

Writing Assessment Perspective: Too Simplistic or too Sophisticated?

Majid Nemati

Assistant professor, University of Tebran

Masoumeh Ahmadi Shirazi

PhD in TEFL, University of Tebran

Abstract

Writing ability is perhaps one of the most challenging constructs to be defined. For writing assessment to be valid indicator of the ability in question, there is a need for a theoretical basis. The uncertainty about the theory underlying the writing construct has led to the development of rating scales, each of which focuses on a number of writing traits. This study is an attempt to provide a taxonomy of writing traits. The question this taxonomy brings into mind is: how can we limit the nature of such an all-encompassing construct? If purpose and practicality are among the justifying reasons for reducing the construct into a few features, then construct validity of writing assessment will remain under question.

Key words: writing ability, writing assessment, construct validity, rating scale, descriptors

1. Introduction

One of the issues challenging the community of language testers is the validity of writing assessment made according to scale descriptors developed either theoretically or empirically. Scale descriptors which are representative of performance level are guiding lines that help assessors to map these features onto the piece of writing they are to evaluate. Although there have been many attempts to present descriptors and rubrics suitable for the purpose of assessment, there is no universally agreed template for the assessment of writing. Classification of descriptors into a taxonomy can help, since it shows how extensive the sphere of writing assessment is while the current look is a reduced snapshot of the construct. The objective this study seeks to put forward is presenting the current descriptors as a taxonomy to serve two purposes: (1) to question the theoretical basis of these rating descriptors, and (2) to challenge the current notion of construct validity of writing assessment.

2. Review of Related Literature

The intricate nature of writing ability stands in the way of defining the construct. Searching for a conclusive definition seems to be futile for there is little agreement on what constitutes writing or what theories support these definitions.

White (1998) cites the controversies present in the literature about what writing is. He believes that attempts should be directed towards bringing graders to agreement or else grades are devoid of meaning and value. The problem may be rooted in the absence of a comprehensive theory of writing about which all can agree; this point is underscored by Johns (1990)

when she asserts that no single theory of ESL composition exists about which all are positive; the reason she presents to justify her proposition is the discrepancy of world views among researchers, theorists and teachers which gives rise to different theories underlying ESL composition.

How can we assess the construct of writing when the literature can provide neither the definition nor the theory for that construct? Faigley, Cherry, Jolliffe, and Skinner (1985, as cited in Huot, 1990) contend that “of necessity, practice has far outrun theory in writing assessment” (p. 257). This point was supported by Hulstijn (1985):

It should be obvious that syllabus writers, teachers, and testers cannot wait for full-fledged theories of language proficiency to emerge from research laboratories; in the absence of theories, they have to work with taxonomies which seem to make sense even if they cannot be fully supported by a theoretical description. (p. 277)

If this is true, then what has been neglected is the validity of the writing construct on the one hand, and its superficial subjective assessment on the other.) It is, after all, the theory on which all else rests; it is from there that the construct is set up and it is on the construct that validity, of the content and predictive kinds, is based (Davies, 1977: 63).

This idea is further supported by Kelly (1978) who comments that “the systematic development of tests requires some theory, even an informal inexplicit one, to guide the initial selection of item content and the division of the domain interest into appropriate sub-areas” (p. 8). Weir (1990) substantiates the need for a theory to define the construct

asserting that “the more fully we are able to describe the theoretical construct we are attempting to measure, at the a priori stage, the more meaningful might be the statistical procedures contributing to construct validation that can subsequently be applied to the result of the test” (p. 24).

In testing writing, from Weigle’s (2002) point of view:

Construct validity must be demonstrated in at least three ways: (1) the task must elicit the type of writing that we want to test; (2) the scoring criteria must take into account those components of writing that are included in the definition of the construct; and (3) the readers must actually adhere to those criteria when scoring writing samples. (p. 51)

It is worth clarifying the second concern stated by Weigle that the scoring criteria have to consider the components of writing as included in the definition of the construct. The question is: what are these components which the definition of the writing construct embraces?

Verbal descriptors are the bases of most rating scales. In fact, descriptors, according to Davies, Brown, Elder, Hill, Lumley, and McNamara (1999: 43), are statements which describe the level of performance required of candidates at each point on a proficiency scale. The descriptors therefore represent the writing components as they are specified in the construct definition. Descriptors are further specified by rubrics which consist of a number of criteria used to rate written works. These rubrics reflect the same components that are included as the bases of the writing construct. Rubrics are made by paying attention to the text features to which raters

attend. As is obvious, the writing samples are rated according to some features specified by descriptors and rubrics. These features enjoy different names: criteria (Turner & Upshur, 2002; Marby, 1999; Luft, 1998; Hart, 1994; Herman, Aschbacher, & Winters, 1992, Brindley, 1991), components (Weigle, 2002), aspects/features (Currier, 2005; Bachman & Palmer, 1996; Pollitt & Hutchinson, 1987), traits (Herman, Aschbacher & Winters, 1992) , qualities (Shaw, 2004), skills and/or sub-skills (White, 1998), categories and/or subcategories (Matthews, 1990), and domains (Engelhard, 1996; Gyagenda & Engelhard, 1998); if they comprise the construct of writing and if they measure the same construct, then one of the above-mentioned names can be sufficient to explicate the writing construct. What is more, different rating scales have their own interpretation of these features. The result, therefore, would be an extensive list of features which do not necessarily assess a different aspect of writing ability. In fact, rating scales limit the skills or sub-skills whose existence in writing samples would be reflector of language ability. White (1998) underscores that there is no agreement on which skills or sub-skills exist in writing. The current study aims: (1) to gather a set of writing features stipulated by literature, (2) to establish a taxonomy of these features, (3) to show the vast picture of writing that has been unfairly abridged, and (4) to encourage another look at the extensive nature of the writing construct that has been reduced to a few criteria in the available rating scales.

3. Collection of Writing Features

The first step in gathering writing features was to find the rating scale descriptors available in the literature. It seems

necessary to state that the versions of the descriptors used in this study are the ones provided by testing organizations for public use. The search began by finding major internationally well known writing tests. The researchers made the most of the available resources, however, the effort was sometimes in vain since not even the public test manual was at hand.

These writing features are not just confined to those specified by the manuals of writing tests, but the researchers gathered almost all the statements of the scholars that can be applied to evaluate a writing sample. In doing so, approximately hundred sources were taken into account. Tables 1 and 2 show these references by test types and authors respectively.

Table 1: Writing Features Specified by Tests

American College Test (ACT)	International English Language Testing System Task 1 Writing Band Descriptors
American Council on the Teaching of Foreign Languages (ACTFL)	International English Language Testing System Task 2 Writing Band Descriptors (IELTS)
Advanced Placement (AP)	Illinois State Board of Education (ISBE)
The Academic Profile (A. Profile)	Key English Test (KET)
Collegiate Assessment of Academic Proficiency (CAAP)	LanguEdge Courseware
Certificate in Advanced English (CAE)	Measure of Academic Proficiency and Progress (MAPP)
Canadian Academic English Language Assessment (CAEL)	Medical College Admission Test (MCAT)
Center for Applied Language Studies (CALs)	Georgia Department of Education Middle Grades Writing Assessment
Common European Framework Writing Scale Descriptors (CEFR)	Michigan Writing Assessment Scoring Guide (MELAB)
Common Educational Proficiency Assessment (CEPA)	National Assessment of Educational Progress (NAEP)
College-Level Academic Skills Test Essay (CLAST)	New Jersey College Basic Skills Placement Test (NJCBSPT)
College-Level Examination Program (CLEP)	Northwest Regional Educational Laboratory: Six Trait Model (NWREL)

College Basic Academic Subjects Examination (College BASE)	Performance Assessment for Language Students: Level 1
College Outcome Measures Program (COMP)	Performance Assessment for Language students: Upper Level
Computerized Adaptive Placement Assessment and Support System (COMPASS)	Preliminary English Test: Part 2 Mark Scheme (PET)
Contextualized Writing Assessment (CoWa)	Preliminary English Test: Part 3 Mark Scheme (PET)
Certificate of Proficiency in English (CPE)	Pre-Professional Skills Test (Praxis I)
College of Education and Human Resources: Utah University Writing Examination	Scholastic Aptitude Test (SAT)
English as a Second Language Placement Examination (ESLPE)	Scale for Evaluating Expository Writing (SEEW)
Florida Writing Assessment Program	Skills for Life (SfL)
First Certificate in English (FCE)	Southeast Missouri State University Writing Proficiency Exam
Graduate Management Test (GMAT)	Special Test of English Proficiency (STEP)
General Record Exam (GRE)	Texas Assessment of Knowledge and Skill (TAKS)
Illinois Inventory of Educational Progress (IIEP)	Texas Academic Skills Program (TASP)
Interagency Language Roundtable (ILR)	TestDaf Scoring Criteria
Internet-Based TOEFL (TOEFL iBT)	Testing English for Educational Purposes (TEEP)
International English Language Testing System Handbook 2005 (IELTS)	Test of Written English (TWE)
International English Language Testing System: General Training Writing Skills	Wissahickon High School World Language Department

Table 2: Writing Features Specified by Authors

L. G. Alexander (1965)	D. P. Harris
L. G. Alexander (1967)	J. B. Heaton
H. D. Brown (2000)	J. L. Herman, M. Gearhart, & E. Baker
H. D. Brown (2004)	J. Hedgcock & N. Lefkowitz
J. D. Brown & K. M. Bailey	A. Hughes
E. P. Bailey & P. A. Powell	K. Hyland
M. Celce-Murcia	M. Inaam & S. Grant
F. Chaplen	S. L. Issacson
K. Chastain	H. L. Jacobs, S. A. Zinkgraf, D. R. Wormuth, V. F. Hartfiel, & J. B. Hughey

A. D. Cohen	B. Kroll
A. M. Cohen	H. S. Madsen
A. D. Cohen & M. C. Cavalcanti	W. E. Messenger & P. A. Taylor
U. Connor & M. Farmer	J. Munby
A. Cumming, R. Kantor, & D. E. Powers	C. Polio (1997)
A. Cumming, R. Kantor, D. E. Powers, T. Santos, & C. Taylor	C. Polio (2001)
P. Davidson & D. Lloyd	W. M. Rivers
P. B. Diederich	D. Soles
A. K. Fathman & E. Whalley	J. M. Swales & C. B. Feak
D. Ferris & J. S. Hedgcock	C. Vaughan
K. Glasswell, J. Parr, & M. Aickman	M. Wesche
B. Q. Gray & V. B. Slaughter	E. M. White
C. Gregory	H. Yukio
L. Hamp-Lyons & G. Henning	

The second step was to establish a taxonomy of features found. To do so, the researchers read the descriptors to identify the features in focus. Then these features were separated with regard to their negative or positive effect on the scoring of writing samples. For example, the test of First Certificate in English (FCE) was thoroughly reviewed for distinguishing features. In fact, for any obtained test rubric, there exists a double in which features have been specified and classified in terms of their positive or negative effect upon scoring results. Tables 3 and 4 indicate these two versions of test rubrics. As can be seen in Table 4, those features which improve the score have been named Gains while those leading to lower marks have been named Losses. Gains and Losses have been extracted from the original rubrics, but for the purpose of this study they have been separated to make the process of setting the taxonomy easier. Looking at rubrics analytically, there are nine important features within the test of FCE: (1) realization of the task set, (2) the kind of effect on the target reader, (3) range of vocabulary and structure, (4) language control, (5)

organization and the role of linking devices in setting up a well-organized essay, (6) appropriacy in register and format, (7) length of the essay in relation to standards, (8) relevance of the content to the task set, and finally (9) legibility of the piece of writing. These analyses were performed for the 56 test rubrics shown in Table 1. The other features which the taxonomy considers came from the statements of the 45 scholars cited in Table 2.

Table 3: Test of First Certificate in English (Original)

5	<p>Full realization of the task set</p> <ul style="list-style-type: none"> • All content points included with appropriate expansion • Wide range of structure and vocabulary within the task set • Minimal errors, perhaps due to ambition; well-developed control of language • Ideas effectively organized, with a variety of linking devices • Register and format consistently appropriate to purpose and audience <p>Fully achieves the desired effect on the target reader.</p>
4	<p>Good realization of the task set</p> <ul style="list-style-type: none"> • All major content points included; possibly one or two minor omissions • Good range of structure and vocabulary within the task set • Generally accurate, errors occur mainly when attempting more complex language • Ideas clearly organized, with suitable linking devices • Register and format on the whole appropriate to purpose and audience <p>Achieves the desired effect on the target reader.</p>
3	<p>Reasonable achievement of the task set</p> <ul style="list-style-type: none"> • All major content points included; some minor omissions • Adequate range of structure and vocabulary, which fulfills the requirement of the task • A number of errors may be present, but they do not impede communication • Ideas adequately organized, with simple linking devices • Reasonable, if not always successful attempt at register and format appropriate to purpose and audience

	Achieves, on the whole, the desired effect on the target reader.
2	<p>Task set attempted but not adequately achieved</p> <ul style="list-style-type: none"> • Some major content points inadequately covered or omitted, and/or some irrelevant material • Limited range of structure and vocabulary • A number of errors, which distract the reader and may obscure communication at times • Ideas inadequately organized; linking devices rarely used • Unsuccessful/inconsistent attempts at appropriate register and format <p>Message not clearly communicated to the target reader.</p>
1	<p>Poor attempt at the task set</p> <ul style="list-style-type: none"> • Notable content omissions and/or considerable irrelevance, possibly due to misinterpretation of task set • Narrow range of vocabulary and structure • Frequent errors which obscure communication; little evidence of language control • Lack of organization, or linking devices • Little or no awareness of appropriate register and format <p>Very negative effect on the target reader.</p>
0	Achieves nothing; too little language for assessment (fewer than 50 words) or totally irrelevant or totally illegible.

Table 4: Test of First Certificate in English (Classified)

Gains	Losses
Full coverage of the task set	Poor coverage of the task set
Wide range of vocabulary and structure within the task set	Limited range of structures and vocabulary; penalty for "lifting" of input
Well-developed control of language	Little or no evidence of language control
Effective organization with a variety of linking devices	Lack of organization or linking devices
Consistent appropriacy in register and format	Little or no awareness of appropriate register and format

Length of the essay in pace with standard (120-180 words)	A too short (fewer than 50 words) essay
Positive effect on the target reader	Very negative effect on the target reader
Relevance of content to the task set	Totally irrelevant content
Legible	Illegible

When all these features were specified, they were put into different macro categories to form a taxonomy of writing features. Some of these features could be classified under the category *content*; others are associated more with *organization*; *grammatical*, *lexical*, and *conventional* features comprised the next macro categories in the taxonomy. Some features were called *miscellaneous* since they bore either little or no relation to the main categories or they could be classified under almost all the macro categories. The references, by both the 56 testing manuals and 45 authors, have been coded to facilitate finding the relevant source of features.

After reducing the names to codes, the basis of taxonomy was set up. The number of writing features came to 288. It should be emphasized that some features are common among macro categories. This came as the ancillary result of classification showing that the macro categories (e.g., content) can hardly present a reasonable justification to embrace the feature/s.

Discussion

This taxonomy has been designed to zoom in on the picture of the construct whose credibility lies within its distinguishing features. These features, as was previously mentioned, bear confusingly various names: criteria, components, aspects, traits, qualities, skills and/or sub-skills, categories and/or sub-

categories, and domains. To the extent that variability exists to name these features, there is as much variation in their classification. There is no hard and fast way of putting one micro feature within a macro one, for example, the presence or absence of topic sentence or thesis statement can be put into both content and organization. Both seem plausible since, on the one hand, topic sentence shows that the writer has fully grasped the main idea proposed by the prompt and, on the other hand, without the leading topic sentence, the supporting sentences in the paragraphs go astray since they do not bear relevance to the main idea expressed by the topic sentence. This fact taps another shortcoming of this classification.

As rubrics of different tests were delved into for this study, several facts came out. First of all, some of the descriptors were in the form of sentences without differentiating the macro or micro features. For example ACT scoring guidelines have a scale ranging from 1 to 6. If we look at just the band 6, we will see statements which, on the whole, do not give labels to the features cited within the lines:

Essays within this score range demonstrate effective skill in responding to the task

<p>The essay shows a clear understanding of the task. The essay takes a position on the issue and may offer a critical context for discussion. The essay addresses complexity by examining different perspectives on the issue, or by evaluating the implications and/or complications of the issue, or by fully responding to counterarguments to the writer's position. Development of ideas is ample, specific, and logical. Most ideas are fully elaborated. A clear focus on the specific issue in the prompt is maintained. The organization of the essay is clear: the organization may be somewhat predictable or it may grow from the writer's purpose. Ideas are logically sequenced. Most transitions reflect the writer's logic and are usually integrated into the essay. The introduction and conclusion are effective, clear, and well developed. The essay shows a good command of language. Sentences are varied and word choice is varied and precise. There are few, if any, errors to distract the reader.</p>
--

As is clear, the first few lines deal with fulfilling the task set; then organization comes into focus followed by sentence structure and vocabulary choice. The question that is formed here is: why not having separate macro and micro features? The reason can be due to the second issue we shortly switch to.

The next point is that some descriptors are given different labels while they are similar in that they are all trying to provide some means to measure the construct of writing, but why do they not have the same labels? Here we are to examine a few labels to show the inconsistency in terminologies used. Brown and Bailey (1984) give the following labels to the macro and micro features:

- I. Organization (Introduction, body and Conclusion)
- II. Logical development of ideas (Content)
- III. Grammar (Correct use of Relative clauses, prepositions, modals, articles, verb forms, and tense sequencing and no fragments or run-on sentences)
- IV. Punctuation (Spelling and mechanics)
- V. Style and quality of expression (Precise vocabulary usage, use of parallel structures, concise, register good)

Jacobs, Zinkgraf, Wormuth, Hartfiel, and Hughey (1981) propose the following labels in their ESL Composition Profile: content, organization, vocabulary, language use, and mechanics. It is worth considering these five macro features of ESL Composition Profile.

Content

- | | |
|-------|---|
| 30-27 | Excellent to Very Good: knowledgeable; substantive; thorough development of thesis; relevant assigned topic |
| 26-22 | Good to Average: some knowledge of subject; adequate range; limited development of thesis; mostly relevant to topic, but lacks detail |

40 Writing Assessment Perspective

- 21-17 Fair to Poor: limited knowledge of subject; little substance; inadequate development of topic
- 16-13 Very Poor: does not show knowledge of subject; non-substantive; not pertinent; or not enough to evaluate

Organization

- 20-18 Excellent to Very Good: fluent expression; ideas clearly stated/supported; succinct; well-organized; logical sequencing; cohesive
- 17-14 Good to Average: somewhat choppy; loosely organized but main ideas stand out; limited support; logical but incomplete sequencing
- 13-10 Fair to Poor: non-fluent; ideas confused or disconnected; lacks logical sequencing and development
- 9-7 Very Poor: does not communicate; no organization; or not enough to evaluate Vocabulary
- 20-18 Excellent to Very Good: sophisticated range; effective word/idiom choice and usage; word form mastery; appropriate register
- 17-14 Good to Average: adequate range; occasional errors of word/idiom form, choice, usage *but meaning not obscured*
- 13-10 Fair to Poor: limited range; frequent errors of word/idiom form, choice, usage; *meaning confused or obscured*
- 9-7 Very Poor: essentially translation; little knowledge of English vocabulary, idioms, word form; Or not enough to evaluate

Language Use

- 25-22 Excellent to Very Good: effective complex constructions; few errors of agreement, tense, number, word order/function, articles, pronouns, prepositions
- 21-18 Good to Average: effective but simple constructions; minor problems in complex constructions; several errors of agreement, tense, number, word order/function, articles, pronouns, prepositions *but meaning seldom obscured*
- 17-11 Fair to Poor: major problems in simple/complex constructions; frequent errors of negation, agreement, tense, number, word order/function; articles, pronouns, prepositions and/or fragments, run-ons, deletions; *meaning confused or obscured*

10-5 Very Poor: virtually no mastery of sentence construction rules; dominated by errors; does not communicate; Or not enough to evaluate

Mechanics

- 5 Excellent to Very Good: demonstrates mastery of conventions; few errors of spelling, punctuation, capitalization, paragraphing
- 4 Good to Average: occasional errors of spelling, punctuation, capitalization, paragraphing but *meaning not obscured*
- 3 Fair to Poor: frequent errors of spelling, punctuation, capitalization, paragraphing; poor hand writing; *meaning confused or obscured*
- 2 Very Poor: no mastery of conventions; dominated by errors of spelling, punctuation, capitalization, paragraphing; handwriting illegible; Or not enough to evaluate

IELTS considers the following features: task fulfillment, coherence and cohesion, lexical resources, and grammatical range and accuracy. What follows is the top band descriptors in IELTS (public version):

Band 9 Task achievement: fully satisfies all the requirements of the task; clearly presents a fully developed response
 Coherence and cohesion: uses cohesion in such a way that it attracts no attention; skillfully manages paragraphing
 Lexical resource: uses a wide range of vocabulary with very natural and sophisticated control of lexical features; rare minor errors occur only as slips
 Grammatical range and accuracy: uses a wide range of structures with full flexibility and accuracy, rare minor errors occur only as slips

Michigan Writing Assessment Scoring Guide takes into account three features: ideas and arguments, rhetorical features, and language control. Here the details of these three features within scale 6 are presented.

Ideas and Arguments: The essay deals with the issues centrally and fully. The position is clear, and strongly and substantially argued. The complexity

of the issues is treated seriously and the viewpoints of other people are taken into account very well.

Rhetorical features: The essay has rhetorical control at the highest level, showing unity and subtle management. Ideas are balanced with support and the whole essay show strong control of organization appropriate to the content. Textual elements are well connected through logical or linguistic transitions and there is no repetition or redundancy.

Language control: The essay has excellent language control with elegance of diction and style. Grammatical structures and vocabulary are well chosen to express the ideas and to carry out the intentions.

Skills for Life (SfL) takes the following features into consideration: content and task realization, audience awareness and impact on reader, organization and coherence, sentence structure, word order, punctuation, capitalization, grammatical range and accuracy, spelling, and handwriting. What Level 2 of SfL credits are as follows:

Content and task organization: communicate information, ideas and opinions, clearly and effectively using length, format and style appropriate to purpose, context and audience in a wide range of documents, ability to convey message effectively and include appropriate information

Audience awareness and impact on reader: ability to adapt text to intended audience in terms of register, layout and/or other text features and engage reader's interest by choice of content matter, vocabulary and/or style features

Organization and coherence: ability to organize text coherently and use text features such as opening and closing formulae and appropriate scene-setting or overview statement

Sentence structure: ability to construct compound sentences and use linking devices and some complex structures

Word order: ability to control word order in statements, negatives, and questions in a variety of verb tenses

Punctuation: sentences are punctuated correctly so that meaning is clear

Capitalization: ability to use capital letters correctly

Grammatical range and accuracy: controlled, natural use of language. Correct and consistent use of tense. Complex language is attempted with some success

Spelling: ability to spell correctly words used in work, studies and daily life, and technical words

Handwriting: is always legible

Special Test of English Proficiency (STEP) focuses on: task fulfillment and appropriacy, conventions of presentation, cohesion and organization, grammatical control. Band 5 of STEP suggests the following features:

Task fulfillment and appropriacy: text relates well to given context. It is thoroughly appropriate and easily understood. Vocabulary choices are appropriate and effective

Conventions of presentation: all aspects of presentation conventions (spelling, punctuation, script or layout) are handled skillfully

Cohesion and organization: text is cohesive and organization is clear and appropriate to task

Grammatical control: competent control of grammatical structures appropriate to the context with only unobtrusive errors

Test of English for Educational Purposes (TEEP), proposed by Weir (1990), is concerned with: relevance and adequacy of content, compositional organization, cohesion, adequacy of vocabulary for purpose, grammar, mechanical accuracy I (punctuation), mechanical accuracy II (spelling). One level of these features indicates:

- A. Relevance and adequacy of content
 - 3. Relevant and adequate answer to the task set
- B. Compositional organization
 - 3. Overall shape and internal pattern clear. Organizational skills adequately controlled
- C. Cohesion
 - 3. Satisfactory use of cohesion resulting in effective communication
- D. Adequacy of vocabulary for purpose
 - 1. Frequent inadequacies in vocabulary for the task. Perhaps frequent lexical inappropriacies and/or repetition
- E. Grammar

- 3. Almost no grammatical inaccuracies
- F. Mechanical accuracy I (punctuation)
- 3. Almost no inaccuracies in punctuation
- G. Mechanical accuracy II (spelling)
- 3. Almost no inaccuracies in spelling

Regarding the above-mentioned descriptors, the term *content* has different meanings in these seven writing rubrics. Brown and Bailey (1984) define content as logical development of ideas. This interpretation of content is not shared and supported by other descriptors provided here. Jacobs et al. (1981) think of content as the amount of information that fully and substantively develop the topic and is relevant to the topic assigned. IELTS does not have a macro feature named content; perhaps task fulfillment can be taken as a similar concept. It cannot be told from public version of IELTS band descriptors which micro features exist within task fulfillment; therefore, further comment seems inappropriate and may be unfair. Ideas and arguments in Michigan Writing Assessment Scoring Guide probably denote the content. Skills for Life (SfL) has a feature called content and task organization the basis of which is the ability of the writer to communicate the message with adequate amount of information which is in tune with context, audience, and purpose of writing. STEP also deals with the issue of content, but puts it another way; content is called task fulfillment and appropriacy. However the difference between STEP and other rubrics lies in the fact that in this rubric vocabulary is not a macro feature by itself but a micro feature within task fulfillment and appropriacy. The first macro feature explicated by TEEP is the relevance and adequacy of content. Obviously, the underlying concept

underscored by all is the same however there is no consensus on the name given to this macro feature.

The same problem applies to *organization* as another main feature influencing the assessment of writing samples. Brown and Bailey (1984) confine it to three concepts: *introduction, body and conclusion*. Jacobs et al. (1981) refer to ideas being stated clearly and expressed fluently following a logical sequence. Perhaps what Brown and Bailey meant by content is defined as organization by Jacobs et al. Interestingly IELTS does not use the term organization, instead it goes for cohesion and coherence; it says that in top band the writer is able to *use* cohesion in such a way that attracts no attention; at the same time the writer is able to skillfully manage paragraphing. Two questions can be raised here:

1. What is cohesion? Can we *use* cohesion or cohesion is a concept the realization of which depends on using cohesive links?
2. If the macro feature is made up of two parts (i.e., cohesion as well as coherence), then should some allowance be made for describing them?

Perhaps the non-public version of band descriptors of IELTS can be the answer to these questions; however the public version cannot help us know what these concepts mean and how they can affect scores.

Michigan Writing Assessment Scoring Guide opts for rhetorical features instead of using organization as a macro feature. Rhetorical features are sought in the unity of the essay as a whole, well supported ideas, the use of logical and linguistic transitions to well connect the elements of the text, and no repetition and/or redundancy. It would sound rational to

have the reason for substituting organization with rhetorical features. Are they similar or different?

One of the macro features specified by SfL is organization and coherence. If we look at the main features stipulated by STEP, we come across cohesion and organization. The story continues into what features in TEEP have the same say in writing assessment: compositional organization and cohesion as two separate main features. As can be seen, SfL differentiates between organization and coherence, STEP separates organization from cohesion while both consider these two as a single main feature, and TEEP puts cohesion and compositional organization into two different main features. If we support SfL, then what main feature in STEP handles coherence and if we agree with the categorization of STEP and SfL which regard cohesion and coherence as inextricable linked to the concept of organization, why does TEEP make them distinct? In fact, we should ask experts why this much variation exists while one term can be representative of all specifications hidden in the concept. De Jong (1990 as cited in North & Schneider, 1998) suggests that “the acceptability of these levels, grades, and frameworks seems to rely primarily on the authority of the scholars involved in their definition, or on the political status of the bodies that control and promote them” (p. 221).

Other main features are not as challenging as content and organization are. Grammar, vocabulary, and mechanics enjoy more or less the same terminologies in different rubrics. However, due to shortage of space, discussion of them will be pursued another time.

Towards a new approach towards construct validity

Given the complex nature of writing as is illustrated in the categorization of writing features of this study, we are faced with the following problems:

1. To find an appropriate name for the different features of writing skill and pertinent categories
2. To categorize these writing features
3. To determine the priority of these features that is, which one is deemed more important when it comes to assessment
4. To group similar features under one class or category
5. To subcategorize features which can be grouped under two or more categories
6. To determine on their inclusion in the grading of written essays
7. To decide on their effect on the construct validity of writing assessment

Perhaps it seems difficult to find solutions to these problems. There is no hard and fast solution for the above-mentioned issues which would affect the construct validity of writing assessment. However, there are some suggestions which may pave the way for enhancing the question of construct validity.

To solve the issue of naming categories, there is the possibility of finding the frequency of occurrence of these categories in the table which this study suggests. Then, to categorize features, we can go through qualitative analysis, for example, to get help from the classifications which present the writing features as the one depicted in this study, we can sort out these features as to their similarity and differences grouping them together regarding their importance in writing assessment.

If we would like the categorization to enjoy level of difficulty, that is, the priority of one category and its related subcategories over the rest, then there is a need to do research to find out the priority of these assessment features over the others. The present study is the basis for another research which would suggest that we can limit the number of features to one quarter of the existing one, for instance, of 288 features, we can choose 68 highly frequent microcategories which belong to five most frequent macrocategories. If we assess the essays with regard to the existence of these categories, then item response theory especially Rasch model of analysis can inform us of the priority of one category over another. With replicating this type of research, we may be able to categorize these features as to their priority, therefore, those with higher status deserve a higher score than the one appearing less difficult and accordingly less important. Furthermore, item response theory provides a rationale for the construct validity of the features which make up the construct of writing. Through Rasch model, we will be able to put forth the unidimensionality of the suggested items or features, then we may claim that the features rightly measure the same construct, that is, writing ability.

Conclusion

Writing assessment perspective creates a paradox where the construct to be assessed bears a sophisticated nature whereas when it comes to its assessment, the perspective is simplistic enough to bring the validity of judgments into question.

All issues pertinent to writing assessment can be value-laden after the construct can be defined. Those binary terminologies used for teaching and assessing writing (form/content, accuracy/fluency, product/process) cannot stand

the complex nature of writing construct. Even different views on writing assessment namely quantitative versus qualitative, direct and/or indirect, holistic or analytic would take a part to play when they are first oriented by the definition presented for the construct in question. To this date, assessment of writing has been confined to traits or features that are either deemed important by teachers or substantiated by scholars intuitively. This can be vividly observed in the taxonomy this study provides. Just as the number of these features changes under different circumstances, their names and importance are also prone to variation.

Now if we suppose that the construct can be defined and there is a universally-agreed theory to support such a definition, then how can we solve the problem of terminologies used in this definition? Can content and organization become unified? How different are these two terminologies: content and task fulfillment? Can cohesion and coherence be classified under organization or are they part and parcel of it?

If we suppose that the perspective is not simplistic but simple, again no headway has been made since it can be simplistic if our sophisticated construct is reduced to several features in accordance with the purpose of the test or practicality of its assessment. If this is true, then it is not a robust theoretical basis to support the validity of the writing assessment, but it is the context of situation which renders value to judgments made about writing ability of individual learners. Strangely enough, there is agreement among scholars that the construct has not been attended to as it should, but no change happens to their criteria for assessment. As a way of conclusion, a quotation by Cumming, Kantor, Powers, Santos

and Taylor (2000) will put an end to our discussion; they say that “although educators around the world regularly work with implicit understandings of what constitutes effective English writing, no existing research or testing programs have proposed or verified a specific model of this, such as would be universally accepted” (p. 27).

References

- Aaron, S., Bullock, G., Doctrie, G., Lewis, M., Mickish, V. L., Mills, P., Scamihorn, D., Smith, C., and Young, D. (2002). Assessment and instructional guide for the Georgia Middle Grades Writing Assessment. (ERIC Document Reproduction Service No. ED471626).
- ACT (2006). Writing test: Scoring guidelines. Retrieved March 25, 2006, from <http://www.actstudent.org/writing/scores/guidelines.html>
- ACTFL (2001). ACTFL Proficiency Guidelines: Writing. Retrieved March 25, 2006, from <http://www.actfl.org/files/public/writingguidelines.pdf>
- Alexander, L. G. (1965). *Essay and letter writing*. Longmans.
- Alexander, L. G. (1967). *Fluency in English*. Longman.
- American College Testing. (1976). College Outcome Measures Program (COMP). ACT: Iowa City, IA.
- American College Testing Program (1991). CAAP Technical Handbook. Iowa City, IA: American College Testing Program.
- American College Testing Program (1992). *Clarifying and Assessing General Education Outcomes of College, College Outcome Measures Program: Technical Report 1982- 1991*. Iowa City, IA:

- American College Testing (n.d.). COMPASS writing skills placement test. Iowa City, IA: American College Testing Program.
- Advanced Placement Development Committee (n.d.). AP: English language and composition. Princeton, NJ: The College Board.
- Association of American Medical Colleges. (1985). Medical College Admission Test: MCAT Essay. NW Washington, DC: Association of American Medical Colleges.
- Bachman, L. F., & Palmer, A. S. (1996). Language testing in practice. Oxford: Oxford University Press.
- Bailey, E. P., & Powell, P. A. (1989). *The practical writer with reading*. 2nd (ed.). Holt, Rinehart and Winston.
- Brindley, G. (1991). Defining language ability: The criteria for criteria. Unpublished manuscript.
- Brown, H. D. (2000). *Principles of language learning and teaching*. 4th (Ed.). Longman.
- Brown, H. D. (2004). *Language assessment principles and classroom practices*. Longman.
- Brown, J. D., & Bailey, K. M. (1984). A categorical instrument for scoring second language writing skills. *Language Learning*, 34, 21-42.
- CAE (2007). General impression mark scheme. Retrieved March 30, 2006, from http://www.cambridgeesol.org/teach/cae/writing/aboutthepaper/assessment/general_imp_mark_scheme.htm
- CAEL (2002). Writing performance band score criteria. Retrieved March 30, 2006, from <http://www.cael.ca/pdf/C4.pdf>

- CARLA (2001). Center for Advanced Research on Language Acquisition. *CoWa (Contextualized Writing Assessment)*. Retrieved June 25, 2008, from <http://www.carla.umn.edu/assessment/MLPA/CoWA.html>
- CEFR (1993). Writing scale descriptors. Retrieved March 30, 2006, from <http://www.coe.int/T/DG4/Portfolio/documents/Manual>
- Center for Applied Language Studies (n.d.). CALS suggested marking scheme (writing skills). CALS, University of Reading.
- Celce-Murcia, M. (2001). *Teaching English as a second or foreign language*. 3rd (Ed).. Heinle & Heinle.
- CEPA (2006). Writing descriptors. UAE: National Admissions and Placement Office.
- Chaplen, F. (1970). *Paragraph writing*. Oxford University Press.
- Chastain, K. (1988). *Developing second language skills: Theory and practice*. (3rd Ed.), Harcourt Brace Jovanovich.
- CLAST Assessment Resources (n.d.). Canadian Association of Second Language Teachers: Second language acquisition rubrics. Retrieved September 13, 2007 from <http://www.medievia.com/whsfl/writing.html>
- CLEP Test Development Committee (1993). CLEP General Exam in English Composition. Princeton, NJ: The College Board.
- Cohen, A. D. (1994). *Assessing language ability in the classroom*. Boston, Massachusetts: Heinle and Heinle Publishers.

- Cohen, A. M. (1973). Assessing college students' ability to write compositions. *Research in the Teaching of English* 7, 356-371.
- Cohen, A. D., & Cavalcanti, M. C. (1990). Feedback on compositions: Teacher and student verbal reports. In B. Kroll (Ed.) *Second language writing: Research insights for the classroom* (pp.155- 177).
- College Board (2002). SAT: How the essay is scored? Retrieved December 8, 2007, from http://www.collegeboard.com/student/testing/sat/about/sat/essay_scoring.html
- College of Education and Human Resources Writing Examination (2008). Definitions of assessment criteria. Utah State University. Retrieved June 24, 2008, from http://www.coe.usu.edu/deansoffice/tel/writing_exam.php
- Connor, U., & Farmer, M. (1990). The teaching of topical structure analysis as a revision strategy for ESL writers. In B. Kroll (Ed.) *Second language writing: Research insights for the classroom* (pp. 126-139).
- CPE (2007). General mark scheme. Retrieved March 30, 2006 from <http://www.cambridgeesol.org/teach/cpe/writing/data/Pg30%20%20cpe%20hb.pdf>
- Cumming, A., Kantor, R., & Powers, D. E. (2001). Scoring TOEFL essays and TOEFL 2000 prototype written tasks: An investigation into raters' decision making and development of a preliminary analytic framework. *TOEFL Monograph Series* 22. Princeton, NJ: Educational Testing Service.

- Cumming, A., Kantor, R., Powers, D., Santos, T., & Taylor, C. (2000). TOEFL 2000 writing framework: A working paper. *TOEFL Monograph Series 18*. Princeton, NJ: Educational Testing Service.
- Currier, W. M. (2005, March). Suggestions for evaluating ESL writing holistically. *The Internet TESL Journal*, XI (3). Retrieved December 24, 2006, from <http://iteslj.org/Techniques/Currier-EvaluatingWriting.html>
- Davidson, P., & Lloyd, D. (2005). Guidelines for developing a writing test. In D. Lloyd, P. Davidson, and C. Coombe (Eds.) *Fundamentals of language assessment: A practical guide for teachers in the Gulf* (pp. 87-104). Dubai: TESOL Arabia Publications.
- Davies, A. (1977). The construction of language tests. In J. P. B. Allen, and A. Davies (Eds.), *Testing and Experimental Method* (pp. 38-104). London: Oxford University Press.
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T and McNamara, T. (1999). *Dictionary of language testing*. Cambridge: Cambridge University Press.
- Diederich, P. B. (1974). *Measuring growth in English*. Urbana, Ill.: National Council of Teachers of English.
- Educational Testing Service (1986). *TOEFL Test of Written English Guide*. Princeton, NJ: Educational Testing Service.
- Educational Testing Service. (1990). *The Academic Profile: A user's guide, part 1*. Princeton, NJ: Educational Testing Service.
- Educational Testing Service. (2002). *Language courseware: Handbook for scoring speaking and writing*. Princeton, NJ: Educational Testing Service.

- Educational Testing Service (2003). *GRE scoring guides*. Princeton, NJ: Educational Testing Service.
- Educational Testing Service (2004). *iBT/Next Generation TOEFL Test: Integrated writing rubrics*. Princeton, NJ: Educational Testing Service.
- Educational Testing Service (n.d.). *Graduate Management Admissions Test (GMAT)*. Princeton, NJ: Educational Testing Service.
- Educational Testing Service (n.d.). *Measure of Academic Proficiency and Progress (MAPP)*. Princeton, NJ: Educational Testing Service.
- Educational Testing Service (n.d.). *SAT: Scholastic reasoning test writing section*. Princeton, NJ: Educational Testing Service.
- Engelhard, G. (1996). Evaluating rater accuracy in performance assessments. *Journal of Educational Measurement*, 33 (1), 56-70.
- ESLPE (2000). ESLPE Composition Rating Scale. LA: UCLA.
- Fathman, A. K., & Whalley, E. (1990). Teacher response to student writing: Focus on form versus content. In B. Kroll (Ed.) *Second language writing: Research insights for the classroom* (pp. 178-190). New York: Cambridge University Press.
- FCE (2007). General mark scheme. Retrieved August 7, 2007, from <http://www.cambridgeesol.org/teach/fce/writing/data/generalms.pdf>
- Ferris, D., & Hedgcock, J. S. (1998). *Teaching ESL composition: Purposes, process, and practice*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

- FCPS (1999). Fairfax County Public Schools Performance Assessment for Language Students (PALS): Holistic rubric for Upper Level writing task. Retrieved September 13, 2007 from http://www.fcps.edu/u/DIS/OHSICS/forlang/PALS/rubrics/uwrt_hol.htm
- FCPS (1999). Performance Assessment for Language Students (PALS): Analytic rubric for Level 1 writing task. Retrieved September 13, 2007 from http://www.fcps.edu/DIS/OHSICS/forlang/PALS/rubrics/1wrt_an.htm
- Florida State Dept. of Education. (1994). *The College-Level Academic Skills Test, Technical Report 1993-1994*. (ERIC Document Reproduction Service No. ED381570.) Tallahassee, FL.
- Florida Writing Assessment Program (2005). Score points in rubric. Retrieved June 25, 2008, from <http://www.fldoe.org/asp/fw/fwaprubr.asp>
- Georgia Department of Education (n.d.). New grade 8 writing assessment rubric (MGWA). Retrieved June 25, 2008, from <http://www.gadoe.org/DMGetDocument.aspx/Grade%208%20Writing%20Rubrics.pdf?p=4BE1EECF99CD364EA5554055463F1FBBF5D074D5FB1F2CAEB3B63B3ECB220CDD26C2114F3C57D8D249B8287AD9AE7FA6&Type=D>
- Glasswell, K., Parr, J., & Aikman, M. (2001). *Development of the asTTle writing assessment rubrics for scoring extended writing tasks* (Tech. Rep. No. 6). University of Auckland.

- Gray, B. Q., & Slaughter, V. B. (1980). Writing. In A. S. Trillin and Associates (Eds.), *Teaching basic skills in college, 12-90*. San Francisco: Jossey-Bass Publishers.
- Gregory, C. (n.d.). The tools you need to assess writing. Retrieved April 3, 2006, from <http://content.scholastic.com/browse/article.jsp?id=4019>
- Gyagenda, I.S., & Engelhard, G. Jr. (1998). Rater, domain, and gender influences on the assessed quality of student writing using weighted and unweighted scoring. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA. (ERIC Document Reproduction Service No. ED422349).
- Hamp-Lyons, L., & Henning, G. (1991). Communicative writing profiles: An investigation of the transferability of a multiple trait scoring instrument across ESL writing assessment contexts. *Language Learning*, 41 (3), 337-373.
- Harris, D. P. (1968). *Testing English as a second language*. New York: McGraw Hill.
- Hart, D. (1994). *Authentic assessment: A handbook for educators*. Menlo Park: Addison-Wesley Publishing Company.
- Heaton, J. B. (1975). *Writing English language tests*. London: Longman Group, Ltd.
- Heaton, J. B. (1988). *Writing English language tests*. Longman.
- Hecht, L. W. (1980). Validation of the New Jersey College Basic Skills Placement Test (NJCBSPT). Technical Report. (ERIC Document Reproduction Service No. ED214945.) Princeton, NJ: Educational Testing Service.

- Hedgecock, J., & Lefkowitz, N. (1992). Collaborative oral/aural revision in foreign language writing instruction. *Journal of Second Language Writing, 1* (2), 255-276.
- Herman, J., Aschbacher, P., & Winters, L. (1992). *A Practical Guide to Alternative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Herman, J. L., Gearhart, M., & Baker, E. (1994). Assessing writing portfolios: Issues in the validity and meaning of scores. *Educational Assessment, 1* (3), 201-224.
- Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Hughes, A. (2003). *Testing for language teacher. 2nd* (Ed.). Cambridge: Cambridge University Press.
- Hulstijn, J. H. (1985). Testing second language proficiency with direct procedures. A response to Ingram. In K. Hyltensstam and M. Pienemann (Eds.) *Modelling and assessing second language development*. Clevedon: Multilingual Matters, (pp. 277-82).
- Huot, B. (1990). The literature of direct writing assessment: Major concerns and prevailing trends. *Review of Educational Research 60* (2), 237-263.
- Huot, B. (1990). Reliability, validity, and holistic scoring: What we know and what we need to know. *College Composition and Communication, 41*, 201-213.
- Hyland, K. (2003). *Second language writing*. Cambridge: Cambridge University Press.
- IELTS (2006). Handbook. Retrieved December 9, 2007, from http://www.ielts.org/_lib/pdf/1649_IELTSbk_2005.pdf

- IELTS (n.d.). Developing general training writing skills. Retrieved September 12, 2007 from http://cambridgeesol.org/teach/ielts/general_training_writing/aboutthepaper/develop_gen_tw_skills.htm
- IELTS (n.d.). Writing band descriptors. Retrieved September 12, 2007 from http://www.cambridgeesol.org/teach/ielts/general_training_writing/data/public_writing_band_descriptors.pdf
- Illinois State Board of Education (1983). Illinois Inventory of Educational Progress Writing Assessment (IIEP). Illinois: Illinois State Board of Education.
- Illinois State Board of Education. (1984). *Write on Illinois! A User's Guide to Scoring Student Essays*. (ERIC Document Reproduction Service No. ED 264298.) Springfield, IL.
- ILR (n.d.). Interagency Language Roundtable Language Skill Level Descriptions: Writing. Retrieved March 3, 2006, from <http://www.govtilr.org/ILRscale5.htm>
- Inaam, M., & Grant, S. (2002). A writing rubric to assess student performance. *Adventures in Assessment*, 41, 33-39.
- Isaacson, S. L. (1996). Simple ways to assess writing: Skills of students with learning disabilities. *The Volta Review*, 98 (1), 183-199.
- Jacobs, H. L. , Zinkgraf, S. A., Wormuth, D. R., Hartfiel, V. F., & Hughey, J. B. (1981). *Testing ESL compositions: A practical approach*. Rowley, MA: Newbury House.

- Johns, A. M. (1990). L1 composition theories: Implications for developing theories of L2 composition. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 24-36). New York: Cambridge University Press.
- Kelly, R. (1978). On the construct validation of comprehension tests: An exercise in applied linguistics. PhD thesis, University of Queensland.
- KET (2004). Mark scheme. Retrieved March 30, 2006, from http://www.cambridgeesol.org/teach/ket/read_write/aboutthepaper/focus/index.htm
- Kroll, B. (1990). Introduction. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 1-5). New York: Cambridge University Press.
- Kroll, B. (1990). What does time buy? ESL student performance on home versus class compositions. In B. Kroll (Ed.) *Second language writing: Research insights for the classroom* (pp.140-154).
- Luft, A. J. (1998). Alternatively assessing an in-service program. *School Science and Mathematics*.
- Madsen, H. S. (1983). *Techniques in testing*. Oxford: Oxford University Press.
- Marby, L. (1999). Writing to the rubric: Lingered effects of traditional standardized testing on direct writing assessment. *Phi Delta Kappa*, 80 (9), 673-680.
- Matthews, M. (1990). The measurement of productive skills: Doubts concerning the assessment criteria of certain public examinations. *ELT*, 44 (2), 117-121.
- MELAB Technical Manual (1996). Ann Arbor, MI: University of Michigan Press.

- Messenger, W. E., & Taylor, P. A. (1989). *Essentials of writing*. Prentice Hall, Canada.
- Munby, J. (1978). *Communicative syllabus design*. Cambridge: Cambridge University Press.
- National Assessment of Education Progress (2007). Testing writing Grade 8. Georgia: Georgia Department of Education.
- North West Regional Educational Laboratory (n.d.). 6+1 Trait writing: Scoring guide. Retrieved December 9, 2007 from <http://www.nwrel.org/assessment/pdfRubrics/6plus1traits.PDF>
- Osterlin, S. (1989,1990). College Basic Academic Subjects Examination (College BASE). The Riverside Publishing Co: Columbia, MO.
- PET (2004). Part 2 mark scheme. Retrieved March 30, 2006, from http://www.cambridgeesol.org/teach/pet/read_write/aboutthepaper/writing_part2/mark_scheme.htm
- PET (2004). Part 3 mark scheme. Retrieved March 30, 2006, from: http://www.cambridgeesol.org/teach/pet/read_write/aboutthepaper/writing_part3/mark_scheme.htm
- Polio, C. (2001). Research methodology in second language writing: The case of text-based studies. In T. Silva and P. Matsuda (Eds.), *On Second Language Writing*. (pp. 91-116). Mahwah, NJ: Erlbaum.
- Polio, C. (1997). Measures of linguistic accuracy in second language writing research. *Language Learning*, 47, 101-143.
- Pollitt, A., & Hutchinson, C. (1987). Calibrated graded assessment: Rasch partial credit analysis of performance in writing. *Language Testing*, 4, 72-92.

- Quellmalz, E. (1982). *Scale for Evaluating Expository Writing (SEEW)*. (ERIC Document Reproduction Service No. ED 236670.) Los Angeles: University of California, Center for the Study of Evaluation.
- Rivers, W. M. (1981). *Teaching foreign language skills*. 2nd (Ed.). Chicago: The University of Chicago Press.
- SfL (n.d.). Writing: Assessment focus. Retrieved September 24, 2006, from <http://www.cambridgeesol.org/sfl/writeass.htm>
- Shaw, S. (2004). IELTS writing: Revising assessment criteria and scales (Phase 3). *IELTS Research Notes 16*, 3-7.
- Soles, D. (2005). *The academic essay: How to plan, draft, write, and revise*. 2nd. Studymates.
- Southeast Missouri State University. (1997). *Writing proficiency exam: Student essays evaluated and annotated*. (ERIC Document Reproduction Services.) Cape Girardeau, MO.
- STEP Test Development Committee (1994). STEP writing skills test specifications. Sydney: Macquarie University.
- Swales, J. M., & Feak, C. B. (1994). *Academic writing for graduate students: Essential tasks and skills*. Ann Arbor: The University of Michigan Press.
- Teacher Programs and Services. (1996). *Praxis I: Academic Skills Assessment*. Princeton, NJ: Educational Testing Service.
- Texas Assessment of Knowledge and Skills (2004). Information booklet: writing grade 7 revised. Retrieved October 8, 2007 from <http://www.tea.state.tx.us/student.assessment/taks/booklets/write/g7.pdf>

- Texas Higher Education Coordinating Board. (1995). *TASP and the Effectiveness of Remediation*. Annual report. (ERIC Document Reproduction Services No. ED394478.) Austin, TX.
- Turner, C. E., & Upshur, J. A. (2002). Rating scales derived from student samples: Effects of the scale maker and the student sample on scale content and student scores. *TESOL Quarterly* 36 (1), 49-70.
- Vaughan. C. (1991). Holistic assessment: What goes on in the rater's mind? In L. Hamp-Lyons (Ed.) *Assessing second language writing in academic contexts*, (pp. 111-125). Norwood, NJ: Ablex.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Weir, J. C. (1990). *Communicative language testing*. NJ: Prentice Hall International.
- Wesche, M. (1987). Second language performance testing: The Ontario test of ESL. *Language Testing*, 37, 28-47.
- White, E. M. (1998). *Teaching and assessing writing: recent advances in understanding, evaluating, and improving student performance*. 2nd (Ed.). Maine: Calendar Islands Publishers.
- Yukio, H. (1998). Writing to improve analytical and organizational skills. *The Language Teacher online*, 22 (12). Retrieved June 25, 2008 from <http://jalt-publications.org/tlt/files/98/dec/hirayanagi.html>