

5-2019

Holistic Assessment of Narrative Discourse: A Progress Monitoring Tool

Teigan Beck
Utah State University

Follow this and additional works at: <https://digitalcommons.usu.edu/gradreports>

 Part of the [Education Commons](#)

Recommended Citation

Beck, Teigan, "Holistic Assessment of Narrative Discourse: A Progress Monitoring Tool" (2019). *All Graduate Plan B and other Reports*. 1348.

<https://digitalcommons.usu.edu/gradreports/1348>

This Report is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Plan B and other Reports by an authorized administrator of DigitalCommons@USU. For more information, please contact rebecca.nelson@usu.edu.



Holistic Assessment of Narrative Discourse: A Progress Monitoring Tool

Teigan Beck

Utah State University

Abstract

Purpose: The purpose of this study was to design and evaluate the validity, developmental sensitivity, and scoring reliability of a rubric for measuring the holistic quality of stories for preschool and school age children.

Method: 224 children, ages 4-7, told stories in a single structured context which were then scored by a group of trained raters using the Holistic Assessment of Narrative Discourse (HAND) to evaluate aspects of its psychometric adequacy, specifically validity, developmental sensitivity, and inter-rater reliability.

Results: The results showed that the scores on the HAND were a valid measure of narrative quality that were sensitive to developmental changes over age and could be scored reliably.

Conclusion: This tool may be a useful adjunct to traditional measures for characterizing children's narratives.

Introduction

Narration is an important type of classroom discourse that has distinct links to socialization and literacy (Fujiki, Spackan, Brinton & Illig, 2008; Miller & Pennycuff, 2008). Preschool and school-age children with language disorders often demonstrate poor narrative skills which puts them at a disadvantage both socially and educationally. Therefore, a number of recent studies have been conducted to determine the efficacy of narrative-based interventions for children with language disorders (Nicolopoulou & Trapp, 2018). Most of the studies to date have examined how well students learned story grammar elements such as setting, initiating event, attempts, and consequences. However, another important aspect of narrative proficiency is the extent to which stories are aesthetically pleasing, or holistically judged to be “good stories.” Holistic scoring is looking at an entire story and judging its organization, use of interesting vocabulary, and so much more. Holistic scoring is used because it identifies holistic errors that may not be recognized when solely assessing for microstructure or macrostructure story elements.

McFadden and Gillam (1996) conducted one of the first studies to compare the holistic quality of narratives told by school-age students with language impairments to those of their typically developing peers. In their study, 40 school-age students between the ages of 9:0 and 11:7 (years:months), including 10 who had been diagnosed with a language disorder, produced two spoken ($n = 80$) and two written narratives ($n = 80$) elicited from picture stimuli. Participants were shown sets of three 7” × 10” color pictures containing pictures depicting a nature scene, an outdoor action scene, and a portrait. Using a set of 25 narratives, randomly selected from all groups, four categories were created, and anchor stories were selected for each

qualitative scores. A team of scorers, including a speech-language pathologist, two teachers, and a college English instructor, designed the rubric and rated the anchor stories.

The categories were numbered 1-4 and were described as follows. Score one stories (weak) consisted of descriptions and poorly organized, uncaptivating stories. Score two stories (adequate) consisted of stories that took one of four forms: (1) an event recount, without a central climax; (2) a bare-bones narrative, with no elaboration; (3) a narrative without an ending; or (4) a confusing narrative with strong descriptive segments. Score three stories (good) were captivating stories that contained problems and resolutions. These narratives may have had some organizational difficulties. Score four stories (strong) were easily understood narratives with a clear, integrated storyline, elaboration, interesting word choices, and some captivating features such as a climax, an ending twist, or a compelling personal voice.

The rubric was used to compare stories told by typical children and those with learning disorders. The results suggested that students with learning disorders (LD) were more likely to tell stories that were of poorer holistic quality than their typically developing peers. There were more holistic scores in the weak and adequate categories for students in the LD group (83%). In contrast, students in the AGE-M group showed the greatest concentration of scores in the good and strong categories (71%). These findings were interpreted to indicate that students in the LD group produced spoken and written stories that were judged to be significantly lower in overall quality than the scores produced by students in the age-matched group.

A more recent study, Newman and McGregor (2006) again questioned whether holistic ratings of narrative quality were effective in differentiating children with SLI from their typically developing (ND) peers. Twenty-seven laypersons and twenty-one teachers used an interval scale to rate the holistic quality of narratives produced by 20 5–7-year-olds. A total of 10 students

were diagnosed as having Specific Language Impairment (SLI) and 10 were designated as Normally Developing peers (ND). The 7-point scale was used to rate all of the samples: 1 represented a very low quality story and 7 represented a very high quality story. Although information on what each rating (1-7) meant was not available from Newman and McGregor, the scorers looked closely at the amount of information conveyed, the ease of understanding of the story, and finally, the ease of presentation on the part of the speaker. The rating procedure lasted 30–45 min, depending on the length of the narrative. Laypersons and teachers judged the narration of children with SLI to be significantly poorer than the narration of normally developing, age-matched peers. This study was used to help format the Holistic Assessment of Narrative Discourse (HAND) rubric and determine what holistic qualities were most important to measure.

A holistic scoring scheme was developed by Canfield, Eigsti, de Marchena, and Fein (2016) to examine the quality of stories told by adolescents with Autism Spectrum Disorder (ASD), called a “story goodness” system. This story goodness system was examined through two different studies. In each study, stories created by adolescents with ASD were elicited by showing six black-and white drawings which depicted a story about two monkeys. The stories were then rated by naïve readers. The first study had 20 adolescent participants with HFA and 16 with TD. The groups did not differ on chronological age, gender, or full-scale IQ. Narrative samples were transcribed and analyzed for the following distinct aspects of story goodness: “How good a story is this?” (i.e., goodness); “How well were you able to follow this story?” (i.e., cohesiveness); “How well does this story reflect the actual story in the cards?” (i.e., accuracy); and “How odd/unusual did you find this story?” (i.e., oddness). Ratings were provided using a 1–5 Likert scale (e.g., for goodness, a score of 1 corresponded to a rating of not good and a score of

5 corresponded to a score of very good). A composite measure of story quality showed a clear group difference, suggesting a deficit in pragmatic language abilities in this sample of youth with HFA.

Canfield et al.'s (2016) second study compared narratives from adolescents across three groups: ASD, TD, and youths with "optimal outcomes," who were diagnosed with ASD early in development but no longer met criteria. Fifteen children and adolescents were selected from a larger such that groups did not differ on age or IQ. Participants from Study 1 were not included in Study 2, creating a distinct sample. Participants completed the same experimental task as in Study 1, with one methodological difference: In Study 1, the picture cards were removed from view before the participants told the story, whereas in Study 2 they remained in view while students created their stories. The ASD group's narratives had lower composite quality scores compared with peers with typical development. Group differences were detectable by naïve raters from reading transcribed narrations, without the influence of prosodic or other personal characteristics that might affect perceptions of narrative quality. Results indicated subtle differences in pragmatic language skills for individuals with optimal outcomes despite otherwise typical language skills. This findings assisted in determining that a naïve rater could use a holistic or story goodness rubric rating system.

The rubrics that have been utilized up to this point have been useful in highlighting holistic quality differences between school-age and adolescents with language disorders, including those with ASD and their typically developing peers. They may also be useful for monitoring progress in narrative intervention as aspects of holistic quality may be addressed along with aspects of story grammar and microstructure. None of the rubrics that have been studied have been examined for psychometric properties which an important aspect of a progress

monitoring tool. Specifically, a progress-monitoring tool should be valid, developmentally sensitive and yield reliable scores (American Institutes for Research, 2015ab). The purpose of the current project was to draw upon past research to design a valid, sensitive and reliable rubric for use in monitoring progress as a result of narrative intervention. There are many rubrics available that measure macrostructural and microstructural elements, but not for measuring overall holistic quality. This rubric was specifically created to look at a child's narrative in terms of overall quality.

This rubric drew upon different aspects of the rubrics reviewed. From McFadden and Gillam (1996), categories of stories from strong to weak were included in the HAND rubric. However, in the HAND rubric, a wider variety of scores were created so that all stories would fit into a specific score instead of receiving a score between two (for example: 2.5, 3.5). From Newman and McGregor (2006), specific holistic qualities were measured that were important to be included in the HAND rubric, including the amount of information conveyed, the ease of understanding of the story, and the ease of presentation on the part of the speaker. From Canfield et al. (2016), it was identified that even a naive rater could determine if a story was holistically good or not. The HAND rubric adapted these rubrics to be useful for naiver raters to use as well as more skilled raters.

The psychometric properties examined in the current study included validity, developmental sensitivity, and inter-rater reliability. Validity is evidence that scores from an instrument measure what the instrument is intended to measure (Briesch et al., 2007; Lueger & Barkham, 2010; Overington & Ionita, 2012). If one wished to measure narrative discourse ability, the tool that is used should be correlated with a measure known to reflect narrative proficiency, such as the *Test of Narrative Language-2* (TNL-2; Gillam & Pearson, 2017). The

HAND scores were compared to the scores obtained on the TNL-2 in order to determine validity.

A progress monitoring tool should be sensitive to differences in skill, such as one might expect when children gain experience over time. Therefore, if a measure of holistic quality is developmentally sensitive, 7 year olds should produce more holistically adequate stories than 4 year-olds. The scores on the HAND were obtained for different age levels of children (4-7) to determine if it was associated with increases in scores over age.

An important measure of reliability is inter-rater reliability. Inter-rater reliability refers to the degree to which different raters reach the same conclusions in scoring. For formal tests, reliability coefficients should be equal to or above .80 to be considered minimally reliable. Coefficients of .90 or above are considered ideal (Aiken & Groth-Marnat, 2006; Salvia, Ysseldyke, & Bolt, 2013). The inter-rater reliability scores for the HAND were calculated in the current study.

The following research questions were posed:

1. Does the HAND correlate with a measure of oral narrative proficiency? It was hypothesized that the HAND scores would correlate with the oral narrative raw scores (ON) and/or the TNLAI (total narrative language ability index) on the TNL-2 if it was a valid measure of narrative skill.
2. Does the HAND demonstrate developmental sensitivity? We hypothesized that if the HAND was sensitive to developmental changes over time, older children would earn significantly higher scores than younger children.
3. Does the HAND demonstrate minimally sufficient inter-rater reliability? We hypothesized that if the HAND was a reliable tool that it could be scored with reliability levels at or above .80.

Method

Participants

The narratives collected from 27 4-year-olds, 50 5-year-olds, 59 6-year-olds, and 88 7-year-olds who were part of the normative sample for the *Test of Narrative Language-2* (TNL-2; Gillam & Pearson, 2017) were used in this study. There were 117 females and 107 males for 224 total participants. There are three subtests on the TNL-2 that require children to retell or create stories. The first is a retell, the second is a set of sequenced pictures and the third is a single scene. The stories elicited from the single scene picture prompt were used for the holistic scoring. To elicit this story, the examiner showed the child a picture of two children who witnessed a family of aliens walking out of a spaceship that had landed in a park and asked him or her to tell a story about it.

All of the stories from the children were transcribed using Systematic Analysis of Language Transcripts (SALT; Miller & Iglesias, 2012). After transcription, the stories were coded for holistic quality using the Holistic Assessment of Narrative Discourse (HAND). The HAND is described below.

Holistic Assessment of Narrative Discourse

Scores on the HAND ranged from 0 to 5. A score of 0 was given when a child did not tell a story or responded that he or she did not know how to tell a story. A score of 1 was given when the story was judged to be weak. The following story is an example of a story that was determined to be weak (rambling, etc):

He left in that ship with their ship. They walked out. And the girl started running to them. They had a dog. And I don't know. He waved his hand in their ship.

A score of 2 was given when a story was determined to be limited or incomplete because of its lack of causal relationships and key story events, such as a beginning. The following story is an example of a story that earned a score of 2:

There's an alien spaceship with a family in it. And they go down on the spaceship. Then a boy and a girl see the alien. And the aliens come out of the spaceship. And they're going to have a picnic. And the girl and boy were going to go um talk to the aliens. And um one of them said no. Because I don't want to go talk to the aliens. And so um when they when they, when they were all off the spaceship they unloaded all their food. And then they put it on the picnic table. And then that's all.

This story is missing the key element of a good ending and is poorly organized in many areas. A score of 3 was given to signify that a story was judged to be mediocre but complete. For example, one child told the following score three story:

One Saturday morning the kids just woke up. And they hurried and got dressed and ran outside. And they found a park. And they saw a thing landing. And there was three octopuses coming out. And one dog. And they're gonna play in the park. And and one of the kids wanted to go and the other kid ran. He said no don't do that. And then and then and then the one a the girl she took his hand and dragged him out. And then they made friends with the kid octopus. The end.

This story scored a score of three because it contained a complete episode, however, was confusing at times. A score of 4 indicated that a story was good story but suffered from some organizational problems. For example the following excerpts are examples of good stories.

The dad and son went to the park. Aliens landed. The son was like cool. I want to go over

there. And the dad was like no. Um the aliens looked around and like here's our vacation. And they were. They went to the park and looked around. The dad and the son had to go home. The boy was sad. And they told the mom. And the dad said that too. And they went back to the park. And they were gone. And the dad and the son felt bad. They went back home. And then that's all I got.

This story that scored a four had many good parts to make it into this scoring, for example: contained a complete episode, a few examples of elaboration, and some interesting word choices. Finally, a score of 5 represented a strong, imaginative, clear story. For example this child tells a strong story about aliens, when they said:

Billy Mandy and Zarnock were looking for a new place to live. They were in the planet Zonnapooa. And they really did not like their home. They wanted to go. They got into their spaceship. They started traveling the galaxy. First they found a planet called Jupiter. You could just fall right through it. Then they found a planet called Monacooacoo. It had many rings. They tried to land there. But it was all made of rocks! And they got their ship damaged. They went and traveled through many more planets. They really couldn't find a great home. Then one came who was called Earth. They had went down but the atmosphere had stopped them. They used their laser gun to shoot through it. They went right through it. A magnetic field had stopped them. But they just used solar power for their ship. When they got to the surface they landed in a park. They saw many people there. And they thought wow this must be a great place! Many aliens have moved here also! So they got out of their spaceship. And people started to run like crazy. They didn't know why! But eventually they made friends with a girl named Emmy and her brother Luke. They went to their house. They thought wow this is a nice home! I wish we had

some on our planet! It would make it much nicer they said to Emma and Luke. Emma and Luke's parents weren't very happy with this plan. They tried to make the alien/s go away! But the aliens just shot them with their laser eyes. The alien/s also bought a dog there. They named it Newacoo. After a while humans had come and tried to destroy them. But they fought them off with their laser eyes. Then later on this planet was called Emma and Lukeness. Because Emma and Luke were the first friends that they had ever had. They had destroyed all of the other humans. But now there were just animals. Now they could roam free and have the best live/s ever. But soon they would die. So they tried to start a new civilization. It is now called Earth again. Now it is the present day. The aliens and Emma and Luke are now gone. But we have great technology now. It is very great.

The end.

This story received a score of five because of its many examples of elaboration, no unreferenced pronouns, and clear causal framework. See Appendix A for further details and operational definitions for each holistic score.

Inter-rater reliability

Five research assistants who were trained in the use of the HAND independently scored all of the stories produced by participants. The assistants had previously participated in two hour-long trainings to learn how to use the HAND rubric using stories not included in the current study. During preliminary training coders were asked to score 10 stories and to ask clarifying questions. The author discussed scoring scenarios with them and answered their questions about scoring the stories according to the rubric. Adjustments were made to the rubric to clarify scoring, and the coders were asked to score 10 more stories. The coders were cleared to begin

scoring stories for this project after they had attained 90% or higher inter-rater reliability with the first author on ten consecutive stories.

The coders were asked to score 10 stories and then meet to calculate their levels of agreement. Discrepancies were resolved through consensus and confirmed by the first author. Then, coders were instructed to score approximately 30 additional stories and to meet again to calculate their agreement scores. This procedure of coding approximately 30 stories, meeting to resolve discrepancies, and oversight by the author was incorporated to control for coder drift (Gillam, Gillam, Fargo, Olszewski, & Segura, 2017).

Coder drift is a phenomenon in which reliability decreases over time due to a lack of calibration. The author kept a record of all discrepant scores at each coder drift prevention meeting. After all coding was completed, the author calculated the overall reliability for all stories by subtracting the number of discrepant stories from the total number of stories to obtain the total number of stories without discrepancies. She then divided the total number of stories without discrepancies by the total number of stories and multiplied the result by 100 to achieve the overall percent reliable. The final overall reliability prior to consensus was 91%. The importance of this score for inter-rater reliability is discussed in the following results section.

Results

Table 1 shows the means and standard deviations for the HAND scores, total raw scores and the oral narration sub-scale on the TNL-2 by age group. The first research question asked whether or not the HAND correlated with a measure of oral narrative proficiency? It was hypothesized that the HAND scores would correlate with the oral narrative raw score and/or the Narrative Language Ability Index on the TNL-2 if it was a valid measure of narrative skill. Pearson Product Moment Correlation coefficients were computed among the holistic scores and

the TNLAI (total score) and ON (oral narration) raw scores (See Table 2). The results of the correlational analyses showed that the scores on the HAND were strongly correlated with the Oral Narration score $.66$ ($p = .001$) and to the TNLAI raw scores on the TNL-2, $.61$ ($p = .001$; Hopkins, 2002). This strong correlation suggested that the HAND was a valid measure of holistic quality of a narrative composed from a single scene picture.

The second research question asked whether the HAND demonstrated developmental sensitivity. We hypothesized that if the HAND was sensitive to developmental changes over time, older children would earn significantly higher scores than younger children. A one-way, multivariate analysis of variance (MANOVA) was conducted to determine the effect of age on the measure of narrative quality (HAND). Significant differences were found among the age groups.

Post hoc analyses revealed that the seven year olds created stories that earned higher holistic scores than stories told by four and five year olds. Six year olds told stories that earned higher scores than four and five year olds. Five year olds told stories that earned higher HAND scores than four year olds. There was no significant difference between scores earned by six and seven year olds ($p = .067$).

The final research question asked whether the HAND demonstrated minimally sufficient inter-rater reliability? We hypothesized that if the HAND was a reliable tool that it could be scored with reliability levels at or above $.80$. The final overall reliability prior to consensus was 91%. This high level of inter-rater reliability indicates that a trained scorer could use the HAND to achieve a strong reliability throughout their caseload or group of students.

Discussion

Narration is an important type of classroom discourse that has important implications for how well students perform in school and interact with others (Fujiki, Spackan, Brinton & Illig, 2008; Miller & Pennycuff, 2008). For many children, particularly those with language disorders, insufficient narrative skills places them at a disadvantage both socially and educationally. Therefore, a number of recent studies have been conducted to determine the efficacy of narrative-based interventions for children with language disorders (Nicolopoulou & Trapp, 2018). All of the studies to date have examined how well students learned story grammar elements. However, another important aspect of narrative proficiency is the extent to which stories are aesthetically pleasing, or holistically judged to be “good stories.”

We wanted to know whether the HAND was a valid measure of narrative ability. If it is, we would see a strong, consistent relationship between the scores from the HAND and the scores from a well-established measurement procedure, specifically the TNL-2. Our findings showed that the HAND was strongly correlated with the TNL-2, suggesting that it is a valid measure of narrative quality.

We hypothesized that if the HAND was sensitive to developmental changes over time, that older children would earn significantly higher scores than younger children. Our findings supported this hypothesis. In this study, the quality scores for students increased with age. The HAND scores for children who were 4 were lower than those of 5, 6 and 7 year olds. Similarly, the scores of 5 year olds were lower than those for 6 and 7 year olds. The children who were 6 and 7 in this study told stories that were judged to be of equal quality using the HAND. Interestingly, our 6 and 7 year olds told stories that earned scores that ranged from 2 which contained part of an episode but were limited, incomplete, containing few pronoun references, to 3 which included complete episodes but were mediocre and disorganized; while the 4 and 5

years olds stories were mostly descriptive sequences containing few temporal connections.

McFadden & Gillam, (1996) had a mean age of 10;7, therefore many of their scores are difficult to apply to our age ranges of 4 to 7. However, similar to their study, we saw that scores improved with age. This suggests that the HAND is developmentally sensitive for describing the quality of stories for children ages 4-7.

We hypothesized that if the HAND was a reliable tool, it could be scored with reliability levels at or above 85%. We were able to score the HAND with 91% reliability. The reliability of other studies similar to ours also looked for a high inter rater reliability. McFadden & Gillam, (1996) followed Myers procedure, which determined if narratives differed by one point (e.g., received a “1” by one scorer and a “2” by another), they were considered to fall between scores (better than “1” but not quite “2”) and were not considered to be disagreements. Narrative scores that differed by more than one point (e.g., received a “1” by one scorer and a “3” by another) were considered to represent true disagreements concerning quality. On this basis, 5 of 160, or 3%, of the corpus were rescored by the group leader, who adjusted one of the discrepant scores. However, in this current study, the scoring system was altered and all stories could not fall between categories and needed to fit into 1 single score. Only 9% of the stories did not receive the same score, point for point, and therefore, had to be re-scored. This created the high score of 91% for inter-rater reliability.

We believe that the inclusion of a valid, sensitive, and reliable measure of narrative quality is an important part of a progress monitoring plan for tracking growth in narrative proficiency. This is because stories can contain complete episodes and be “sufficient” in terms of structure but be disorganized, illogical and hard to follow. To illustrate, the following story, told by a student with a language impairment, contains a complete episode (i.e., initiating event,

attempt, and consequence); but is very disorganized and hard to follow:

There was aliens at the park. The aliens wanted to go to the park with their dog and play fetch. But when they got off the bus, they started walking to the park. And the boy couldn't play fetch. He had an owie on his leg. He couldn't play fetch. Then a girl and a boy and another boy went over. One of the boys thought it would be cool to play fetch with the alien dog. The other one did not. He tried to hold him back. But he went anyways. He played fetch with the dog. The bigger brother was wrong. It was fun to play fetch with the alien dog. That's all.

This story contains characters (aliens, boy, girl), a setting (park), and actions (playing fetch) however, there was no clear episode with characters motivated by goal directed actions, and it is very disorganized. Using the HAND rubric, this story might be scored as follows: There is an apparent initiating event which qualifies it as a score 2 story or higher. After reading the entire story, it is apparent that the story is poorly organized with missing key elements, including an action that moves the story along. The story is holistically limited and would therefore be scored a 2. All six scores are considered for each story until holistic factors are found that qualify it for a certain score; in this case, a score of 2. This score helps a reader know that the storyteller is missing key elements and may need help with organization and clarity of the story.

Clinical Implications

Overall, the HAND demonstrates strong proof of validity, developmental sensitivity and reliability as a tool to measure the holistic quality of narratives for children ages 4-7. The use of a holistic rubric, along with other measures of macrostructure (story structure) and microstructure (coordinating conjunctions, adverbs, elaborated noun phrases) may be a more

comprehensive way to characterize the stories created by students who are learning to compose coherent, logical, and complete narratives. Further research is needed to determine whether the rubric is useful for measuring holistic quality across broader age ranges (students 8 and older) and for children with language impairments.

Table 1

Means and Standard Deviations for the Holistic Assessment of Narrative Discourse (HAND) scores, total raw scores, and oral narration subscale on the measure of narrative proficiency (Test of Narrative Language-2) by age group

Age	HAND	Oral Narration Raw Scores	Narrative Language Index Raw Scores
4 (n=27)	1.19 (.557)	15.56 (8.84)	27.15 (14.13)
5 (n=50)	1.58 (.758)	23.46 (10.21)	42.46 (15.39)
6 (n=59)	1.98 (.799)	29.85 (10.11)	52.97 (16.63)
7 (n=88)	2.22 (.765)	38.18 (10.88)	65.61 (15.89)

Table 2

Pearson Product Moment Correlations between the raw scores on the measures of oral narrative proficiency and the Holistic Assessment of Narrative Discourse (HAND)

	Oral Narration Raw Score	Narrative Language Index Raw Scores (total score)	HAND scores
Oral Narration Raw Score	1	0.959**	0.656**
Narrative Language Index Raw Scores	0.959**	1	0.611**
HAND scores	0.656**	0.611**	1

**Correlation is significant at the 0.01 level (2-tailed).

Appendix A

Holistic Assessment of Narrative Discourse Rating Scale (HAND) Scoring Rubric

Score 0: Unrelated

- No response
 - Child responds with “I don’t know” or “I don’t remember”
 - Entirety of response is unrelated to topic/task
-

Score 1: Weak

- Conveys description instead of telling a narrative
 - Few to no temporally related sequences
 - Rambling
-

Score 2: Limited - Incomplete

- An event recount (may have a complete episode) with weak causal relationships
 - Lack of because and/or relationships to events
 - May be missing key elements (such as: action that moves the story along, consequence or ending), however, needs an initiating event
 - Most pronouns lack reference
 - Poorly organized, uncaptivating stories
-

Score 3: Mediocre - Boring or Disorganized

- Must have a complete episode: initiating event, action, consequence
 - Confusing narrative with strong, descriptive segments
 - Narrative with too much detail or bare-bones narrative with no elaboration
 - Contains a few unreferenced pronouns
-

Score 4: Good

- Must have a complete episode: initiating event, action, consequence
 - Captivating stories that contain initiating events that motivate a character to action and end with resolution
 - May have some organizational difficulties
 - Contain elaboration, some interesting word choices, and/or some unexpected events
 - Most pronouns have easily understood reference
-

Score 5: Strong

- Must have a complete episode: initiating event, action, consequence
- Contains elaboration, such as: intonation, attention getter, onomatopoeia, dialogue, exaggeration

- No unreferenced pronouns and temporal sequence is clear
- Clear causal framework and may contain some complex syntax

References

- Aiken, L. W., & Groth-Marnat, G. (2006). *Psychological testing and assessment* (12th ed.). Boston, MA: Pearson.
- American Institutes for Research (2015a). The essential components of RTI. www.rti4success.org. Accessed on 3/3/2015.
- American Institutes for Research (2015b). Technical review committees process. <http://www.rti4success.org/technical-review-committees-process>. Accessed on 3/3/2015.
- Canfield, A. R., Eigsti, I. M., de Marchena, A., & Fein, D. (2016). Story Goodness in Adolescents With Autism Spectrum Disorder (ASD) and in Optimal Outcomes From ASD. *Journal of Speech, Language, and Hearing Research*, 59(3), 533-545.
- Fujiki, M., Spackman, M. P., Brinton, B., & Illig, T. (2008). Ability of children with language impairment to understand emotion conveyed by prosody in a narrative passage. *International Journal of Language & Communication Disorders*, 43(3), 330-345.
- Gillam, R. B., & Pearson, N. A. (2017). *Test of narrative language--second edition*. Austin, Tex: Pro-Ed.
- Gillam, S. L., Gillam, R. B., Fargo, J. D., Olszewski, A., & Segura, H. (2017). Monitoring indicators of scholarly language: A progress-monitoring instrument for measuring narrative discourse skills. *Communication Disorders Quarterly*, 38(2), 96-106.
- Holbrook, S., Beck, T., Reische, D., Froerer, C., Mumford, S., & Gillam, S. (2018). *Holistic assessment of narrative discourse*. Manuscript in preparation.
- Hopkins, (2002). A scale of magnitudes for the effect statistics. In a new view of statistics. Retrieved from <http://www.sportsci.org/resource/stats/effectmag.html>.
- Liles, B. Z. (1985). Cohesion in the narratives of normal and language-disordered children.

- Journal of Speech, Language, and Hearing Research*, 28(1), 123-133.
- Liles, B. Z., Duffy, R. J., Merritt, D. D., & Purcell, S. L. (1995). Measurement of narrative discourse ability in children with language disorders. *Journal of Speech, Language, and Hearing Research*, 38(2), 415-425.
- McFadden, T. U., & Gillam, R. B. (1996). An examination of the quality of narratives produced by children with language disorders. *Language, Speech, and Hearing Services in Schools*, 27(1), 48-56.
- Miller, J. & Iglesias, A. (2012). Systematic Analysis of Language Transcripts (SALT), Research Version 2012 [Computer Software]. Middleton, WI: SALT Software, LLC.
- Miller, S., & Pennycuff, L. (2008). The power of story: Using storytelling to improve literacy learning. *Journal of Cross-Disciplinary Perspectives in Education*, 1(1), 36-43.
- Newman, R. M., & McGregor, K. K. (2006). Teachers and laypersons discern quality differences between narratives produced by children with or without SLI. *Journal of Speech, Language, and Hearing Research*, 49(5), 1022-1036.
- Nicolopoulou, A., & Trapp, S. (2018). 17 Narrative interventions for children with language disorders: A review of practices and findings. *Handbook of Communications Disorders: Theoretical, Empirical, and Applied Linguistic Perspectives*, 15.
- Petersen, D., & Spencer, T. (2011). Progress monitoring of language: The narrative language measures. Presentation given at The American Speech Language and Hearing Convention, San Diego, California.
- Reische, D., Froerer, C., Mumford, S., Beck, T., Holbrook, S., & Gillam, S. (February 2018). *Give us a HAND: Holistic narrative quality rating of stories told by typically developing*

children. Poster presented at the National Conference on Undergraduate Research, Edmond, OK.

Salvia, J., Ysseldyke, J. E., & Bolt, S. (2013). Assessment in special and inclusive education, (pp. 47-53). *Belmont, CA: Wadsworth.*