Options for Written Language Assessment

Mary Ross Moran

"Since writing is a learned set of skills, not an innate ability, comprehensive assessment of writing skills must be understood in terms of the educational goals of a society and the chief expression of those goals—the schools" (Faigley, Cherry, Jolliffe, & Skinner, 1985, p. 99).

"Writing is more likely to be assessed than to be taught" (Applebee, 1984, p. 184).

"Time and again, the methods that have been employed to measure writing have been criticized as either unreliable or invalid" (Charney, 1984, p. 65).

Concerns about writing assessment are summarized in the above quotations. Why do educators want students to learn to write? Should schools teach learners merely to imitate, in answers to essay questions, the ideas and organization previously presented by lectures or textbooks? Is writing to be conceptualized as communicative and interpersonal in purpose? Should writing also be taught as a noncommunicative means of self-expression, discovery of meaning, application of inductive and deductive reasoning to the development of a personal set of values?

Do adopted assessment practices reflect the instructional program, or has measurement lagged behind teaching progress? Do schools indeed assess more than they teach writing? Might available assessment instruments and practices dictate which writing skills are taught? What are prevailing assessment procedures for writing? How do reported reliabilities and validities differ across options?

Only the last two of these questions are addressed in this article. The other questions raise policy issues and implications for empirical research extending well beyond the scope of this discussion. Any consideration of writing assessment, however, must acknowledge the context in which assessment occurs. Administrators and instructors, including special educators, select assessment practices in accord with desired educational or social outcomes of writing instruction.

If, for example, the goal of writing proficiency for handicapped learners is merely to meet the demands of classroom assignments and to be able to take notes and written tests in mainstream classes, it is enough to measure skills in textbook organization of referential material with attention to surface conventions of grammar, punctuation, and spelling. If, on the other hand, writing instruction for handicapped learners is to encompass all six language functions described by Jakobson (1960)—referential, persuasive, emotive, metalingual, poetic, and interpersonal—assessment practices must be broadened to cover more types of writing for varied audiences and purposes.

Dr. Moran is Associate Professor in the Department of Special Education, University of Kansas, Medical Center, Kansas City.

TERMINOLOGY

For purposes of this discussion, terms are defined as follows.

Writing encompasses both composition and transcription, both process and product.

Composition refers to the substantive activities of selecting appropriate relationships among form, audience, and purpose, generating original text, and organizing text into a variety of styles in accord with conventions of literary forms.

Transcription skills are those superimposed upon composition when language is in written form; that is, writing requires transcribing language into another medium requiring a literary, rather than a colloquial, language register. Conventions include end punctuation and capitalization to mark sentences, commas and other internal marks for clarity, spelling, handwriting, segmenting with subheadings, and paragraphing.

Indirect measures are those that do not require composing skills—no decisions about audience or purpose, no original text; tests measure knowledge of transcribing rules.

Direct measures apply evaluation criteria to writing samples to examine both composition and transcription skills across varied tasks.

Process is the act of writing while it is in progress, typically including recursive stages of preplanning, drafting, revising, and editing.

Product is the completed version of a writing task as submitted to meet an assignment or conveyed to a reader as a final form of a communication.

Holistic evaluation is based on an impressionistic reaction to a product on first reading without attempting to account for the relative contributions of individual components; raters ranks products in accord with anchor papers and assign a single score for overall merit.

Analytic evaluation involves examination of selected components of the writing task to isolate skills for instruction; instruments may include checklists, numerical scores, or qualitative judgments based on sets of decision criteria.

ASSESSMENT SELECTIONS

Researchers, administrators of writing programs, and individual instructors have developed assessment procedures for written language. A review of mainstream composition literature discloses a number of options, along with expert opinion on their respective merits. The search for reliable procedures—those yielding consistent results—and valid measures—those that comprehensively assess the real tasks in which writers engage—is ongoing.

Current choices permit relative emphasis upon (a) *composition* or *transcription* proficiencies, (b) *indirect* or *direct* measurement, (c) *process* or *product* direct evaluation, and (d) *holistic* or *analytic* product examination. Members of these pairs are not polar opposites; rather, they are extreme points on a continuum with some degree of overlap at the center. Though the pairs are useful to organize information, they are not parallel concepts; some pairs are subsumed under others. *Process* and *product* are types of *direct* measurement; *holistic* and *analytic* are subtypes of *product* evaluation.

Each procedure has advantages and disadvantages for different purposes. Choices are made on the basis of cost-benefit ratio according to the way results are to be used.

For *administrative* purposes—determining which students are assigned to remedial or honors English sections, for example, or estimating district students' writing status in

comparison to national peers—*indirect* methods limited to *transcription* skills or *direct* evaluation of *products* using *holistic* procedures have been considered appropriate. When assessment is conducted for *instructional* planning, however, emphasis on *composition* through *direct* measurement, and use of *analytic* procedures have been recommended.

For special educators, the three purposes of writing assessment are:

—to establish whether a performance problem contributes to eligibility for special services.
—to gain diagnostic information to plan appropriate instruction.
—to monitor progress through and outcomes of instruction.

Indirect norm-referenced measures limited to *transcription* skills offer relative data useful for documenting an academic performance problem in general English proficiency. Special educators who wish to teach *composition* as well as *transcription* skills, however, employ *direct* measurement of multiple writing samples to derive short-term objectives. The following discussion offers options to be considered before selecting writing assessment procedures to be incorporated into the process of developing individualized education programs.

COMPOSITION EVALUATION

Composing is the act of creating novel connected discourse, either oral or written. The task imposes substantive demands on formulation and arrangement of language.

Composition acknowledges three discourse variables—writer, text, and audience. Closely associated with this communicative triangle is the notion of classes of discourse, typically identified as *expressive, narrative, descriptive, expository,* and *argumentative* prose, along with types of poetry and drama (D'Angelo, 1984). Discourse aims (Jakobson's six as listed earlier) overlap with, but offer concepts broader than, discourse classes. Composing further subsumes recursive activities called *prewriting, drafting, revising,* and *editing* (Flower & Hayes, 1981).

The communicative interaction of writer, text, and audience is based on classical rhetorical theory. Composing demands decisions about appropriate relationships among the three. Voice or tone, relative formality or informality, and selection of informative, emotive, or logical content depend upon a writer's perception of the intended audience and the desired effect on that audience. Bases for decisions are multiple—familiarity with the topic, how well a writer knows the readers, which genre or medium is conventionally used for specific purposes, and the importance of the task to personal goals (Selzer, 1984).

The assessment question is the extent to which writers have integrated their purposes with the needs or interests of their audiences by selections of content and form. For example, a mismatch occurs when a writer chooses slang vocabulary for logical and ethical arguments in an essay designed to persuade school administrators to delete a dress code.

Though the five discourse types listed above are frequently mentioned, numbers of other taxonomies have been suggested (D'Angelo, 1984). Discourse modes have been criticized as inadequate conceptual frameworks for writing or as secondary to the aim of the writing task (Lloyd-Jones, 1977). Nevertheless, the notion that identifiable combinations of purpose and audience give rise to specific requirements of content and form has persisted in writing assessment. Evaluators attempt to determine the extent of compliance with those constraints. For instance, a writer embarking on a narrative is expected to establish characters, setting, and chronological sequence of events; a writer developing an argument must provide a conclusion drawn from premises and support. Furthermore, writing for the referential purpose of explaining a procedure requires less emotive and poetic vocabulary but more rigid sequence than does an imaginative work.

Cognitive processes involved in composing have been described in working hypotheses resting on four key points: (a) writing is a set of distinctive thinking processes; (b) these processes have a hierarchical, embedded organization; (c) writing is guided by the writer's developing goals; (d) writers create and alter goals throughout the composing process (Flower & Hayes, 1981, p. 366). "Composing theory has moved beyond the old linear 'stage models' of composition (prewriting, writing, revising) to a recognition that composing activities are more recursive than linear, more flexible than constrained" (Selzer, 1984).

When Flower and Hayes conceptualized stages as recurring throughout the composing process (p. 367), they also identified a need for a novel type of assessment procedure to investigate the nature of rhetorical and stylistic decisions. Insight into stages is dependent upon *process* assessment. The assessment question is whether writers appropriately incorporate into their composing strategies, in accordance with the communicative triangle and constraints of discourse type and purpose, elements of planning, drafting, revision, and editing. For example, an argumentative term paper on nuclear disarmament demands more planning, organization, and revision than does a weekly one-page formatted laboratory report for biology class.

TRANSCRIPTION EVALUATION

As distinct from substantive elements of the writing task, components discussed under this category are variously addressed in the literature as *mechanics, conventions,* or *surface features*. The concept of *transcribing* is used here because the student writer's task is analogous to transcribing music into another key; oral and written language codes differ, and a developing writer must learn to transcribe from a known, relatively informal code to a sometimes more formal one with different cues.

To give an obvious example, oral features of falling or rising intonation are marked in writing by a period or a question mark—an adjustment that may not be difficult to master; but speaking permits *phrases* to be so marked whereas most school writing allows only independent *clauses* to be marked as sentences. More complex relationships between oral and written codes, described by Elbow (1985), include many similarities but enough differences to demand greater clarity in "final drafts . . . that could survive outside the context of local author and local audience" (p. 291).

Six features might be subsumed under *transcription*— literary language register, spelling, punctuation , capitalization, handwriting, and spatial conventions such as paragraph indentations, margins, headings, and footnotes. With the exception of personal notes or letters, written English in schools is to be in literary register—as opposed to the colloquial version appropriate for most spoken language. Spelling, punctuation, and capitalization can obscure messages of writers who are unfamiliar with these conventions (Shaughnessy, 1977). Handwriting is less critical in an age of word processing, but some studies have shown that it affects judgments of writing quality (Charney, 1984). Conventional spatial features clearly are related to overall coherence of text (Bamberg, 1983).

Evaluation of transcribing skills has been characterized as unrelated to language competence, as "a test of social conformity, of how well a person recognizes the language forms most commonly used by those in authority . . . but that does not make it a test of skills in discourse" (Lloyd-Jones, 1977, p. 34). Nevertheless, transcribing skills that have not reached an automatic level cannot be ignored in writing assessment because, "For children and inexperienced writers, this extra burden may overwhelm the limited capacity of short-term memory. If the writer must devote conscious attention to demands such as spelling and grammar, the task of translating can interfere with the more global process of planning what one wants to say" (Flower & Hayes, 1981, p. 373).

INDIRECT MEASURES

Indirect procedures require students to make judgments about correctness of contrived items, to correct another writer's errors, or to spell or punctuate dictated material. Occasionally an item requires rewording a sentence or paragraph, but testees do not engage in formulation of original connected text. Instead, tasks follow a multiple-choice format, a cloze procedure, or, infrequently, sentence-combining activities. The focus of measurement is *knowledge* about writing convention, rather than *application* of those principles to real or simulated writing purposes.

Indirect measures carry the advantages of ready availability, quick administration, and objective scoring requiring limited professional time. Indirect methods far outperform direct procedures in terms of interrater agreement (Charney, 1984). Indirect tests typically are norm-referenced, offering advantages of known reliability and standard error of measurement, and they establish testees' status relative to national age and grade peers. This information can be useful when the assessment question is the extent of conformity with conventions.

One disadvantage of indirect measures has been associated with the cost-benefit ratio. Because they are quick to score, objective tests are said to save examiner time, offsetting the limited data they provide (Stiggins, 1982). As White (1984) pointed out, however, expenses associated with developing a standardized indirect measure are considerable; costs to consumers are not reflected in professional scoring time, but expenses of field testing and norming are passed on to schools in the purchase price of indirect instruments.

A more critical disadvantage is inadequate validity because of failure to tap *composition* components of the writing task. Since indirect items are selected to permit scoring ease or to provide variability across groups of testees rather than to cover essential skills, indirect measures are extremely limited in coverage (Cooper & Odell, 1977). They typically measure only *transcription* skills of literary usage, punctuation, capitalization, and spelling.

The very qualities that make indirect measures reliable therefore result in inadequate validity. In particular, *content* validity is questionable because tasks are so far removed from the desired outcome of writing instruction—the ability to generate written text. Results depend on inferences about a complex set of skills that are never demonstrated. Indirect tests "are insensitive to a students' ability to write cogent, coherent and fluent prose" (Charney, 1984, p. 67).

Criterion-related validity has been estimated by a number of studies correlating objective usage scores with evaluation

of direct writing samples. Six studies over subjects ranging from third grade through college reported correlations as high as .75 (Stiggins, 1982). Writing samples for these studies, however, were evaluated holistically; since some evidence indicates that holistic scores may be unduly influenced by surface features such as grammatical usage, punctuation, capitalization, spelling, and even handwriting (Charney, 1984), criterion-related validity studies may be contaminated by limitations of the direct measures.

Studies of *predictive* validity of specific indirect measures typically compare test scores to grades assigned for language arts or English classes. Again, studies have produced correlations in satisfactory ranges (Cooper & Odell, 1977). Class grades, however, are based on reading and interpretation of literature, and many activities other than writing. Indeed, writing of paragraph length or more occupies a minor role in English classes (Applebee, 1984). Therefore, critics have argued that indirect measures correlate with *general language* proficiency rather than with *writing* skill (Lloyd-Jones, 1977).

Examples of individually administered indirect measures illustrate some of the validity problems. For instance, the Written Language Cluster of Part Two of the Woodcock-Johnson Psychoeducational Battery (Woodcock & Johnson, 1977) includes spelling to dictation and proofreading of sentences to locate spelling, punctuation, or usage errors and provide an oral correction. This battery is used as a survey of academic achievement of students referred to special services (McLaughlin & Lewis, 1986).

Beyond the central issue of testing knowledge rather than application, several questions arise about whether the test measures what could be called a *written language cluster*. Testees are asked to spell to dictation a series of single words (not a real-life writing task) that have not first been tested to establish that the words are in testees' oral or reading vocabularies. Sentences presented for proofreading not only have been written by someone else (not a real editing situation) but also are read silently so an examiner cannot judge whether testees have read accurately, and thus cannot determine possible contamination attributable to reading disability.

Other examples of indirect individual tests identified by publishers as measures of written language skills are the Spelling, Word Usage, and Style subtests of the Test of Written Language (TOWL) (Hammill & Larsen, 1983). Again, spelling words are dictated without first being presented for vocabulary knowledge or sight reading, and sentences to be completed for Word Usage and those rewritten for Style are read silently. Although examiners are instructed

to decode words for testees on request, students may not ask for help when needed. These subtests thus share with the Woodcock-Johnson subtests the problem of unknown contribution of reading skills to the outcomes. The TOWL however, includes an examination of a single narrative direct product—though far from comprehensive—along with indirect measures.

The Brigance set of criterion-referenced Writing measures, like the Woodcock-Johnson norm-referenced instrument, is limited to indirect items. In the instrument designed for older subjects, the Diagnostic Inventory of Essential Skills (Brigance, 1981), transcription skills of letter formation, handwriting quality, capitalization, punctuation, and eight subtests of spelling are covered. A contrived task involves composing sentences using required words; other items require filling in blanks on standard forms. Generating a letter, which is analyzed for formal elements, constitutes the single composing task. Most of the same items occur in two other Brigance volumes. Inferences about *composing* cannot be drawn from these subtests, though they have some utility as measures of *transcribing* proficiency.

In summary, indirect measures offer advantages as screening instruments to estimate general knowledge of formal, literary language and written conventions. These tests, however, are inadequate for purposes of planning writing instruction as they cannot predict how an individual writer will apply knowledge when presented with a task requiring *composition* skills (Cooper & Odell, 1977).

DIRECT MEASURES

Direct procedures require testees to engage in production of original, connected discourse. Examiners may observe writers in action, videotape them, or ask them to verbalize the *process* to create a protocol. When the task is completed, direct *products* in the form of written language samples can be evaluated.

Typically an examiner presents testees with a brief writing prompt, such as a question or a statement that invites writers to express feelings, recount an event, describe an object, person or place, explain a procedure, express and support an opinion, or any combination thereof. Well stated prompts identify the form of the writing to be produced, the specific audience to be addressed, and the purpose of the communication. Writers respond by preparing original written discourse of at least paragraph length, in accordance with directions accompanying the prompts.

Direct measures offer advantages of original connected text available for examination, a simulated communicative

purpose for the exercise, and at least potential application of a broad variety of composition and transcription skills (Cooper & Odell, 1977). Writers are required to interpret the prompt, to plan, generate, draft, revise, and edit original text in accord with the demands of a specific communicative task (Flower & Hayes, 1981). Testees have considerably more control over the material than over indirect items.

Direct measures generally are conceded to provide much higher content or face validity than do indirect tests. Since testees are engaging in the criterion, real-life task, evaluation is applied to the very skills that most experts would call *writing* (Cooper & Odell, 1977). The major disadvantage of direct measures is that they are time-consuming. The cost in professional staff commitment is high (Stiggins, 1982).

Acceptable intrarater and interrater reliabilities, as well as content validity, depend upon controlling variables associated with elicitation, process protocols, and scoring. But most experts have argued that *only* direct procedures can be said to measure *writing* as defined to include both *composition* and *transcription*. Therefore, the many precautions that must be taken to ensure reliable and valid outcomes are considered to be balanced by the utility of the data provided by direct measurement.

Process Evaluation

Scardamalia and Bereiter (1986) have argued that the distinction between process and product is somewhat artificial because "the more realistically learners are aware of what product characteristics they should be striving for and of how successful they are in achieving them, the more likely it is that the attendant cognitive processes will develop" (p. 798). These two types of assessment become better integrated as research on process continues to uncover writers' strategies (Gorrell, 1983).

The rationale for process analysis is that developing writers can benefit from bringing to consciousness the strategies used in composing. Process is a set of problem-solving sequences through recursive stages of planning, drafting, revising, and editing. Bases for choices become accessible by verbalizing and studying decision points.

Several researchers (Flower & Hayes, 1981; Perl, 1979; Sommers, 1980) have compared inexperienced writers with experts to study similarities and differences in strategies. Research has shown that some writers plan and revise extensively for most purposes, while others engage in relatively planless single-draft writing for many tasks (Selzer, 1984). "While nothing approaching a comprehensive theory of development of composing processes has yet been suggested, researchers have identified certain cognitive strategies that seem to be requisite of mature writing ability. Moreover, the work of some researchers has suggested that writers' increasing awareness of these strategies may signal their having achieved greater control of composing processes" (Faigley, Cherry, Jolliffe, & Skinner, 1985, p. 165).

Two methods, which may be combined in various ways, have been associated with assessing writing process—(a) thinking-aloud protocols, and (b) interviews, both following and during the task. Thinking-aloud protocols, the most widely reported procedure, have been described by Emig (1971), Flower and Hayes (1981), and Perl (1979). Examiners ask writers to verbalize musings during the composing act. Writers' talk is tape-recorded; then transcripts are prepared and subjected to analysis. Protocols offer the advantage of immediate access to what is going through writers' minds rather than relying on recall. A disadvantage is the intrusiveness of the procedure, which distracts writers from the task by requiring them to verbalize.

Interviews usually follow completion of a written product. Researchers observe writers composing, then question them about their practices. This approach has the advantage of introducing no intrusive procedures during the task, but it is potentially hampered by failures of recall or inferences drawn by writers about processes of which they were not conscious during composition.

Instead of asking writers to wait until the process has been completed, Faigley et al. (1985) have developed a *process log* to present questions at various points prior to and between drafting, revision, and editing stages. Intermittent interviews reduce the likelihood of losing data through memory lapse. If, however, writers engage in processes not covered by the specific questions and the accompanying scales, examiners are not likely to learn about them.

Interview methods sometimes make use of videotapes that record writing in progress for later playback. Examiners ask open-ended questions with the tape serving as a prompt. Though Rose (1984) reported that his use of this procedure improved recall, he also concluded that reports of strategies were less accurate than thinking-aloud procedures about the order of activities and the time spent on each.

Research methodologies to document writers' strategies are not yet sufficiently operationalized to provide immediate assessment information to instructors, nor have reliability and validity studies received attention. Nevertheless, procedures are described in the literature for those who would like to investigate process and are willing to translate research methods into classroom assessment tools. To date,

methods of assessing process strategies have been limited to single-subject procedures rather than systems applicable to groups (Faigley et al., 1985). Therefore, process methods can be expected to serve instructional, rather than administrative, purposes.

Product Evaluation

Direct *product* evaluation involves examining samples collected in natural settings, such as classroom assignments, or samples elicited specifically for assessment. Because classroom writing samples may be contaminated by inappropriate assignments, inadequate or misunderstood directions, and other unfavorable conditions (Applebee, 1984), elicited samples are more widely used.

Score stability within raters, across raters, and over time has been satisfactorily established for some types of scoring systems applied to *direct products* (White, 1984). But a list of variables affects the statistical consistency of ratings. "The list includes: the number of separate readings of each writing sample, the number of writing samples evaluated per student, the writing topic, the size of the writing scale, the consistency with which the readers are trained, the conditions under which the papers are read and so on" (Charney, 1984, p. 70).

Though content validity is superior to that of indirect measures, some types of direct-product evaluation nevertheless have been criticized on the basis of validity. Bias has been ascribed to the nature of writing prompts, including the extent to which they tap prior topic knowledge (Langer, 1984). Different discourse modes (Lloyd-Jones, 1977), varying cognitive demands (Flower & Hayes, 1981), and specific audiences to whom writing is addressed (Applebee, 1984) alter writing quality to such an extent that experts judge content validity on a case-by-case basis. "Thus it is possible to disagree on the validity of any given test, depending upon its design" (Charney, 1984, p. 71).

When the decision has been made to employ direct product measurement, evaluators may choose *holistic* or *analytic* scoring procedures. Holistic scoring has achieved prominence over the past 20 years or so as a means of permitting direct measurement of writing in programs involving large numbers of students. Analytic evaluation has evolved more recently to supplement holistic impressions of quality with detailed information for instructional planning.

Holistic Rating

Through the period of development of holistic rating, several practices have become standard. White (1984) sum-marized prevailing procedures. Trained raters read papers at the same time and place to reduce distractions and to standardize conditions. Scoring criteria are stated, defining points on a rating scale. Anchor papers are provided as concrete examples of points on the scale. Table leaders check intrarater consistency for five to seven readers and may retrain raters who demonstrate scoring drift. Two readers from two tables score each essay independently for interrater agreement.

Charney (1984) also listed conditions prerequisite to reliable and valid assessments. Precautions include employing qualified readers from similar backgrounds and training readers to conform to agreed criteria.

Noting that few studies of consistency across various types of writing have been reported, Mishler and Hogan (1982) estimated interscorer reliabilities from published studies at .85 and intrascorer reliabilities at .90. The consensus seems to be that the procedure demonstrates acceptable reliabilities if standard practices and precautions are followed rigorously (Stiggins, 1982).

According to White (1984), content validity of holistic ratings is higher than that of either *indirect* or direct *analytic* methods. According to White, writing is "an artistic and human whole, not merely . . . a collection of scraps and parts" (p. 409). By this reasoning, holistic scoring is valid because it provides an overall human response to a human activity that resists reductionism.

Charney (1984), however, questioned the validity of holistic rating on two grounds. First, whether criteria are set up in advance by experts or are derived by a homogeneous group of readers, the standards of what constitutes the "best" papers are disputable as subject to bias (p. 73). Second, the extent to which readers use standards set in advance has been called into question by studies showing that separable traits correlate with holistic ratings. "It is disconcerting to find holistic scores, which are supposed to be a *qualitative* measure, so directly predictable by such mundane quantitative measures as the length of the sample, the number of errors and the number of unusual vocabulary items" (p. 75).

In addition to advantages of satisfactory reliabilities and face validity superior to other types of assessment, holistic scoring is generally considered to be cost-efficient for scoring writing in quantity. Though two readers are involved (three in cases of disagreement), each spends an average of only 30 seconds to 1 minute (Mishler & Hogan, 1982) or 2 to 3 minutes (Cooper & Odell, 1977) on each paper.

Disadvantages center on the inability of holistic evaluation to provide any information for instructional planning. Holistic scoring merely assigns values to writing samples and

yields judgments of quality relative to other papers in the same group. "It gives no meaningful diagnostic information beyond the comparative ranking it represents. . . . All we have is a single score, where we might wish to have a profile" (White, 1984, p. 406).

An example of a specialized type of holistic scoring developed for the National Assessment of Educational Progress has been described by Lloyd-Jones (1977). Primary Trait Scoring relies upon elicited samples in response to writing prompts that specify the form, audience, and purpose of the writing. Traits associated with those specifics are identified in advance and applied to the written product to determine the extent of compliance. Development of differing criteria for varied writing assignments sets this procedure apart from holistic systems that attempt to use the same criteria for several writing assignments. Descriptions of a variety of other holistic scoring procedures were presented by Myers (1980), along with examples of scored papers.

Analytic Scoring

Analytic methods yield a series of scores, checklist marks, or descriptions for presumably separable subskills derivable from writing samples. Because this is the most recent type of measurement to develop, there is some inconsistency in terminology. Cooper (1977) used *analytic scale* to describe what he called a type of *holistic* scoring that considers separate features of the sample without requiring readers "to stop and count or tally incidents of the feature" (p. 4). In contrast, Mishler and Hogan (1982) clearly distinguished *holistic* from *analytic* scoring, which they described as a system "in which papers are rated on any number of different facets" (p. 12). A resolution to the conflict is to consider systems that result in a single score as *holistic* and those incorporating separable components that do not sum up to one score as *analytic*.

Analytic scoring often has been equated with counting or, in even narrower interpretations, with counts of *errors* only, without acknowledgment of strengths (Ney, 1986). But the concept appears to be expanding to include qualitative judgments as well as quantitative measures, strengths as well as weaknesses.

Analytic scoring involves selecting, from the many possible ways of categorizing writing subskills, some characteristics that will be measured. For example, a developer of an analytic system might consult Cooper and Odell (1977), Myers (1985), or Purves (1984) for lists of components that could be included. Under superordinate categories such as *content, organization, vocabulary,* and *syntax,* for instance, subskills might be judged according to criteria.

In contrast to holistic rating, which compares products with those of peers in the same group, analytic scoring uses only *intrastudent* comparisons across subskills. By studying a profile, examiners determine proficiencies and limitations not apparent in an impressionistic reading of papers.

Advantages of analysis have been variously described. A complex task is rendered manageable so that teachable components can be isolated (Gorrell, 1983). Identified subskills then can be taught with some concentration of effort so instructors and students do not have to try to deal with all aspects of a complex task simultaneously (Simmons, 1984). Writers' tendency to become overloaded if too many skills are attempted at once (Ammon & Keech, 1985) is avoided. Forcing evaluators to go beyond surface conventions reduces that potential source of invalidity and may lead to discovery of composition strengths obscured by transcription errors (Moran, in press).

Disadvantages of analytic scoring have been summarized by White (1984), who argued that although "analytic scoring should provide the diagnostic information that holistic scoring fails to provide," it has thus far failed to do so (p. 407). Experts have not agreed that subskills of the complex writing task are in any way separable, so artificial designation of discrete components may involve overlap and interdependence of skills. Intrarater and interrater reliabilities are difficult to establish for multiple characteristics. Because scoring is complicated and involves a number of passes through several papers, the procedure is time-consuming and, therefore, costly in professional hours. Mishler and Hogan (1982) estimated 15 minutes per paper for analytic scoring, but considerable training may be required to reach that level of efficiency.

Examples of analytic evaluations include Shaughnessy's (1977) procedures for inexperienced writers. Three systems that have been applied to writing by samples of handicapped learners illustrate varying approaches.

The Diagnostic Evaluation of Writing Skills (DEWS) (Weiner, 1980) has been applied to one direct product per writer elicited in timed writing sessions that include a segment set aside for revision. "The 41 criteria for assessment are divided into six categories to include most of the common writing errors" (p. 54).

Graphic features subsume handwriting and spacing; 18 *Orthographic* and *Phonologic* items focus on spelling; *Syntactic* subskills include six covering standard literary usage and four concerned with variety and productivity at the sentence level. The *Semantic* category includes composition items that tap vocabulary, coherence, sequencing, transi-

tions, and focus. *Self-monitoring* is credited if the writer engages in revision and editing. Scoring consists of "simple tabulation of the number of errors" (p. 55). The measure was described by its developer as "providing a permanent record of the student's deficiencies, which subsequently become the basis for remedial teaching" (p. 59).

Poteet's (1980) Inventory of Written Expression and Spelling features a Checklist of Written Expression, which the developer recommends applying to multiple samples. Elicitation procedures are not specified. Transcription items are grouped under "Penmanship" and "Spelling." Literary register and conventions (capitalization, punctuation, and syntax) appear under "Grammar." An "Ideation" section combines composition skills of comprehensibility, level of abstraction, tone, word choice, and sentence variety with tallies of total words produced and numbers of complete, run-on, and fragmented sentences. The checklist is marked to indicate whether the subskill under consideration is too advanced or adequate for the writer's age and grade level or whether it should be introduced. The developer described the system as yielding "results which suggest instructional strategies" (p. 98).

The Analytic Profile of Composing and Transcribing Skills (Moran, 1987) is applied to multiple samples representing at least three relationships among writer, text, and purpose. Samples are elicited under controlled conditions. Following an oral reading of papers for impressionistic reactions by the writer and at least one peer, a scoring guide presents decision criteria under four major categories, three covering *composition* and one devoted to *transcription*.

Under "Complexity," the system measures percentage of complex T-units (one independent clause plus embedded or attached dependent clauses) (Hunt, 1970), mean words, morphemes and free modifiers per T-unit, and varieties of clause construction. The "Continuity" section identifies five types of relationships among writer, text, and audience and provides weighted scores for degrees of compliance with constraints of each type. "Connections" cover some semantic and syntactic ties associated with unifying written discourse. "Conventions" include basic T-unit construction, sentence boundaries, capitalization, punctuation, spelling, and formal language register. Each section yields numerical data, some in percentages and others in arbitrary weighted figures.

The described analytic systems were not designed to be normed; therefore, profiles of normally developing, randomly selected writers are unknown. These measures can be called criterion-referenced because the components identified and subjected to analysis are drawn from a general domain described by writing experts in the professional literature, and results are referenced only to *intra*student rather than *inter*student comparisons. Validity decisions depend upon which list of subskills might be considered representative of *writing* by the specific experts selected to review a given analytic system.

These criterion-referenced measures are appropriately used, if applied to multiple samples written for different purposes, to describe writing characteristics of individuals for instructional planning. Criterion levels, which determine whether a given component is subject to intervention, are set by users of these systems.

Or an analytic system can be developed by school district personnel directly from the writing curriculum to which their own students are exposed. Because curriculum-based assessment takes into account the methods and materials employed in direct writing instruction, content validity is satisfied. A major advantage is that analytic systems based on district instructional objectives are specific enough to call attention to writing components that might be overlooked by more global methods of evaluation.

CONCLUSION

Consideration of similarities and differences among the options available for writing assessment leads to a conclusion that no method is inherently superior to another. Relative usefulness or relevance varies as a function of the purpose of assessment.

Therefore, indirect methods, yielding screening data on general proficiency in recognizing standard English, can contribute to evidence for or against a performance problem in language arts, though they may not distinguish *reading* from *writing* problems. Direct methods, capable of describing application of a variety of skills to a complex task, are preferable for developing preintervention and progress data for instruction, despite time-consuming precautions necessary to safeguard reliability and validity.

In turn, purpose is related to the goals of writing instruction. Assessment procedures must reflect the expectations, writing assignments, and focus of classroom lessons. Writing instruction in the mainstream has undergone shifts away from transcription skills toward increasing attention to writing as a thinking process (Faigley et al., 1985). For educators planning instruction for handicapped learners, assessment decisions will be formed by the extent to which writing programs attempt to incorporate the higher cognitive and linguistic functions of composition.

REFERENCES

Ammon, P., & Keech, C. (1985). Longitudinal research on writing assessment. In K. Greenberg & V. Slaughter (Eds.), *Notes from the National Testing Network in Writing* (pp. 13-22). New York: City University of New York.

Applebee, A. (1984). *Contexts for learning to write.* Norwood, NJ: Ablex.

Bamberg, B. (1983). What makes a text coherent? *College Composition & Communication, 34,* 417-429.

Brigance, A. (1981). *Diagnostic inventory of essential skills.* North Billerica, MA: Curriculum Associates.

Charney, D. (1984). The validity of using holistic scoring to evaluate writing: A critical overview. *Research in the Teaching of English, 18,* 65-81.

Cooper, C. (1977). Holistic evaluation of writing. In C. Cooper & L. Odell (Eds.), *Evaluating writing: Describing, measuring, judging* (pp. 3-32). Urbana, IL: National Council of Teachers of English.

Cooper, C. & Odell, L. (Eds.) (1977). *Evaluating writing: Describing, measuring, judging.* Urbana, IL: National Council of Teachers of English.

D'Angelo, F. (1984). Nineteenth-century forms/models of discourse. *College Composition & Communication, 35,* 31-42.

Elbow, P. (1985). The shifting relationships between speech and writing. *College Composition & Communication, 36,* 283-303.

Emig, J. (1971). *The composing processes of twelfth graders* (Research Report No. 13). Urbana, IL: National Council of Teachers of English.

Faigley, L., Cherry, R., Jolliffe, D., & Skinner, A. (1985). *Assessing writers' knowledge and processes of composing.* Norwood, NJ: Ablex.

Flower, L., & Hayes, J. (1981). A cognitive process theory of writing. *College Composition & Communication, 32,* 365-387.

Gorrell, R. (1983). How to make Mulligan stew: Process and product again. *College Composition & Communication, 34,* 272-277.

Hammill, D., & Larsen, S. (1983). *Test of written language.* Austin, TX: Pro-Ed.

Hunt, K. (1970). Syntactic maturity in school children and adults. *Monographs of the Society for Research in Child Development, 35*(Serial No. 134), 1-44.

Jakobson, R. (1960). Linguistics and poetics. In T. Sebeok (Ed.), *Style in language* (pp. 350-376). New York: John Wiley.

Langer, J. (1984). The effects of available information on responses to school writing tasks. *Research in the Teaching of English, 18,* 27-44.

Lloyd-Jones, R. (1977). Primary trait scoring. In C. Cooper & L. Odell (Eds.), *Evaluating writing* (pp. 33-68). Urbana, IL: National Council of Teachers of English.

McLoughlin, J., & Lewis, R. (1986). *Assessing special students.* Columbus, OH: Charles E. Merrill.

Mishler, C., & Hogan, T. (1982). Holistic scoring of essays. *Diagnostique, 8,* 4-16.

Moran, M. (1986). *Analytic profile of composing and transcribing skills.* Kansas City, MO: MAPCATS.

Moran, M. (1987). *Analytic profile of composing and transcribing skills.* Kansas City, MO: Mss. Press.

Myers, M. (1980). *A procedure for writing assessment and holistic scoring.* Urbana, IL: National Council of Teachers of English.

Myers, M. (1985). *The teacher researcher: How to study writing in the classroom.* Urbana, IL: National Council of Teachers of English.

Ney, J. (1986). Error analysis, theories of language, and the teaching of writing. *Written Communication, 3,* 15-30.

Perl, S. (1979). The composing processes of unskilled college writers. *Research in the Teaching of English, 13,* 317-336.

Poteet, J. (1980). Informal assessment of written expression. *Learning Disability Quarterly, 3,* 88-98.

Purves, A. (1984). In search of an internationally valid scheme for scoring compositions. *College Composition & Communication, 35,* 426-438.

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

Scardamalia, M., & Bereiter, C. (1986). Research on written composition. In Wittrock, M. (Ed.), *Handbook of research on teaching* (pp. 778-803). New York: Macmillan.

Selzer, J. (1984). Exploring options in composing. *College Composition & Communication, 35,* 276-284.

Shaughnessy, M. (1977). *Errors and expectations.* New York: Oxford University Press.

Simmons, J. (1984). The one-to-one method of teaching composition. *College Composition & Communication, 35,* 222-229.

Sommers, N. (1980). Revision strategies of student writers and experienced adult writers. *College Composition & Communication, 31,* 378-388.

Stiggins, R. (1982). A comparison of direct and indirect writing assessment methods. *Research in the Teaching of English, 16,* 101-114.

Weiner, E. (1980). The diagnostic evaluation of writing skills (DEWS). *Learning Disability Quarterly, 3,* 54-59.

White, E. (1984). Holisticism. *College Composition & Communication, 35,* 400-409.

Woodcock, R., & Johnson, M. (1977). *Woodcock-Johnson psychoeducational battery.* Allen, TX: DLM.
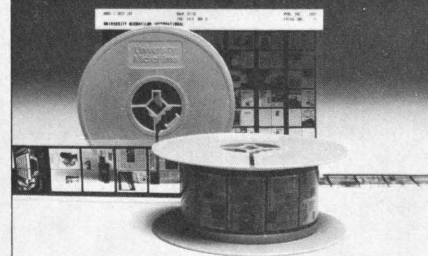
A Horticultural Training Program for Special Students

Douglas Airhart and John Tristan

Horticultural therapy is a dynamic process that incorporates the tasks and activities of horticulture to benefit individuals with special needs (Airhart & Cronin, 1981; Olszowy, 1978; Relf, 1981; Thoday, 1978). Therapists can provide behavior and job skill training through structured programs to help individuals develop an improved self-image and acquire social skills while learning a trade (Wehman, Moon, & McCarthy, 1986; Thoday, 1978).

A horticultural training program for students who are unable to attend regular high schools because of intellectual or social impairments has been initiated at Durfee Conservatory, the University of Massachusetts (Daubert & Rothert, 1981; Hudak & Malloy, 1980; Relf, 1980). This facility has five attached greenhouses containing a tropical plant collection of over 800 plants including full-size papaya, banana, cacao, orange, and coffee trees in fruit and a representative array of economic and ornamental plants. Several outdoor plots showcase annual flower and vegetable varieties new to the commercial trade. A selection of culinary and medicinal herbs and flowering perennials also is included.

Moderately retarded adolescents ranging in age from 16 to 21 are selected for training at Durfee Conservatory. The horticultural training program, incorporating instruction in all phases of greenhouse management and horticultural activities, also exposes the students to social activities (taking public transportation to and from the worksite, becoming familiar with work routines, providing the opportunity to interact with age-appropriate peers who visit the Conservatory, etc.) and activities of daily living (making change, telling time, etc.). Plant and Soil Sciences undergraduate students enrolled in independent study projects, as well as the faculty, provide the supervision and instruction.

Activities undertaken are in greenhouse and outdoor gardening relating primarily to greenhouse management, sanitation, plant propagation, and plant culture. More specifically, tasks may consist of maintaining sanitation in the greenhouse; preparing and amending ground beds; propagating plants; transplanting and repotting specimens; mulching; watering; fertilizing; preparing labels and signs; identifying and displaying plants; controlling weeds; and building a yearly compost heap. Use of standard garden tools is required.

Weekly activities in addition to regular work tasks are incorporated to stimulate creative processes and as a motivating reward. These include making flower arrangements, corsages, terraria, dish gardens, flower boxes, wreaths, pine cone decorations, bird feeders, and bonsai, and propagating special plants that capture the adolescents' interest.

Once accepted into the program, a complete profile summary sheet of each student is obtained. Pertinent information includes an assessment of learning and physical disabilities, personal interests, work experience, family background, and general comments. This material is reviewed before tasks are initiated, to adequately approach and evaluate the individual needs of each. During an initial adjustment period (usually 3-4 weeks) job compatibility, adjustment, and strengths and weaknesses in ability are carefully monitored. A general plan or goal, defined by specific work duties, then is developed for each student.

A solid structural format for the work process is provided (Hudak and Malloy, 1980; Popovich, 1981). A timetable of each day's activities, with scheduled days and working hours, is assigned to each student. Time slips are signed daily for a token payment scheme arranged with students' parents. This immediately reinforces good working effort and fixes in a student's mind the concept of monetary reward for a job well done. One rest break is allowed during the work period, and permission is required to leave the work area at all other times. Progress is check with a base time schedule allotted for each task (Hudak & Malloy, 1980).

A clear statement of a work task in defined steps is essential. Frequent demonstrations (modeling) and graduated guidance while working side by side are the most effective aids in task acquisition and comprehension (Popovich, 1981). Verbal instructions are short and direct and are repeated as necessary to teach a skill. All prompts are faded gradually, with intermittent reinforcement contingent upon successful effort. Supervision on a one-to-one basis is preferred.

Benefits of the program are twofold. University students receive educational field experience in horticultural training and exposure to the vicissitudes of a new career option. The opportunity to work with special needs students—to teach them a trade that one day may lead to sheltered or competitive employment—provides rewarding, beneficial, and well-earned scholastic credit. Daily personal contacts in one-to-one situa-

Dr. Airhart is a Master Horticultural Therapist, School of Agriculture, Tennessee Technological University; and John Tristan is a Senior Technical Assistant, Department of Plant and Soil Science, University of Massachusetts.

tions with a client bring the practices of vocational training, personnel management, and horticulture directly into their lives in a way no classroom could.

The vocational students acquire new skills that improve their own lives. Constantly reinforced in a positive way, they learn new tasks and a sense of job responsibility to bolster their growing self-reliance (Thoday, 1978). Marked improvement in ability, from near total dependence to a degree of self-autonomy, has been noted in some clients. For them, an improved degree of self-sufficiency in adult society is a worthy achievement.

Responses of the adolescents, their parents, the university students, and visitors have been favorable. Interest in horticultural therapy has grown steadily as university students consider this new career option. In all, the future looks bright for continuing successes in this endeavor, combining people and plants in a horticultural therapy program designed for the betterment of both adolescents with special needs and the university students who help them work toward this end.

ALERT

Association for Children and Adults With Learning Disabilities
February 25-28, 1987
San Antonio, Texas

National Association of School Psychologists
March 4-8, 1987
New Orleans, Louisiana

American Occupational Therapy Association
April 5-8, 1987
Indianapolis, Indiana

American Educational Research Association
April 20-24, 1987
Washington, DC

Council for Exceptional Children
April 20-24, 1987
Chicago, Illinois

World Conference on Gifted and Talented Children
August 4-6, 1987
Salt Lake City, Utah

REFERENCES

Airhart, D.L., & Cronin, D. (1981). Gardening ideas for physically handicapped individuals. *HortScience, 16*(3), 461.

Daubert, J.R., & Rothert, E.A. (1981). *Horticultural therapy for the mentally retarded.* Glencoe, IL: Chicago Horticultural Society.

Hudak, J.L., Jr., & Malloy, D. (1980). *The Melwood manual.* Menomonie: University of Wisconsin-Stout, Stout Rehabilitation Institute.

Olszowy, D.R. (1978). *Horticulture for the disabled and disadvantaged.* Springfield, IL: Charles C Thomas.

Popovich, D. (1981). *Effective educational and behavioral programming for severely and profoundly handicapped students.* Baltimore: Paul H. Brookes.

Relf, P.D. (1980). *The establishment of horticultural training centers for the mentally retarded.* Gaithersburg, MD: National Horticultural Therapy Association.

Relf, P.D. (1981, May-June). Dynamics of horticultural therapy. *Rehabilitation Literature, 42*(5-6).

Thoday, P.R. (1978). *Therapy through horticulture.* British Association for Advancement of Science Annual Meeting, Section M-Agriculture.

Wehman, P., Moon, M.S., & McCarthy, P. (1986). Transition from school to adulthood for youth with severe handicaps. *Focus on Exceptional Children, 18*(5), 1-12.

NEW BOOK

Related Services for Handicapped Children
by Morton M. Esterson and Linda F. Bluth

Here is a new book that capsulizes services related to special education, including the roles of audiology, counseling, early identification, medical services, occupational therapy, parent counseling, psychological services, recreation, school health services, social work services, speech pathology, and transportation. If the introductory material seems basic and simplistic, skip to the chapters on fields with which you are not so familiar. The chapters are each written by applicable specialists, with a practical orientation.

With the growing emphasis on multidisciplinary/transdisciplinary involvement, this handy little resource will be most helpful to members of an IEP team. Each chapter gives definitions, options for service delivery, and implications for IEP development. Information and references are up-to-date, including the AIDS issue, new technology in identification, interpretation of test results, therapeutic procedures, and the like. A glossary of terms and a list of suggested readings are appendix items.

This 158-page paperback is available from the publisher, Little, Brown, 34 Beacon St., Boston, MA 02108.