 (an inverse relationship). We then note which relations are significant, expressed in probabilities (*p*): \**p* <.05 identifies a 95% confidence interval (the range of scores likely to include the mean, or average, score), and \*\**p* <.01 designates a 99% confidence interval. Probability estimates signify that the results are not an artifact of chance. The use of a weighted kappa (a measure of interrater consistency) adds additional validation to our efforts. We report both nonadjudicated and adjudicated scores because it is important not to mask the initial reaction of raters to the observed student ePortfolios. In outcomes assessment, transparency must be always and everywhere apparent.



**Table 2. Reliability analysis: Pearson r and weighted kappa, spring 2009**

Indicators	Non-Adj. Pearson r (2-tailed)	Adj. Pearson r (2-tailed)	Non-Adj. Weighted Kappa	Adj. Weighted Kappa
1. ePortfolio	.14	.53**	.035	.299**
2. Clear Style	.32*	.64**	.256**	.441**
3. Accurate Usage	.199	.586**	.089	.348**
4. Task Knowledge	.294*	.566**	.142	.339**
5. Relevant Content	.3*	.639**	.181*	.411**
6. Adapted Tone	.27*	.66**	.193*	.428**
7. Graphic Cohesion	.364**	.618**	.180*	.333**
8. Citation	.2	.831**	.129	.672**
9. Overall Score	.251	.581**	.127	.348**

\* $p < .05$ \*\* $p < .01$ 

We used the Pearson's correlation to gain a sense of the strength of the relationships in the variable model as well. Using the same function, we produce numbers that indicated how well the elements in the model were correlated. As shown in Table 3, all the variables are significantly related to each other at a 99% confidence level.

**Table 3. Correlation analysis of the NJIT model, fall 2004 to spring 2009 (n=636)**

	Clear Style	Accurate Usage	Task Knowledge	Relevant Content	Adapted Tone	Graphic Cohesion	Overall Score
Clear Style	-	.713**	.623**	.677**	.632**	.574**	.7**
Accurate Usage		-	.55**	.584**	.588**	.532**	.626**
Task Knowledge			-	.789**	.691**	.62**	.794**
Relevant Content				-	.731**	.669**	.796**
Adapted Tone					-	.633**	.734**
Graphic Cohesion						-	.707**
Overall Score							-

\* $p < .05$ \*\* $p < .01$ 

Another important test to judge the integrity of the model is a linear regression analysis, a tool that allows us to compare the connections between

the eight predictor (*X*, or independent) variables and the outcome (*Y*, or dependent) variable shown in Figure 1. As Table 4 shows, the model is coherent indeed, accounting in the spring of 2009 for 68% of the variance within the model—the degree to which the predictor variables are related to the outcome variable at a 99% confidence interval.

**Table 4. Regression analysis of the NJIT model, fall 2004 through spring 2009**

	$R^2$	$F(df)$	$p$
Fall 2004 (n=61)	.681	19.21(6, 60)	.01
Spring 2005 (n=50)	.756	22.18(6, 49)	.01
Fall 2005 (n=124)	.853	112.74(6, 123)	.01
Spring 2006 <sup>1</sup> (n=140)	.795	73 (7, 139)	.01
Fall 2006 <sup>2</sup> (n=92)	.729	31.57(8, 91)	.01
Spring 2007 (n=88)	.836	50.22(8, 87)	.01
Spring 2008 (n=25)	.915	21.48(8, 24)	.01
Spring 2009 (n=56)	.68	12.5(8,55)	.01

<sup>1</sup> Citation variable added to the model

<sup>2</sup> ePortfolio design added to the model

Because the ePortfolios were read reliably after adjudication and the model was cohesive, we were then able to report our means (average scores), standard deviations (a measure of score dispersion from the mean), and range (a measure revealing use of the entire scoring scale). Our descriptive statistics are presented in Appendix 2 as we follow the reporting guidelines of the American Psychological Association (2010, pp. 21–59).

Among our inferential statistics (used to draw relational evidence), an independent sample *t*-test was employed to give us a sense if, from year to year, our scores were rising, or falling, at statistically significant levels. We will turn to the score difference later when we analyze Table 5. Among our final tests, a correlation analysis allowed us to see if our ePortfolio scores were related to course grades and cumulative grade point averages. The answer, shown in Appendix 3, will also be discussed later.

The task of learning measurement concepts and statistical terminology is complicated, but the language and the processes can be mastered, bit by bit. The process of mounting a validation argument, as we will demonstrate throughout the rest of the article, is different from a way of thinking writers often use: educational measurement is not only analytic, but makes use of mental functions that cannot easily be described with the Euclidean

geometry learned in school. It is best to understand, at the beginning, that all statistics cannot be mastered in one concerted effort. Working piece by piece, the system eventually comes together. Although technical writing teachers will not become psychometricians, we can bring something to the educational measurement that has not been brought by others: an understanding of creativity formed in the service of technical communication.

## **Designing the Variable Model: Construct Validation**

Our work began in the fall of 2004 as we sought to define our landscape with a single question: What is it that we wanted to teach? In previous articles, we described the process of using an online Delphi, a formal email exchange between faculty, to describe, and then agree on, the aspects of technical writing we wanted to teach (Johnson, 2006a, 2006b). The results were originally under five headings: writing and editing, substance and content, audience awareness, document design, and textual attribution. Over time, we modified the variables and designed the present set, represented in Figure 1. We omitted some independent variables that were slightly repetitive to make the reading session more manageable, thus decreasing the number of predictor variables while retaining the outcome variable of the overall ePortfolio score. With interaction from our resource librarians, we revised the criteria for textual attribution to express our construct of information literacy in a single variable that reflected our defined goals of textual attribution. Because information literacy was rapidly becoming an important part of the university curriculum and had been examined in other undergraduate ePortfolios, we wished to introduce this important instructional element into our technical writing course. We will return to this important new variable of technical writing later.

The model represented in Figure 1, then, affords our assessment a sense of defined constructs, or traits, that we associate with proficiency in technical writing. In educational measurement terms, we thus take Figure 1 to be our construct of technical writing. Again the *Standards* (AERA, APA, & NCME, 1999) is conceptually helpful with the definition of a construct as “a theoretical variable inferred from multiple types of evidence” (p. 174), a model that can be validated by the very processes we describe. Figure 1 depicts both the construct of technical writing—the relationship of the eight predictor variables to the outcome variable, expressed in the overall ePortfolio score—and the process of validation. As Michael Kane (2006) defined this process, we are thus able to develop evidence to support the proposed use of the

assessment to create curricular goals and to examine the plausibility of our claims (p. 17). To develop evidence, the form of assessment we created is more analytic than holistic. We needed more than a single holistic score if we were to truly understand, teach, and improve the components of technical writing. Although the overall ePortfolio score was holistically scored according to classical methods (Godshalk, Coffman, & Swineford, 1966), the predictor variables were designed to be analytically scored (Purves, Gorman, & Takala, 1988). The model finds its origin in our earlier assessment research (Elliot, Briller, & Joshi, 2007).

We began to embrace and score ePortfolios because we believed that writing for audiences beyond the classroom was central to the technical writing experience. In addition, we embraced ePortfolios as a way to gain a more robust sense of the ways each instructor captures the variable model within each course and how each student responds to it. We can see student work in multiple drafts, including the teacher's comments, the visual elements, and the design of the ePortfolio itself. We see an entire semester's worth of effort, how the students developed, and sometimes, how they felt. Technically, the move to ePortfolios—the only way to truly capture such transactions efficiently—required that we include basic HTML instruction in the early years of the project for students to post ePortfolios housed on the university servers with links to their work for the semester. It is important that the ePortfolio is open to the Web rather than in a closed database because it gives the student a stake in the assessment process: Their ePortfolios are visible to other students, professors, their families, and their friends. When the assessment is over, students can revise the website for personal or professional reasons. With Jason Swarts and Loel Kim (2009), we hold that the possibilities for rhetorical action are “being reshaped by information and communication technologies, by near ubiquitous connectivity, and by more robust networking capabilities that have facilitated the creation of an expansive information stance that frequently meshes with the material places in which we live,” what they term *hybrid spaces*, which is not only a commodity, but also “a frame on the world around us” (p. 212). A seemingly technical consideration—teaching traditional HTML—thus becomes a part of construct validation process within our model, itself framed by and framing the world in which students live.

Within this environment, we further strengthened the construct validity of the model—our theory of the variables of technical writing and the processes by which we would support its use in curriculum design—by creating assignments (constructed response tasks) to address the variables and, during the assessment, sharing our ideas with each other, a process

leading to further modifications and new assignments. This set of common elements ensured that the class would work through five separate discourse tasks aimed to enhance their instructional, persuasive, visual, oral, and online communicative abilities. These additional efforts at construct validity increased the meaning of the assessment and demystified the contents of the course for students. After we held our first successful analytic online assessment of technical writing in the fall of 2004, we sequenced our assessment with other similar departmental programs and repeated it every semester until the spring of 2007; from that point forward we have held the assessment reading once a year in the spring.

Name	Type	Width	Decim	Label	Values	Missing	Columns	Align	Measure
1 Name	String	12	0	Student Name	None	None	10	Left	Nominal
2 Var1	Numeric	2	0	Clear Style - now Var2 on SSS	None	None	4	Center	Scale
3 Var2	Numeric	2	0	Accurate Usage - now Var 3 on SSS	None	None	4	Center	Scale
4 Var3	Numeric	2	0	Understanding Assignment - now Var 4	None	None	4	Center	Scale
5 Var4	Numeric	2	0	Relevant Content - now Var 5 on SSS	None	None	4	Center	Scale
6 Var5	Numeric	2	0	Rhetorical Response - gone on SSS	None	None	4	Center	Scale
7 Var6	Numeric	2	0	Adapt Content - gone on SSS	None	None	4	Center	Scale
8 Var7	Numeric	2	0	Adapt Tone - now Var 6 on SSS	None	None	4	Center	Scale
9 Var8	Numeric	2	0	Graphic Cohesion - now Var 7 on SSS	None	None	4	Center	Scale
10 Var9	Numeric	2	0	Parallel Structure - gone on SSS	None	None	4	Center	Scale
11 Var10	Numeric	2	0	Basic Graphics - gone on SSS	None	None	5	Center	Scale
12 Var11	Numeric	2	0	Citation - now Var 8 on SSS	None	None	5	Center	Scale
13 Var12	Numeric	2	0	Research - gone on SSS	None	None	5	Center	Scale
14 Var13	Numeric	2	0	Overall Portfolio Score - now Var 9 on S	None	None	5	Center	Scale
15 Var14	Numeric	2	0	Web Page - new Var1 on SSS	None	None	5	Center	Scale
16 Cl	Numeric	1	0	Classroom	{1, Traditional Classroom}...	None	4	Center	Scale
17 TS	Numeric	1	0	Transfer Status	{1, First-time, full-time freshman}...	None	4	Center	Scale
18 Semes	Numeric	1	0	Semester	{1, fall2004}...	None	4	Center	Scale
19 Grade	Numeric	8	2	Course Grade	None	None	5	Center	Scale
20 GPA	Numeric	8	2	Overall GPA	None	None	5	Center	Scale
21 SN	String	10	0	Student ID#	None	None	8	Left	Nominal
22 Major	String	10	0	College, Year, Major	None	None	8	Left	Nominal
23 Sect	String	8	0	Section	None	None	4	Left	Nominal
24 DL	Numeric	1	0	Distance Learning	{1, Face to Face}...	None	5	Right	Scale

Figure 2. Original codebook showing variable view

In addition, following the ECD orientation, we developed our codebook and database as we were discussing the variable model itself. Although such work is traditionally associated with data analysis, a later step in the process, we selected the SPSS program early because of its codebook properties and data analysis qualities. Figure 2 shows our original codebook in the SPSS variable view. That codebook, in relation to Figure 1, shows the history of our assessment project—constructs we tried and later abandoned (such as rhetorical response and parallel structure), those included from the beginning (such as style and usage), and those added in the journey (such as citation and webpage design). Ever capacious, SPSS allows the user to switch back and forth between Variable View and Data View. In Variable View, researchers can see and define the qualitative contents of each column. In Data View, researchers see the quantitative contents.

We kept separate SPSS files for each semester we held the assessment, including nonadjudicated scores and adjudicated scores. After entering the data, we transferred each semester's data to the total database, where we added more data, such as course grade and semester cumulative grade point average (GPA)—standard elements of criterion validation (the process of relating the phenomena under analysis to other performance measures).

The history of our research project is captured in Figure 2. At the beginning, one instructor asked that we analyze the difference in ePortfolio outcomes between traditional and computer classrooms. (We hypothesized that the computer classroom scores would be higher, but they were not). Thus, there is an entry "1" for the traditional classroom and an entry "2" for the PC lab. The next item on the list is the transfer status of students, an ongoing concern in which the relationship between instructional origin and outcomes is examined. Our colleagues asked if upper division technical writing instructors found that students matriculating from community colleges had skills similar to full-time, first-time first-year students. We confirmed that no significant difference existed, thereby replicating the results of a study conducted 13 years earlier (Elliot, Kilduff, & Lynch, 1994). The other rows are used to collect data on the course grade and cumulative GPA. Gathering such information allows us to compare many different elements about the students, the course, and university environment in which they are situated.

Figures 3 and 4 show what the database looks like at the present time. Because we have maintained this assessment for eight semesters, we have a total of 636 students in the database. After the ePortfolios are scored, the nonadjudicated and adjudicated scores are entered in a different database. The rest of the data—the grades and GPA—is gathered from the university's student information database. The result is a database that can be queried for a variety of information on student performance: It can be queried to assess the components of the course itself and also to address questions that are outside the purview of the course, such as the classroom and transfer issues. It is possible in SPSS, of course, to add variables for other data, such as available SAT® scores, as further studies are undertaken. In sum: Figures 1–4 serve as symbolic representation of our efforts to design a model and to examine its efficacy. How, then, do we operationalize that model into a scoring methodology?

## Undergraduate Technical Writing Assessment: A Model

	Name	Type	Width	Deci	Label	Values	Missing	Columns	Align	Measure
1	Name	String	12	0	Student Name	None	None	10	Left	Nominal
2	Var1	Numeric	2	0	EPortfolio Design	None	None	4	Center	Scale
3	Var2	Numeric	2	0	Clear Style	None	None	4	Center	Scale
4	Var3	Numeric	2	0	Accurate Usage	None	None	4	Center	Scale
5	Var4	Numeric	2	0	Task Knowledge	None	None	4	Center	Scale
6	Var5	Numeric	2	0	Relevant Content	None	None	4	Center	Scale
7	Var6	Numeric	2	0	Adapted Tone	None	None	4	Center	Scale
8	Var7	Numeric	2	0	Graphic Cohesion	None	None	4	Center	Scale
9	Var8	Numeric	2	0	Citation	None	None	4	Center	Scale
10	Var9	Numeric	2	0	Overall Portfolio Score	None	None	4	Center	Scale
11	Grade	Numeric	8	2	Course Grade	None	None	5	Center	Scale
12	GPA	Numeric	8	2	Overall GPA	None	None	5	Center	Scale
13	Major	String	10	0	College, Year, Major	None	None	8	Left	Nominal
14	Sect	String	8	0	Section	None	None	4	Left	Nominal

**Figure 3. Codebook reflecting current variables**

	Name	Var1	Var2	Var3	Var4	Var5	Var6	Var7	Var8	Var9	Grade	GPA	Major	Sect
1	Alvaros	5	6	7	7	7	7	7	7	5	2.50	2.75	NCE JR	001
2	Angolas	9	9	11	11	11	9	10	10	10	4.00	3.50	NCE JR	001
3	Atalah	8	9	9	9	8	9	9	5	6	3.50	2.41	NCE SR	005
4	Aadoyas	6	4	8	7	8	8	7	9	8	3.00	2.96	NCE JR	001
5	Boldusa	9	9	9	9	9	8	8	9	9	4.00	3.79	NCE FR	107
6	Curzias	8	8	8	8	8	8	8	8	8	3.00	2.93	CCS SO	107
7	Cabalas	7	7	9	9	9	10	10	9	9	3.50	2.92	NCE JR	107
8	Cisnowal	8	8	7	7	4	7	7	7	7	3.50	2.04	NCE JR	103
9	Chmielars	9	8	9	9	9	9	9	9	9	2.00	2.31	SOM JR	103
10	Dhajnacki	10	7	9	8	8	9	9	9	9	4.00	3.36	CCS SR	005
11	Desai, Dh	7	7	8	9	9	9	9	11	10	3.50	3.56	NCE JR	001
12	Desai, Do	9	8	7	7	9	6	5	5	5	3.50	3.03	NCE JR	S101
13	Dorcacki	4	4	8	8	8	8	7	7	7	3.50	3.33	NCE SO	001
14	Douglass	9	9	5	6	9	6	6	7	9	3.50	2.78	NCE SR	103

**Figure 4. Data view reflecting current variables**

## Designing the Assessment Methodology: Content Validation

Once the ePortfolio is created, the student submits a link. The random sampling described below is taken from a list of all students. That subset sampling is then made into an Microsoft® WORD document and/or an HTML page with links to the selected ePortfolios. We gather in a room, calibrate ourselves by scoring three sample ePortfolios (with superior, medium, or poor scores) and discuss our initial scoring reactions for about an hour. Using the rubric shown in Figure 5, each rater scores the sample ePortfolios individually, keeping notes on why each decision was made. We then tabulate the group results on a whiteboard.

The raters, especially the outliers, explain their reasoning. This discussion brings the group into a closer consensus. It is a normative discussion in which we align the ePortfolios from the semester with the criteria—the levels of scores from 1 (very strongly disagree) to 6 (very strongly agree)—that exist across semesters. After discussion, we distribute the rubric and the cards shown in Figure 6, and each instructor independently scores the ePortfolios for the eight predictor variables and the overall holistic score. The assessment leader collects first and second readings to check for discrepancies, highlights discrepant criteria on a third rubric, and distributes it, if necessary, to a third rater.

English 352 (Technical Communication) Portfolio Assessment – Spring 2009						
Student's Name _____						
Reader's Name _____						
<b>EPortfolio Design</b>						
1. The web page is clear and navigable.						
Very Strongly Agree	Strongly Agree	Agree	Disagree	Strongly Disagree	Very Strongly Disagree	
<b>Clear Style</b>						
2. The contents of this portfolio exhibit clear style (readable, concise, cohesive).						
Very Strongly Agree	Strongly Agree	Agree	Disagree	Strongly Disagree	Very Strongly Disagree	
<b>Accurate Usage</b>						
3. The contents of this portfolio demonstrate accurate language usage (grammar, punctuation, spelling).						
Very Strongly Agree	Strongly Agree	Agree	Disagree	Strongly Disagree	Very Strongly Disagree	
<b>Task Knowledge</b>						
4. The contents of this portfolio exhibit clear understanding of assignments.						
Very Strongly Agree	Strongly Agree	Agree	Disagree	Strongly Disagree	Very Strongly Disagree	
<b>Relevant Content</b>						
5. The contents of this portfolio demonstrate accurate, thorough, relevant, and coherent content and ideas.						
Very Strongly Agree	Strongly Agree	Agree	Disagree	Strongly Disagree	Very Strongly Disagree	
<b>Adapted Tone</b>						
6. The contents of this portfolio demonstrate that the author can adapt tone for audience.						
Very Strongly Agree	Strongly Agree	Agree	Disagree	Strongly Disagree	Very Strongly Disagree	
<b>Graphic Cohesion</b>						
7. The contents of this portfolio demonstrate cohesion by graphic means (headings, white space) in documents.						
Very Strongly Agree	Strongly Agree	Agree	Disagree	Strongly Disagree	Very Strongly Disagree	
<b>Citation</b>						
8. This portfolio includes sources that are documented so that the original source can easily be found.						
Very Strongly Agree	Strongly Agree	Agree	Disagree	Strongly Disagree	Very Strongly Disagree	
<b>Overall Portfolio Score:</b>						
9. The materials in this portfolio demonstrate class work that is:						
Superior	Very Good	Average	Below Average	Near Failure	At Failure	

Figure 5. Score sheet for the NJIT model



Engl. 352 Assessment S '09		
Name Hanlen, Glen D	1 <sup>st</sup> Reading	After Reading, Initial Below
URL: <a href="http://web.njit.edu/~ps13/">http://web.njit.edu/~ps13/</a>	2 <sup>nd</sup> Reading	
Discrepancy? Yes <input type="checkbox"/> No <input type="checkbox"/>		
Discrepancy in section:		
1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 <input type="checkbox"/>		
Done <input type="checkbox"/>	3 <sup>rd</sup> Reading	

**Figure 6. Index card (4" by 6") used to process and document reading**

With this process, we operationalize our construct of technical writing through the student samples (produced under naturalistic classroom conditions) and the rubric (designed to allow a range of scores). Thus, the content domain of our model (Kane, 2006, p. 19)—the desired interpretation of scores based on a performance activity as an estimate of the overall level of skill in technical writing—is articulated in the ePortfolios and the rubric. Even the basic HTML training thus becomes part of the context that allows for full construct emergence that can be captured fully in our evaluative setting: Because students can present their entire repertory of coursework, completed over a 15-week semester, we lessen the chance of construct underrepresentation (Kane, 2006, pp. 38–39), the major obstacle in all writing assessment (Elliot, 2005, pp. 270–277). As we explain later in our discussion of consequential validity as related to overall ePortfolio score, course grade, and cumulative GPA; however, we do not claim to have captured all that exists in the phenomenon identified as technical writing.

Our model is thus set; we have allowed student performance to emerge in a robust fashion. We have designed a scoring methodology that allows judgment based on both a limited rubric (allowing range) and a variety of samples (allowing depth). How can we ensure that we are not overwhelmed by the information we have collected? How can we make this process possible within a limited amount of instructor time?

## Designing the Sampling Plan: Economic Constraint

We have developed a specific formula to achieve the lowest possible number of ePortfolios to score to represent the course (Johnson, 2006a, 2006b). In the spring of 2009, 216 students enrolled in our technical writing course. Raters that semester included two adjuncts, one instructor, and two faculty members. In addition, we were fortunate to have two representatives from our information literacy initiative—seven raters in all. Although neither the words *efficiency* nor *economy* appear in the *Standards* (AERA, APA, & NCME, 1999), it is clear that resource allocation is closely tied to construct underrepresentation, “the extent to which a test fails to capture important aspects of the construct that test is intended to measure” (p. 174). If an assessment of writing is captured by a multiple choice test, that item type would be said to underrepresent the construct of writing; nevertheless, the test would meet the goal of efficiency. How do we then capture the assessment of technical writing by an ePortfolio and still meet the goal of efficiency with only seven raters on hand, with only a day to volunteer? To address efficiency, we have become adept at sampling plan design. That is, we have become determined to assess the smallest number of students possible with the greatest possible confidence in our results. We describe the formula we developed below.

We begin with a standard formula (Kerlinger & Lee, 1999, pp. 297–298) modified to address our sampling plan design:

$$n = \frac{Z^2 \sigma^2}{d^2} \quad (1)$$

Where:

$Z^2 = 1.96$ , the  $Z$ -value associated with a 95% confidence interval (for a .10 confidence level, the  $Z$  score = 1.645; for a .01 confidence level, the  $Z$  score = 2.575)

$\sigma^2$  = the standard deviation of the population

$d^2$  = the specified deviation defined as the deviation that we can tolerate between the sample mean and the true mean.

We then apply the correction for a finite sample:

$$n' = \frac{n}{1 + n/N} \quad (2)$$

Where:

$n'$  = estimated sample size

$n$  = sample size estimated using formula 1 described previously

$N$  = sample size of the population

Here is the step-by-step calculation we make.

### **Step 1. Calculate the specified deviation**

We begin with a conceptualization of the specified deviation—the deviation that we can tolerate between the sample mean and the true mean. In our program, we have defined the specified deviation as the mean score of the overall ePortfolio score from the previous semester's reading (the outcome variable of our model) plus or minus the Z-score (the standard score corresponding to the specified probability for risk) multiplied by standard error of the overall ePortfolio score. Calculations based on the previous semester's readings ensure that we use the information we gained to make our next set of decisions; the Z-score allows us to address the standard 95% confidence interval for decision making, although we have used lower confidence intervals when we have been unable to read all the ePortfolios in other NJIT programs (Elliot, Briller, & Johsi, 2007, p. 7). The standard error of the overall ePortfolio score used in this calculation is easily obtained from the descriptive statistics in SPSS.

Hence,

8.19 (mean score of the overall ePortfolio score from the previous semester's reading)

$\pm 1.96$  (the standard score corresponding to the specified probability for risk)

$\times 1.65$  (the standard error of the overall ePortfolio mean score)

Now,  $1.96 \times .165 = .32$ . Thus, .32 is the specified deviation. For the upper range of scores, we can be 95% confident that the scores will be  $8.19 + .32 = 8.51$ ; for the lower range, we can be 95% confident that the scores will be  $8.19 - .32 = 7.87$ . In sum, the specified deviation allows us to be 95% certain that the range of scores from 8.51 to 7.87 will include an individual student's true mean score.

### **Step 2. Calculate the sample size**

Now that we are certain of the specified deviation, we use equation 1 in Step 1.

Hence,

$$n = 1.96^2 \times 1.65^2 / .355^2$$

$$n = 3.84 \times 2.72 / .126$$

$$n = 10.44 / .126$$

$$n = 82.89$$

Therefore, to achieve a 95% confidence interval, we would need to read 83 ePortfolios. However, equation 1 is designed for an infinite sample—a sample in which the total number of students in the sample is unknown. Formula 2 described in Step 3 allows us to make the correction for a finite sample—in the spring of 2009, the 216 students enrolled in all sections of the course.

### **Step 3. Make the correction for a finite sample**

We now use equation 2.

Hence,

$$n' = 83 / 1 + (83 / 216)$$

$$n' = 83 / 1 + .38$$

$$n' = 83 / 1.38$$

$$n = 60$$

Therefore, our target is to take a random sample of 60 ePortfolios. To choose the random sampling, we obtain a list of all students taking the course from the Student Information System (SIS) database. We put that list in an alphabetized Microsoft® WORD table with columns for student identification number, student name, and website URL. Using a list of random numbers generated from a table of random numbers generated on the internet, we select students sequentially according to the random numbers until we have the requisite number. We then make a separate list, either in WORD or in HTML, with the URLs for student websites so that the raters can easily access them during the reading. With attrition, we scored 56 ePortfolios in spring of 2009, as Appendix 2 shows. The variations in the number of student ePortfolios assessed each semester were due to the evolution of our sampling methodology and the contingencies of rater participation.

Such sampling plan calculations rest at the heart of outcomes assessment. If the idea is to study and refine the curriculum, there is no need that each student be examined; rather, a well defined sampling plan with random selection allows a confidence level to be established that will allow administrators and

instructors to allocate resources efficiently under conditions of scarcity. Each semester, seven instructors can handle the designated sampling plan in one long morning, and the results we present enable us to have time to manage the assessment into our busy instructional and research lives. Rather than a burden—it would take two exhausting days to review the student ePortfolios from each section of the course—the end-of-term assessment episodes become yet another community-building task for the instructional group.

## **Designing the Data Analysis: Validation Argument**

After the scores from the ePortfolios are entered and checked, the first test to run is to establish the interrater agreement and the interrater reliability of the scores (Stemler, 2004). Although we view reliability as integral to a unified theory of validity (Messick, 1989), we also believe that establishing reliability is a precondition of validity. If the raters cannot agree on what has been observed, there can be no pursuit of additional analyses.

### **Interrater Agreement Analysis, Spring 2009**

Interrater agreement is based on the extent to which the raters agree on an ePortfolio score for one of the predictor variables or the outcome variable. The most straightforward way to judge the amount of interrater agreement is to count how many discrepancies had to be resolved. With eight variables and the overall ePortfolio score all read concurrently, there are usually discrepancies—this is how analytic assessment differs from holistic assessment—because more judgmental variety is recorded. As Table 1 from the spring of 2009 illustrates, the percent of agreement—the consensus estimate—for each predictor variable is quite high; nevertheless, very few ePortfolios require no adjudication whatsoever. When a new variable is added, such as citation (introduced in the spring of 2006), the percent of agreement is often low.

### **Reliability Analysis, Spring 2009**

Constructs such as those we use as representing technical writing cannot be entirely captured due to their complexity; raters do not agree on what they are viewing with the same precision they would if observing presence or absence of an infiltrate on a chest radiograph (Viera & Garnett, 2005). Thus, we used two tests to analyze the probability of the precision of raters. As is the case of the economy, quantum mechanics, and archaeology, assessment data about writing can be expressed in estimates rather than certainties.

We used two tests with the results of both for spring 2009—the consistency estimate—presented in Table 2. The first test, Pearson's correlation coefficient (Pearson's  $r$ ), shows meaningful association in values between

1.0 (a perfect positive relationship), 0 (no relationship) and -1.0 (a negative relationship).<sup>3</sup> As shown in Figure 7, to run these measures of linear relationships in SPSS we selected the column of nonadjudicated scores of first raters and the column of the nonadjudicated scores of second raters. This correlation produced the nonadjudicated scores in Table 2. We then adjudicated any discrepant scores by taking the third score, which most closely approximated the first or second score. So, for example, a first rater awarding an overall ePortfolio score of 6 and a second rater awarding a score of 4 would be discrepant, resolved by a third rater who might award a score of 4, thus confirming a final score of 8. If that third rater awarded a score of 5, then our “tie goes to the (student) runner” rule applies, and a total score of 11 is awarded.

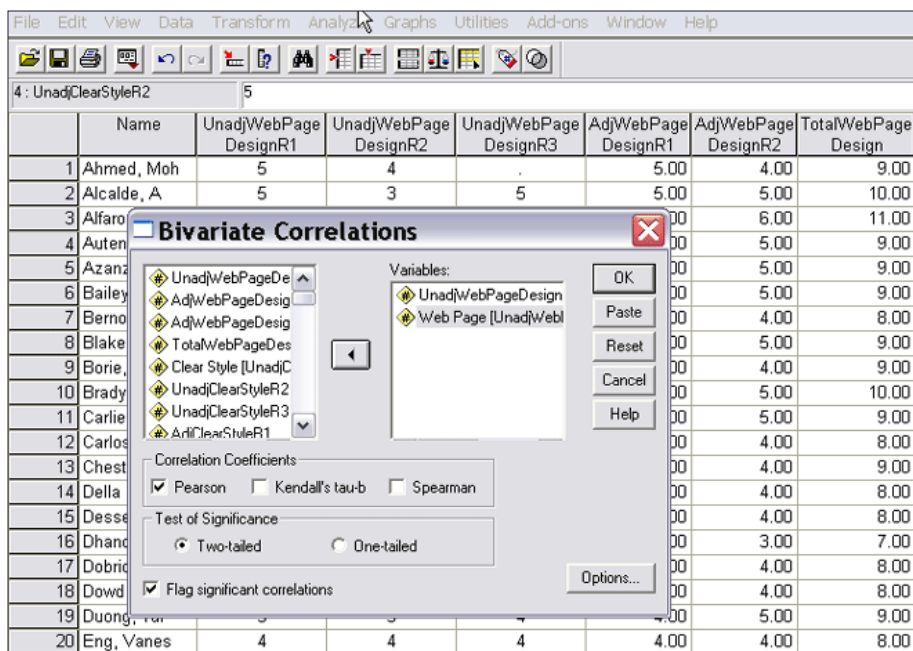


Figure 7. Dialogue box used in SPSS to calculate Pearson's r for nonadjudicated scores

As can be seen in Table 2, the nonadjudicated scores for ePortfolio design, accurate usage, citation, and the overall score did not reach the 95% confidence interval on the first reading, but did meet and exceed that level after adjudication, reaching the higher .01 confidence interval. The low correlation in the nonadjudicated scores is likely due to the number of variables that must be judged and rater inattention as the reading period progresses. This type of

<sup>3</sup> To run the Pearson *r* in SPSS, select Analyze/Correlate/Bivariate. Select the desired variables and click OK. A video demonstration of this process can be found at the NJIT site for iTunesU. See <<http://deimos3.apple.com/WebObjects/Core.woa/Browse/njit.edu.1302671158.01302671168.1303126577?i=1408077901>>.

analytic (multivariable) assessment is cognitively difficult for raters, especially at the end of a busy semester after grading student papers. Although Edward Haertel (2006) pointed out that when adjudication is used, the assumptions for many statistical models are violated (p. 120), it is also important to point out that adjudication is a necessity if shareholders are to be assured that discrepancies were resolved by a rater, rather than buried by an average.

Treating the data categorically, we also use Jacob Cohen's (1968) weighted kappa ( $k$ ), as shown in Table 2.<sup>4</sup> Again, we see that ePortfolio design, accurate usage, citation, and the overall score did not reach the 95% confidence interval; task knowledge also failed to meet the confidence interval. Under adjudication, the agreement substantially improved with each variable reaching the .01 significance level. According to strength of agreement levels established by J. Richard Landis and Gary Koch (1977, p. 165), the levels of agreement are fair (above .2) to substantial (above .61). If we return to the health analogy offered by Anthony Viera and Joanne Garnett (2005), we might compare our observations on these complex variables as similar to an observation on tactile fremitus, a rare observation of the chest wall vibrating during speech. As Viera and Garnett reminded us, "For rare findings, very low values of kappa may not necessarily reflect low rates of overall agreement" (p. 362). An observation of a rare occurrence in any field will not be recognized, and we need to take care to understand fully the complexities involved in any observation before setting a standard.

### **Correlation Matrix (Associative Analysis)**

Table 3 presents the correlations among the permanent variables that endured from the fall of 2004 to the spring of 2009, a variable set tested over 636 students. Each correlation is significant, at the .01 level, and the correlations range from .55 to .796. The relationship among the six permanent variables (clear style, accurate usage, task knowledge, relevant content, adapted tone, and graphic cohesion) and the overall ePortfolio is especially strong, with five of the six variables above .7. Clearly, the relationships among the variables are solid, with very high correlations of the variables with the overall ePortfolio score.

### **Regression Analysis (Predictive Analysis)**

Another important aspect of validation is the regression analysis. Because the model we have created is relational—that is, a predictor-outcome variable model in which every variable is related to the overall ePortfolio score—it is

<sup>4</sup> The weighted kappa statistic cannot be run in SPSS. However, our colleague, Kamal Joshi, has written a statistical analysis system (SAS) program for use with that software, and the program may be obtained by contacting the authors.

important to understand the extent to which the individual variables predict the overall, holistic, ePortfolio score.<sup>5</sup> Figure 8 provides a visual display of the method to perform a regression analysis in SPSS with the six permanent predictor variables used as the independent variables and the overall ePortfolio score used as the outcome variable. Table 4 demonstrates the strength of the model.

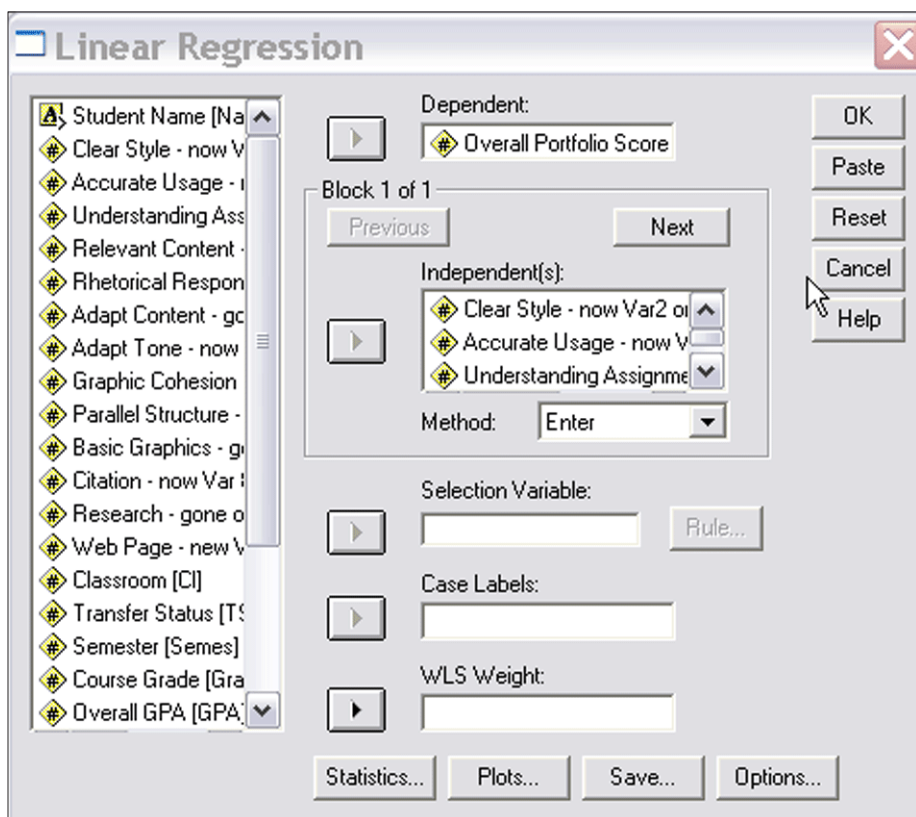


Figure 8. Dialogue box used in SPSS to calculate a linear regression

From the very first reading, our model was strong. At no time has the model fallen below an  $R^2$  of .68. That is, 68% of the variability of the model is accounted by the relationship of the six permanent predictor variables to the outcome variable of the overall ePortfolio score. At its highest, in the spring of 2008, 91% of the variability of the model is accounted for by the predictor-outcome model. Such model strength for the undergraduate technical writing program is comparable to other regression studies

<sup>5</sup> To run this test in SPSS, Analyze/Regression/Linear, select the dependent variable and the independent variables and click OK. A video demonstration of this process can be found at the NJIT site for iTunesU. See <<http://deimos3.apple.com/WebObjects/Core.woa/Browse/njit.edu.1302671158.01302671168.1303126584?i=2070933018>>.



performed on our graduate program model (Coppola & Elliot, 2007, p. 464; Coppola & Elliot, 2010, p. 150).

Hence, we may conclude that our construct of technical writing has been well designed and articulated. Because the construct has emerged in a performance-based environment that reviewers can understand and judge, the ePortfolios have been read reliably. Each variable is related in a statistically significant fashion, and the model is internally consistent. Our validation processes concerning the model have thus been well articulated. How can we investigate the model to keep it from being solipsistic in nature? That is, what gains can be demonstrated as a result of our program assessment effort, and how do these gains relate to other measures of student performance?

## **Using Assessment Results: Consequential Validation**

Robert L. Brennan (2006) observed that “perhaps the most contentious topic in validity is the role of consequences” (p. 8). How may we come to terms with the impact of our program assessment model? To judge the impact of our model, in this section of the article we describe performance across time and study the effects of our efforts to build community through attention to outcomes.

### **Differences in Mean Scores over Time**

Our main measure of success in achieving goals is that on a scale of 2 to 12, a score of 7 or above is acceptable, an indication of earned proficiency.<sup>6</sup> As shown in Appendix 2, we achieved those goals in the six permanent predictor variables and the overall ePortfolio score. Creating a model to measure learning outcomes and successfully meeting the minimum score means that, in the most straightforward way, we are teaching what we want to teach. We have imagined a concept of technical writing that we can deliver to students, encourage their responses, and then measure their abilities.

Yet is proficiency sufficient in the increasing environment of global competitiveness that students must face? Once the concept is in hand on the part of the instructional staff, do the scores change over the semesters? Although numerical differences are immediately visually evident, as seen in the Appendix 2, the differences may not be statistically significant. Here we apply an independent

---

<sup>6</sup> To find the means for each semester, use the semester’s database and select Analyze/ Descriptive Statistics/ Descriptives. In the dialogue box that appears, select Options and then select Range to get a report on the range as well. Also select S.E. Mean for the sampling plan calculation. Click OK. A video demonstration of this process can be found at the NJIT site for iTunesU. See <<http://deimos3.apple.com/WebObjects/Core.woa/Browse/njit.edu.1302671158.01302671168.1303077417?i=1664416832>>.

sample *t*-test to measure whether differences are statistically significant.<sup>7</sup> If the *t*-test indicates that the differences across semesters are statistically significant, and if a researcher can demonstrate that the student population has not shifted, then she can seriously marshal evidence that the curriculum and its instructors are making a difference in the lives of students. Because the NJIT student population has remained consistent in terms of SAT® performance, we have found that undergraduate population, presently resting at 5,248, is consistent across the period of this presented assessment with an SAT® Critical Reading mean score of 538 and an SAT® Mathematics score of 604. Although a new strategic plan aims to raise both the enrollment and the SAT® scores—and we hope to capture those efforts in our program assessment efforts—student gains in technical writing since 2004 may be attributed to the curriculum we have designed and assessed.

As Table 5 shows, five of the six predictor variables have statistically significant gains made since the beginning of our program in 2004 as compared to our most recent assessment.

**Table 5. Comparison of common variables, fall 2004 and spring 2009**

	t(df = 115)	p
Clear Style	-2.76	.01
Accurate Usage	-2.34	.05
Task Knowledge	-1.99	.05
Relevant Content	-1.92	.05
Adapted Tone	-.988	.3
Graphic Cohesion	2.19	.05
Overall ePortfolio Score	-2.86	.01

The scores on the following variables have been raised: clear style (fall 2004:  $M = 7.77$ ,  $SD = 1.28$ ; spring 2009:  $M = 8.46$ ,  $SD = 1.44$ ;  $t = -2.76$ ,  $p < .01$ ); accurate usage (fall 2004:  $M = 7.54$ ,  $SD = 1.4$ ; spring 2009:  $M = 8.18$ ,  $SD = 1.54$ ;  $t = -2.34$ ,  $p < .05$ ); task knowledge (fall 2004:  $M = 8$ ,  $SD = 1.22$ ; spring 2009:  $M = 8.5$ ,  $SD = 1.48$ ;  $t = -1.99$ ,  $p < .05$ ); graphic cohesion (fall 2004:  $M = 7.84$ ,  $SD = 1.52$ ; spring 2009:  $M = 8.48$ ,  $SD = 1.66$ ;  $t = 2.19$ ,  $p < .05$ ); and the overall ePortfolio score (fall 2004:  $M = 7.82$ ,  $SD = 1.34$ ; spring 2009:  $M = 8.63$ ,  $SD = 1.65$ ;  $t = -2.86$ ,  $p < .01$ ).

<sup>7</sup> To run an independent sample *t*-test, use the semester's database and select Analyze/Compare Means/Independent Sample *t*-Test. Because score comparisons are run across semesters, be sure to set the Grouping Variable in the codebook. A video demonstration of this process can be found at the NJIT site for iTunesU. See <<http://deimos3.apple.com/WebObjects/Core.woa/Browse/njit.edu.1302671158.01302671168.1303093807?i=1251486064>>.

Although no statistically significant difference can be observed of the scores on adapted tone, they consistently exceed the 7.82 score level in the comparative period.

Yet such improvement and consistency is not at all the case with the citation variable, as Appendix 2 clearly demonstrates. We had comfortably ignored the issue of attribution in technical writing instruction until our university research librarian challenged us to address and solve the problem. Based on her other information literacy work with university colleagues (Sharf, Elliot, Briller, Huey, & Joshi, 2007; Katz, et al., 2008), she focused her efforts on the ability of students to cite sources in a standard way (e.g., APA or MLA format) so that the original might easily be found. In the spring of 2006, this variable received the lowest scores we had ever witnessed ( $M = 5.12, SD = 2.77$ ) because the instructors were not yet including basic information literacy instruction in the curriculum. As Appendix 2 shows, our experience with this variable has been tenuous, though at present we appear to have greater control over its instruction. The present score for the citation variable ( $M = 7.66, SD = 2.84$ ) is statistically higher than it was during the fall of 2004 ( $t(df) = 194, p < .01$ ). Thus we introduced a new element into the curriculum and, due to persistent instructional efforts, it statistically rose over time. These scores rose because our librarian defined it and an instructor began requiring annotated bibliographies in a proposal assignment. During the assessment itself, other instructors saw how this variable was being introduced and imitated the assignment; thus, the assessment enabled the emergence of a new variable for our model. We take such studies as evidence of the sensitivity of our model to context and its ability to facilitate the emergence of group knowledge.

### **Test-Criterion Relationships:**

#### **Overall Score, Course Grade and Cumulative GPA**

Even though the variables and the model are highly correlated, the same is not true of the overall ePortfolio score, the course grade, and the cumulative GPA. As any keen reader has no doubt noted by now, we are assessing a program, not an individual student. We firmly believe that no assessment effort, however well-designed and executed, can ever capture a complex construct such as technical writing. The efficiency limits of such an assessment would themselves result in construct underrepresentation. As such, we hold that only the classroom instructor, present with a student for 15 weeks, can evaluate an individual student performance. Because our ePortfolios are always read near or after final grades are assigned, it is clear that our efforts are programmatic, not individualistic. It is therefore logical to take the course grade as a relevant criterion of technical writing. Hence, with the *Standards*

(AERA, APA, & NCME, 1999), we ask how accurately do the “test” scores predict the criterion performance (p. 14)?

Although the internal consistency of the model is very strong, correlations between the overall ePortfolio score—the best proxy for our model—and course grades are often absent, as Appendix 3 demonstrates. In the spring of 2009, only the citation variable has a statistically significant relationship to the course grade ( $r = .29$ ;  $p = .05$ ). A regression model taking all nine present variables shown in Figure 1 as the predictor and the course grade as the outcome yields a stunningly low relationship that lacks statistical significance:  $R^2 = .162$ ,  $F(9, 55) = .987$ ;  $p = .464$ ). There is a relationship, however, between course grade and cumulative GPA, ( $r = .527$ ,  $p < .01$ ). A regression model taking the course grade as the predictor variable and the cumulative GPA as the outcome has a degree of prediction as well as statistical significance:  $R^2 = .277$ ,  $F(1, 55) = 20.72$ ,  $p < .01$ .

Although the lack of a relationship between our model and the course grades might upset some, we hold both that the model was not established to control grades and that the model does not incorporate all that has value in a classroom. Although we believe that our model is robust, we would never claim that it encompasses *all* that is present in the trait termed technical writing and its teaching. From persistence in revision to poise in oral presentations, there are a host of elements present in classrooms that will never be part of our assessment model. One element that our model cannot take into account is the diversity of the student body—students at NJIT are extraordinarily multicultural and international, often having fluency in multiple languages, so that each one begins from a different place. As well, relationships between course grades and cumulative GPAs are expected in students who have traditionally earned over 60 credits before enrollment, who have cumulative GPAs of 3.13 ( $SD = .448$ ), and whose technical writing course grades are part of the GPA for that semester.

## **Conclusion**

The model we have described in this article works for the purposes it was devised: to ensure construct validation by means of an articulated model, to design an assessment methodology to ensure the content validation of that model, to design a sampling plan to ensure wise use of time, to plan data analysis techniques to demonstrate our validation argument, and to use the assessment results to assure positive consequences. The assessment has had a positive effect on students, their instructors, our program, and our institution. This modest sense of program assessment, one that locates the students and their curriculum at the center of our efforts, makes the program as a whole

stronger and makes the goals of the program clearer to the university community. Because the process is embedded in our program, we can easily document our outcomes for accreditation agencies such as ABET and MSCHE without creating new work.

Ongoing cycles of assessment can provide a basis for collaboration and intellectual exchange to help us review and revise criteria, to look at ourselves and our programs critically, to make changes, and to query those changes. It is within our power to use assessment to help us adjust to change in a continually changing world. It is our hope that the model described in this article will provide a way for others to replicate and refine our efforts for their unique institutional sites.

## References

- Accreditation Board for Engineering and Technology. (2008). *Criteria for accrediting engineering programs effective for evaluations during the 2009-2010 accreditation cycle*. Retrieved August 7, 2010, from <<http://www.abet.org/forms.shtml>>
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington: American Educational Research Association.
- American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.
- Baldwin, Doug; Fowles, Mary; & Livingston, Skip. (2005). *Guidelines for constructed response and other performance assessments*. Princeton: Educational Testing Service. Retrieved August 7, 2010, from <<http://www.abet.org/forms.shtml>>
- Bazerman, Charles, & Rogers, Paul. (2008). Writing and secular knowledge outside modern European institutions. In Charles Bazerman (Ed.), *Handbook of research on writing: History, society, school, individual, text* (pp. 143–156). New York: Lawrence Erlbaum.
- Beaufort, Anne. (2007). *College writing and beyond: A new framework for university writing instruction*. Logan: Utah State University Press.
- Beaufort, Anne. (2008). *Writing in the professions*. In Charles Bazerman (Ed.), *Handbook of research on writing: History, society, school, individual, text* (pp. 221–235). New York: Lawrence Erlbaum.
- Bolter, Jay D. (1999). *Remediation: Understanding new media*. Cambridge: MIT Press.
- Brennan, Robert L. (2006). Perspectives on the evolution and future of educational measurement. In Robert L. Brennan (Ed.), *Educational measurement* (4<sup>th</sup> ed.) (pp. 1–16). Westport, CT: American Council on Education/Praeger.
- Broad, Robert. (2003). *What we really value: Beyond rubrics in teaching and assessing writing*. Logan, Utah: Utah State University Press.

- Camilli, Gregory. (2006). Test fairness. In Robert L. Brennan (Ed.). *Educational measurement* (4<sup>th</sup> ed.) (pp. 221–256). Westport, CT: American Council on Education/Praeger.
- Chicago Public Art Group. (2009). *Collaborative design*. Retrieved August 7, 2010, from <[http://www.cpag.net/home/wwd\\_collaborativedesign.html](http://www.cpag.net/home/wwd_collaborativedesign.html)>
- Cohen, Jacob. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement of partial credit. *Psychological Bulletin*, 79, 213–220.
- Condon, William. (Ed.) (2009). Writing assessment as a generative practice. [Special issue]. *Assessing Writing*, 14.
- Coppola, Nancy, & Elliot, Norbert. (2007). A technology transfer model for program assessment in technical communication. *Technical Communication* 54, 459–474.
- Coppola, Nancy, & Elliot, Norbert. (2010). Assessment of graduate programs in technical communication. In Margaret Hundleby & Jo Allen (Eds.), *Assessment in technical and professional communication* (pp. 127–160). New York: Baywood.
- Council for Programs in Technical and Scientific Communication (CPTSC). (2008). *Research assessment project*. Retrieved August 7, 2010, from <<http://assessment-cptsc.njit.edu/>>
- Elliot, Norbert. (2005). *On a scale: a social history of writing assessment in America*. New York: Peter Lang.
- Elliot, Norbert; Kilduff, Margaret; & Lynch, Robert. (1994) The assessment of technical writing: A case study. *Journal of Technical Writing and Communication* 24, 19–36.
- Elliot, Norbert; Briller, Vladimir; & Joshi, Kamal. (2007). Portfolio assessment: Quantification and community. *Journal of Writing Assessment*, 3, 5–30.
- Field, Andy. (2005). *Discovering Statistics Using SPSS*. London: Sage Publications.
- Godshalk, Fred I.; Swineford, Frances; & Coffman, William E. (1966). *The measurement of writing ability*. New York: College Entrance Examination Board, 1966.
- Haadr, Mads (2010). Random.org. Dublin: Trinity College. Retrieved August 7, 2010, from <<http://www.random.org/mads/>>
- Haertel, Edward H. (2006). Reliability. In Robert L. Brennan (Ed.). *Educational measurement* (4<sup>th</sup> ed.) (pp. 65–110). Westport, CT: American Council on Education/Praeger.
- Huot, Brian. (2002). *(Re)articulating writing assessment for teaching and learning*. Logan, Utah: Utah State University Press.
- Johnson, Carol Siri. (2006a). The analytic assessment of online portfolios in undergraduate technical communication: a model. *Journal of Engineering Education*, 95, 279–287.
- Johnson, Carol Siri. (2006b). A decade of research: Assessing change in the technical communication classroom using online portfolios. *Journal of Technical Writing and Communication*, 36, 413–431.
- Johnson, Carol Siri, & Elliot, Norbert (2004, October). *Beyond cut scores: Portfolio assessment in program evaluation*. Paper presented at the meeting of the American Society for Engineering Education. Washington, DC.

- Kane, Michael T. (2006). Validation. In Robert L. Brennan (Ed.). *Educational measurement* (4<sup>th</sup> ed.) (pp. 17–64). Westport, CT: American Council on Education/Praeger.
- Katz, Irvin R.; Elliot, Norbert; Scharf, Davida; Attali, Yigal; Powers, Donald; Huey, Heather; Joshi, Kamal; & Briller, Vladimir . (2008). *Information literacy assessment: Case study of a multi-method approach* (ETS Research Report RR-08-33). Princeton: ETS.
- Kerlinger, Frederick N., & Lee, Howard B. (1999) *Foundations of behavioral research* (4th ed.). New York: Wadsworth.
- Landis, J. Richard, & Koch, Gary G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Lane, Suzanne, & Stone, Clement A. (2006). Performance assessment. In Robert L. Brennan (Ed.). *Educational measurement* (4<sup>th</sup> ed.) (pp. 387–431). Westport, CT: American Council on Education/Praeger.
- Lockhart, Robert S. (1998). *Introduction to statistics and data analysis for the behavioral sciences*. New York: W. H. Freeman.
- Mazzeo, John; Lazer, Stephen; & Zieky, Michael J. (2006). Monitoring educational progress with group-score assessments. In Robert L. Brennan (Ed.), *Educational measurement* (4<sup>th</sup> ed.) (pp. 681–699). New York: American Council on Education and Macmillan.
- Messick, Samuel. (1989). Validity. In Robert L. Linn (Ed.), *Educational measurement* (3<sup>rd</sup> ed.) (pp. 13–103). New York: American Council on Education and Macmillan.
- Messick, Samuel. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23, 13–23.
- Middaugh, Michael. (2010). *Planning and assessment in higher education: Demonstrating institutional effectiveness*. San Francisco: Jossey-Bass.
- Middle States Commission on Higher Education (MSCHE). (2006). *Policies, guidelines, procedures & best practices*. Retrieved August 7, 2010, from <<http://www.msche.org/publications.asp>>
- Mislevy, Robert J.; Almond, Russell G.; & Lukas, Janice F. (2003). *A brief introduction to evidence-centered design* (ETS RR-03-13). Princeton: ETS. Retrieved August 7, 2010, from <<http://www.ets.org/Media/Research/pdf/RR-03-16.pdf>>
- Mislevy, Robert J. (2007). Validity by design. *Educational Researcher* 36, 463–469.
- Murray, Joddy. (2009). *Non-discursive rhetoric: Image and artifact in multimodal composition*. New York: State University of New York Press.
- Norušis, Marija J. (2011). *PASW statistics 18 guide to data analysis*. Upper Saddle River, NJ: Prentice Hall.
- O'Neill, Peggy; Moore, Cindy; & Huot, Brian. (2009). *A guide to college writing assessment*. Logan: Utah State University Press.
- Popper, Karl. (1963). *Conjectures and refutations*, London: Routledge/Keagan Paul.
- Purves, Alan C.; Gorman, Thomas P.; & Takala, Sauli. (1988). The development of the scoring scheme and scales. In Thomas P. Gorman, Alan C. Purves, & R. Elaine Degenhart (Eds.), *The IEA study of written composition I: The international writing tasks and scoring scales* (pp. 41–58). Oxford: Pergamon.

- Rosco, John T. (1969). *Fundamental research statistics for the behavioral sciences*. New York: Holt.
- Rosnow, Ralph L., & Rosenthal, Robert. (2008). *Beginning behavioral research: A conceptual primer*. (6th ed.). Upper Saddle River, NJ: Prentice Hall.
- Scharf, Davida; Elliot Norbert; Huey, Heather; Briller, Vladimir; & Joshi, Kamal. (2007). Direct assessment of information literacy using writing portfolios. *The Journal of Academic Librarianship*, 33, 462–478.
- Simon, Julian L., & Bruce, Peter. (1991). Resampling: A tool for everyday statistical work. *Chance: New directions for statistics and computing*, 4, Retrieved August 7, 2010, from <<http://resample.com/content/teaching/texts/chance.txt>>
- SPSS (PASW, Version 18). [Computer software]. Chicago, IL: SPSS Headquarters.
- Stemler, Steven E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation*, 9(4). Retrieved August 7, 2010, <<http://PAREonline.net/getvn.asp?v=9&n=4>>
- Swarts, Jason, & Kim, Loel. (Eds.). (2009). New technological spaces. [Special issue]. *Technical Communication Quarterly*, 18.
- Toulmin, Stephen. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
- Viera, Anthony J., & Garrett, Joanne M. (2005). Understanding interobserver agreement: The Kappa statistic. *Family Medicine*, 37, 360–363.
- White, Edward M. (2005). The scoring of writing portfolios: Phase 2. *College Composition and Communication*, 56, 581–600.
- Yancey, Kathleen B. (2004). Postmodernism, palimpsest, and portfolios: Theoretical issues in the representation of student work. *College Composition and Communication*, 55, 738–761.

## **Author's Note**

The undergraduate technical communication assessment model presented in this article was created by Norbert Elliot and Carol Johnson. The sampling plan was designed by Vladimir Briller and Kamil Joshi, with the support of Perry Deess. Contributors to the effort include James Lipuma, Nina Pardi, Robert Lynch, John Lyczko, Michael Kerley, Susan Fowler, Frank Casale, Brenda Moore, Michele Fields—and research librarians Davida Scharf and Heather Huey.



## Appendix 1

### Measurement Concepts and Statistical Terms: A Critical Vocabulary for Researchers

Term	Definition	Use	Key Source
<i>Measurement Concepts</i>			
Bias identification	The identification of bias is a process by which performance is observed to be different in defined groups due to systematic error.	Identification of difference in performance among groups is an important part of assuring fairness in assessment.	AERA, APA, NCME (1999, p. 172); Camilli (2006)
Consensus estimate	A consensus estimate is a measure of agreement between two raters.	A simple percent of agreement provides evidence of interrater agreement.	Stemler (2004)
Consequential validation	The consequences of assessment, both positive and negative, are key to the validation process.		AERA, APA, NCME (1999, pp. 23–24); Brennan (2006); Messick (1989)
Consistency estimate	Beyond a simple count of agreement, a consistency estimate provides correlations as evidence of reliability.	A Pearson product moment correlation and a weighted kappa both provide evidence of interrater reliability.	Haertal (2006, pp. 101–102); Stemler (2004)
Construct	The construct is the phenomenon that is under examination.	A combination of a well-articulated scoring rubric and samples of levels of student performance allows an expression of the construct to be measured.	AERA, APA, NCME (1999, pp. 17–18); Brennan (2006, pp. 22–23); O’Neill, P., Moore, C., & Huot, B. (2009, p. 198); White (2005)

## Appendix 1

Continued

Term	Definition	Use	Key Source
Construct underrepresentation	If an assessment fails to capture the targeted construct, or provide evidence that a key aspect of the construct has been measured, then the meaning of the assessment is limited.	Because construct underrepresentation has been a perennial problem in the assessment of written communication (the overuse of tests of grammar, for example), validity argument assures that the construct, or a key aspect of the construct, has been captured.	AERA, APA, NCME (1999, p. 174); Brennan (2006, p. 31).
Construct validation	The process by which evidence is gathered in the service of the validity argument.	Construct validation may be achieved by three methods: specification of the proposed interpretation of scores during the assessment design; dedication to an extended research activity; and examination of plausible rival score interpretations.	Brennan (2006, p. 22); Messick (1989); Popper (1963)
Constructed response assessment	As a performance assessment, a constructed response task requires that students perform (that is, construct) a response.	A constructed response assessment holds the potential to allow the construct to be measured.	Baldwin, Fowles, & Livingston (2005); Lane & Stone (2006)
Content validation	The detailed statement of the construct to be measured.	If a rubric is well designed, it will serve as evidence that the construct has been fully defined.	AERA, APA, NCME (1999, pp. 18–19)

## Appendix 1

Continued

Term	Definition	Use	Key Source
Criterion validation	The process by which a performance is related to the construct under examination.	If criterion scores on the assessment are related to performance levels on related measures—the relationship between, for example, ePortfolios and SAT® Writing scores—evidence of criterion validation is present.	AERA, APA, NCME (1999, pp. 56–57); Haertel (2006, pp. 66–67); Kane (2006, pp. 18–19)
Error in sampling	Defined as the difference between the sample and the given population, error exists when the outcome of the research fails due to sampling plan design.	Type I error (blindness) may be controlled by specifying a confidence interval for the sample; Type II error (gullibility) may be controlled by sample size.	Rosco (1968, pp. 152–158)
Evidence-centered design (ECD)	The evidence-centered design model focuses on assessment as an activity based on evidence.	Adherence to an evidence-centered design model allows researchers to anticipate the validation argument that will be offered in the design stage of the assessment.	Mislevy (2007); Mislevy, Almond, & Lukas (2003)
Mediated communication	The transactional nature of communication is transformed—that is, mediated—in digital environments.	As researchers recognize that communication is made complex in multi-modal environments, they will be better able to achieve construct validation.	Bolter (1999); Coppola & Elliot (2010); Murray (2009); Yancey (2004)

## Appendix 1

Continued

Term	Definition	Use	Key Source
Sampling plan	A sampling plan is a designated sub-set of the larger specified population.	Because constructed response assessments are complex to design and difficult to evaluate, a randomly designed sampling plan allows the performance of the sub-sample to be representative of the specified population.	Mazzo, Lazer, & Zieky (2006, pp. 684–688)
Validation	Validation is a process by which the targeted construct, or a key aspect of that construct, is measured.	Attention to both evidence-centered design and consequential validation will help to ensure that an assessment will serve its shareholders.	Brennan (2006); Huot (2010, pp. 23–31)
Validation argument	A rhetorical term emphasizing process and audience, the validation argument presents the claim that targeted construct, or a key aspect of that construct, has been measured.	The Toulmin model of logic is well suited to the presentation of validity arguments.	Kane (2006, pp. 27–31); Toulmin (1958)
Variable model	A variable model is the construct to be measured expressed in terms of relationship between the predictor ( <i>X</i> , or independent) variables and the outcome ( <i>Y</i> , or dependent) variable.	A variable model allows the construct to be expressed in terms of its component elements.	Coppola & Elliot (2007, 2010); Elliot, Briller, & Joshi (2007); Johnson (2006a, 2006b); White (2005)

## Appendix 1

Continued

Term	Definition	Use	Key Source
<i>Statistical Terms</i>			
Confidence interval	A confidence interval is the range of scores thought to include the mean score of the specified population.	A confidence interval allows the researcher to provide a validity argument that the sampling plan is representative of the specified population.	Lockhart (1998, pp. 224–234)
Correlation	A correlation coefficient is a number that ranges from 1 (perfect) to 0 (no relationship) expressing the relationship between two variables.	Used in consistency estimates, a correlation coefficient (such as Pearson $r$ ) provides an estimate of inter-rater reliability and a probability of the relationship occurring by chance. Correlations can also be used to gain information on variable models and establish criterion validation.	Lockhart (1998, pp. 485–486)
Descriptive statistics	The use of descriptive statistics—the mean, mode, median, and range—allows basic analysis.	The use of descriptive statistics allows a basic sense of patterns, often displayed graphically.	Lockhart (1998, pp. 51–80)
Inferential statistics	Inferential statistics—probability, regression, and tests of significance—allow relational evidence to be drawn.	Correlation and regression yield associative and predictive evidence.	Rosco (1968, pp. 144–262, 273–284)
Mean	The mean is the sum of scores divided by the number of scores.	The balance point of the scores, or the average, is a central feature of descriptive statistics.	Lockhart (1998, pp. 74–75)

## Appendix 1

Continued

Term	Definition	Use	Key Source
Median	The median divides a set of scores into two halves.	Defining the middle score allows a description of the lower and upper half of the scores.	Rosco (1968, pp. 40–41)
Mode	The mode is the most frequently occurring score.	Analysis of the mode of scores allows examination of distribution.	Lockhart (1998, pp. 73–74)
Probability	The probability of a behavior occurring, such as a score, is equal to the relative frequency of the score occurring in the larger population.	Expressed in terms of a confidence interval, the probability estimate provides evidence of certainty that the sub-population is representative of the larger specified population.	Rosco (1968, p. 117)
Range	The range allows a description of score dispersion.	Analysis of a range of scores demonstrates the extent to which scores distributed.	Rosco (1968, pp. 45–46)
Regression	Regression analysis, indicated by the coefficient of determination, allows strength of models to be analyzed and their probability estimates to be drawn.	A regression analysis demonstrates the prediction of the relationship between the predictor (X, or independent) variables and the outcome (Y, or dependent) variable.	AERA, APA, NCME (1999, p. 21); Lockhart (1998, pp. 448–507)
Specified deviation	The specified deviation is defined as the deviation that the researcher can tolerate between the sample mean of the sub-population and the true mean of the larger population.	The specified deviation is a measure that allows the researcher to be confident, at a designated level, that the mean score of a sub-group is representative of the scores of the total population.	Kerlinger & Lee (1999, pp. 297–298)

## Appendix 1

Continued

Term	Definition	Use	Key Source
Standard deviation	The standard deviation, the square root of the variance, is a measure of score dispersion.	As a descriptive measure, the standard deviation allows determination that the percentage of scores will lie within certain intervals from the mean score.	Lockhart (1998, pp. 80–82)
Standard error of the mean	The standard error of the mean is calculated by dividing the standard deviations by the square root of the population under investigation.	The standard error of the mean allows researchers to estimate how much the sample size means may vary if different samples are taken from the same population.	Norusis (2011, p. 98)
Tests of significance	Built on a family of distribution curves with the single parameter as degrees of freedom (the number of observations on which an estimate is based), tests of significance allow the researcher to determine if score differences are statistically significant and are unlikely to have occurred by chance.	The independent sample <i>t</i> -test allows examination of the degree of difference of the scores of two groups.	Lockhart (1998, pp. 230–233)
Weighted kappa	A measure of consistency, the weighted kappa allows benchmarks for strength of agreement.	Cohen's weighted kappa ( $\kappa$ ) allows interrater reliability to be determined.	Landis, J. R., & Koch, G. G. (1977).
Z-score	The Z-score, or standard score, allows scores to be transformed so that that they have the same mean and standard deviation.	In sampling plan design, the use of a designated Z-score allows a confidence interval to be established for the representativeness of the scores in the sub-sample.	Kerlinger & Lee (1999, pp. 297–298)

## Appendix 2

### Descriptive statistics for the NJIT Model (n = 636)

	Mean Scores								Standard Deviations								Range							
	f04	s05	f05	s06	f06	s07	f08	s09	f04	s05	f05	s06	f06	s07	f08	s09	f04	s05	f05	s06	f06	s07	f08 <sup>1</sup>	s09
n	61	50	124	140	92	88	25	56	61	50	124	140	92	88	25	56	61	50	124	140	92	88	25	56
ePortfolio	---	---	---	---	8.79	8.59	8.24	8.71	----	---	---	---	2.41	1.64	1.20	1.22	---	---	---	---	2,13	3,12	5,10	7,12
Clear Style	7.77	7.94	8.19	7.58	8.15	8.08	8.16	8.46	1.28	1.48	1.95	1.84	1.87	1.66	.99	1.44	4,10	4,11	2,12	2,12	3,13	3,11	6,10	5,12
Accurate Usage	7.54	7.76	7.77	7.31	7.52	7.75	7.64	8.18	1.40	1.57	2.16	1.70	1.72	1.60	1.19	1.54	4,9	4,11	2,12	2,12	3,12	3,11	6,9	2,12
Task Knowledge	8.00	8.32	8.20	7.76	8.04	8.18	8.08	8.50	1.22	1.38	2.05	1.75	1.79	1.81	1.29	1.48	5,11	6,11	2,12	2,12	3,12	2,11	4,10	4,12
Relevant Content	7.93	7.84	8.25	7.45	7.97	8.18	7.80	8.43	1.20	1.53	1.89	1.73	1.69	1.81	1.32	1.57	5,11	4,11	2,12	2,11	4,12	2,11	4,10	4,12
Adapted Tone	7.82	7.94	8.19	7.41	7.78	8.07	7.68	8.11	1.27	1.43	1.92	1.73	1.47	1.57	.99	1.85	5,10	5,11	2,12	2,12	4,12	3,11	5,9	3,12
Graphic Cohesion	7.84	8.00	8.10	7.59	8.13	8.02	8.24	8.48	1.52	1.67	1.98	1.84	1.70	1.75	.90	1.66	3,11	5,11	2,12	3,12	4,12	3,11	6,10	5,12
Citation	----	---	---	5.12	6.50	7.22	6.44	7.66	----	---	---	2.77	3.23	3.49	1.85	2.84	---	---	---	2,11	2,12	2,12	3,11	2,12
Overall Portfolio	7.82	8.08	8.45	7.66	8.13	8.19	8.04	8.63	1.34	1.44	1.98	1.83	1.79	1.70	1.34	1.65	4,10	4,11	2,12	2,12	5,12	2,11	4,10	7,12

<sup>1</sup>This semester the range of scores was limited because the only analyzed portfolios were those with no discrepancies.



## Appendix 3

### Correlation analysis of the present NJIT Model with criterion variables of course grade and cumulative GPA, Spring 2009 (n= 56)

	ePortfolio Design	Clear Style	Accurate Usage	Task Knowledge	Relevant Content	Adapted Tone	Graphic Cohesion	Citation	Overall Score	Course Grade	Cumulative GPA
ePortfolio Design	-	.607**	.387**	.506**	.522**	.475**	.654**	.203	.563**	.059	.162
Clear Style		-	.742**	.607**	.666**	.659**	.52**	.217	.658**	.135	.283*
Accurate Usage			-	.64*	.705*	.697*	.463*	.338*	.637**	.119	.161
Task Knowledge				-	.768**	.734**	.589**	.232	.706**	-.063	.127
Relevant Content					-	.762**	.657**	.257	.682**	.038	.22
Adapted Tone						-	.629**	.257	.731**	.063	.135
Graphic Cohesion							-	.27*	.599**	.031	.009
Citation								-	.4**	.29*	.208
Overall Score									-	.009	.111
Course Grade										-	.527**
Cumulative GPA											-

\*p<.05

\*\*p<.01

**Author information**

Carol Siri Johnson is associate professor of English at New Jersey Institute of Technology (NJIT). She is author of *The Language of Work: Technical Communication at Lukens Steel, 1810 to 1925* (Baywood, 2009).

Norbert Elliot is professor of English at NJIT. He is author of "Assessing Technical Communication: A Conceptual History" in *Assessment in Technical and Professional Communication*, edited by Margaret N. Hundleby and Jo Allen (Baywood, 2010).