

Research, Consultation, & Teaching Program
Training Module No. 8

RCTP

Analytic Scoring
of Writing

Clarice Jentsch
Gerald Tindal

Published by
Research, Consultation, & Teaching Program
Behavioral Research and Teaching
College of Education
University of Oregon

Staff
Gerald Tindal, Program Director
Jerry Marr, Editor
Denise Styer
Donna Jost
Clarice Skeen
Mike Rebar

Copyright ©1991 University of Oregon. All rights reserved. This publication, or parts thereof, may not be used or reproduced in any manner without written permission. For information, write University of Oregon, College of Education, Research, Consultation, & Teaching Program, 237 Education, Eugene, OR 97403-1215.

Jentsch, Clarice; Tindal, Gerald
Analytic Scoring of Writing
Training Module No. 8

This project funded in part by the U.S. Department of Education, grant number H029B00053-90. However, the opinions expressed herein do not necessarily represent the position of the U.S. Department of Education or the College of Education, University of Oregon, and no official endorsement by the Department, College of Education, or University should be inferred.

Cover design: George Beltran

*to be printed on
reverse side of
cover.*

*inside
front cover*

Analytic Scoring of Writing

Clarice Jentsch
Gerald Tindal
University of Oregon

INTRODUCTION

Determining the best method for measuring writing ability is a common concern for educators and researchers. Questions regarding the validity of indirect assessments because the student doesn't write has prompted a call for the use of direct assessments to make educational decisions (Breland & Gaynor, 1979; Charney, 1984). In fact, in a review of published tests, Stiggins and Bridgeford (1983) report that the use of student writing samples for measuring writing proficiency increased from 8% in 1973 to 58% in 1981.

But using student writing samples to measure writing competence does not guarantee scores on those samples accurately reflect student skills. Often, inadequate subjective rating systems are used to score writing samples, producing information that is questionable for making educational decisions (Willis, 1990). This training module is designed to help increase reliability with one of three most frequently used subjective scoring systems. And although reliability issues are important from test creation through data entry, this module focuses only on reliability issues associated with this one form of subjective scoring.

SUBJECTIVE EVALUATION SYSTEMS:

AN OVERVIEW

Cooper (1977) lists as many as seven types of subjective evaluation procedures, but three have been used most frequently: (a) holistic, which pro-

vides general or global information (Diederich, 1974); (b) primary trait, which provides discourse-defined information (Lloyd-Jones, 1977); and (c) analytic, which provides specific characteristic information (Mullis, 1984).

Holistic Scoring

Holistic evaluation is the subjective procedure used most often in schools. With holistic evaluation raters make quick judgments on writing samples and assign each one an overall score. Two strategies can be used to score the writing samples. One strategy is to match individual writing samples to anchor papers, which are student writing samples that represent particular ratings. When using anchor papers, raters score their first impression of a writing sample and then make sure it is of similar quality to the anchor with the same score, insuring that reliability may be obtained. Another strategy is to use scoring guides, which outline the features the rater should address when scoring the writing samples. When using scoring guides, raters match features listed on the guides to features on the writing sample and then assign the sample the corresponding rating.

One advantage with holistic evaluation is that it is the quickest method for scoring. Writing samples are read only once or twice, unlike other subjective scoring systems that require many readings. It also provides general or global information, making it especially useful for someone who needs a quick report on a large number of writing samples.

Holistic evaluation takes the least amount of time in actual scoring and creation of the scoring guides but also has the most difficulty with rater consistency. Anchor papers and scoring guides help with consistency, but often, the scoring guides are too general causing, extraneous issues to influence scoring. Carefully written scoring guides can help, but again because the holistic score is a general or global score, the scoring guides reflect this, making them general as well. Vacc (1989) found a lack of clarity in specific features used during holistic evaluation, making it possible that one characteristic such as handwriting was a greater influence than what was intended. General guidelines make it possible for personal bias to influence raters' scores. Anchor papers offer the most guidance for holistic scoring, but using anchors requires that someone read through a group of writing samples and choose the anchors, which can be inconvenient and sometimes not possible.

Another problem with holistic scoring is that the rating or grade offers no specific feedback for the student or teacher. The rating consists of one score, which encompasses all the important qualities that comprise the writing sample, making it difficult to address specific problems in each student's writing. "There is no attempt to analyze the factors that constitute good writing" (Stein, 1983, p. 105) So students who get an A on a paper do not know why. Similarly, students who get a C do not know how to improve their writing to B or A quality.

Primary Trait Scoring

Unlike holistic scoring, primary trait scoring provides the opportunity for giving specific feedback on writing sample. Currently used by the National Association for Educational Progress (NAEP), primary trait scoring involves subjectively scoring writing samples, according to a particular audience or purpose (Tindal & Marston, 1990). Specific scoring guides outline the trait (e.g., persuasive, informative reporting, narrative, expository, etc.) that will be scored in each writing sample.

One advantage of primary trait scoring is that the rating provides specific information that can be used for planning instruction or for student feedback. Also, excessive correction is avoided because the feedback is focused on one area or trait of the composition. Another advantage with primary trait scoring is that, because the guides are so specific, the possibility for judgment consistency across raters is increased, even when raters come from different backgrounds.

But the specific guides also can be problematic if raters let extraneous issues influence them during scoring. Some raters may include characteristics that

aren't on the scoring guide, decreasing the possibility of obtaining a reliable score. Raters must be trained to make decisions that directly relate only to the features outlined on the scoring guide (Mullis, 1984). Spandel and Stiggins (1980) caution that "raters must accept the fact that they will be looking for specific, well-defined traits, and be cautious about allowing extraneous criteria to influence scoring" (p. 24). Understanding the scoring guides is a necessity for obtaining reliability in primary trait scoring.

Another problem with the specificity of the scoring guides is that the writing tasks used with primary trait scoring may be somewhat restrictive and, thus, less attractive to some examinees, possibly decreasing the motivation for writing (Lloyd-Jones, 1977). And creating the scoring guides can be time-consuming. Because each guide is specific to a particular writing task, a new guide must be created for each new task. The time commitment involved with primary trait scoring makes it inappropriate when someone needs a quick judgment on a large number of writing samples.

Analytic Scoring

Analytic scoring is a subjective scoring method in which separate scores are assigned to different aspects of writing. Specific scoring guides outline how many dimensions (e.g. Story Idea, Mechanics, Voice, etc.) will be scored and the specific features of those dimensions that the rater should address. Each dimension is given its own rating, independent of the other dimensions. The Oregon Department of Education (1989), which uses analytic scoring for statewide assessment, assigns writing samples a rating from 1 to 5, according to the following six dimensions: Ideas and Content, Organization, Voice, Word Choice, Sentence Structure, and Conventions.

Because specific features are rated independently, the analytic scale can provide precise feedback and is often used to gain prescriptive and diagnostic information, unlike holistic scales, which provide a general, overall picture (Vacc, 1989). "Focusing on specific characteristics of the student's product ... can assist the teacher in providing instruction that will improve the composition as a whole." (Isaacson, 1987, p. 529). Not only can analytic scoring denote specific strengths and weaknesses in writing, it also can be used to monitor student progress. "Analytical scoring has one great advantage: it provides potential for trait-by-trait analysis of students' writing proficiency" (Spandel & Stiggins, 1980, p. 23).

The opportunity to evaluate specific features of writing may be an advantage, but it also poses some problems. Because each feature is scored independently, analytic scoring may be time-consuming. To

evaluate six features, each writing sample must be read six times, making it difficult to score under time constraints. Also, reliability may be more difficult to establish when multiple features are addressed, especially if the guides for scoring are not delineated clearly (Moran, 1987, p. 8). For example, raters might let mechanical issues influence their score when they should be focusing only on the story idea. "Descriptions of the categories on analytic scales must be written to remove any inevitable relationship between them" (Freedman, 1981, p. 254). Otherwise, a paper might be unjustifiably assigned a score because of some extraneous issue, thus decreasing the likelihood of obtaining an accurate score.

All three evaluation systems are important because they provide evaluation of actual writing samples. Some researchers believe that the only way to truly evaluate the complexities of writing is by direct evaluation because it is more valid than indirect evaluation systems like multiple-choice tests. But, to use each scoring method responsibly involves addressing the reliability issues that are associated with each method. Holistic raters must pay close attention to the anchor papers, which serve as checks for keeping scoring reliable; primary trait raters must pay specific attention to the scoring guides and ignore features that might be important when scoring analytically or holistically; and analytic raters should adhere to the scoring guides to diminish problems associated with scoring on multiple dimensions.

DIRECTIONS FOR ANALYTIC SCORING

This training module provides directions for analytically scoring writing samples. The Oregon Department of Education (1989) outlines six dimensions for scoring, but in this training module, student compositions are assigned ratings on three dimensions: (a) Story Idea, the quality of characters and ideas, (b) Organization and Cohesion, the effectiveness of ordering events, and (c) Writing Conventions, the quality of mechanics and word usage. For scoring, the rater reads through each paper in a selected sample and focuses on one of the dimensions (i.e. Story Idea), assigning a score from 1 to 5 based on the pre-existing scoring guides. After all writing samples are scored on one dimension, the rater reads the samples again to assign each paper a score on one of the remaining two dimensions. When the rating task is completed, each writing sample has three independent scores, one for Story Idea, one for Organization and Cohesion, and one for Writing Conventions.

A list of the 7 steps for analytic scoring is given in Figure 1 to help outline the directions. Tear out directions are included in Appendix A.

Steps 1 & 2: Understand criteria sheets

The dimensions that must be scored are outlined on three scoring guides called criteria sheets. The first step in scoring the written expression samples is to read all three criteria sheets and become familiar with the different dimensions in order to insure judgment consistency across the group of writing samples. When reading the criteria sheets, the rater should take note of the differences between the dimensions, asking questions like, "What makes

Directions for Analytic Scoring

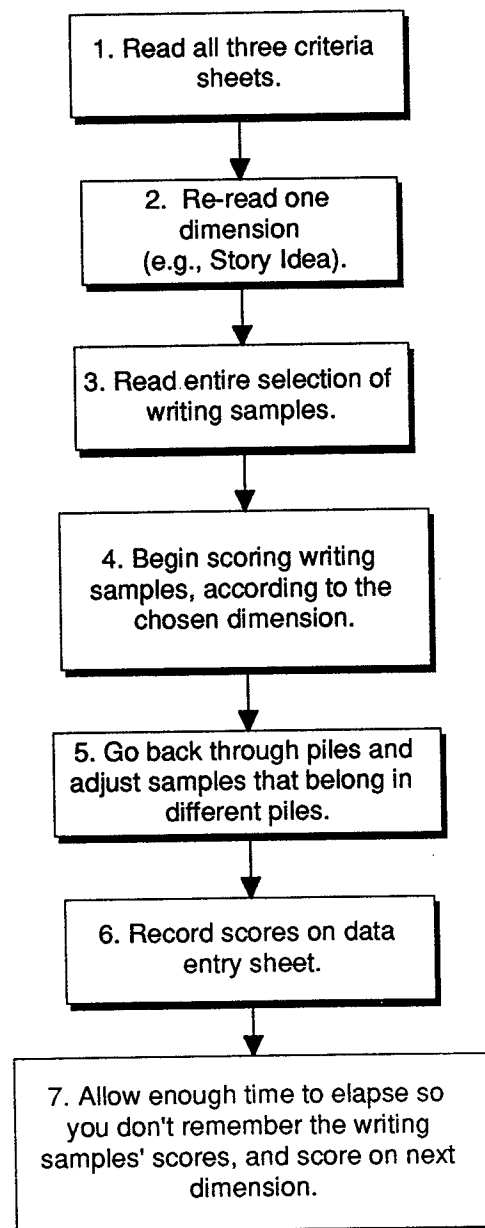


Figure 1. Seven steps for analytic scoring.

story idea different from organization and cohesion?" and "What makes organization and cohesion different from writing conventions." The questions should direct the rater's attention to specific differences between the dimensions so that the characteristics for only one dimension at a time will be judged. Figure 2 contains the criteria sheet for scoring

Story Idea Criteria Sheet

- 1
 - includes few if any characters
 - contains no original ideas
 - events are few and predictable
 - detail is significantly absent
 - lacks descriptors (adverbs and adjectives)
- 2
 - includes few if any characters
 - contains virtually no original ideas
 - events are very predictable
 - detail is significantly absent
 - word choice is predictable, lacking descriptor (adverbs and adjectives)
- 3
 - characters are predictable and undeveloped
 - may or may not contain original ideas
 - events are somewhat predictable
 - lacks detail
 - word choice is somewhat predictable
 - only sometimes contains descriptors (adverbs and adjectives)
- 4
 - includes characters, but they are not original, often coming from movies
 - contains some original ideas but is fairly predictable
 - events are somewhat predictable
 - word choice contains some descriptors (adverbs and adjectives) and some colorful, infrequently used words
- 5
 - includes characters
 - contains original ideas
 - events are not predictable
 - contains some detail
 - word choice: contains descriptors (adverbs and adjectives) and colorful, infrequently used words

Figure 2. Example of Story Idea Criteria Sheet.

Organization & Cohesion. Criteria sheets for all three dimensions are included in Appendix B.

The criteria sheets were designed for use with a story starter writing task, a 10-minute, timed writing task that provides the first sentence of a story to be completed by the student. Figure 1.3 gives an example of a story starter and a student response. During writing, the student marks the writing sample with a star to denote that 3-minutes has passed, after which 7 more minutes are allowed to finish the writing task. The star helps during scoring because the whole writing sample is scored when it is subjectively evaluated, but only the portion to the star is scored during objective evaluation [e.g., counting the number of words, number of correctly spelled words, and the number of correct word sequences (Hasbrouck, 1989)]. A 3-minute sample is adequate to objectively score written expression, but a 10-minute sample is needed for subjective evaluation. This training module focuses on analytic scoring, a subjective evaluation system, and does not provide information on objective scoring. Appendix C contains more story starter examples and directions for its administration.

After reading all three criteria sheets and becoming familiar with their specific characteristics (Step 1), the rater should re-read the criteria sheet for Story Idea in preparation for scoring that dimension (Step 2).

Step 3: Read the writing samples

Next, the rater should read the entire selection of writing samples to become familiar with the range of writing content within the population. At this point, scores are not assigned because the first reading is only to deduce their overall quality. After reading the entire selection of writing samples, the rater reads the Story Idea rating criteria again to help anchor judgments to the 1 to 5 rating scale.

Step 4: Score and place into piles

At Step 4, scoring of the writing samples begins. The rater reads each sample and focuses on the characteristics outlined on the criteria sheet, allowing only about 1 to 2 minutes for each writing sample. This will facilitate efficiency and help the rater maintain a focus on only one dimension at a time. The rater should refer to the criteria sheet frequently during the rating process, making certain that each writing sample is judged according to the same criteria. Also, the rater should refer to previously judged writing samples, ensuring that the ratings remain consistent throughout the judgment process. An easy way to keep track of the ratings previously made is to place each writing sample in an appropriate pile and make it visible at all times. Papers should

Name _____ Date _____
Yesterday, a monkey climbed through the window at school and...
jumped on the teacher's desk. It jumped on my desk. Then it
jumped on the floor. The principal came * and took it away. I was
sad to see it go.

Figure 3. Example of a story starter writing task and a student response.

be aligned within the piles so that, at a glance, as many samples as possible can be viewed at one time, reminders of previously assigned scores.

Steps 5 & 6: Adjust samples and record scores

After all papers have been rated and placed in piles, a quick, third reading, comparing samples *within* each score, should be done to make certain that each writing sample fits within the given rating (Step 5). This reading should not be as careful or comprehensive as the second reading because it is not an opportunity to re-rate across writing samples. It is designed to allow the rater to adjust those writing samples that obviously belong in different piles. Next, the scores assigned to each writing sample are recorded on data entry sheets (Step 6). An example of a data entry sheet is included in Appendix D. Blank data entry sheets for each dimension are included in Appendix E.

Step 7: Wait and score on next dimension

After all writing samples are judged on one dimension, they must be judged on the next one until all three dimensions have been scored. Raters should allow enough time to elapse between the ratings of each dimension so they are not familiar with the writing samples, ensuring each dimension is rated independent of the other two. Remembering a writing sample's score on Story Idea might influence the score given to the same writing sample on Organization-Cohesion or Writing Conventions.

When the rater is finished with the task of scoring, three scores should be recorded for each written expression sample: one for Story Idea, one for Organization-Cohesion, and one for Writing Conventions. Appendix F & G contain ratings of second- and fifth-grade writing samples and explanations for the ratings they received.

RELIABILITY

Importance of Reliability

To be useful, scores or ratings of student work must accurately reflect their skills. They must be reliable, that is it must be possible to produce a comparable outcome or score on a similar test at a different time or on the same test when scored by someone else. With objective tests, like multiple-choice formats, ensuring reliability during scoring is easy because each question has distinctly correct and incorrect answers. No subjectivity is involved in scoring the test so anyone, including a computer, can produce the same score.

Reliability is more difficult to establish when using subjective evaluation systems like analytic scoring of written expression. For example, two teachers read the same writing sample. One gives it a "B" and the other gives it a "C," and both justify their grades based on their own set of guidelines. So which is the most accurate rating for the writing sample, the "B" or the "C"? The answer is neither. The reason the two teachers assigned different grades to the same writing sample is that they didn't use the same criteria, causing their scores to be inconsistent. But test scores must be consistent to be useful. Huot (1990) explains the need for reliability in subjective writing evaluations by stating the following:

In testing theory, the importance of reliability is related to the need to generalize test outcomes, to be able to say that success on a particular test insures success on future tests or demonstrations of the tested skill....The more reliable a test, the more we can generalize about its outcomes. Unless we can generalize about test results, they have no meaning beyond the specific testing situation. In other words, we must be able to generalize scores if we wish to claim that holistic scoring results reflect writing quality and ability. (p. 203)

Without reliability, teachers can't be certain that scores truly reflect writing quality, making educational decisions based on those test scores inappropriate. Only when scoring procedures "yield scores that are stable over time, across exercises and across independent evaluators, [can] those scores can be confidently used for educational decisions" (Spandel & Stiggins, 1980, p. 19). The ability to use test scores for making educational decisions depends on their reliability.

Strategies for Insuring Reliability

Reliability issues are important from creation of the test to data entry, but once the examinee relinquishes a test or writing sample to the proctor, testing reliability issues become the responsibility of the scorer or rater. Ensuring that a rater scores reliably is possible if issues are addressed and problems are minimized. Two variables may influence the reliability of raters' scores: extraneous influences and personal bias.

Avoiding extraneous influences

Extraneous influences are environmental issues that may distract the rater during scoring. To avoid problems with extraneous influences, noise or discomfort should be limited so the rater can concentrate on the rating task. Limiting discomfort requires the rater to be comfortable, including sitting position, room temperature, noise level, etc. The rater must be attentive to the scoring process and not be bothered by extraneous environmental issues.

Fatigue also can cause the rater discomfort, possibly affecting scoring. Fatigue occurs when scoring is done for hours at a time, increasing the possibility that the rater won't attend to the scoring criteria or carefully read the writing samples. Myers (1980) suggests that the rater should take frequent breaks, at least one per hour to ensure that fatigue does not affect the rating process.

Both extraneous issues and fatigue can be especially difficult when judging papers in the middle range of examples because features sometimes overlap in this area. This overlap can make reliability a problem. On a scale of 1 to 5, raters should especially note the differences between a 2 and a 3 paper and between a 3 and a 4 paper. The key to understanding the differences in the ratings is to keep fatigue and extraneous issues from affecting the rater.

Minimizing Personal Bias

Subjective methods of scoring often lack reliability because raters differ in their perceptions of good writing (Charney 1984, p. 67), that is they may be influenced by personal biases. Some characteristics that influence subjective ratings include essay physical appearance, word choice, length, vocabulary, and spelling errors (Grobe, 1981; McColly, 1970). For example, one rater may believe that word choice is more important than spelling and another rater may believe the opposite. Their personal biases influence their judgment of the writing sample. Even if raters agree that a particular feature is important, they may not agree on how important. For instance, a group of raters may agree that handwriting is important when judging writing, but they may not

agree on its relative importance. Some raters may argue that handwriting is only important in issues of legibility, that is handwriting becomes an issue when the composition is no longer legible and is not as important as punctuation. But other raters may say that handwriting is as important as punctuation. Each group of raters' scores would reflect the differing beliefs or personal biases. As Marshall & Powers (1969) wrote, written composition "characteristics such as handwriting quality and composition errors are usually apparent to teachers, but in differing degrees, and thus they could logically be factors affecting the grades of an essay exam" (p. 97).

The best way to avoid personal biases is to adhere to scoring guidelines or criteria sheets, which outline the features the rater should judge during the scoring task. The rater's personal view of what constitutes good writing is not important because each person's opinion will differ, making it impossible to produce reliable scores. They must focus on the scoring guidelines to limit the influence of personal bias. Besides frequent use of the guides, scorers must read papers fast, not spending more than 2 minutes per sample (Cooper, 1977). If the rater thinks too much about a paper, personal bias might enter into the judgment process.

Even if raters adhere to the scoring guidelines, it is possible that they will unknowingly be influenced by essay characteristics like word choice, essay length, vocabulary, and spelling errors. For example, raters may be influenced by the appearance of the essay even if they receive instructions that neatness is not important. Marshall and Powers (1969) found appearance to be a major influence on ratings. In their study, teachers rated four handwriting forms of the same essay: a typed essay, a neat handwritten essay, a poor handwritten essay, and a fair handwritten essay. They found that, even with explicit instructions to disregard everything but content, teachers were still influenced by both quality of composition and handwriting neatness. To limit the affects, raters should frequently refer to the scoring guidelines to insure that personal bias is not affecting their rating. Raters must recognize that essay characteristics can

Table 1. Five suggestions for increasing reliability during scoring.

1. Spend 1 to 2 minutes per writing sample.
2. Take frequent breaks (at least one per hour).
3. Refer to the criteria sheets frequently.
4. Attend to the ratings in the middle range.
5. Score in a comfortable environment.

influence scores when they are not aware of it. Table 1 lists 5 suggestions for increasing reliability during scoring.

Even if a raters do everything they can to minimize personal bias and avoid extraneous influences, there is no guarantee that the scores they produce will be reliable; therefore, it is important to calculate. Reliability does not need to be calculated for each writing task, but it should be checked one time per project. For example, during a schoolwide assessment of writing, reliability should be checked on at least one group of papers, like the 4th-Grade writing samples, but not on the whole school. Also, if different raters are used, reliability should be checked for each rater.

Inter- and Intra-rater Reliability

Inter-rater Reliability

Although there are many types of reliability, only the two that affect raters are important for subjective evaluations: inter- and intra-rater reliability. Inter-rater reliability refers to how much score agreement there is between two judges on the same scoring task and is checked by having two raters score the same writing samples. For example, a teacher rates a group of writing samples, according to a set of scoring criteria. That same teacher then asks another teacher to rate the same papers, using the same criteria. After the second teacher has scored the writing samples, the two teachers' scores are compared and a percentage of agreement, the reliability coefficient, is calculated to see how similarly they scored the set of writing samples. Inter-rater reliability involves two raters independently scoring the same set of writing samples.

Intra-rater Reliability

Because of workloads and time constraints, often it isn't possible for teachers to have a colleague score papers. Intra-rater reliability allows the teacher to check reliability by scoring the samples him or herself. Intra-rater reliability refers to the level of score agreement for the same task and rater at two separate occasions. For example, a teacher rates a group of writing samples from her class and records the scores. She then waits 3 weeks, long enough for her to forget the scores she gave on the first rating, and evaluates the papers again, using the same criteria. Then, the teacher calculates the reliability coefficient, the level of agreement between the two separate scorings. Intra-rater reliability involves the same rater scoring a set of writing samples on two, separate occasions.

Calculating Reliability

Both inter- and intra-rater reliability for analytic scoring can be calculated through three methods.

Table 2. Summary of how to assign values for the total number of hits.

Scoring Criteria	Method 1	Method 2	Method 3
Same Rating (hit)	1	1	1
+ or - 1 in the middle range (2-3 or 3-4)	0	.5	.5
Off by 1 at the end (1-2 or 4-5)	0	0	.5
Off by more than 1	0	0	0

The method chosen should reflect the type of decision that will be made from the score. For example, for screening and eligibility decisions, the most stringent method for calculating reliability would be the best choice because the decision should be based on the most accurate assessment. But for parent conferences with actual protocols or other instances where more general scores are needed, a less stringent method might be the best choice. A summary of the criteria for tallying the total number of hits for Methods 1, 2, and 3 is included in Table 2.

Method 1

The first method is the one most used in research. For Method 1, a value of 1 is given for each "hit," a writing sample that receives the same score by Rater 1 and Rater 2. If the raters disagree on a score, 0 points are given. The total number of hits are divided by the total hits possible, producing the reliability coefficient. Method 1 is the most stringent method for calculating reliability because the two raters must get exactly the same score on writing samples in order to be in agreement.

Method 2

Method 2 is based on the idea that errors are less important in the middle than at the extremes. Also, it is much more difficult to judge samples in the middle range, which consists of the ratings 2, 3, and 4. A value of 1 is given for each hit (each writing sample Rater 1 and Rater 2 judge the same) just as in Method 1. But when raters disagree by one rating in the middle range of values, a value of .5 is given. For example, if Rater 1 gives a sample a 3 and Rater 2 gives the same sample a 4, then .5 is added to the total number of hits. No points are assigned to those scores that differ on the ends of the scale (e.g. those scores that differ from 4 to 5 and those that differ from 1 to 2). For example, if Rater 1 gives a sample a 2 and Rater 2 gives a sample a 1, then 0 points are given because the scores are off at the end of the scale. A score of 0 also is given if the scores are off by

Table 3. Example of two raters' scores on the same writing sample and the calculation of reliability for each method.

Story Idea			Hit Tally		
Student	Rater 1	Rater 2	Meth. 1	Meth. 2	Meth. 3
Adams, J.	3	2	0	.5	.5
Barker, L.	2	2	1	1	1
Edmand, T.	4	5	0	0	.5
Markham, S.	3	3	1	1	1
Sanders, J.	1	1	1	1	1
Taylor, A.	4	4	1	1	1
Wilson, R.	3	3	1	1	1

5	5.5	6	Hit Totals
7	7	7	Possible Hits
.71	.79	.86	Rel. Coef.

Hit Total/Possible Hits =
Reliability Coefficient

more than one rating (e.g. if Rater 1 assigns a sample a 2 and Rater 2 assigns the same sample a 4). After tallying the total number of hits, the reliability coefficient is calculated as in Method 1. Method 2 is less stringent than Method 1 for calculating reliability because some credit still is given for those scores Rater 1 and Rater 2 judge differently (at least in the middle range).

Method 3

Method 3 is useful when rough ordering is important but an exact score is not needed. For Method 3, a value of 1 is given for each hit just as in Methods 1 and 2. But when raters disagree by one rating whether it be at the middle or the end of the scale, a value of .5 is given. For example, if Rater 1 gives a sample a 5 and Rater 2 gives the same sample a 4, then .5 is included in the total number of hits. If the scores are off by more than one rating (e.g. if Rater 1 assigns a sample a 5 and Rater 2 assigns the same sample a 3), a 0 is given. The total number of hits are divided by the total hits possible, just as in Methods 1 and 2, producing the reliability coefficient. Table 3 lists two raters scores on the same writing samples and the reliability coefficients for each calculation method. Method 3 is the least stringent method for calculating reliability. Practice for calculating reliability is included in Appendix H. A blank reliability calculation sheet is included in Appendix I.

What is a Reliable Test Score?

The reliability coefficient yields a number between 0 and 1. To be reliable, the coefficient for a

scoring task must be compared to guidelines established by researchers who have studied reliability. Table 4 lists ranges of coefficients and their meanings, from weak to almost perfect reliability.

In order to make high-stakes decisions, the reliability coefficient should be at least average (.85 or greater). High-stakes decisions include those for eligibility, placement, and permanent behavioral records. If reliability is moderate (.81 to .84), the ratings are more appropriate for use in making low-stakes decisions (which can be reversed) like instructional planning and evaluation, but should not be used for making high-stakes decisions (which are relatively permanent). If reliability is weak (.80 or less), the raters should refer back to the rating criteria, discussing the guides until agreement is

Table 4. Ranges of reliability coefficients, their meanings, and decision uses.

Reliability Coefficient	Meaning	Decision Use
.80 and less	Weak	Don't use for making decisions
.81 to .84	Moderate	Group Decisions
.85 to .90	Average	High-Stakes Decisions
.91 to .93	Strong	
.94 to .99	Almost perfect	

Webb, 1983

reached on their interpretation. Then they should rescore the writing samples based on the new criteria.

The calculation method (Method 1, 2, or 3) influences the reliability coefficient. In Table 2, the difference in reliability between Method 1 and 3 makes the scoring look reliable for Method 3 and not reliable for Method 1. Method 1 is appropriate for

making high-stakes decisions on all students. Method 2 should be used for high stakes decisions with students at the extremes (students at the extremely high or low end). Method 3 should be used when making low-stakes decisions. It is important to determine the decision that will be made based on the rating so that an appropriate calculation method can be chosen and the usefulness of the rating can be determined. If scores are unreliable, it might be tempting to calculate the coefficient using a less stringent method. However, that solution attacks the symptom, not the problem. Instead, reconsideration should be given to the scoring guidelines in the module.

The sample size, the total number of scored papers, also is important. When scoring a small number of writing samples, the reliability coefficient may be skewed because it can be changed dramatically by 1 differing rating. For this reason, the sample size should be at least 25. The six steps for calculating reliability are included in Figure 4.

Reliability can be increased during subjective evaluation if current procedures that help make judgments reliable are followed and reliability is monitored. Reliable test results insure scores are useful for making educational decisions. When a score is unreliable, "the assessment results can lead to erroneous conclusions or decisions" (Spandel & Stiggins, 1980, p. 21). Students also are affected when reliability is lacking because, without reliable scores, students have problems interpreting how to improve their skills based on the rater's interpretation of their writing proficiency. Vacc (1989) states that students, "whose writing is evaluated holistically by teachers and testing services, will not be able to develop writing skills to their maximum potential as long as the interpretation of teachers' holistic scores of writing samples remains in question." In order for scores to be useful, teachers as well as students must be able to trust that scores on a particular writing task are reliable and, therefore, accurately reflect the student's ability to write.

CONCLUSION

As Stiggins and Bridgeford (1983) report, the use of direct writing samples is increasing, creating a need for adequate subjective systems for scoring written expression. This training module provides criteria and practice for analytic scoring. It also includes information for increasing reliability during scoring, which is important for insuring ratings truly reflect writing ability. If scores are not reliable, they are not useful for making educational decisions. The most important issue is that students are adequately served in the school setting and that educational

Six Steps for Calculating Reliability

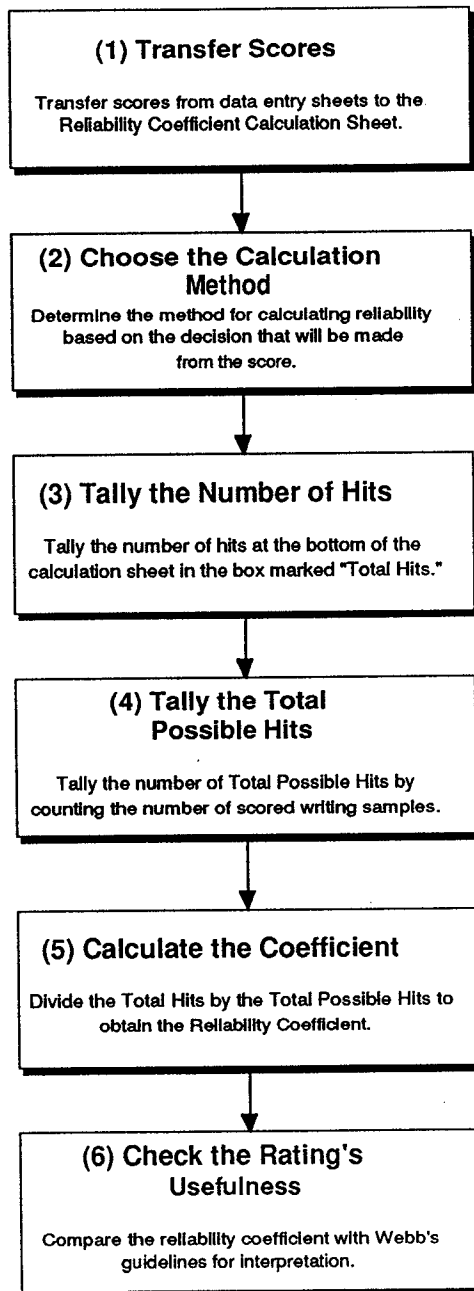


Figure 4. Directions for Calculating Reliability.

decisions are made from reliable information. A complete training section, from scoring writing samples to calculating reliability, is included in Appendix J.

LIST OF APPENDICES

- Appendix A: Analytic Scoring Directions.
- Appendix B: Criteria Sheets for Analytic Scoring.
- Appendix C: Administration Directions for the Story Starter and Sample Story Starters.
- Appendix D: Data Entry Sheet Example.
- Appendix E: Blank Data Entry Sheets
- Appendix F: Sample Ratings and Explanations for Grade 2.
- Appendix G: Sample Ratings and Explanations for Grade 5.
- Appendix H: Reliability Practice.
- Appendix I: Blank Reliability Calculation Sheet.
- Appendix J: Analytic Scoring Practice.

ABOUT THE AUTHORS

Clarice Jentzsch is a Master's student in Special Education at the University of Oregon. She has a B.A. in Journalism and a Secondary Teaching Certificate in Language Arts and Journalism for the state of Oregon.

Gerald Tindal is an Associate Professor at the University of Oregon in the Division of Teacher Education. A specialist in assessment and consultation, he is the Director of the Resource Consultant Training Program.

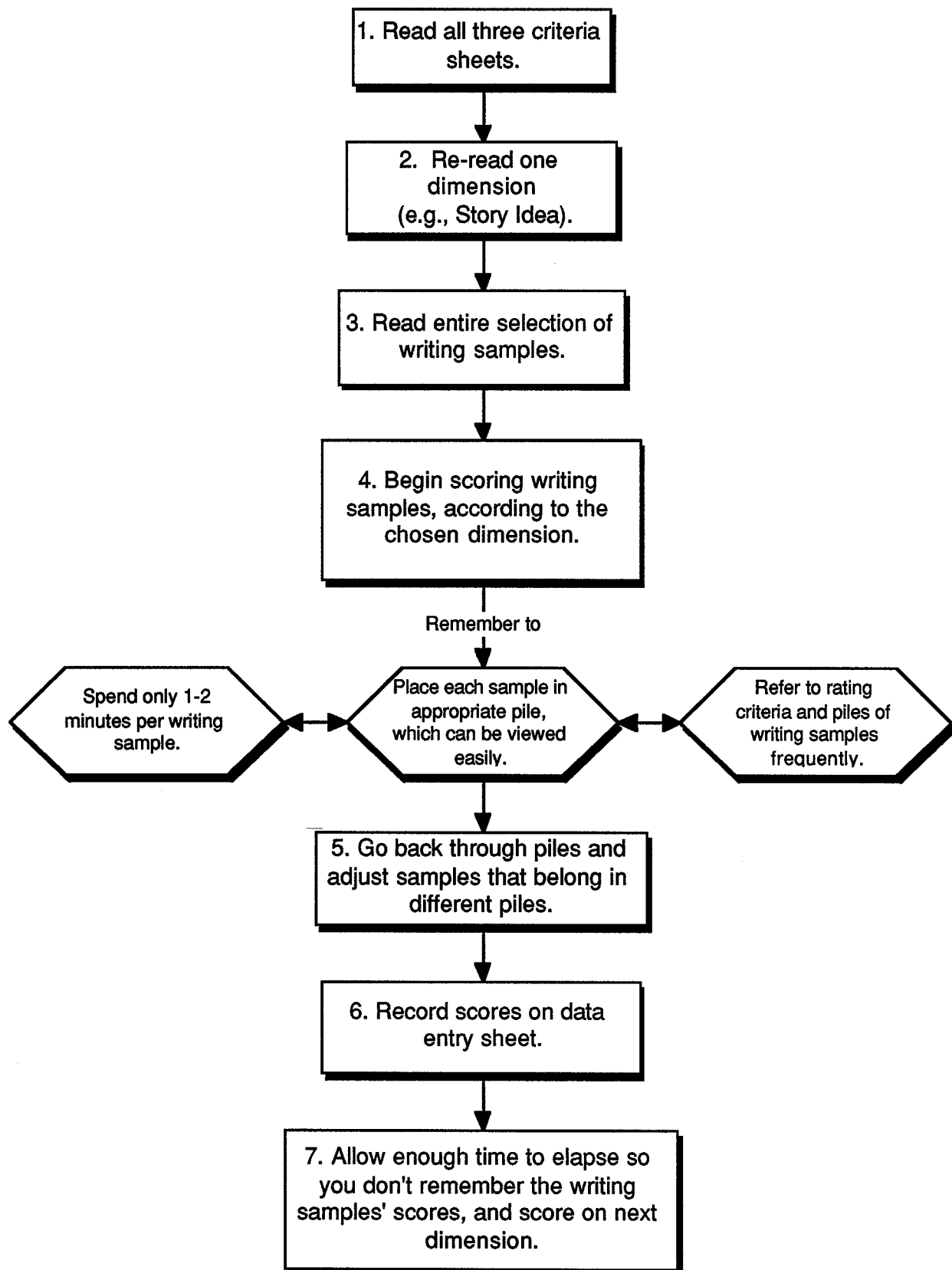
REFERENCES

- Breland, H. M., & Gaynor, J. L. (1979). A comparison of direct and indirect assessments of writing skill. *Journal of Educational Measurement*, 16(2), 119-128.
- Charney, D. (1984). The validity of using holistic scoring to evaluate writing: A critical overview. *Research in the Teaching of English*, 18(1), 65-81.
- Cooper, C. R. (1977). Holistic evaluation of writing. In C. R. Cooper, & L. Odell (Eds.), *Evaluating writing*. (p. 3-31). Buffalo, NY: National Council of Teachers of English.
- Diederich, P. (1974). *Measuring growth in English*. Urbana, IL: National Council of Teachers of English.
- Freedman, S. (1981). Influences on evaluation of expository essays: Beyond the text. *Research in the Teaching of English*, 15(3), 245-255.
- Grobe, C. (1981). Syntactic maturity, mechanics, and vocabulary as predictors of quality ratings. *Research in the Teaching of English*, 15(1), 75-86.
- Hasbrouck, J. (1989). *Objective and Holistic Scoring of Writing* (Training Module No. 2). Eugene, OR: University of Oregon, Resource Consultant Training Program.
- Huot, B. (1990). Reliability, validity, and holistic scoring: What we know and what we need to know. *College Composition and Communication*, 41(2), 201-213.
- Isaacson, S. I. (1987). Assessing the writing product: Qualitative and Quantitative Measures. *Exceptional Children*, Vol. 54(6), 528-534.
- Lloyd-Jones, R. (1977). Primary trait scoring. In C. R. Cooper, & L. Odell (Eds.), *Evaluating writing*. (p. 33-66). Buffalo, NY: National Council of Teachers of English.
- Marshall, J. C., & Powers, J. M. (1969). Writing neatness, composition errors, and essay grades. *Journal of Educational Measurement*, 6, 97-101.
- McColly, W. (1970). What does educational research say about the judging of writing ability. *The Journal of Educational Research*, 64(4) 148-156.
- Moran, (1987). Options for written language assessment. *Focus on Exceptional Children*, 19(5), 1-12.
- Mullis, I. V. S. (1984). Scoring direct writing assessments: What are the alternatives? *Educational Measurement: Issues and Practice*, 3(1), 16-18.
- Myers, M. (1980). *A procedure for writing assessment and holistic scoring*. Urbana, IL: ERIC Clearinghouse on Reading & Communication Skills.
- Oregon Department of Education (1989). *Results and Analysis of the 1989 Oregon Statewide Writing Assessment*. Salem, OR: Oregon Department of Education.
- Spandel V., & Stiggins, R. J. (1980). *Direct measures of writing skill: Issues and applications*. (rev. ed.). Portland, OR: Clearinghouse for Applied Performance Testing (CAPT) of the Northern Regional Educational Laboratory.
- Stein, N. L. (1983). Methodological and conceptual issues in writing research. *The Elementary School Journal*, 84(1), 100-108.
- Stiggins, R. J., & Bridgeford, N. J. (1983). An analysis of published tests of writing proficiency. *Educational Measurement: Issues and Practice*. (Spring). 6-10, 26.
- Tindal, G. A., & Marston, D. B. (1990). Written expression assessment. In Tindal, G. A., & Marston, D. B. (Eds.) *Classroom-based assessment: Evaluating instructional outcomes*. (p. 203-232). Columbus: Merrill Publishing Company.
- Vacc, N. N. (1989). Writing evaluation: Examining four teachers' holistic and analytic scores. *The Elementary School Journal*, 90(1), 87-95.

- Webb II, M. W. (1983). A scale for evaluating standardized reading tests, with results for Nelson-Denny, Iowa, and Stanford. *Journal of Reading*, 26, 424-429.
- Willis, S. (1990). Transforming the test. *Update*, 32(7), 2-6.

Appendix A:
Analytic Scoring Directions

Seven Steps for Analytic Scoring





Appendix B:
Criteria Sheets for Analytic Scoring

Story Idea

Scoring Story Idea involves analyzing the following questions:

- *Are characters in the story developed?*
- *How unique or original is the story idea?*
- *How predictable are the story's events?*
- *How much detail is included?*
- *Are descriptors used to add detail?*

Story Idea Criteria Sheet

1

- includes few if any characters
- contains no original ideas
- events are few and predictable
- detail is significantly absent
- lacks descriptors (adverbs and adjectives)

2

- includes few if any characters
- contains virtually no original ideas
- events are very predictable
- detail is significantly absent
- word choice is predictable, lacking descriptors (adverbs and adjectives)

3

- characters are predictable and undeveloped
- may or may not contain original ideas
- events are somewhat predictable
- lacks detail
- word choice is somewhat predictable
- only sometimes contains descriptors (adverbs and adjectives)

4

- includes characters, but they are not original, often coming from movies
- contains some original ideas but is fairly predictable
- events are somewhat predictable
- word choice
- contains some descriptors (adverbs and adjectives) and some colorful, infrequently used words

5

- includes characters
- contains original ideas
- events are not predictable
- contains some detail
- word choice: contains descriptors (adverbs and adjectives) and colorful, infrequently used words

Organization and Cohesion

Scoring Organization and Cohesion involves analyzing three questions:

- *How effective is the general framework or plot of the story--does it have a beginning, a middle, and an end?*
- *How cohesive are the events in the story?*
- *How linked are the sentences?*

Organization and Cohesion Criteria Sheet

1

- plot is virtually nonexistent
- events are few and random
- lacks transitions
- lacks referents

2

- plot lacks organization into a beginning, middle, and an end
- events are random, lacking in cohesion
- lacks transitions
- often lacks referents

3

- somewhat of a plot exists but story may still lack a beginning, middle, or an end.
- events are somewhat random
- often lacks transitions
- sometimes lacks referents

4

- story has somewhat of a beginning, middle, and an end.
- events appear somewhat random, but some organization exists
- sample may contain some transitions to help with organization (finally, then, next, etc.)
- story often contains too many events, disrupting cohesion

5

- overall story is organized into a beginning, middle, and an end.
- events are linked and cohesive
- sentences are linked, often containing some transitions to help with organization (finally, then, next, etc.)

Writing Conventions

Scoring Writing Conventions involves analyzing the following questions:

- *How accurate is the sentence structure?*
- *How accurate is the spelling?*
- *How legible is the handwriting?*
- *Is punctuation used correctly?*
- *How accurate is word choice?*

Writing Conventions Criteria Sheet

1

- sentence structure is problematic
- spelling makes it extremely difficult to read
- handwriting is illegible, making it extremely difficult to decode
- punctuation is virtually nonexistent
- word usage is problematic (s,v,o / homophone /s-v agreement)

2

- sentence structure makes story difficult to read
- spelling makes it difficult to read
- handwriting is not very legible
- punctuation is inconsistent and problematic
- word usage is problematic (s,v,o / homophone /s-v agreement)

3

- sentence structure has a few problems
- spelling is somewhat of a problem
- handwriting is legible
- punctuation is fair
- dialogue often is not punctuated correctly
- problems sometimes occur with word usage (s,v,o / homophone /s-v agreement)

4

- sentence structure generally is accurate but not as good as a 5
- spelling does not hinder readability too much
- handwriting is legible
- punctuation does not effect readability too much
- dialogue may be punctuated correctly
- word usage generally is correct (s,v,o / homophone / s-v agreement)

5

- sentence structure generally is accurate
- spelling does not hinder readability
- handwriting is legible
- punctuation does not effect readability too much
- dialogue usually is punctuated correctly
- word usage generally is correct (s,v,o / homophone / s-v agreement)

Appendix C:

Administration Directions for the Story Starter and Sample Story Starters

Administration Directions for the Story Starter Writing Task

Grade 2

"I want you to write a story. I will be giving you a piece of paper. It will have the start of the story written at the top. Please leave your pencils down while I tell you what to do. First, I am going to read the start of the story. Then I want you to finish writing the story about what will happen next. You will have 30 seconds to think about the story that you will write. Then you will have 10 minutes to finish the story. Use your best sentences. Write as many things as you can think of. If you cannot spell a word, spell it the best you can. Are there any questions?"

Read the story starter and then say: "Do not begin writing yet. Take 30 seconds to plan your writing and when I say 'begin,' start writing. If you use up the lined sheet, please turn it over and continue writing on the back. Begin." Start the stop watch and time their writing for 3 minutes.

After 3 minutes, say, " Everyone, stop writing. Please make a star after the last word you wrote. Now you have 7 more minutes to finish your story."

Grades 3-5

"I want you to create a story. I will be handing out a lined paper with the beginning of a story written at the top. Please leave your pencils down and listen to me as I explain the task.

I am going to read the beginning of the story first and then I want you to compose a short story about what happens. You will have 30 seconds to think about the story that you plan to write and then have ten minutes to write it. Use your best and most creative writing: Try to write an original ending to the story that is well organized and uses your best sentences. If you do not know how to spell a word, spell it the best you can. Are there any questions?"

Read the story starter and then say: "Do not begin writing yet. Take 30 seconds to plan your writing and when I say 'begin,' start writing. If you use up the lined sheet, please turn it over and continue writing on the back. Begin. Start the stop watch and time their writing for 3 minutes.

After 3 minutes, say, " Everyone, stop writing. Please make a star after the last word you wrote. Now you have 7 more minutes to finish your story."

Story Starters

1. I was playing outside when a spaceship landed and...
2. One very dark, spooky night I was camping in the woods. I heard a strange...
3. I was walking down the street and found a key in front of a big gold door. I took the key, opened the door and...
4. The mother Grizzly Bear had been sleeping all winter but now it was springtime. She poked her head out of her hole and saw...
5. A little boy looked up and saw a very large giant. He started to climb it and...
6. One day I was out flying my airplane. I saw a dangerous thunderstorm approaching so I...
7. Yesterday, a monkey climbed through the window at school and...
8. Once upon a time there was a haunted house...
9. Our sailboat ran into some rocks and crashed. We were stranded on an island....
10. I was walking my dog and all of a sudden a pack of wolves came running out and...
11. The day was dark and misty as Mr....
12. One day we went to school and the teacher had turned into a...
13. The door latch clicked softly, and I knew I was...
14. The class bully was picking on my best friend and I...
15. My friend and I went to the fair last summer and...
16. One day a burglar climbed through our living room window and...
17. Yesterday, I went to the park and...
18. As I went to the store, I saw a large glowing object in the sky. It was a...
19. I was fishing in the river when I felt a terrific tug on the line and...
20. It was a hot, dry day and I had been walking for hours without food or water when...
21. I waved out the window at my family as...
22. I once had a magic pencil and...
23. Walking slowly down the stairs, Greg felt the hair stand up on the back of his neck and...
24. I decided to follow the oversized footprints, as I was...
25. Working madly in my laboratory, I suddenly realized...
26. My friend and I were walking by an old deserted house and...
27. On Tuesday, a big wind came up and...
28. I was chewing a piece of bubble gum when...
29. Out of the darkness came the sound of large flapping wings and...
30. The sun was just coming up over the horizon, and then, in the middle of the lake...

Appendix D:
Data Entry Sheet Example

Appendix E:
Blank Data Entry Sheets

Appendix F:
Sample Ratings and Explanations
for Grade 2

Grade 2: Story Idea

Scoring Story Idea involves analyzing the following questions:

- *Are characters in the story developed?*
- *How unique or original is the story idea?*
- *How predictable are the story's events?*
- *How much detail is included?*
- *Are descriptors used to add detail?*

Story Idea Criteria Sheet

1

- includes few if any characters
- contains no original ideas
- events are few and predictable
- detail is significantly absent
- lacks descriptors (adverbs and adjectives)

2

- includes few if any characters
- contains virtually no original ideas
- events are very predictable
- detail is significantly absent
- word choice is predictable, lacking descriptors (adverbs and adjectives)

3

- characters are predictable and undeveloped
- may or may not contain original ideas
- events are somewhat predictable
- lacks detail
- word choice is somewhat predictable
- only sometimes contains descriptors (adverbs and adjectives)

4

- includes characters, but they are not original, often coming from movies
- contains some original ideas but is fairly predictable
- events are somewhat predictable
- word choice
- contains some descriptors (adverbs and adjectives) and some colorful, infrequently used words

5

- includes characters
- contains original ideas
- events are not predictable
- contains some detail
- word choice: contains descriptors (adverbs and adjectives) and colorful, infrequently used words

Rating: 1

Student samples that rate a 1 have few if any characters. The story idea is not unique or original, and the events are predictable and few, causing the story to lack detail and descriptors.

Grade 2 Story Starter

Yesterday, a monkey climbed through the window at school and. . .

Rating: 1

G2-G-07	He got in trabl by the Teacher
---------	--------------------------------

This writing sample is a 1 because events are so few that it is difficult to judge the characteristics of story idea.

Rating: High 1

G2-T-01	and the macy cist the ticr and the cis and at a dnana and clim dac at the in
---------	--

This is a high 1 because of the monkey kissing the teacher ("and the macy cist the ticr"), an original event when compared to the writing samples.

This is not a 2 because the story is not developed enough to make judgments about the story idea, and the monkey eating a banana is not an original event.

Rating: 2

Student samples that rate a 2 have few characters but often will have more than stories that rate a 1. The story idea still is not unique or original, and the events generally are predictable. Often, it is possible to judge detail and word choice because the writing sample is longer than a 1.

Rating: Low 2

G2-T-11	the monkey ran down the hely and the monkey ran too the store and he at all the bnanus and he ran too the nekst store and ther were know.
---------	---

This is a low 2 because the events are more detailed than with a 1. For example, the monkey eats bananas in this story, but the author included where he ate the bananas (*"the monkey ran too the store and he at all the bnanus"*).

This is not a high 2 because the story idea is not original, and word choice is not unique. Also, the story only contains one character, the monkey, who isn't developed.

Rating: High 2

G2-T-10	the teachir goat wread that wy I was goan a stealir came in and tuck hem untayl the childin toald the teachir that he klimd owt the winoae oly the childrin sad wy cant we goat to the muce is goaeng to git in trubl childin.
---------	--

This is a high 2 because more characters are included. Also, the teacher is developed through inclusion of the teacher's thoughts about the monkey coming in, *"the teachir goat wread that wyl was goan a stealir came in and tuck hem"* (the teacher got worried that while was gone a stealir came in and took him).

This is not a 3 because the story idea and events are not unique. Also, dialogue makes the characters in the story confusing because no attribution is given for the last statement even though it can be assumed it is the teacher.

Rating: 3

Student samples that rate a 3 have characters but they are predictable and undeveloped. The story idea still is not unique or original, and the events are somewhat predictable and lack detail. The story more often contains descriptors than with a 2.

Rating: Low 3

G2-T-13	mast ap the school we hried to stop hem but he was to fast he had in Brandon Brandn sad lev hem alon or I wont cam bak to school apan. the tehr sad you can Kep hem in ntl you go hom ok said Brandon Yas Yas said Brandon
---------	--

This is a low 3 because the characters are more developed than with a 2. For example, Brandon received a name rather than being referred to as a boy. The dialogue also adds more detail to the characters' identities.

This is not a high 3 because the story idea is not original, and word choice is not unique. Events lack enough detail to help the reader understand the story.

Rating: High 3

G2-G-23	he looked meen at first. but then he huged me. I like him. he is kind to me. I brohgt him home with me. he liked my sister and my brother and my perens very much. I could keep him. thanks Mom and Dad I LOVE you and the monkey. I Shall Call him Jack The Monkey.
---------	--

This is a high 3 because the events are more detailed than with a low 3. It also provides more detail by giving the the monkey a name (*Jack*).

This is not a 4 because the story idea and word choice are not unique. Also, the story introduces many characters (*mom, dad, sister, and brother*), but they are not developed.

Rating: 4

Student samples that rate a 4 have characters that may be developed but who often come from movies. The story idea may contain some original idea, and the events are somewhat predictable but contain more detail than with a 3. Descriptors may be used to add detail to the story.

Rating: Low 4

G2-M-13	got into ar descs and ran away. But he came back and rote on the chock bord. he made a mese and hart one of ore stodints. But he din't hrat me I ran our of the windo and ran home he chast me in my bedroom. But I shut the door. my Mom want into her bedroom too and shut the door The End
---------	---

This is a low 4 because the events are more detailed and elaborate than with a 3. For example, the process through which the monkey chases the boy home includes details of climbing through windows (*"I ran our of the windo"*) and writing on the chalkboard (*"he came back and rote on the chock bord"*) instead of just saying that the person left the classroom.

This is not a high 4 because the story idea is not unique, and the characters are not named or developed. The story focuses on the events, not encompassing the events and the characters together.

Rating: High 4

G2-T-07	he made a very bad mese he trnd over desks. and chers I grabd him and ran bake to my. hows and hid him onder my bed and Jost in time. mom kame rite then I siad hiy mom she siad I smel sothing I siad I dont smel ine theing. and boy did he blow it. he gigld. so I said kan we kep him she said.
---------	---

This is a high 4 because the events of the monkey giggling (*"and boy did he blow it. he gigld."*) is unique, making the story stand out amongst the other stories. The mother saying she smelled something (*"she siad I smel sothing"*) also is unique. The story also provides somewhat detailed events of how the monkey got under the bed.

This is not a 5 because word choice is not unique and the characters are not developed. Also, other than the one event of the monkey giggling and the mother smelling something, the whole story idea is not very original.

Rating: 5

Student samples that rate a 5 have characters that usually are developed. The story idea usually contains some original ideas, and the events are more detailed than with a 4. Descriptors add detail to the story.

Rating: Low 5

G2-M-14	brok his nose, his arm at school, , , went to the hospitll they fixt him up, he went home in a week latter he got so sick he throo up ants so back to the hospitll he went, wen he went back home he got so sick a gene he throo up wrms, ants so he went back to the hospitll and they hoded him doun, gave gave hem a shot, he had to stay in th hospitll for a yeer, he dide so if you ever di mabe you will see him THE END!
---------	--

This is a low 5 because the events are more unique than with a 4. For example, the monkey threw up worms (*"he got so sick a gene he throo up wrms"*) and ants (*"he throo up ants"*), which are both unique events.

This is not higher than a low 5 because the only character (the monkey) is not named, limiting detail about that character.

Rating: 5

G2-M-07	Brok the Lite and Licked me on the cheek he got into our desks and took all the stuff out. He became our frend he ate Lunch with us he helped us on our work and he Plaid with us. he became the hole schools frend. the monkey was a girl a boy monkey came in and we all were verry happy that she had a boyfriend and we were all happy because she and he were our frend and the hole schools frend to but he and she went for a walk and rigt after the monkeys Left we neded them Because we needed them to go down to the offes to take the Penut Pass and he and she diad the End
---------	---

This is higher than a low 5 because the monkey had a boyfriend (*"and we all were verry happy that whe had a boyfriend"*), which is a unique idea. The story also contains some descriptive verbs (*"brok the Lite and Licked me on the cheek"*), adding detail to the story. The concept of the *"penut pass,"* although not explained gives detail to the story as well.

Grade 2: Organization and Cohesion

Scoring Organization and Cohesion involves analyzing three questions:

- *How effective is the general framework or plot of the story--does it have a beginning, a middle, and an end?*
- *How cohesive are the events in the story?*
- *How linked are the sentences?*

Organization and Cohesion Criteria Sheet

1

- plot is virtually nonexistent
- events are few and random
- lacks transitions
- lacks referents

2

- plot lacks organization into a beginning, middle, and an end
- events are random, lacking in cohesion
- lacks transitions
- often lacks referents

3

- somewhat of a plot exists but story may still lack a beginning, middle, or an end.
- events are somewhat random
- often lacks transitions
- sometimes lacks referents

4

- story has somewhat of a beginning, middle, and an end.
- events appear somewhat random, but some organization exists
- sample may contain some transitions to help with organization (finally, then, next, etc.)
- story often contains too many events, disrupting cohesion

5

- overall story is organized into a beginning, middle, and an end.
- events are linked and cohesive
- sentences are linked, often containing some transitions to help with organization (finally, then, next, etc.)

Rating: 1

Student samples that rate a 1 have no organized plot mainly because events are so few. Sometimes stories are only one or two sentences long. Events are random, and the story often lacks referents, causing confusion in sentence structure.

Grade 2 Story Starter

Yesterday, a monkey climbed through the window at school and. . .

Rating: 1

G2-G-13	he had to. go to the zoo with the ethr anamolts and hav fan
---------	---

This is a 1 because events are so few it is difficult to judge their cohesiveness. Likewise, it is impossible to judge how linked the sentences are because the story contains only one sentence. Also, the story has only a beginning, no middle or end.

Rating: High 1

G2-G-02	and he wit to his hideout wer his monkey firlns wrer at thay sed did you git a.
---------	---

This is a 1 mainly because the plot is not developed. The student seems to be trying to develop some sort of a plot but the amount of writing isn't sufficient enough to make a judgment on this characteristic. The story has only a beginning, no middle or end. The events are so few that transitions can't be used, another characteristic of a 1.

This paper is a high 1 because the events in the story, although few, are linked together. For example, the use of the word "where" links the two clauses "*he wit to his hideout*" (he went to his hideout) and "*his monkey firlns wrer at*" (his monkey friends were at).

Rating: 2

Student samples that rate a 2 have somewhat of an organized plot but still are not organized into a beginning, middle, or an end. Events often are random. Transitions are used occasionally but not often enough, and a lack of referents makes identifying characters problematic.

Rating: Low 2

G2-M-17	he foud a ruban and he prote his hand and evree anml laft at him they cold him name like dambo he is fany with no hand hes funy he is so funy. the end
---------	--

This is a 2 because the plot is more developed than with a 1. Also, the word "and" connects clauses, making transitions between ideas smoother.

This is not a high 2 because the plot still doesn't have a developed beginning, middle, or end. In fact, the story is based on one event (the fact that the monkey has no hand), which is the basis for the animals ("and evree anml") laughing at him. Also, notice that the reader can only guess that the animals are the ones saying, "he is fany with no hand" (he is funny with no hand) because no reference to the speaker is made by way of attribution. It isn't even obvious that the writer intended it to be dialogue or if it is a statement made by the writer.

Rating: High 2

G2-G-16	He went home to tel His mom and His Dad 'He yeld owt dad' mom I got Big news I went to Schlool and I met my rerind ta Schlool. One bay ti Hpin He got kit out of The School.
---------	--

This is a high 2 because the dialogue is obviously attributed to someone, thus making it possible to identify it as dialogue and who is speaking. Also, the transition "One bay" (one day) helps move the reader into the middle, thus making the plot more developed giving the story a beginning and the start of a middle.

This is not a 3 though because of the lack of plot development and the fact that the only referent for the word "He" is the monkey, making it questionable who is the character in the story. A student might be the character in the story but it is not apparent.

Rating: 3

Student samples that rate a 3 have more of an organized plot than a 2, although they still may lack a beginning, middle, or an end. Events often are somewhat random, but seem to be more linked than 2. Transitions are used more often, helping with sentence organization, and referents may be used to limit problems with identifying characters.

Rating: Low 3

G2-G-09	And we liked in we lafte and giglde. it was fun to and we plaed with it and it was fun and we paste it around and it was fun. We took trnse leting it on ar desk and it was fun. and we took trnse throing the monkey up in the aer.
---------	--

This is a low 3 because there is no confusion about who the "we" and "it" is in the story. Also, the writer has tried to create some sequence of events, even though transitional words other than "and" weren't used.

This is not a high 3 because of the lack of transitions other than the word "and" and the randomness of events.

Rating: High 3

G2-G-10	The monkey went in the caFuterea and he ate all uf the bnanus and he went in room 21 and he drac all uv the woor then he went dac to the zoo and he livd hapie-evr aFtr.
---------	--

This is a high 3 because it is organized into a beginning, middle and an end. Also, the use of the word "then" in the sentence "then he went dac to the zoo" (then he went back to the zoo) provides a sufficient pause to lead the reader into a transition of events and into the conclusion in which the monkey lives "hape evr aFtr" (happy ever after).

This is not a 4 because the plot is just a framework and is not developed enough to need transitions between sentences or events. Also, even though it is obvious that the "he" at the beginning of each clause ("he ate all uf the bnanus") refers to the monkey, the story would read better if the writer had used "the monkey" occasionally in place of the word "he."

Rating: 4

Student samples that rate a 4 have an organized plot, unlike a 3, which still may lack a beginning, middle, or an end. Events with a 4 most often are not random and seem to be linked and sequential. Transitions are used when needed, helping with sentence organization, and referents are used to limit problems with identifying characters.

Rating: Low 4

G2-G-22	he climbed on my desck and then he jumpt on the teachers desck. The teacher screamd AAAA! Everiebady laft at the monkey. The monkey went to lanch with the clas and ate a bnana. the end
---------	--

This is a low 4 because of the inclusion of the beginning, middle, and somewhat of an end, making it a cohesive unit. The use of the word "then" ("*and then he jumpt on the teachers desck*") helps organize the events. There is no lack of referents either, thus helping the reader discern which character is doing what action.

This is not a high 4 because it lacks the use of sophisticated transitions. The plot isn't developed enough to need or warrant using them. Also, the plot's ending is not obvious. It is possible that the last event (the monkey going to eat the banana) is just another event in a sequence of events. Only the use of the words "*the end*" tells the reader that the story is over.

Rating: High 4

G2-M-03	he was arested by the polees they brot him to jale. He broke out of jale and left the kunchry on a trane. He was tired then the trane stoped some polees walked in they founed the monkey and arested him he never ecscapt from jale. the end
---------	---

This is a high 4 because the plot is organized into an obvious sequence of events, ranging from the beginning to the end. The writer also uses the referent "*monkey*" where it is needed to avoid confusion ("*they founed the monkey and arested him*").

This is not a 5 because not enough transitions are used to help the reader move through the many events in the story. In fact, there are almost too many events in the story, detracting from its cohesiveness. The writer too frequently uses "*and*" yet only once uses the word "*then*," making the writing choppy and stilted.

Rating: 5

Student samples that rate a 5 have an organized plot like a 4 but also have more detail in the plot, making transitions and referents necessary. Events are not random and are more linked and sequential than with a 4. Transitions, including some sophisticated transitions, are used when needed, helping with sentence organization. Identifying characters is not a problem.

Rating: Low 5

G2-M-21	<p>He mest up evrey thing it sprised evry body the next day. Wen we did calinder it sit down rigt by me. He played with me out on recess. We hade a good time. We played with Riched's kite. Then he ate all of the penut's in the penut mushing. He was vary vary stind arter that. It was the end of school. So Ms. Morin said, "I think we better let it stay here and slep" We all said, "yes we'v better are he mite get to cold and dide." So they left him ther and the next day all of the children said, "duy it's summer vekashin." So he went away and lived happaly ever arter.</p>
---------	---

This is a low 5 because it is organized into a beginning, middle, and an end and contains detail in the plot. For example, not only does the story say that the monkey messed up everything, it tells what happened as a consequence of that ("it sprised evry body the next day.") Also, events are sequential, making it easy for the reader to understand the events in the story.

This is not a high 5 because it lacks enough transitions, making the writing somewhat choppy.

Rating: 5

G2-M-04	<p>he was cute. I liked him alot. He was friendly and nice. Everybody liked him. We would play with him. One day he broke his leg. So we took him to the hospital. We always would go and visit him. He got out after a while. He was O.k. after he got out of the hospital. We were glad he got out of the hospital. Then one day he left. He went to another school. We missed him alot. In fact the hole school missed him!</p>
---------	--

This is a 5 because it is organized into a beginning, middle and an end and because it is developed enough to need transitions. The writer correctly uses words like "then," "so," and "one day" to help the reader see the connection between events. Also, the use of the phrase "in fact" ("In fact the hole school missed him") is a sophisticated transitional device for this group of students.

Grade 2: Writing Conventions

Scoring Writing Conventions involves analyzing the following questions:

- *How accurate is the sentence structure?*
- *How accurate is the spelling?*
- *How legible is the handwriting?*
- *Is punctuation used correctly?*
- *How accurate is word choice?*

Writing Conventions Criteria Sheet

1

- sentence structure is problematic
- spelling makes it extremely difficult to read
- handwriting is illegible, making it extremely difficult to decode
- punctuation is virtually nonexistent
- word usage is problematic (s,v,o / homophone / s-v agreement)

2

- sentence structure makes story difficult to read
- spelling makes it difficult to read
- handwriting is not very legible
- punctuation is inconsistent and problematic
- word usage is problematic (s,v,o / homophone / s-v agreement)

3

- sentence structure has a few problems
- spelling is somewhat of a problem
- handwriting is legible
- punctuation is fair
- dialogue often is not punctuated correctly
- problems sometimes occur with word usage (s,v,o / homophone / s-v agreement)

4

- sentence structure generally is accurate but not as good as a 5
- spelling does not hinder readability too much
- handwriting is legible
- punctuation does not effect readability too much
- dialogue may be punctuated correctly
- word usage generally is correct (s,v,o / homophone / s-v agreement)

5

- sentence structure generally is accurate
- spelling does not hinder readability
- handwriting is legible
- punctuation does not effect readability too much
- dialogue usually is punctuated correctly
- word usage generally is correct (s,v,o / homophone / s-v agreement)

Rating: 1

With student samples that rate a 1, sentence structure and spelling are problematic. Handwriting is illegible, making it difficult to read the story. Punctuation is rarely used and word usage is problematic as well.

Grade 2 Story Starter

Yesterday, a monkey climbed through the window at school and . . .

Rating: 1

G2-T-03

1 an Money was a kesan Therewas
 2 a did Money.
 3 an Money was did but Money was a
 4 did* an Money hed a bok nee.
 5 in They hafad a gronup in he was
 6 a Monsr. in nana gat skarae in
 7 There was Monsr.

This is a 1 and not a 2 because the sentence structure makes it difficult to read. The spelling also is problematic; the student didn't spell monkey (*money*) correctly even though it was included in the story starter. Attempts were made to spell words that he or she didn't know, making it difficult to read. The writing sample also lacks punctuation.

Rating: 2

With student samples that rate a 2, sentence structure and spelling still are problematic but not as problematic as with a 1. Handwriting is not very legible but reading the story is also not as difficult as with a 1. Punctuation is used but is inconsistent and problematic, sometimes confusing instead of helping the reader. Word usage is problematic as well.

Rating: 2

G2-G-15

and ^{He} ~~he~~ ^{er}reck ^{er}very thing and ^{He} ~~he~~ was
 cut and funny. The monkey we haed.
 In ~~S~~ school He was tara bl* today
 the monkey ~~we~~ was good today
 and we cep hin for a wile we ~~tooth~~ took
 Him to the zoo ~~to~~ today the end

This is a 2 and not a 1 because the sentence structure makes it possible to understand the story. Also, even though spelling is can be problematic in places ("cep hin" meaning kept him) closer approximations are accomplished, making it easier to understand than with a 1 ("cuet" for cute "tarabl" for terrible).

This is not a 3 because punctuation is inconsistent and often lacking, which can be somewhat confusing. Capital letters are randomly used, falsely indicating the beginning of a sentence.

Rating: 3

With student samples that rate a 3, sentence structure has a few problems but is not problematic like with a 2. Handwriting does not impede reading the story, and punctuation is used more consistently than with a 2. Dialogue often is not punctuated correctly. Sometimes problems occur with word usage, but those problems do not hinder readability.

Rating: 3

G2-M-16

he yelled eeeeeee!
 Then he yelled BOOB
 It scared me hafe to deTh!!!
 The teacher cept him in
 a cage. The teacher yelled
 eeeeeeeek she sawe the
 mun ky geting out of his
 cage. She got so tierd she
 told me to tech the clas

This is a 3 and not a 2 because the sentence structure and punctuation make the story very understandable. The author even uses exclamation marks to show excitement.

This is not a 4 because spelling, although not very problematic, is somewhat incorrect. Also, dialogue is not punctuated correctly.

Rating: 4

With student samples that rate a 4, sentence structure has fewer problems than with a 3. Handwriting does not impede reading the story, and punctuation is consistent and helpful. Dialogue sometimes will be punctuated correctly. Occasionally problems occur with word usage.

Rating: 4

G2-M-19

ate the teachers apple, and (liked)
 Someones nose he went to the call-
 bord and started to light. He
 (rot) "Me" the class said "Me!"
 The Class was trying to (ignor) Him
 But he cept* (liking) people. So
 it was hard to do work sheets
 The Monkey went back out
 and never came back again.

This is better than a 3 because punctuation is fairly consistent and helpful. The writer also punctuated dialogue correctly. Sentence structure also is more correct than with a 3.

This is not a 5 because the spelling can make the story difficult to read ("rot" instead of wrote for example, and "cept" for kept).

Rating: 5

With student samples that rate a 5, sentence structure is accurate. Handwriting is not an issue when trying to read the story, and punctuation is consistent and helpful. Dialogue usually is punctuated correctly. Word usage poses no problems in reading the story.

Rating: 5

G2-M-11

We were so suprised evrybody was so aife and then the teacher spoke up and said "Will some one go and get the ganeter please?" Tomy said "yes" so he went ~~as~~ as fast as he could go to get the ganeter. I didn't say a word until the monkey was out and we started to work again. The End

This is 5 and not a 4 because punctuation is more correct. Dialogue also is punctuated correctly for the long quote by the teacher. The sentence structure makes it easy to read, and the spelling is generally correct.

Appendix G:
Sample Ratings and Explanations
for Grade 5

Grade 5: Story Idea

Scoring Story Idea involves analyzing the following questions:

- *Are characters in the story developed?*
- *How unique or original is the story idea?*
- *How predictable are the story's events?*
- *How much detail is included?*
- *Are descriptors used to add detail?*

Story Idea Criteria Sheet

1

- includes few if any characters
- contains no original ideas
- events are few and predictable
- detail is significantly absent
- lacks descriptors (adverbs and adjectives)

2

- includes few if any characters
- contains virtually no original ideas
- events are very predictable
- detail is significantly absent
- word choice is predictable, lacking descriptors (adverbs and adjectives)

3

- characters are predictable and undeveloped
- may or may not contain original ideas
- events are somewhat predictable
- lacks detail
- word choice is somewhat predictable
- only sometimes contains descriptors (adverbs and adjectives)

4

- includes characters, but they are not original, often coming from movies
- contains some original ideas but is fairly predictable
- events are somewhat predictable
- word choice
- contains some descriptors (adverbs and adjectives) and some colorful, infrequently used words

5

- includes characters
- contains original ideas
- events are not predictable
- contains some detail
- word choice: contains descriptors (adverbs and adjectives) and colorful, infrequently used words

Rating: 1

Student samples that rate a 1 have few if any characters. The story idea is not unique or original, and the events are predictable and few, causing the story to lack detail and descriptors.

Grade 5 Story Starter

Our sailboat ran into some rocks and crashed.
We were stranded on an island. . .

Rating: High 1

G5-H-09	a we were stranded for seven days and had to eat grass and berries to live. Then on the seventh day a ship came passing by and rescued us.
---------	--

This is a high 1 because it contains some detail ("*seven days*" and "*grass and berries*").

This is not a 2 because the characters are not developed. A reference is made to "*we*" but it is not discernable who the "*we*" are. Also, the events are few and predictable, and the story lacks descriptors.

Rating: 2

Student samples that rate a 2 have few characters but often will have more than stories that rate a 1. The story idea still is not unique or original, and the events generally are predictable. Often, it is possible to judge detail and word choice because the writing sample is longer than a 1.

Rating: Low 2

G5-H-04	it was hot, but I knew it would get colder as night came, so I set out looking for shelter and brought back some small logs and big leaves. Night came and went. The next morning we saw a ship and we built a fire and it came and picked us up and the next time we were in a boat we were more carfull.
---------	--

This is a low 2 because the events are more detailed than with a 1. Descriptors are used to help create a picture for the reader. For example, the use of the words small and big ("*small logs and big leaves*") provide detail about size.

This is not a high 2 because the story idea is not original, and word choice is not unique. Also, the story contains characters, but it isn't apparent who the characters are because they are not developed.

Rating: High 2

G5-H-05	I was really scared, it was just me and two others. I kept saying to myself. I shouldn't have went on the sailboat in the first place. How do I get home? Where am I? Who are these two people I'm with? Those thoughts raced through my mind until I just screamed and then fainted. Then I woke up in my bed. It had just been a nightmare
---------	--

This is a high 2 because in the last two sentences the writer reveals that the story had been a nightmare, which is a unique twist to being stranded on an island.

This is not a 3 because the story idea and events are not developed. The characters aren't developed. In fact, no one is identified even though a reference is made to two people being with the main character on the island ("*Who are these two people I'm with?*"). Also, the story is just a series of questions, lacking development in events.

Rating: 3

Student samples that rate a 3 have characters but they are predictable and undeveloped. The story idea still is not unique or original, and the events are somewhat predictable and lack detail. The story more often contains descriptors than with a 2.

Rating: Low 3

G5-L-28	That had not been used for over 100 years! There were trees and weird stuff! After a day we became very hungry, so we split up in 2 groups there were 4 people in each group, after that we searched for food. My friend got bite by a snake and died we all felt sad... After 3 weeks we got rescued and we were taken home and cared for. I'll never forget my friend and that adventure.
---------	---

This is a low 3 because the story contains more detail than with a 2. For example, numeral descriptors often are used to add detail to the story ("100 years," "2 groups," "4 people," and "3 weeks"). The inclusion of the snake adds some interest to the story.

This is not a high 3 because the story idea is not unique, and the characters are not developed or identified. Events lack sufficient detail to help the reader understand the story. For example, it is not known how they were rescued or who rescued them.

Rating: High 3

G5-B-18	And there was no way to get off this island. So we made a house to keep warm in. One day we seen a pirates ship going toward us. We had to make weapons to protect our selfs. The pirates were now getting off there ship, we got in our fighting stations and got ready to fight. The pirates found our house and went in side. We left our food in there because we were getting ready to eat. The pirates left the house and went searching for the people who lived in the house. We seen them get there swords and guns out and get ready to fight. They found us and we fought tell every pirate was dead. We won so we took there ship and seld home. Then we heard
---------	--

This is a high 3 because the story idea is carried out in more detail than with a low 3. For example, the pirate theme is carried throughout the story, and all events provide detail about that theme.

This is not a 4 because the characters are not developed. It is uncertain who the "we" are. Also, word choice is not unique, and descriptors are not used.

Rating: 4

Student samples that rate a 4 have characters that may be developed but who often come from movies. The story idea may contain some original idea, and the events are somewhat predictable but contain more detail than with a 3. Descriptors may be used to add detail to the story.

Rating: Low 4

G5-L-07	for two days. We had no food or shelter. It was scary. That night it rained hard. It was very cold and windy. My dad told us not to worry, but I still was. I thought we'd be stranded here forever, but we weren't. I was so hungry. The next morning I woke up to the sound and smell of fish getting cooked above a crackling fire. It was my Dad. I asked him how he got it. He said he made a spear with palm frons, and then went spear fishing. Then he told me to go wake up mom and my older brother. I did. Then we had breakfast. We made a shelter, and went fishing again.
---------	---

This is a low 4 because the events are more detailed and elaborate than with a 3. For example, the story reveals that "it rained hard" and that it was "cold and windy," adding to the dilemma of not having any shelter. Also, descriptors are used ("*crackling fire,*" *spear out of "palm frons," "older brother"*), adding detail to the events in the story.

This is not a high 4 because the story idea is not unique, and the characters are not developed enough.

Rating: High 4

G5-L-08	and we had to escape. So we walked around the island to look for some logs to make a raft. We didn't find much, just a little bit of driftwood. So we to some strong vines that were lying on the the beach, and tied the logs together. We pushed out into the blue sea. we were going really good for about two and a half hours until our vines started to break on the raft. And then suddenly the clouds parted, and down out of the heavens came a giant pillar of darkness and we found ourselves getting pulled toward it, closer, closer until we could all see it really closely now. There was a whirlpool spinning around really fast, and then we got sucked right into it, going down through water. It was
---------	---

This is a high 4 because descriptors add detail to the story ("*strong vines,*" "*blue sea,*" "*giant pillar,*"). Word choice adds to the sophistication of the story. For example, active verbs often are used, making the writing more interesting ("*we pushed out into the blue sea,*" "*the clouds parted,*" "*we got sucked right into it*"). The "*pillar of darkness*" also is a unique idea.

This is not a 5 because the story lacks sufficient detail to explain how the building of the raft and the pillar of darkness are related.

Rating: 5

Student samples that rate a 5 have characters that usually are developed. The story idea usually contains some original ideas, and the events are more detailed than with a 4. Descriptors add detail to the story.

Rating: Low 5

G5-H-08	the island was small about a mile long. We were very scared to be stranded on the island. We decided to go and see if anybody lived on the island when we got into a forest we saw a cottage it was very small made out of wood, mud, and leaves. We went and knocked on the door to see who lived there it was and old man. He asked us how we got on the island and we told him the whole story he said his boat was crashed on the rocks too. We ate dinner at his house and my friend that was on land in town got worried about us and came and found us and took us home. The old man thanked us for getting him the island.
---------	--

This is a low 5 because the story idea, finding an old man on the island, is unique. The story also also contains detail, helping create a picture for the reader. For example, the size of the island ("*small about a mile long*"), the composition of the house ("*wood, mud, and leaves*"), and the age of the man ("*old man*") all help to create images in the story. The story idea also is carried through to the end.

This is not higher than a low 5 because the characters are not named, which could add a great deal to their development. Also, the story refers to "*we*" but never really explains who the "*we*" are.

Grade 5: Organization and Cohesion

Scoring Organization and Cohesion involves analyzing three questions:

- *How effective is the general framework or plot of the story--does it have a beginning, a middle, and an end?*
- *How cohesive are the events in the story?*
- *How linked are the sentences?*

Organization and Cohesion Criteria Sheet

1

- plot is virtually nonexistent
- events are few and random
- lacks transitions
- lacks referents

2

- plot lacks organization into a beginning, middle, and an end
- events are random, lacking in cohesion
- lacks transitions
- often lacks referents

3

- somewhat of a plot exists but story may still lack a beginning, middle, or an end.
- events are somewhat random
- often lacks transitions
- sometimes lacks referents

4

- story has somewhat of a beginning, middle, and an end.
- events appear somewhat random, but some organization exists
- sample may contain some transitions to help with organization (finally, then, next, etc.)
- story often contains too many events, disrupting cohesion

5

- overall story is organized into a beginning, middle, and an end.
- events are linked and cohesive
- sentences are linked, often containing some transitions to help with organization (finally, then, next, etc.)

Rating: 1

Student samples that rate a 1 have no organized plot mainly because events are so few. Sometimes stories are only one or two sentences long. Events are random, and the story often lacks referents, causing confusion in sentence structure.

Grade 5 Story Starter

Once upon a time, there was a castle. . .

Rating: 1

G5-H-03	and I lived there, the castle was full of gold
---------	--

This is a 1 because events are so few it is difficult to judge their cohesiveness. Likewise, it is impossible to judge how linked the sentences are because the story contains only two sentences. Also, the story has only a beginning, no middle or end.

Rating: High 1

G5-H-21	and new kids on the block wanted it they couldn't have it unless the told a verey good joke. here is the joke "knok-knok-whos there-just-just who-justin time for super!! and they got it and they sang haply ever after.
---------	---

This is a 1 mainly because the plot is not developed. The student seems to be trying to develop some sort of a plot but the amount of writing isn't sufficient enough to make a judgment on this characteristic. The story has a beginning and an end, but the end is not explained other than they "*sang haply ever after.*" The events are random, another characteristic of a 1.

This paper is a high 1 because, even though the ending is insufficiently detailed, at least some sort of an ending was attempted.

Rating: 2

Student samples that rate a 2 have somewhat of an organized plot but still are not organized into a beginning, middle, or an end. Events often are random. Transitions are used occasionally but not often enough, and a lack of referents makes identifying characters problematic.

Rating: Low 2

G5-B-20	Once upon a time there was a big castle and it was made out of candy and what lived in the castle was a candy Queen and a candy King and every day kids would come up to there castle and take a pice of candy off the castle of course the king and the Queen didn't care because in side the castle they had lots of candy so when ever the kids wanted candy they went to candy castle. The End
---------	--

This is a 2 because the plot is more developed than with a 1. Also, transitions between ideas is smoother because the author attempted to use transitions like "and," "when," and "so."

This is not a high 2 because the plot still doesn't have a developed beginning, middle, or end. The author includes a beginning and a middle, but it is unclear from the passage if the last sentence is actually the end. Only the words "The End" clue the reader that the story is over.

Rating: High 2

G5-H-09	on a mountain. Dracula lived there. Any person that went within 15 miles from the castel would die. One day my friend went within 15 miles from it and saw dracula fly overhead and land behind hin. dracula bit his neck and he died.
---------	--

This is a high 2 because the story is organized into a beginning, middle, and an end. Also, the author uses the transition to move from the beginning of the story to the middle.

This is not a 3 though because not enough transitions are used, causing transitions between events to be choppy. Also, because events are so few, the story, especially the ending, seems rushed.

Rating: 3

Student samples that rate a 3 have more of an organized plot than a 2, although they still may lack a beginning, middle, or an end. Events often are somewhat random, but seem to be more linked than 2. Transitions are used more often, helping with sentence organization, and referents may be used to limit problems with identifying characters.

Rating: Low 3

G5-B-18	in the deep hills far away, that lived a man named Pee wee he was a scientist that made lots of robots. Once he made a robot with 6 hands and 4 legs. That was one of his best robots. So he desided to make one of the strongest robots, so that he can roul the world and be for the rest his life. But one day someone found the castle while he was haunting for deer. Then he went to door and nocked, one of the robots the door. The robot let him in. Then the scientist came out and the man tried to kill the scientist and robot killed him and the robot lived happily ever after
---------	---

This is a low 3 because the author used transitions to connect ideas. Also, the story has a beginning, middle, and an end.

This is not a high 3 because events seem to happen randomly. Because so many different events occur in the story, the author must use many transitions to link the events, disrupting cohesion.

Rating: High 3

G5-H-04	a big castle. A very wealthy King, Queen and princess lived there, and since they were rich they had ar treasure room all full of money. One day they found out two very important things, one, someone was trying to steel his money, and two, a dragon was in town. The king was so smart, he thought that if he could capture the dragon his money wouldn't get stolen and thats exactly what he did and they lived happily ever after.
---------	--

This is a high 3 because it is organized into a beginning, middle and an end. Also, the author uses transitions when they are needed, but transitions are not used too often, which would disrupt cohesion. Also, the author uses a sophisticated organizational device, a number sequence ("one," and "two"), to introduce "two very important things."

This is not a 4 because the plot is just a framework and is not developed enough to need very many transitions between sentences or events. Also, the ending is rushed, delineated mainly by "and they lived happily ever after."

Rating: 4

Student samples that rate a 4 have an organized plot, unlike a 3, which still may lack a beginning, middle, or an end. Events with a 4 most often are not random and seem to be linked and sequential. Transitions are used when needed, helping with sentence organization, and referents are used to limit problems with identifying characters.

Rating: Low 4

G5-L-16	on a hill, where there lived a mad scientist. One day he was cooking up a spell and the spell was for him to be able to turn people into zombies. He tried to potion on himself and it worked. He was planning on telling the people he put the the spell on the go around and rob people of their jewels and money. No one knew about the castel on the hill because it was actually behind the hill. So one day he went to town dressed as a rich person and got about ten people to come to the castel on the hill, and turned them into zombies, and they stold for him and in the end he was so rich he died. The End
---------	--

This is a low 4 because of the inclusion of the beginning, middle, and somewhat of an end, making it a cohesive unit. The use of the transition "*one day*" (*One day he was cooking up a spell*) moves the reader from the beginning of the story to the middle. Other transitions are used where needed (*So one day he went to town*).

This is not a high 4 because it lacks the use of sophisticated transitions. The plot is developed enough to combine sentences for cohesion, but the writing lacks adequate sentence combination, making the writing choppy. Where sentences are combined, the author tends to overuse the conjunction "*and*." Also, the referent "*he*" for the mad scientist is overused.

Rating: High 4

G5-B-19	Whit a hidin Chocolate Hershey bar in a locked up treshire chest so I The wonderfull good for nothing afull dirty roten scondrel walked in the door and said to the king I would like to know were the hidin Hershy bar is he said sese him so I got took to the dungen were all these skelotons wer laying dead and my eyes saw a treshire chest but it was locked with golden master locks so I found a ax by one of the skelotons and I broke The lock with the ax and in the chest was not just one Hershy bar but a life suply of Hershey bars so I lived Happly ever after piging out on Hershey bars.
---------	--

This is a high 4 because the plot is organized into an obvious sequence of events, that are connected with varied transitions. Also, the author skillfully reminds the reader that one Hershey bar was what was anticipated by including the statement "*not just one Hershy bar but a life suply*," thus connecting the beginning and the end.

This is not a 5 because too many sentences are combined together. In fact, there are almost too many events in the story, detracting from its cohesiveness.

Rating: 5

Student samples that rate a 5 have an organized plot like a 4 but also have more detail in the plot, making transitions and referents necessary. Events are not random and are more linked and sequential than with a 4. Transitions, including some sophisticated transitions, are used when needed, helping with sentence organization. Identifying characters is not a problem.

Rating: Low 5

G5-B-02	with a Vampire it only came out at night like most vampires. Kids always dared each other to go into the castle. One night there wher kids playing outside they dared one kid to go into the castle. He went in because if he didn't he would not be in their club. He went up the long staircase up to the attic. He found a cophen that was open. He looked around he didn't see anything but a wooden steak. Then he turned around and it was a huge vampire. He tried to stick the steak in it's heart but he missed. He tried again and missed. He tried one more time and struck the steak into it's heart. Then the vampire died. THE END
---------	--

This is a low 5 because it is organized into a beginning, middle, and an end, which all are detailed, making the reader feel more complete and less hurried. Also, events are sequential, making it easy for the reader to understand the events in the story.

This is not a high 5 because the sentences mostly contain single events, making the writing somewhat choppy. Also, the referent "he" is overused.

Rating: 5

G5-B-17	which had a treasure so big you couldn't it all in a million years. But the problem to getting the treasue was that a fieresome dragon gaurded it, and if anybody tryed to get it he scorch them with his fire. Pretty much no one tried getting it but sometimes after a long time someone would try but would always die. But one day a warrior with a flying pegasus came to this castle to try to get the treasure. When he got to the dragon he drew his sword and called to his pegusus fly. The pegusus flew around the room barely missing the flames. On one time he flew down close enough without being hit and the warrior threw his sword and it plunged right through the heart of the dragon. It fell to the groud dead. The warrior had did it. The treasure provided him and his family for years and he passed it on for along time. The End
---------	--

This is a 5 because it is organized into a beginning, middle and an end and because it is developed enough to need transitions. The writer correctly uses words like "when," "but," and "one time" to help the reader see the connection between events. The author also combines names with referents, making it easy to identify characters.

Grade 5: Writing Conventions

Scoring Writing Conventions involves analyzing the following questions:

- *How accurate is the sentence structure?*
- *How accurate is the spelling?*
- *How legible is the handwriting?*
- *Is punctuation used correctly?*
- *How accurate is word choice?*

Writing Conventions Criteria Sheet

1

- sentence structure is problematic
- spelling makes it extremely difficult to read
- handwriting is illegible, making it extremely difficult to decode
- punctuation is virtually nonexistent
- word usage is problematic (s,v,o / homophone /s-v agreement)

2

- sentence structure makes story difficult to read
- spelling makes it difficult to read
- handwriting is not very legible
- punctuation is inconsistent and problematic
- word usage is problematic (s,v,o / homophone /s-v agreement)

3

- sentence structure has a few problems
- spelling is somewhat of a problem
- handwriting is legible
- punctuation is fair
- dialogue often is not punctuated correctly
- problems sometimes occur with word usage (s,v,o / homophone /s-v agreement)

4

- sentence structure generally is accurate but not as good as a 5
- spelling does not hinder readability too much
- handwriting is legible
- punctuation does not effect readability too much
- dialogue may be punctuated correctly
- word usage generally is correct (s,v,o / homophone / s-v agreement)

5

- sentence structure generally is accurate
- spelling does not hinder readability
- handwriting is legible
- punctuation does not effect readability too much
- dialogue usually is punctuated correctly
- word usage generally is correct (s,v,o / homophone / s-v agreement)

Rating: 1

With student samples that rate a 1, sentence structure and spelling are problematic. Handwriting is illegible, making it difficult to read the story. Punctuation is rarely used and word usage is problematic as well.

Grade 5 Story Starter

Our sailboat ran into some rocks and crashed. We were stranded on an island . . .

Because of space limitations, only a portion of the writing sample has been reproduced. Use the reproduction to judge handwriting, and the typed version to judge the other features.

Rating: 1

G5-H-01	that has litle Hermet crabs all over the ground and at might they pinch you and wak you up in the mirdle of the night ther is teh people stranded on this Island help.
---------	--

that has Little Hermet crabs
 all over the ground and at
 might they pinch you and wak
 you up in the mirdle of the night
 ther is teh people stranded
 ON THIS ISLAND help.

This is a 1 and not a 2 because the writing sample lacks punctuation. The only punctuation that is included is at the end, making the story difficult to read. Also, the handwriting is illegible, and spelling is problematic.

Rating: 2

With student samples that rate a 2, sentence structure and spelling still are problematic but not as problematic as with a 1. Handwriting is not very legible but reading the story is also not as difficult as with a 1. Punctuation is used but is inconsistent and problematic, sometimes confusing instead of helping the reader. Word usage is problematic as well.

Rating: 2

G5-B-14	After we got all the people from the shpip to gether we just sat in silence for about 18 mineuts all stuned two had died we just left the bodes one still a baby. One of the men said that we should go in the ship and get our belongs and food. It wase'nt sank yet the Hole on land there were only forty four ofus now forty two we all climed in the ship and got food blankets and plenty of fresh water we were forced to make shelters and fires after two dreadful monthes drgged by a ship sailed by and recued us from our soon to be cematarys.
---------	---

After we got all the people from the shpip to gether we just sat in silence for about 18 mineuts all stuned two had died we just left the bodes one still a baby. One

This is a 2 and not a 1 because the spelling and handwriting are better than with a 1. It also contains some punctuation, making it easier to read.

This is not a 3 because it lacks enough punctuation to make it easy to read; the story contains so many run-on sentences that it is difficult to understand. Also, spelling is somewhat inaccurate.

Rating: 3

With student samples that rate a 3, sentence structure has a few problems but is not problematic like with a 2. Handwriting does not impede reading the story, and punctuation is used more consistently than with a 2. Dialogue often is not punctuated correctly. Sometimes problems occur with word usage, but those problems do not hinder readability.

Rating: 3

G5-H-02	There was no food in site so my dad and I went to cut some wood for a place to stay during the night. The next morening my dad got up erty and was thingking how to get of the iland so he thought for a long time. Untill he thought of some thing he said we had to bild a boat so we got bisy geting some wood until we herd a horn we ran out on the beach and it was a boat full of life gards. They saw the boat reck and came quick as they could and so we were picked up and saild home in a life gaurd boat.
---------	--

There was no food in site so my dad and I went to cut some wood for a place to stay during the night. The next morening my dad got up erty and was thingking how to get off the iland so he thought

This is a 3 and not a 2 because the sentence structure and punctuation make the story understandable. The handwriting also is legible.

This is not a 4 because spelling, although not very problematic, is somewhat incorrect. Also, the story needs more punctuation that what it contains to help readability.

Rating: 4

With student samples that rate a 4, sentence structure has fewer problems than with a 3. Handwriting does not impede reading the story, and punctuation is consistent and helpful. Dialogue sometimes will be punctuated correctly. Occasionally problems occur with word usage.

Rating: 4

G5-H-11	Me and my friend kept on wating for a boat to come but it never hapend so we just sat there. We only had a little food. It started to get dark so we set up the tent we had. We were hungry but we could not help it so we went to sleep. The next morning we woke up we were on a boat. So we got up and looked around. There was nobods on the boat so we started up the boat and drove home. And the next day at school we had a very long story to tell. But one thing I imete it was very very fun.
---------	--

Me and my friend kept on wating for a boat to come but it never hapend so we just sat there. We only had a little food. It started to get dark so we set up the tent we had. We

This is a 4 and not a 3 because punctuation is consistent and helpful. Sentence structure also is more correct than with a 3.

This is not a 5 because the spelling can make the story difficult to read ("hapend" for happened "imete" for admit).

Rating: 5

With student samples that rate a 5, sentence structure is accurate. Handwriting is not an issue when trying to read the story, and punctuation is consistent and helpful. Dialogue usually is punctuated correctly. Word usage poses no problems in reading the story.

Rating: 5

G5-H-10	The island was very tropical and very strange. I tried to wake up the other people but they wouldn't buge. I tried for hour's to wake them up but they wouldn't come through. I gave up on them and tried to find some shelter. But all I could find was tropical plant's and nerotic flower's. I heard a voice I looked up and there in a tree was a talking bird. I opened my mouth to speak but nothing came out of my mouth. I stood there looking at the bird when finally I said something. "What island is this?" I asked. The bird looked down and replied "this is the island of nubumbu." "The island of what?" "The island of nubumbu." he anserwed. The
---------	---

The island was very tropical and very strange.
 I tried to wake up the other people but they
 wouldn't buge. I tried for hour's to wake them
 up but they wouldn't come through. I gave up
 on them and tried to find some* shelter.
 But all I could find was tropical plant's

This is 5 and not a 4 because punctuation is more correct. Dialogue also is punctuated correctly. The sentence structure makes it easy to read, and the spelling is generally correct.

Appendix H:
Reliability Practice

Practice for Calculating Reliability

The following page lists the steps for calculating reliability. Use the steps to complete the practice worksheets for calculating reliability. A completed example is given first, and then 3 problems provide practice. Answers to the problems are included on the last page in Appendix H. Student codes have been used in place of names for anonymity.

The following page lists the steps for calculating reliability. In the problems, the first two steps have already been done for you. Complete Steps 3 to 6 and then check your answer.

Six Steps for Calculating Reliability

(1) Transfer Scores
Transfer scores from data entry sheets to the Reliability Coefficient Calculation Sheet.

(2) Choose the Calculation Method
Determine the method for calculating reliability based on the decision that will be made from the score.

Method	Decision Type	Decision Example
Method 1	High-Stakes Decisions	Screening/Eligibility
Method 2	High-Stakes Decisions w/ students at the extremes	Permanent Records Program Placement
Method 3	Low-Stakes Decisions	Parent Conferences Instructional Planning

(3) Tally the Number of Hits
Tally the number of hits at the bottom of the calculation sheet in the box marked "Total Hits."

Scoring Criteria	Method 1	Method 2	Method 3
Same Rating (Hit)	1	1	1
+ or - 1 in the middle range (2-3 or 3-4)	0	.5	.5
Off by 1 at the end (1-2 or 4-5)	0	0	.5
Off by more than 1	0	0	0

(4) Tally the Total Possible Hits
Tally the number of Total Possible Hits by counting the number of scored writing samples.

(5) Calculate the Coefficient
Divide the Total Hits by the Total Possible Hits to obtain the Reliability Coefficient.

(6) Check the Rating's Meaning and Decision Use
Compare the reliability coefficient with Webb's guidelines for interpretation.

Reliability Coefficient	Meaning	Decision Use
.80 and less	Weak	Don't use for making decisions
.81 to .84	Moderate	Group Decisions
.85 to .90	Average	High-Stakes Decisions
.91 to .93	Strong	
.94 to .99	Almost Perfect	

(Webb, 1983)

Answers for Reliability Calculation Practice

Organization-Cohesion Method 2

Total Hits	23
Total Possible Hits	26
Reliability Coefficient	.89

Writing Conventions Method 3

Total Hits	23.5
Total Possible Hits	26
Reliability Coefficient	.90

Story Idea Method 2

Total Hits	25.5
Total Possible Hits	30
Reliability Coefficient	.85

Appendix I:
Blank Reliability Calculation Sheet



Appendix J:
Analytic Scoring Practice

Analytic Scoring Practice: Grade 4 Written Expression

This Appendix provides an opportunity to practice the information given in this training module. Before completing this section, become familiar with Appendices F and G, which contain written expression samples and ratings they received. Also, before completing the reliability portion of this Appendix, do the reliability practice provided in Appendix H.

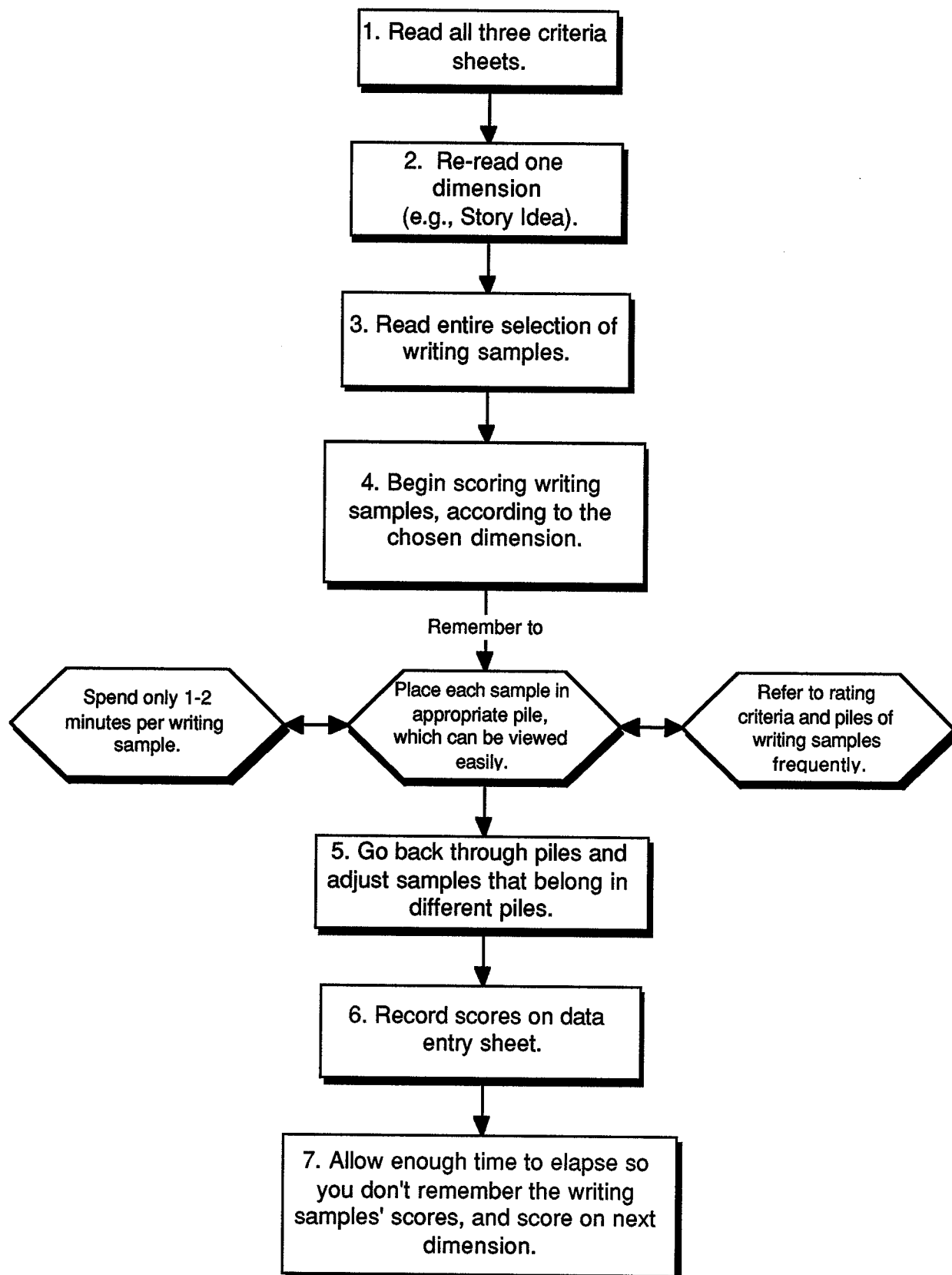
Thirty writing samples from a fourth-grade class are included for practice. Because of space limitations, student compositions were converted to electronic copy. In transferring to electronic copy, no error corrections were made to the student's composition. Also, it is important to note that handwriting can no longer be a factor in judging Writing Conventions. Each writing sample contains a student code (for anonymity) followed by the student's composition.

Follow the directions provided. Remove the writing samples from the module and cut them apart so they can be placed in appropriate piles during scoring. To check reliability, use the inter- and/or intra-rater method, defined in this Training Module.

The following is the story starter for the Grade 4 writing samples in this Practice Section:

Name _____ Date _____
I was walking down the street and found a key in front of a big gold door. I took the key, opened the door and. . .

Seven Steps for Analytic Scoring



Story Idea

Scoring Story Idea involves analyzing the following questions:

- *Are characters in the story developed?*
- *How unique or original is the story idea?*
- *How predictable are the story's events?*
- *How much detail is included?*
- *Are descriptors used to add detail?*

Story Idea Criteria Sheet

1

- includes few if any characters
- contains no original ideas
- events are few and predictable
- detail is significantly absent
- lacks descriptors (adverbs and adjectives)

2

- includes few if any characters
- contains virtually no original ideas
- events are very predictable
- detail is significantly absent
- word choice is predictable, lacking descriptors (adverbs and adjectives)

3

- characters are predictable and undeveloped
- may or may not contain original ideas
- events are somewhat predictable
- lacks detail
- word choice is somewhat predictable
- only sometimes contains descriptors (adverbs and adjectives)

4

- includes characters, but they are not original, often coming from movies
- contains some original ideas but is fairly predictable
- events are somewhat predictable
- word choice
- contains some descriptors (adverbs and adjectives) and some colorful, infrequently used words

5

- includes characters
- contains original ideas
- events are not predictable
- contains some detail
- word choice: contains descriptors (adverbs and adjectives) and colorful, infrequently used words

Organization and Cohesion

Scoring Organization and Cohesion involves analyzing three questions:

- *How effective is the general framework or plot of the story--does it have a beginning, a middle, and an end?*
- *How cohesive are the events in the story?*
- *How linked are the sentences?*

Organization and Cohesion Criteria Sheet

1

- plot is virtually nonexistent
- events are few and random
- lacks transitions
- lacks referents

2

- plot lacks organization into a beginning, middle, and an end
- events are random, lacking in cohesion
- lacks transitions
- often lacks referents

3

- somewhat of a plot exists but story may still lack a beginning, middle, or an end.
- events are somewhat random
- often lacks transitions
- sometimes lacks referents

4

- story has somewhat of a beginning, middle, and an end.
- events appear somewhat random, but some organization exists
- sample may contain some transitions to help with organization (finally, then, next, etc.)
- story often contains too many events, disrupting cohesion

5

- overall story is organized into a beginning, middle, and an end.
- events are linked and cohesive
- sentences are linked, often containing some transitions to help with organization (finally, then, next, etc.)

Writing Conventions

Scoring Writing Conventions involves analyzing the following questions:

- *How accurate is the sentence structure?*
- *How accurate is the spelling?*
- *How legible is the handwriting?*
- *Is punctuation used correctly?*
- *How accurate is word choice?*

Writing Conventions Criteria Sheet

1

- sentence structure is problematic
- spelling makes it extremely difficult to read
- handwriting is illegible, making it extremely difficult to decode
- punctuation is virtually nonexistent
- word usage is problematic (s,v,o / homophone /s-v agreement)

2

- sentence structure makes story difficult to read
- spelling makes it difficult to read
- handwriting is not very legible
- punctuation is inconsistent and problematic
- word usage is problematic (s,v,o / homophone /s-v agreement)

3

- sentence structure has a few problems
- spelling is somewhat of a problem
- handwriting is legible
- punctuation is fair
- dialogue often is not punctuated correctly
- problems sometimes occur with word usage (s,v,o / homophone /s-v agreement)

4

- sentence structure generally is accurate but not as good as a 5
- spelling does not hinder readability too much
- handwriting is legible
- punctuation does not effect readability too much
- dialogue may be punctuated correctly
- word usage generally is correct (s,v,o / homophone / s-v agreement)

5

- sentence structure generally is accurate
- spelling does not hinder readability
- handwriting is legible
- punctuation does not effect readability too much
- dialogue usually is punctuated correctly
- word usage generally is correct (s,v,o / homophone / s-v agreement)

Student Writing Samples: Grade 4

G4-C-01 I found a bottle laying on the flour that said put this in water and the rest of it was torn off. so I put the powder in a cup and filled the cup with water then I drank it. I didn't feel any different. I went home and took the bottle with me and then I look in the mirror and all I saw was my clothes hang in mid air I was.... INVISIBLE and if I took off my clothes you couldn't see anything. So I took off my clothes and I saw nothing I put my clothes back on and went down stairs and got some goodies out of the fridge. I went up to my room again and looked in the mirror and I saw my hole body again I guess it only lasts for an hour or two. I showed my friend Casy she said that she wanted some so I gave her a couple little drops and told her not to tell anyone so she didn't but she stole some things out of peoples houses and then she came to me and asked for more and I said NO!

G4-C-02 There it was the golden eagle sitting on the nasty villans desk "Boo" yelled craig the crook I jumped and ran out of the room with the key and boted home in my corvete going 70 mph the whole way home. Later on that night me and my buddy Mike we went back over there and I took out the key put it in the door and I opended very softly then its gone

G4-C-03 ther were ghost and monsters they said get out But I DiD no lisen I when in and said get out of my why and the DiD and I went in a room there was gold per gold I to it and put some on me and smelled it then I took it out of the Bedroom and whent out side and when to the place e and they said you can Have I Bought lots of stuff I was rich

G4-C-04 It is a boy woh live here I fond big gold coin that covrd the enten it had a sun on one sid but no sun on ather side there was a island in was hot the island was big enuf to put 10 elufer a but it wood sink because there was bats to palt it wit so if i wonid to go some were I wood have food but be adle to came bake to. The was a 2 yea susip of food so don't wariy. biy I haf to go naw. P.S. I rit so by.

G4-C-05 a gold map was laying on the floor. I picked up the map and folowed it. I said "To go 5 ft left. Then 10 ft nort then go to a green horse step in the closes pile then 2 ft. south. Then find a x on the rock then move the rock on another rock then dig 50 ft.

G4-C-06 I saw the most beautifulest mansion in the world! It has gold everything! I walk in and it was so beautiful. It has crystal chandilears, and dimond vases with roses in them. It had black leather couches and chairs, checkered floors and white walls. I walked out the door ran home with the key in my hand & told my mom. She came down with & I showed her the inside. She loved it! So we lived there from now until we die. Plus no one can break through our gold door! The End!

G4-C-1707 I saw a person sleeping on a chair it was my Grandma I didn't think that she was my Grandma and I went home and asked my mom if Grandma lived on 10th main street in a white & pink house whith a sliver car my mom said yes I said that grandma said I could have a key too her house. I get to go there after school. every day why your at work. and dady to and then I walked home when you get home now one and I get to spend the night there now on firday tell Sunday I get to go to church with her on Sunday. Grandma very nice too me.

G4-C-08 there was a big sters and I started to walk up it when I got to the top I saw a dor rite in front of my and I walked to it and I trid to open it but it was locet so I used the key that I fawned and It worked I started to walk in but when I got in the door shot and I looked bake and there was know door and when I tornd lack I fowl

G4-C-09 I so a lot of gold I tried to pick it up but i was to heavy then I went to the door becuse I wanted to go home to get me wagen but win i ternd a round ther was no door then I went to see if ther was another door dut ther was seant but then I So a trap bor I opend it and ther was my house but wen I. got in anither was my plaing bas ball wen I was a kid then I So the hous that I was in defore then desoperd and my alorm wen off.

G4-C-10 it was a big room with lots of gold and brass in side it. The room looked like a golden room with things like brass beds and golden cobbrds in it. the room was filled with stuff I wanted to take it and go away to France or Germany and buy a castle. Then my mom woke me up for school.

G4-C-11 At first I was sceard to go in so I closed my eyes and walked in I saw my self in side it was pere white in side then I bumped in to a button on the wall and then I opened the door and I was sceard and then I saw me and my family in the past and it was incredibal I went back in and I gese I hit what was call the controle borod and I close my eyes verey titely and when I opened the door and I was home it was amasing. It must of been some kind of trick to scare me I never told any one my secreat and I lived happily ever after.

G4-C-12 I walked in side and there was gold all over then I jumped in to it and threw money all over and I told all of my friends first sean came over then danial came over then kreg came over then jona came over then josh came ove then kevin came over and we had fun then my mom woke me up but it was a dream.

G4-C-13 I found a mew world it didn't have any electric thing it was very beauliful and pretty. I looked at my self for a second I was wearing armor and I had a sword at my side I was going. to go back through the door but it wasn't there but I thought was going to live here for the rest of my life I started to cry but then I heard someones voice I looked back and saw a man dressed in a blue satin robe with stars and moons all over it. He said, please don't cry but to get home you must find ahrd kill the black dragon. In his cave you will find a bag of magic powder when get it sprinkle some on you. and you will return to your own world.

G4-C-14 I saw piles and piles of money and there was somuch I could of melted it all down and made a sold gold statuae of liberty for evry day of the year and the first thing I did was grab a 100 dolar bill and go to toys R us and bought a Nintendo and I went back to the door and got some money and went out and bought me a lombergini and my dad and mom a ferri and my little sister a four weller and bouaht a manchon and was the richest man in the world and I saved all the rest of the money for colege for me and my two sisters.

G4-C-15 there was nothing in it I was cross so cross I left the door open when I left. That night when I went to bed a spirit came out of the door that I left open I didn't know it but the spirit saw me open the door and found my house he woke me up firt I thought I was dreaming but I pinched my self and I felt it so I was awake. I felt like screaming but it would probly disapere and they would get mad at me for waking then up so I talked with him and we became friends, Oh by the way his name is Jermy.

G4-C-16 I saw a big pile of gold behind the door. But with the gold I saw a dragon and a bird with fire comming out of the birds mouth. On the wall I saw a spear. I got the spear off the rock wall. I shot the spear through the dranoin and hit the bird with fire comming out off it's mouth, and the bird fell to the groud but missed the gold. After a minute or two I got the gold out of there took it home and became the richest man in the whole wide world!!! The End

G4-C-17 I saw a big pile of but I also saw big giant so I ran to the big gold doo but wasn't there ang and than the stood up and said I snell some craker jack and sat back than the giant may I have some so I said if you give me your gold you can all the cracker jack you want and giant deal so he made come and I went and bought him the crocker he wanted he gav me his The End

G4-C-18 I saw tons and tons of gold! it was unblevible! I shut the door took the key with me ran home to Mom and said. I found tons of gold. she asked me were? Mom said in a gold mine a big unfound mine it haug gold safe whith the key on the karpet. we ran to the untold mines safe opend the door are we where rich for ever.

G4-C-19 there was land with tropical plants I thought this was wered tropical land in the middle of new york I went in the door Then I saw a giant lizard in there. It was so big It was a dinosaur! There wher taridactols triceritops and even trauntasaurus rex he started comeing after me when I was running I saw a humuges dinosaur when I was comeing to the door It started to close I jumped jest in time and the traunasars rex didn't quite get out and then the door vanished and all I had was the key.

G4-C-20 saw a bunch of gold and dimonds. I ran home as fast as I could! My friend Amnie was there I showed her, and we both looked at it. I suggested that we take it to the palice, so we did after 2 months no one clamied it So me and Annie got to keep it all to our selves. Pluse a extra \$100.00 couse we turned it in! That's how we became Rich!

G4-C-21 what did I see I see a night with a sord I gues he wanted to fight so i ask hine do you want to fight he said yes!! so I pict up the soud an foughted and ges ow one to Be conteno For next time

G4-C-22 I saw a big black injaw how gave me his golden sourd and. I said thank you sore and then he gave me a throwing star and I said thank you agan then he gave me his mask and once agen I said thank you and then he gave me his sute and I said thank you and do you know how it was? It was my dad drest up as a injaw the end

G4-C-23 And he seid wow becuse there was a brit light. And he had to colos his eis becuse the light was so brit. And then the light whent out and the door shoot and went slam! and the boy wld ay and then some one seid ha boy thock you for opoming the door I have in that door for 100 yars. io be cortud.

G4-C-24 I took a peak and there was a big room I opened the door more and the room sparkled with gold. I liked it and walked in but the only part I didn't like was the spooky sounds saying "help help" I heard it from upstairs I walked up slowly as it became louder and louder. I stoped at the door and turned the nob and there was me me running and trying to kill I hid behind the door and the other me ran out and down the stairs I ran after her and said you are not me I don't kill poeple and she dissapeared and I all the sudden was in my room laying on my bed.

G4-C-25 saw a ton of gold it was as big as Scrooges money bin on the duck tales. I swam in it I threw it up it the air and then I took some to a store and spent it on baseball cards I got a Ben Mcdonald error and it is worth 60 dollars and a Todd Zeial rookie Fleer then I kept some money I put 300 gold coins in my back pack and went home and gave it to my mom I never told any one I had found the gold tresure. the end

G4-C-26 I saw a Alien with a oozy and I saw another oozy but it was by it's self I grabbed the oozy and shot the alie then 20 more aliens walked into the same room that I was in. They all had bazookas and they started shooting at me I fired backed and killed all of them. A big alien walked threw the wall with a a tomic bom. he threw the bom at me and I cought it and I threw it back at him as fas as I could and he caught it and it blew up in his hands and killed him and I jumped outside threw the window and I was still alive. THE END

G4-C-27 I saw the biggest harriest living creature on the face of this earth it had Feet as big as ronald mc. donald and had the tinniest face is was as small as a normel sized foot. His body was fat he liked to eat alot "I thought" should I go inside and see what was in there! but of corse she did she went in there and took it home it stunk real bad so she gave him a bath then she noticed as it was getting out it was her dad. She said 'dad" Don't do that "Ever Ever Ever" "Again" the END.

G4-C-28 stepped inside causily and then I stepped on something an old button attached to the old wood on which I was standing on. I pushed it and enough water to fill a city pool came rumbling out towards me. Of course I ran and hid behind the nearest store which was across the street. When it was safe I went back in the old house not stepping on the button and saw that the old door was really painted gold because the paint had washed off during my little adventure. I walked through the living room and saw the dusty, cobwebby, dirty old floor and the same for everything around me so I went down to the basement where I fell into a metal cart and went down, down, and down.

G4-C-29 Saw tall man he said he could help me in my problem I was poor I wanted to be rich so the man said he could make him have money the next day he woke up and found 5 millon dollers in his room 5 million in his moms room and found 5 millon in his dads room. and 5 millon in his brothers room I went Back to thank the tall man and I did then the man said wenever I need somthing this gold door and key will always be there the end

G4-C-30 saw a chrystal chandelier it had chrystals hanging down the chairs and tables were gold and silver with chrystals hanging off the back. There was a stage with curtains of real gold. Some one in a real big dress and white gloves with a purse that had gold and diamonds started to sing real pretty. She said "how did you get in here." She said "I found a key and opened the door. "Oh." she said. "Would you like to be in my act." So she said "Yes". They practiced and when people came she sang so beatiful she became famous and got rich. And She Lived Happily Ever After

Six Steps for Calculating Reliability

(1) Transfer Scores
Transfer scores from data entry sheets to the Reliability Coefficient Calculation Sheet.

(2) Choose the Calculation Method
Determine the method for calculating reliability based on the decision that will be made from the score.

Method	Decision Type	Decision Example
Method 1	High-Stakes Decisions	Screening/Eligibility
Method 2	High-Stakes Decisions w/ students at the extremes	Permanent Records Program Placement
Method 3	Low-Stakes Decisions	Parent Conferences Instructional Planning

(3) Tally the Number of Hits
Tally the number of hits at the bottom of the calculation sheet in the box marked "Total Hits."

Scoring Criteria	Method 1	Method 2	Method 3
Same Rating (Hit)	1	1	1
+ or - 1 in the middle range (2-3 or 3-4)	0	.5	.5
Off by 1 at the end (1-2 or 4-5)	0	0	.5
Off by more than 1	0	0	0

(4) Tally the Total Possible Hits
Tally the number of Total Possible Hits by counting the number of scored writing samples.

(5) Calculate the Coefficient
Divide the Total Hits by the Total Possible Hits to obtain the Reliability Coefficient.

(6) Check the Rating's Meaning and Decision Use
Compare the reliability coefficient with Webb's guidelines for interpretation.

Reliability Coefficient	Meaning	Decision Use
.80 and less	Weak	Don't use for making decisions
.81 to .84	Moderate	Group Decisions
.85 to .90	Average	High-Stakes Decisions
.91 to .93	Strong	
.94 to .99	Almost Perfect	

(Webb, 1983)

