

Monitoring Indicators of Scholarly Language (MISL): A Progress-monitoring Instrument
for Measuring Narrative Discourse Skills

Sandra Laing Gillam, Ronald B. Gillam, Jamison D. Fargo
Utah State University

Abbie Olszewski
University of Nevada, Reno

Hugo Segura
Universidad de Talca

Sandra Laing Gillam, PhD.
Communicative Disorders and Deaf Education
Utah State University
Emma Eccles Jones Early Childhood Education and Research Center
2610 Old Main Hill
Utah State University
Logan, UT 84322
sandi.gillam@usu.edu

Ronald B. Gillam, PhD.
Communicative Disorders and Deaf Education
Utah State University
Emma Eccles Jones Early Childhood Education and Research Center
2610 Old Main Hill
Utah State University
Logan, UT 84322
Ron.gillam@usu.edu

Jamison D. Fargo, PhD.
Department of Psychology
Utah State University
Emma Eccles Jones Education
2800 Old Main Hill
Logan, UT 84322

Abbie Olszewski, PhD.
Department of Speech Pathology and Audiology
University of Nevada, Reno
1664 North Virginia Street
Nell J. Redfield Building
Reno, NV 89557

Hugo Segura, MsC-SLP
Carrera de Fonoaudiología
Universidad de Chile sede Talca
5 poniente #1670
Codigo postal: 3460000
Talca, Chile

Author Note

This research was supported in part by a grant from the Institute for Educational Sciences, National Center for Special Education Research (Award Number R324A100063). The SKILL program can be ordered at <https://usuworks.usu.edu> and includes progress monitoring tools and video examples to support implementation. The authors would like to thank Allison Hancock, Natalie Nelson, Julise Nelson, Sara Hegsted, Sara Hicken, Katie Squires, Shannon Davenport, and all of the undergraduate and graduate research assistants who administered tests and analyzed language samples. A special thank you to our former doctoral student, Doug Petersen, who contributed to an earlier version of this instrument.

Abstract

Purpose: The purpose of this study was to assess the basic psychometric properties of a progress-monitoring tool designed to measure narrative discourse skills in school-age children with language impairments (LI).

Method: A sample of 109 children with LI between the ages of 5;7 and 9;9 (years;months) completed the *Test of Narrative Language*. The stories told in response to the alien's picture prompt were transcribed and scored according to the TNL manual's criteria and the criteria established for scoring the progress-monitoring tool, *Monitoring Indicators of Scholarly Language* (MISL).

Results: The MISL total score demonstrated acceptable levels of internal consistency reliability, inter-rater reliability and construct validity for use as a progress-monitoring tool for specific aspects of narrative proficiency.

Conclusions: The MISL holds promise as a tool for tracking growth in overall narrative language proficiency that may be taught as part of an intervention program to support the Common Core State Standards related to literacy.

SLPs are increasingly being called upon to provide evidence that their intervention efforts result in positive educational outcomes for students in school-based settings (American Speech Language and Hearing Association; ASHA, 2000). This involves the provision of educationally relevant instruction and authentic documentation of student outcomes through a process called progress-monitoring (Gillam & Gillam 2006; Gillam & Justice, 2010). The information obtained through progress-monitoring is used to inform clinical decisions about methods and procedures, dosage, service-delivery and to communicate accurate and consistent information about a child's progress to others (Paul & Hasselkus, 2004; Sutherland Cornett, 2006; Warren, Fey, & Yoder, 2007). Ideally, these tools should possess some basic psychometric properties such as inter-rater reliability, internal consistency reliability and construct validity if SLPs are to have some degree of confidence in their ability to capture differences in performance as a result of intervention (American Institutes for Research, 2015a).

One of the roles and responsibilities of speech-language-pathologists (SLPs) employed in educational settings is to design and implement intervention programs that target the language underpinnings that are foundational to curricular content related to literacy development. Then, they should monitor how well students' respond to the instruction (ASHA, 2001; Ehren & Whitmire, 2009). According to Common Core State Standards (CCSS-ELA.Literacy.W.3.3), school-age children must be able to "compose narratives to develop real or imagined experiences or events using effective technique, well chosen details, and well-structured event sequences" (CCSS; National Governors Association and Council of Chief State School Officers, 2011). Component language skills that may be taught in support of this over-arching discourse-level goal may include teaching students to "ask and answer questions about key details in text (CCSS.ELA-Literacy.RL.1.1)", "retell stories including key details (CCSS.ELA-

Literacy.RL.1.2)”, and to “describe the overall structure of a story, including how the beginning introduces the story and the ending concludes the action (CCSS.ELA-Literacy.RL.2.5).” The authors designed a progress-monitoring tool to measure growth in the ability to generate fictional stories consistent with standards outlined in the Common Core State Standards (CCSS, 2010). A brief list of the reading and writing anchor standards that define what students should understand and be able to accomplish by the end of grade 3 that are directly measured on the progress-monitoring tool described in this paper (Monitoring Indicators of Scholarly Language; MISL, Gillam, S., Gillam, R., & Laing, C., 2012) is provided in the supplemental materials (supplemental materials content A). The purpose of this study was to evaluate the psychometric properties of a progress-monitoring tool called, *Monitoring Indicators of Scholarly Language* (MISL, Gillam, S., Gillam, R., & Laing, C., 2012).

Measuring Key Components of Narrative Discourse

In addition to measuring skills that are related to the Common Core, a narrative progress-monitoring tool should contain items that are consistent with models of narration. Narratives are generally characterized according to macrostructure and microstructure components. Macrostructure is usually defined as a setting plus one or more episodes (Stein, 1988; Stein & Glenn, 1979). A *setting* is a reference to the time or place that the story occurred. Children may use fairly simple setting references, such as “outside” or “in the rain,” or more specific, sophisticated setting elements such as “Central Park” or “Washington, D.C.” A basic episode consists of an *initiating event* (IE), which is an incident that motivates actions by the main character(s) goal directed actions known as *attempts*, and a *consequence* (or outcome) that is related to both the initiating event and the actions. By eight years of age, typically-developing children tell complex narratives that contain complicating actions (occurrences that interfere with

the goal directed actions of characters) and/or multiple IEs with associated actions and consequences (Berman, 1988). For story coherence, it is important that the temporal and causal relationships between the IE, character actions related to the IE, and the consequences of those actions are clear to the listener. In fact, the amount of information one can retrieve for use in answering questions and composing retells is related to the number of causal relationships contained in a story (van den Broek, Linzie, Fletcher, & Marsolek, 2000; White, van den Broek, & Kendeou, 2007).

Narrative *microstructure* consists of the words and sentences that comprise a story. A critical part of narrative development during the school age years relates to the increased use of literate or scholarly microstructure forms, sometimes referred to as literate language structures (Greenhalgh & Strong, 2001; Paul, 1995; Westby, 1985). Important aspects of literate language include coordinating and subordinating conjunctions (*for, and, nor, but, or, yet, so*), adverbs (*suddenly, again, now*), and elaborated noun phrases (*the big green monster*). Other literate language features include metacognitive verbs such as *think, believe, and decide* that refer to acts of thinking or feeling, and metalinguistic verbs such as *tell, yell, and argue* that refer to acts of speaking (Westby, 2005).

Measures of microstructure summarize relevant aspects of linguistic proficiency and have been used to differentiate between typically developing children and children with delayed or impaired language abilities (Justice, 2006; Liles et al., 1995). Conjunctions, adverbs, elaborated noun phrases, metacognitive and metalinguistic verbs appear less frequently in the narratives of children with language impairments than their typically developing peers (Greenhalgh & Strong, 2001). A progress-monitoring tool known as the Index of Narrative Microstructure, (INMIS; Justice, Bowles, Kadaravek, Ukrainetz, Eisenberg & Gillam, 2006) was designed to assess

narrative microstructure in children ages 5-12. The measure yields information about language productivity (word output, lexical diversity, T-unit output) and complexity (syntactic organization). Scores on two factors (productivity and complexity) may be compared against field test reference data based on age or grade level.

Some narrative measures have been developed to examine aspects of both macrostructural and microstructural aspects of narratives produced by school-age children (Heilmann, Miller, Nockerts, & Dunaway, 2010). For example, the Narrative Scoring Scheme (NSS, Heilmann, Miller, Nockerts, & Dunaway, 2010) incorporates a Likert scale scoring approach for coding story elements related to introduction (setting, characters), character development (main character, supporting characters, first person), mental states (feelings), referencing (unambiguous pronouns), conflict resolution (clearly stated), cohesion (logical order, smooth transitions), and conclusion (story has clear ending). Story elements are coded as proficient (score of 5), emerging (score of 3), or minimal/immature (score of 1). Normative databases using the NSS to score self-generated stories and retells generated from wordless picture books are included in the Systematic Analysis of Language Transcripts manual (Miller, Andriacchi & Nockerts, 2011).

The *Index of Narrative Complexity* (INC) was also developed for measuring macrostructure and microstructural elements of narration in school age children (Petersen, Gillam & Gillam, 2008). The INC contains scales to measure macrostructure components (character, setting, initiating event, internal response, plan, attempt, consequence) and microstructure features (coordinated & subordinated conjunctions, adverbs, metacognitive and metalinguistic verbs, and elaborated noun phrases) of self-generated stories and retells. We revised the INC into a measure called, *Monitoring Indicators of Scholarly Language (MISL)*, which was designed to track the range of progress from the production of simple descriptions produced by very young children to

more sophisticated multi-episode narratives produced by children in the upper elementary grades (The MISL rubric is available as supplemental Material B).

The MISL is primarily used for assessing self-generated narratives elicited in response to sequenced pictures and single scene prompts, but it has also been used to track progress in story retelling. In the next sections, we describe the psychometric properties that we report for the MISL including estimates of reliability and construct validity.

Characteristics of psychometrically sound progress-monitoring tools

A progress-monitoring tool should yield reliable scores for measuring the component skills that correspond to success in a particular domain (American Institutes for Research, 2015a). According to The National Center on Intensive Intervention technical review committee, progress-monitoring tools should contain estimates of reliability and construct validity (American Institutes for Research, 2015b).

Reliability estimates for performance level scores may include internal consistency reliability and inter-rater reliability. Internal consistency reliability refers to the extent to which responses to the items on a scale correlate with one another. Typically, internal consistency reliability is measured using a statistic called Cronbach's alpha. Inter-rater reliability refers to the degree to which different raters reach the same conclusions in scoring. In order to demonstrate minimum reliability, reliability coefficients should be equal to or greater than .70 (Nunnally & Bernstein, 1994).

In addition to being reliable, progress-monitoring tools should be valid (Briesch et al., 2007; Lueger & Barkham, 2010; Overington & Ionita, 2012). One measure of validity is construct validity, which is an accumulation of evidence indicating that scores from an instrument measure what the instrument is intended to measure. A confirmatory factor analysis

(CFA) may be conducted to establish this construct. In CFA, examiners create factor structures that test whether hypotheses made about the measure correspond to a theoretical notion. For example, if a clinician wished to measure narrative discourse skills, the tool should be composed of items known to reflect knowledge of narrative macrostructure and microstructure. The purpose of this study was to assess the inter-rater reliability, the internal consistency reliability, and the construct validity of the MISL. Our research questions were:

1. To what extent do two raters who score narratives independently agree on the values that are assigned to the MISL items (inter-rater reliability)?
2. To what extent do the items on the MISL correlate with each other (internal consistency reliability)?”
3. Are there two multiple dimensions (macrostructure and microstructure) underlying the items on the MISL (construct validity)?

Method

The participants were 109 children (69 males and 40 females) with identified language impairments (LI) between the ages of 5;7 and 9;9 (years; months). These participants were recruited as part of a series of studies to examine the outcomes of language and narrative instructional approaches. Consistent with the EpiSLI model (Tomblin et al, 1997), children were determined to have a language impairment if they displayed standard scores at or below 81 on two or more composite scores from the *Test of Language Development: Primary: 3rd edition* (TOLD:P:3; Newcomer & Hammill, 1997) or a composite score below 82 on the *Comprehensive Evaluation of Language Fundamentals-4* (CELF-4; Semel, Wiig & Secord, 2004) or the *Comprehensive Assessment of Spoken Language* (CASL; Carrow-Woolfolk, 1999). None of the participants presented with hearing, visual, or gross neurological impairments, oral-structural

anomalies, or emotional/social disorders, but they all demonstrated average to above average nonverbal reasoning skills as measured by the *Brief Kaufmann Intelligence Test* (K-BIT-2: Kaufman & Kaufman, 1990) or the *Universal Nonverbal Intelligence Test* (UNIT: Bracken & McCallum, 1998). Ninety-two of the children were from Texas, and 17 were from Utah. Their demographic characteristics are shown in Table 1.

Procedures

Trained research assistants or certified speech language pathologists administered *The Test of Narrative Language (TNL)* to all of the participants before their respective intervention programs began (pre-test). All of the assistants were graduate students in speech language pathology programs under the direct supervision of certified SLPs. Training was provided by the first and second authors to all of the research team involved in conducting these assessments. The TNL is a standardized test designed to assess narrative comprehension and production in children between the ages of 5 and 12. The TNL utilizes three successively more difficult contexts to assess narrative production proficiency. The first context is a scripted narrative. Children were asked to answer questions about the story and to retell it. In the second context children listened to a story that corresponded to a series of 5 sequenced pictures. They answered questions about the story they heard and then generated their own story that corresponded to a novel set of 5 sequenced pictures. The prompts for the third narrative context were single scene pictures depicting fictional events. Children listened to a story about a dragon guarding a treasure and answered questions about it. Then, children were asked to generate a story that corresponded to a novel scene depicting an alien family landing in a park. The TNL yields an overall narrative language ability index (NLAI) as well as composite scores for narrative comprehension (NC) and oral narration (ON). MISL scoring was conducted on the narratives

generated while children looked at the novel scene depicting an alien family landing in the park.

Transcription

The stories told in response to the alien picture prompt were digitally recorded and transcribed according to Systematic Analysis of Language Transcription (SALT) conventions (Miller & Chapman 2004). Narratives were transcribed verbatim with the inclusion of both child and examiner utterances when applicable. Two research assistants who did not administer the TNL and who were unaware of the purpose of the research project segmented transcripts into communication units (C-units; Loban, 1976) that consisted of an independent main clause and any phrases or clause(s) subordinated to it. Utterances were also coded for the presence of mazes (reformulations, reduplications, and false starts). Accuracy of the transcription and coding process was reviewed by examining 30% of the written transcripts. Percentage of agreement between primary and secondary transcribers/coders was 98% for C-unit segmentation and 95% for mazes.

MISL Description and Scoring Procedures

The MISL has a macrostructure subscale and a microstructure subscale whose scores are combined to reflect an overall narrative proficiency score (total MISL score). The macrostructure subscale consists of 7 story elements (character, setting, initiating event, internal response, plan, action and consequence). Definitions for these story elements and examples for each are provided in Table 2. Scores of 0 are interpreted as evidence that a story does not contain elements that constitute a basic episode. Accounts that earn scores of 0 may contain simple descriptions of objects or actions (*There is a tree. They are running*). Scores of 1 indicate that a story has an emerging episodic structure (*There is a boy. He's at the table eating.*) Scores of 2 are taken as evidence that a story contains the necessary elements to constitute a basic episode

(The boy is eating breakfast and then he is going to school. He likes school, so he is hurrying to finish. He ran to school after breakfast.); and scores of 3 indicate that a story is complex and elaborated *(John and Bill are brothers. They are hurrying to eat breakfast before school. They love going to State Middle School so they are hurrying. All of a sudden, they knocked their cereal bowls over and milk went everywhere. They decided to clean it all up and grab breakfast bars instead. They ate their breakfast bars as they ran to school. They got there just before the bell rang. They were glad they'd gotten to eat breakfast and that they'd made it to school on time).*

The scoring system for character and setting is similar such that items related to the use of *character* earns a score of 0 if no reference to a character is made; a score of 1 if an ambiguous reference is stated *(the boy, in the park)*; a score of 2 if a specific name is used *(Mark, Central Park)*; and a score of 3 if two or more specific references are indicated in the story *(Mark and Mary; Central Park and California)*. Therefore, *Mary and Mark walked through Central Park in California* would receive a score of 3 for character and 3 for setting. Recall that scores of 3 are interpreted as evidence that a story is complex and elaborated. The scoring procedures for initiating event, internal response, plan, action and consequence is based on whether there is clear evidence that the elements are causally linked and is anchored at a score of 2. (See supplemental materials C for more detail regarding macrostructure scoring).

There are seven items on the microstructure scale: five items that relate to literate language, a grammaticality item, and a tense item. Nippold (1998) used the term *literate lexicon* to refer to words that are “important for the literate activities of reading, writing, listening to lectures, talking about language and thought, and mastering school curriculum” (p. 21). More recently, Paul (2007) wrote that literate language is “the style used in written communication and

is typically more complex and less related to the physical context than the language of ordinary conversation” (p. 394). There are five specific linguistic forms that are identified as literate language on the microstructure subscale of the MISL (Benson, 2009): coordinating conjunctions (*for, and, nor, but, or, yet, so*), subordinating conjunctions (*so, that, because*), adverbs (*quickly, slowly, fast*), metacognitive verbs (*thought, planned, decided, said, yelled*), and elaborated noun phrases (*the girl, the happy girl, the sweet happy girl*). The grammaticality item relates to grammatical errors such as improper use of pronouns, lack of subject-verb agreement, or tense and inflection errors. For example, the utterance, *Her went home*, would be judged as ungrammatical because it contains a pronoun use error. The tense item assesses whether sentences produced in students stories contain changes from present to past or future tense or reflect consistent use of one tense. For example, *Yesterday, she **walked** home. She **runs** all the way there. She **will walk** home yesterday*, would be scored as two tense changes. Stories that contained three or more grammatical or tense errors earned scores of 0 in each category. A score of 3 was given for each item if the story contained no grammatical errors or tense changes. Table 3 contains the literate language structures, definitions, and scoring criteria for the microstructure subscale items.

Inter-rater reliability

Two research assistants (coders) who were trained in the use of the MISL and blind to group assignment and the purpose of the study independently scored all of the stories produced by participants. The coders had previously participated in an hour-long training to learn how to use the MISL rubric to score macrostructure and microstructure for stories not included in this study. During preliminary training coders were asked to score four or five stories with the first author, and to ask clarifying questions. The first author discussed scoring scenarios with them

and answered their questions about scoring the stories according to the rubric. 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 five consecutive stories.

The procedure for scoring the stories used in this project was as follows: the coders were asked to score 10 stories that were selected randomly from the total corpus of transcripts and then meet to calculate their levels of agreement. Care was taken to select stories from children at each age-level (5-6 year olds, 7-8 year olds, 9-10 year olds). Discrepancies were resolved through consensus and confirmed by the first author who made the final decision on scoring. Then, coders were instructed to score 10 additional stories and to meet again to calculate their agreement scores. This procedure of coding 10 stories, meeting to resolve discrepancies, and oversight by the first author was incorporated to control for coder drift (Gillam, Olszewski, Fargo, & Gillam, 2014). Coder drift is a phenomenon in which reliability decreases over time due to a lack of calibration. Inter-rater reliability percentages were calculated for 20 stories (20%) that had been scored independently by the two raters. To obtain the percentages, the total number of items that the raters agreed on was divided by the total number of items in each subtests and for the total index, then multiplied by 100. The final inter-rater reliability percentages are presented in Table 4. Scores for inter-rater reliability are discussed in the following results section.

Results

Inter-rater Reliability. The first research question was, “To what extent do two raters who score narratives independently agree on the values that are assigned to the MISL items?” We wanted to know whether the MISL possessed reasonable inter-rater reliability to be useful in measuring narrative discourse skills. As can be seen in Table 4, the inter-rater reliability scores

for items and subscales ranged from 90% to 100%. For the macrostructure subscale, inter-rater reliability ranged from 92% (consequence) to 100% (character) and for microstructure, 90% (elaborated noun phrases) to 100% (coordinating conjunctions). The inter-rater reliability scores for each item, the total score, and the macrostructure and microstructure scores were 90% or higher, indicating acceptable levels of coder reliability. These data represent scores for students who range in age from 5;7 to 9;9.

Internal Consistency Reliability. The second research question was, “To what extent are the items on the MISL internally reliable?” Total, subscale, and item-level descriptive statistics for the MISL are presented in Table 5. Reliability coefficients at or greater than .70, were considered acceptable (Nunnally & Bernstein, 1994). Preliminary analyses suggested that the measure we used to calculate internal consistency (Cronbach’s alpha) for the MISL significantly improved with the removal of the Grammaticality and Tense items. The Cronbach’s alpha improved from .67 to .79 for the total instrument, and from .36 to .67 for the microstructure scale after removal of these two items. In summary, scores obtained from the MISL demonstrated acceptable levels of internal consistency reliability for the total instrument ($\alpha = .79$) and the Macrostructure subscale ($\alpha = .71$), but were slightly lower for the Microstructure subscale ($\alpha = .67$).

Construct Validity. The third research question was, “Are there two multiple dimensions (macrostructure and microstructure) underlying the items on the MISL?” Construct validity was evaluated by conducting a confirmatory factor analysis (CFA) that assessed the extent to which items within each subscale (i.e. Macrostructure or Microstructure) correlated, forming a construct or latent variable. The fit of the CFA was estimated by comparing the observed correlation structure to that obtained through model fitting. Model fitting involved determining

how well the proposed theoretical model (narrative macrostructure and microstructure) captured the covariance between all of the items in the model. If the correlations were low, the results of the CFA would indicate a poor fit, prompting the removal of items.

We conducted a full information CFA with a weighted least square parameter estimator (WLSMV) due to the presence of categorical data to assess the degree of fit between the item properties and the measurement model. Two latent variables (i.e. Macrostructure and Microstructure) were allowed to covary in this model. Latent variables were not directly observed, but were “inferred” from the variables that were directly observed (component items and subscales of the MISL). The following guidelines were used for identifying the characteristics of an “adequately fitting” CFA: composite reliability estimates $\geq .70$ for each latent variable (Fornell & Larcker, 1981; Hatcher, 1994, p. 339); a chi-square (χ^2) statistic to degrees of freedom (df) ratio ≤ 2 (Hatcher, 1994, p. 339); a Comparative Fit Index (CFI) and a Tucker Lewis Index (TLI) $\geq .95$ (Hu & Bentler, 1999); a Root Mean Square Error of Approximation ($RMSEA$) $\leq .06$ (Hu & Bentler, 1999); and a Weighted Root Mean Square Residual ($WRMR$) $\leq .90$ (Yu & Muthén, 2002). After removing items related to Grammaticality and Tense from the Microstructure subscale, the CFA measurement model consisting of two latent factors (Macrostructure and Microstructure subscales) demonstrated an overall model fit with $\chi^2 (df = 53) = 81.27, p = .008, \chi^2/df$ ratio = 1.53; Comparative Fit Index (CFI) = .99; Tucker Lewis Index (TLI) = .98; Root Mean Square Error of Approximation ($RMSEA$) = .06; and the average Weighted Root Mean Square Residual ($WRMR$) = .82.

Estimates of variance accounted for by each item (from the latent variable) in the form of R^2 (variance explained by the model), their standard errors, and p -values are presented in Table 5. A p -value of $< .05$ was judged to be significant. As shown in the Table, items measuring

setting, initiating event, attempt, consequence, coordinating conjunctions, and metacognitive/metalinguistic verbs were highly significant at $p = .01$, and items measuring character, subordinating conjunctions and elaborated noun phrases were moderately significant at $p = \leq .043$. Items that were not significant included internal response ($p = .336$) and plan ($p = .05$). Nonsignificance for internal response and plan reflects floor effects for these elements because the children who participated in this study rarely included them in their stories. Aside from a slightly larger RMSEA, results of the CFA measurement model indicated adequate model fit to support the construct validity of the MISL instrument.

Discussion

The purpose of this study was to assess the inter-rater reliability, internal consistency, and construct validity of the MISL. Our first question was, To what extent do two raters who score narratives independently agree on the values that are assigned to the MISL items (inter-rater reliability)? Our second question was, To what extent are the items on the MISL internally reliable as measured using Cronbach's alpha ($\geq .70$; internal consistency). Our final question was, Are there two multiple dimensions (macrostructure and microstructure) underlying the items on the MISL (construct validity)?

Inter-rater reliability. Recall that inter-rater reliability is the extent to which two raters agree on how to score individual items. This construct is important for a progress-monitoring tool because a determination of progress can only be trustworthy to the extent that another professional would have obtained the same scores. We found relatively high levels of inter-rater reliability (90-100%) across all of the items on the MISL rubric. One potential reason for this high degree of inter-rater reliability was the rigorous training and support the coders received as they learned to use the rubric. Recall that coders were asked to independently score four or five

stories on their own and then turn them in with any questions they had to the first author. The coders reported that as they became more familiar with the rubric, their independent scoring time decreased by as much as 50% depending on the length or complexity of the narrative. Rapid and accurate scoring was related to the amount of experience the coders had with the rubric, meetings among coders to discuss their scores (5-10 minutes per meeting), and group discussion of discrepancies.

School-based SLPs may not have the luxury of meeting with other trained professionals after they score every 5-10 narratives in order to refine and calibrate their skills. Nor will they be able to meet regularly with research staff to obtain final judgments on scoring discrepancies. Therefore, inter-rater reliability estimates among clinicians in authentic contexts may be somewhat lower than those reported here, at least initially. It is expected, however, that clinicians will increase their scoring proficiency and their scoring reliability as they become more familiar with the tool and how to use it to monitor narrative proficiency with their own students.

One additional consideration when using the MISL rubric in authentic, school-based settings is whether or not to orthographically transcribe narratives before attempting to score them. Recall that the stories in this study were orthographically transcribed before they were scored. School-based practitioners may not have the time and resources necessary to use this process to score every narrative obtained from students on their caseload. One way to reduce the amount of transcription that may be necessary for reliable scoring is to digitally record stories told by students and then take abbreviated notes while replaying them. These notes may be used during the scoring process. The use of audio-recordings to score narratives has been shown to have adequate inter-rater reliability using procedures outlined in the manual for the Test of

Narrative Language (TNL). However, more research is necessary to determine whether the MISL may be scored reliably using a similar method.

The most important way to achieve sufficient inter-rater reliability using the MISL rubric is to adhere to the operational definitions of the items included in the measure. The definitions contained in this paper, and the examples provided in the supplemental materials should assist clinicians in achieving sufficient inter-rater reliability to use the rubric for the purpose of progress-monitoring in school-based settings.

Internal consistency reliability. Internal consistency represents the homogeneity of the items that have been selected to measure a particular construct. The MISL rubric was intended to measure narrative proficiency (the construct of interest). Toward that end, the items that were included on the rubric were selected because they have been shown to contribute to narrative skill. Initially, it was thought that grammar and tense may be important items to include in the measurement of narrative proficiency, however, the analysis suggested otherwise. The overall internal consistency reliability of the MISL was sufficient (Cronbach's $\alpha = .79$) only after the removal of the two items related grammatical acceptability and tense change. The data in this study suggest that grammar and tense, while important linguistic skills, may not be critical contributors to overall narrative competence.

Recall that the internal consistency of the macrostructure and microstructure subscales was minimally acceptable when measured independently, particularly the microstructure subscale ($\alpha = .67$). However, the total MISL score, with a Cronbach's α of .79, may be a more meaningful measurement of narrative proficiency than either scale used in isolation. For statistical reasons, we recommend that clinicians base global decisions about intervention progress on the total score as a reliable indicator of change rather than the macrostructure or

microstructure scores separately. That is not to suggest that clinicians should not utilize each of the individual subscale scores to monitor mastery of each of these important skills. We have used individual scores to make decisions about specific targets for intervention sessions and feel that this is a useful tool for planning.

Construct validity. In combination, the nature of the relationships among item scores on the MISL was consistent with the theory that narratives are comprised of macrostructure and microstructure components. The macrostructure items included on the MISL that are consistent with theory were setting, initiating event, internal response, plan, attempt and consequence (Stein and Glenn, 1978). The MISL also included an additional element, character, because many narrative intervention programs often include instruction on this component. Character was shown to load or be consistent with the other macrostructure items on the MISL.

The microstructure elements that were included on the MISL were coordinating and subordinating conjunctions, adverbs, elaborated noun phrases, metacognitive and metalinguistic verbs. The items related to grammaticality and tense were removed from the rubric because model fit statistics indicated that these items were inconsistent with the other microstructure items on the scale. For the type of coding that was used for the MISL, grammatical accuracy and consistency of tense did not correlate well with other aspects of macrostructure or microstructure. Clinicians who work on these aspects of language during intervention would want to use a means other than the MISL to monitor children's progress in these domains.

The data presented in this paper suggest that a progress-monitoring tool designed to measure narrative proficiency may not be improved by adding measures of grammaticality or tense change. Our findings of lower internal consistency when items measuring grammar and tense were included were very important findings relevant to clinical practice. It is possible that

grammaticality and tense are “distinct skills” that are separate from macrostructure and microstructure. If so, including grammar and tense in a narrative rubric may indicate a lack of progress in narrative skills when in fact progress is being made. If clinicians are targeting grammaticality in therapy, it may be important to acknowledge that fluctuations between grammaticality and narrative discourse proficiency may occur as students focus on learning difficult narrative discourse skills, although again, we do not provide data in this study to support this assertion. Tentative data suggests that grammaticality will improve after knowledge of narrative content has become more stable (Crotty & Gillam, 2016a; 2016b). Future research is needed to provide solid evidence for this hypothesis. What is important to note is that we are not saying to clinicians they should not work on tense and grammar. We are suggesting that these items may make a tool for measuring narrative proficiency less reliable in measuring macrostructure and microstructure relative to narrative production.

Limitations

The MISL was designed to measure changes in a very specific set of macrostructure and microstructure features that have been documented to contribute to narrative proficiency and that are aligned with Common Core Curricular Standards (Gillam et al., 2014). If a clinician is not teaching these aspects of narrative macrostructure and/or microstructure in their narrative instruction, the MISL may not be as useful in documenting progress. In addition, the pilot studies we have conducted with versions of this progress-monitoring tool have included fairly small numbers of participants. Therefore, findings related to reliability and validity using larger samples could yield different results from those reported here. Finally, we calculated the psychometric properties based on only one elicitation context (spontaneous generation) using a specific prompt from the TNL. It is possible that findings may differ using different elicitation

contexts (retelling) and prompts (sequenced pictures, story books). Future research may investigate the ways in which the MISL might be modified for use with various other elicitation contexts including sequenced scene pictures and retells.

Summary and Clinical Implications

The purpose of progress-monitoring tools like the MISL is to provide clinicians with information that can inform clinical decisions about the nature of narrative intervention needed to support children's ability to meet the language demands of the classroom curriculum. Valid and reliable outcome measures are crucial for progress-monitoring tools to be useful in driving systematic, data-based decisions about language instruction. Progress-monitoring for narrative discourse poses a unique challenge to researchers, educators and clinicians. This is because measuring narrative proficiency requires tracking multiple sources of macrostructure and microstructure information in increasingly more demanding contexts (Petersen, Gillam & Gillam, 2008).

The data collected in this study suggest that the unified construct score (total MISL score) is the most valid measure for assessing narrative discourse progress using the MISL rubric. Neither the macrostructure nor the microstructure subscales on their own was sufficient to reflect the complexity of narrative discourse proficiency. These assumptions were drawn from the psychometric data reporting lower internal consistency scores for each of these scales when evaluated independently. This is not to say that the individual subscales (macrostructure, microstructure) are not informative for intervention planning. For example, clinicians may use data from the subscales to note macrostructure and microstructure features that are consistently absent from students' stories and target them explicitly during future sessions. When evaluating progress in response to narrative instruction, the total MISL score is the most well-supported of

the three scores that may be obtained using the rubric.

References

American Institutes for Research (2015)a. The essential components of RTI.

www.rti4success.org. Accessed on 3/3/2015.

American Institutes for Research (2015)b. Technical review committees process.

<http://www.rti4success.org/technical-review-committees-process>. Accessed on 3/3/2015.

American Speech-Language-Hearing Association. (2001). *Roles and responsibilities of speech-language pathologists with respect to reading and writing in children and adolescents* [Position Statement]. Available from www.asha.org/policy.

American Speech-Language-Hearing Association. (2000). *Guidelines for the roles and responsibilities of the school-based speech-language pathologist* [Guidelines]. Available from www.asha.org/policy.

Benson, S. (2009). Understanding literate language: Developmental and clinical issues. *Contemporary Issues in Communication Science and Disorders*, 36, 174-178.

Berman, R. (1988). On the ability to relate events in narrative, *Discourse Processes*, 11, 469-497.

Bracken, B. A., & McCallum, R. S. (1998). *The Universal Nonverbal Intelligence Test*. Itasca, IL: Riverside.

Carrow-Woolfolk, E. (1999). *Comprehensive Assessment of Spoken Language*. Circle Pines, MN: American Guidance service.

Colozzo, P., Gillam, R. B., *Wood, M., *Schnell, R. D., & Johnston, J. R. (2011). Content and Form in the Narratives of Children with Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, 54, 1609-1627, DOI: 10.1044/1092-

4388(2011/10-0247).

- Crotty, B., & Gillam, S. (June 3, 2015). Content and form in the narratives of children with Autism Spectrum Disorder in two elicitation contexts: Implications for Assessment and Instruction. Poster presentation presented to the *Symposium on Child Language Disorders*, Madison, Wisconsin.
- Crotty, B., & Gillam, S. (November, 2015). Content and form in the narratives of children with Autism Spectrum Disorder in two elicitation contexts: Implications for Assessment and Instruction. Technical session presented to the *American Speech Language and Hearing Convention*, Denver, Colorado.
- Ehren, B., & Whitmire, K. (2009). Speech-language pathologists as primary contributors to response to intervention at the secondary level. *Seminars in Speech and Language*, 30, 90-104.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18, 39-50.
- Gillam, S. & Gillam, R. (2006). Making evidence-based decisions about child language intervention in schools. *Language, Speech, and Hearing Services in Schools*, 37, 1-12.
- Gillam, S., & Justice, L. (2010). Progress-monitoring Tools for SLPs in Response to Intervention (RTI): Primary Grades. *The ASHA Leader*, 9/21 issue: Feature.
- Gillam, S., Gillam, R., & Laing, C. (2012). Supporting knowledge in language and literacy (SKILL). Narrative Intervention Curriculum. Utah State University.
- Gillam, S., Olszewski, A., Fargo, J., & Gillam, R. (2014). Classroom-based narrative and vocabulary instruction: Results of an early-stage, nonrandomized comparison study. *Language, Speech, and Hearing Services in Schools*, 45, 204-219.

- Gillam, R. B., & Pearson, N. (2004). *Test of Narrative Language*. Austin, TX: PRO-ED.
- Greenhalgh, K., & Strong, C. (2001). Literate language features in spoken narratives of children with typical language and children with language impairments. *Language, Speech, and Hearing Services in Schools, 32*, 114-125.
- Hatcher, L. (1994). *A step-by-step approach to using the SAS® system for factor analysis and structural equation modeling*. Cary, N.C.: SAS Institute Inc.
- Heilmann, J., Miller, J., Nockerts, A., & Dunaway, C. (2010). Properties of the narrative scoring scheme using narrative retells in young school-age children. *American Journal of Speech-Language Pathology, 19*, 154-166.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. (*Structural Equation Modeling, 6*, 1-55.
- Hughes, D., McGillivray, L., & Schmidek, M. (1997). *Guide to narrative language: Procedures for assessment*. Eau Claire, WI.
- Justice, L. M. (Ed) (2006). *Clinical Approaches to Emergent Literacy Intervention*. San Diego, CA: Plural Publishing.
- Justice, L. M., Bowles, R., Eisenberg, S. L., Kaderavek, J. N., Ukrainetz, T. A., & Gillam, R. B. (2006). The index of narrative micro-structure (INMIS): A clinical tool for analyzing school-aged children's narrative performance. *American Journal of Speech-Language Pathology, 15*, 177-191.
- Kaufman, A. S., and Kaufman, N. L. (1990). *Kaufman Brief Intelligence Test*. Circle Pines, MN: American Guidance Service.
- 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 and Hearing*

- Research*, 38, 415–425.
- Loban, W. (1976). *Language development: Kindergarten through grade twelve*. Urbana, IL: National Council of Teachers of English.
- Merritt, D., & Liles, B. (1987). Story grammar ability in children with and without language disorder: Story generation, story retelling, and story comprehension. *Journal of Speech and Hearing Research*, 30, 539-552.
- Miller, J. F., Andriacchi, K., & Nockerts, A. (2011). *Assessing Language Production Using SALT Software: A Clinician's Guide to Language Sample Analysis*. Middleton, WI: SALT Software, LLC.
- Miller, J. F., & Chapman, R. S. (2000). *Systematic Analysis of Language Transcripts (Version 6.1a) [Computer software]*. Madison: University of Wisconsin—Madison, Waisman Research Center, Language Analysis Laboratory.
- Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide* (5th Ed.). Los Angeles, CA: Muthén & Muthén.
- National Governors Association and Council of Chief State School Officers. (2011, October 27). *Common Core State Standards Initiative*. Retrieved from www.corestandards.org
- Newcomer, P. L., & Hammill, D. D. (1997). *Test of Language Development—Primary, Third Edition*. Austin, TX: Pro-Ed.
- Nippold, M. (1998). *Later language development: The school-age and adolescent years* (2nd Ed). Austin, TX: Pro-Ed.
- Nunnally, J. C., & Bernstein, I. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Paul, R. (1995). *Language Disorders from Infancy through Adolescence: Assessment and Intervention*. St. Louis: Mosby-Year Book.

- Paul, R., (2007). *Language disorders from infancy through adolescence: Assessment & Intervention* (3rd ed.). St Louis, MO: Mosby.
- Paul, D. & Hasselkus, A. (2004). Clinical record-keeping in speech-language pathology for speech-language pathology for healthcare and third-party payers. Rockville, MD: ASHA.
- Petersen, D., Gillam, S., & Gillam, R. (2008) Emerging procedures in narrative assessment. The index of narrative complexity. *Topics in Language Disorders*, 28, 2, 111-126
- R Development Core Team. (2012). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Semel, E., Wiig, E., & Secord, W. (2004) Clinical evaluation of language fundamentals (CELF-4). Austin, TX. PROED
- Stein, N.L. (1988). The development of children's storytelling skill. In M.B. Franklin and S.S. Barten (Eds.), *Child language: A reader* (pp. 282-297). NY: Oxford University Press.
- Stein, N.L., & Glenn, C.G. (1979). *An analysis of story comprehension in elementary school children*. In R.O. Freedle Squires, K., Lugo-Neris, M., Pena, E., Bedore, L., Bohman, T., & Gillam, R. (2014). Story retelling by bilingual children with language impairments and typically developing controls. *International Journal of Communication Disorders*, 49, 60-74.
- Sutherland Cornett, B. (2006, September 05). Clinical Documentation in Speech-Language Pathology: Essential Information for Successful Practice. *The ASHA Leader*.
- Swanson, L., Fey, M., & Mills, C., & Hood, L. S. (2005). Use of narrative-based language intervention with children who have specific language impairment. *American Journal of Speech-Language Pathology*, 14, 131-141. doi: 10.1044/1058-360(2005/014).

- Tomblin, B., Records, N., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, E. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*, 40, 1245-1260. (Ed.), *New directions in discourse processing*. Hillsdale, NJ: Erlbaum.
- van den Broek, P., Linzie, B., Fectcher, C., & Marsolek, C. (2000). The role of causal discourse structure in narrative writing. *Memory and Cognition*, 28, 711-721.
- Warren, S., Fey, M., & Yoder, P. Differential treatment intensity research: A missing link to creating optimally effective communication intervention. *Mental Retardation and Developmental Disabilities Research Reviews*, 13, 70-77.
- Westby, C. E. (1985). Learning to talk - talking to learn: Oral - literate language differences. In C.S. Simon (Ed.), *Communication skills and classroom success: Therapy methodologies for language-learning disabled students* (pp. 181-213). San Diego, CA: College-Hill.
- Westby, C. (2005). Assessing and facilitating text comprehension problems. In H. Catts & A. Kamhi (Eds.), *Language and reading disabilities* (pp. 157-232). Boston, MA: Allyn & Bacon.
- Westerveld, M. & Gillon, G. (2008). Oral narrative intervention for children with mixed reading disability. *Child Language Teaching and Therapy*, 24, 31-54. doi: 10.1177/0265659007084567
- White, M., van den Broek, P., Kendeou, P. (April, 2007). Comprehension and basic language skills predict future reading ability: A cross-sectional study of young children. Symposium paper presentation at the *Society for Research on Child Developments Biennial Conference*, Boston, MA.
- Yu, C.-Y. & Muthén, B. (2002, April). *Evaluation of model fit indices for latent variable models*

with categorical and continuous outcomes. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, Louisiana.

Table 1

Demographic Characteristics of Study Samples

Children with Language Impairments	
	Mean (Std. Dev.)
<u>Gender</u>	
Male	69
Female	40
<u>Variable</u>	
<u>Race & Ethnicity</u>	
<u>White, not Hispanic</u>	26
	26 (12 bilingual, English was first
<u>White, Hispanic</u>	language)
<u>African American</u>	21
<u>American Indian</u>	24
<u>Asian</u>	12
Test of Narrative Language (NLAI)	76.67 (11.99)
Comprehensive Assessment of Spoken Language	75.91 (11.10)
Comprehensive Evaluation of Language	
Fundamentals	73.00 (8.66)
Nonverbal Intelligence Quotient	95.54 (8.33)
NLAI –Narrative Language Ability Index; Nonverbal Intelligence Quotient (Universal nonverbal intelligence test; UNIT or Kaufman Brief Intelligence Test; K-BIT)	



Monitoring Indicators of Scholarly Language

Gillam & Gillam (2010)

Story Grammar Element	Description	Examples	Description	Examples	Description	Examples	Description	Examples
Character Salt Code = CH	0 Points: No character is included, or only ambiguous pronouns are used.	<u>They</u> were walking. <u>She</u> and <u>him</u> were walking.	1 Point: Includes at least one character using non-specific labels (pronouns, nouns) WITH a determiner “the” or “a”.	Once there was a boy walking. <u>The boy</u> was walking.	2 Points: Includes at least 1 character using a “ name ” for the character <i>Note: Only code each character one time.</i>	Once there was a boy named <u>Charles</u> .	3 points: Includes more than 1 character using specific name	There was a boy named <u>Charles</u> , a girl named <u>Connie</u> , and a mom named <u>Jody</u> .
Setting Salt Code = S	0 Points: No reference to a specific time or place.	The boy and girl were walking.	1 Point: Includes reference to a general place or time (*not necessarily related to a “story”)	The boy and the girl were <u>outside</u> . The space ship came from <u>outer space</u> .	2 points: 1 reference to a specific place or time in the same story. (*must be related specifically to the story).	Once there was a boy and a girl walking in <u>Central Park</u> .	3 points: Includes 2 or more references to specific places and/or times (in the same story).	Last week there was a boy and a girl walking in <u>Central Park</u> . They lived in <u>Logan</u> .
Initiating Event Salt Code = IE Event that motivates/elicits action “starts the story” <i>*Note: The IE must be explicitly stated by the child, not inferred by the scorer.</i>	0 Points: A problem or “starting” event is not stated.	The girl looked at the boy. The boy and girl were walking in the park. The boy is next to a car. There is a tree.	1 Point: Includes at least one event or problem that does not motivate/elicit an action from the character	<u>A spaceship</u> landed in the park (potential initiating event). There were aliens laughing and a dog running and a table... (no action/attempt related to potential IE)	2 points: Includes at least one event or problem that elicits an active response from the character(s).	A spaceship landed in the park (IE). The girl ran (A) out to say “hi” to the aliens.	3 points: 2 or more IE’s in one story (complex episode)	A spaceship landed in the park (IE) The girl ran (A) out to say “hi” to them. They became friends (C). Then, the spaceship caught on fire (IE). They ran to get some water.

Internal Response	0 Points: There are no feelings, desires or thoughts explicitly stated	The girl and boy saw the aliens land and they ran out to meet them.	1 Point: Words are used that describe feelings that are <u>not</u> directly related to the IE.	The boy saw a spaceship land in the park (IE). There was a happy dog.	2 points: The feelings, desires or thoughts of the character are explicitly stated and <u>relate</u> to the IE.	The spaceship landed (IE). The girl was <u>afraid</u> (IR) of meeting the aliens.	3 points: Two or more feelings, desires or thoughts are explicitly stated and <u>relate</u> to the IE.	The spaceship landed. The girl was excited to meet the aliens. She was <u>happy</u> when they greeted her nicely.
*Note: Adjective or adverb that expresses a mental state related to emotion					(One stated IR)		(2 or more stated IRs)	
Plan	0 Points: No statement or wording that relates to planning to take action that can be directly tied to the IE.	The aliens landed. The girl ran out to meet them.	1 point: Terms are used or statements are made that use “gonna, going to” or a cognitive/mental state verb NOT related to how the character may react to the IE. The statement is NOT directly related to the IE.	The girl decided to have a picnic with her brother.	2 points: There is a statement about planning to act and it is tied directly to the IE. Must be made by the main character.	The spaceship came down (IE). The aliens came out (A). The girl <u>wanted</u> to go (P) meet them.	3 points: There is more than one statement about planning to act and it is tied directly to the IE. Must be made by the main character.	The aliens landed. The girl decided to go meet them. She ran over and said, “Hi.” The boy <u>thought</u> he would sneak away. He went home and no one saw him go.
Salt Code: P <i>Key words:</i> wanted, thought, decided, pondered, considered	Decided, wanted, thought are NOT included		character may react to the IE. The statement is NOT directly related to the IE.					
Action/Attempt	0 Points: No actions are taken by the main character(s) (no action verbs contained in the story). Basically, a series of random descriptions.	There is a girl. There is a boy. It is sunny.	1 point: Actions are taken by the main character(s) that are not directly related to the IE. Descriptive actions	The spaceship landed. The boy and the girl were <u>going</u> to a park.	2 Points: One or more actions is taken by the main character(s) that IS directly related to the IE.	The spaceship of aliens landed in the park (IE). The girl <u>ran</u> out to meet them. She went up and said, “Hi.”	3 Points: The addition of a complicating action that interferes with the character's actions in response to the IE.	The aliens landed in the park (IE). The girl wanted to be their friend (P). She walked over to say hi (A). <u>They</u> <u>sarled</u> at her (Complication). She ran home to tell her parents what happened (C).
Salt Code = A Note: Cognitive state verbs NOT included (thought, decided, wanted, said, saw)								

Consequence	0 Points: No outcome of the action/attempt is explicitly stated.	The spaceship landed (IE). The aliens go out (A). The boy was afraid (IR). * An internal response may not serve as a consequence.	1 point: One consequence with no IE. *The consequence is linked only to an action.	The girl ran over there (A). She fell and got hurt (C).	2 Points: One consequence directly linked to IE.	The spaceship came from space and landed (IE) in the park. The aliens got out to (A) look at the earth (A) and then they flew back to their home (C).	3 Points: Two or more consequences. To get a 3: IE #1 must match up with Conseq #1; IE #2 must match up with Conseq #2	The spaceship landed (IE). The aliens got out (A) and looked at the earth (A) and flew home (C/IE). On the way they hit a meteor (A). They fixed the hole (A) and flew on home (C).
--------------------	---	--	---	---	---	---	---	---

Literate Language	Description	Example	Description	Example	Description	Example	Description	Example
Coordinating Conjunctions FANBOYS (<i>for, and, nor, but, or, yet, so</i>) Can coordinate nouns, verbs, or clauses. 'so excited' = adverb	0 points	No coordinating conjunctions in story For, an, nor, but, or, yet, so NOT included	1 point: One coordinating conjunction used in story.	The girl was afraid and the boy ran away as fast as he could.	2 points: Two coordinating conjunctions used in story.	John walked to the store but it was closed.	3 points: Three or more different coordinating conjunctions used in story.	Sally ran home but their mom wasn't there, so they went back to the park.
Subordinating Conjunctions (<i>when, while, because, after, if, since, before</i>) 'that day' = adjective	0 points	No subordinating conjunctions	1 point: <u>One</u> subordinating conjunction used in the story	<u>When</u> the aliens landed the girl ran.	2 points: <u>Two</u> different subordinating conjunctions used in the story	The girl saw the aliens while she was playing in the park. She ran home because she was <u>afraid</u> .	3 points: <u>Three</u> or more different subordinating conjunctions used in the story	<u>After</u> the aliens landed, they walked out of the spaceship. John said, if they have ray guns they will kill us. Sally said, I don't think they do <u>since</u> they look so nice.

Mental verbs Salt Code: M Mental Verbs: decided, though, wanted	0 points	No mental verbs.	1 point: 1 mental verb.	The boy <u>thought</u> it was hot.	2 points: 2 DIFFERENT mental verbs explicitly stated.	He decided to go and meet the aliens. He planned to get to them.	3 points: 3 or more different mental verbs explicitly stated.	He decided to go and meet the aliens. The girl thought he was brave and he decided to act that way.
Linguistic verbs Salt Code: L Linguistic Verbs: said, told, yelled	0 points	No linguistic verbs.	1 point: 1 linguistic verb.	The boy said, "NO!"	2 points: 2 DIFFERENT linguistic verbs explicitly stated.	The boy said, "no," and the girl yelled, "stop!"	3 points: 3 or more different linguistic verbs explicitly stated.	The girl told him he was brave. He said, "thanks," and she said, "you are welcome."
Adverbs Salt Code: ADV Note: Additional examples below chart.	0 points	No adverbs	1 point: One adverb that conveys tone, attitude, time, or manner, degree or reason and modifies a verb, adjective, negation, or another adverb.	<u>Sometimes</u> , they like to watch aliens. He is <u>very</u> good.	2 points: Two different adverbs	The boy and the girl were <u>very</u> scared. They left <u>quickly</u> .	3 points: 3 or more different adverbs.	The aliens yelled loudly, "Don't come over here." <u>Surprisingly</u> , the kids went anyway. <u>After</u> that, they were all friends.
Elaborated Noun Phrases Salt Code: ENP (articles, possessives, determiners, quantifiers, wh-words, big, black, funny) Note: Additional Examples below chart	0 points	No noun phrase elaboration. He saw spaceship.	1 point: A noun phrase contains one modifier that precedes the noun	Her brother saw the spaceship. The dog is happy. Two aliens came out.	2 points: A noun phrase that contains 2 different modifiers that precedes the noun.	<u>The black dog</u> saw the spaceship.	3 points: Noun phrases in which 3 or more different modifiers precede the noun.	<u>The old, black dog</u> was sick.
Grammaticality	0 points	3 or more grammatical errors	1 point	2 grammatical errors	2 points	1 grammatical error	3 points	No grammatical errors
Tense	0 points	3 or more tense changes	1 point	2 tense changes	2 points	1 tense change	3 points	No tense changes

Based on the research and contributions of many including: Anderson, 2010; Curenton & Justice, 2004; Greenhalgh & Strong, 2001; Hughes, McGilivray & Schmidek, 1997; Petersen, Gilliam & Gilliam, 2008; Pellegrini, 1985.

*Contributions from Michelle Merrill, Karen Turnbow, Britney Lamb, Sara Hegsted, Julie Jager, Allison Hancock, Abbie Olszewski.

Date _____
Story used to elicit narrative _____
Total macrostructure score _____
Total microstructure score _____

Additional Examples of Microstructure elements (not an exhaustive list)

Coordinating conjunctions may include and, and then, then, for, or, yet, but, nor, and so. They are used to coordinate clauses (The boy ran back home but he got there too late). We do not give credit when they are used to coordinate nouns in a noun phrase (The boy and the girl) or verbs in a verb phrase (They were running and playing).

Subordinating conjunctions include *after, although, as, because, if, for, like, once, since, that (but that, in that, in order that, such that), unless, when, where, while*. These words set up a hierarchical relationship between clauses. You must have 2 clauses to have a subordinating conjunction. "That" in the sentence, "I saw that." is not subordinating. "That" in the sentence, "I saw that you really liked him," is subordinating.

Adverbs may relate to time (e.g., *all of a sudden, suddenly, again, now, tomorrow, yesterday, then*), manner (e.g., *somehow, well, slowly, accidentally*), degree (e.g., *very, each, some, almost, barely, much*), number (e.g., *first, second*), affirmation or negation (e.g., *definitely, really, never, not*).

Elaborated Noun Phrases are a group of words comprising of a noun with one or more modifiers providing additional information about the noun. Modifiers may include articles (e.g., *a, an, the*), possessives (e.g., *my, his, their*), demonstratives (e.g., *this, that, those*), quantifiers (e.g., *every, each, some*), wh-words (e.g., *what, which, whichever*), and true adjectives (e.g., *tall, long, ugly*).

Simple Elaborated Noun Phrases consist of a single modifier and a noun. Examples include *one day, big doggy* (adjective + noun), *that girl* (determiner + noun), and *those ones* (demonstrative + noun). **Complex Elaborated Noun Phrase** (CENP) consist of two or more modifiers and a noun. Examples include *big red house* (adjective + adjective + noun), *a tall tree* (article + adjective + noun), and *some mean boys* (quantifier + adjective + noun).

Mental Verbs are a type of verb that are used differently than active verbs and are not used in progressive tenses. Mental verbs may include *think, know, believe, imagine, feel, consider, suppose, decide, forget, see, hear, and remember*.

Linguistic Verbs target the verbs that relate to the acts of writing and speaking which may include *read, write, say, tell, speak, shout, answer, call, reply, whisper, and yell*.

Supplemental Materials A. Sample Literacy & Writing Anchor Standards and their corresponding MISL Items by Grade level (1-3).

Anchor Standards: Key Ideas and Details	MISL Item
<p>CCSS.ELA-Literacy.RL.1.1</p> <p>Ask and answer questions about key details in a text.</p> <p>CCSS.ELA-Literacy.RL.2.1</p> <p>Ask and answer such questions as who, what, where, when, why and how to demonstrate understanding of key details in a text.</p> <p>CCSS.ELA-Literacy.RL.3.1</p> <p>Ask and answer to demonstrate understanding of a text, referring explicitly to the text as the basis for answers</p>	<p>Key details = individual scores for each story element (macrostructure scale)</p> <p>Who = character score; What = take-off score</p> <p>Where = setting score; When = setting score and coordinated conjunction score; Why = plan & action scores and subordinated conjunction score; How = action and adverb scores</p>
<p>CCSS.ELA-Literacy.RL.1.2</p> <p>Retell stories, including key details and demonstrate understanding of their central message or lesson.</p> <p>CCSS.ELA-Literacy.RL.2.2</p> <p>Recount stories, including fables and folktales from diverse cultures, and determine their central message, lesson, or moral</p> <p>CCSS.ELA-Literacy.RL.3.2</p> <p>Recount stories, including fables and folktales, and myths from diverse cultures; determine the central message, lesson, or moral and explain how it is conveyed through key details in the text.</p> <p>CCSS.ELA-Literacy.RL.1.3</p> <p>Describe characters, settings, and major events in a story using key</p>	<p>Retelling/recounting/Key details = total macrostructure score</p> <p>Central message, lesson, moral = take off + action + landing; feeling, subordinated conjunction & mental verb scores</p> <p>Explain how = subordinated conjunction and adverb scores; take-off, internal response, plan, attempt, and landing.</p> <p>Describe characters & settings = character, setting and elaborated noun phrases score</p>

<p>details.</p> <p>CCSS.ELA-Literacy.RL.2.3</p> <p>Describe how characters in a story respond to major events and challenges</p> <p>CCSS.ELA-Literacy.RL.3.3</p> <p>Describe characters in a story (e.g., their traits, motivations, or feelings) and explain how their actions contribute to the sequence of events</p>	<p>Major events = take off + action + landing</p> <p>Key details = individual scores for each story element (macrostructure scale)</p> <p>Response to major events = feeling, plan, mental and linguistic verb scores</p> <p>Describe feelings = feelings score, mental verb score</p> <p>Actions contribution to events = plan, action, landing and subordinating conjunction scores</p>
<p>CCSS.ELA-Literacy.W.1.3</p> <p>Write narratives in which they recount two or more appropriately sequenced events, including some details regarding what happened, use temporal words to signal event order, and provide some sense of closure</p> <p>CCSS.ELA-Literacy.W.2.3</p> <p>Write narratives in which students recount a well-elaborated event or short sequence of events, include details to describe actions, thoughts, and feelings, use temporal words to signal event order, and provide a sense of closure.</p> <p>CCSS.ELA-Literacy.W.3.1</p> <p>Use linking words and phrases (e.g., because, therefore, since, for example) to connect opinion and reasons</p>	<p>Recount 2 or more sequenced events = take off + action + landing (scores of 3)</p> <p>Details = take off + action + landing, individual macrostructure element scores; adverb score</p> <p>Temporal words = coordinating conjunctions score</p> <p>Closure = landing score</p> <p>Thoughts = feelings and mental verb scores</p> <p>Linking words and phrases = coordinating and subordinating conjunctions scores</p>

*Note: there are minimal differences between the anchor standards from grades 1-3.

Supplemental Materials C. Scoring Procedures for the MISL

Macrostructure Scoring Details

The scoring procedures for initiating event, internal response, plan, action and consequence is based on whether there is clear evidence that the elements are causally linked and is anchored at a score of 2.

Consider the following story.

An airplane landed in Central Park one Saturday morning and some creepy creatures jumped out. They were looking for a spot to have a picnic. Bob and Jane had been playing when they saw them and they were afraid because they thought the creatures might eat them. They decided to run and hide because they didn't want to be eaten. They took off running so they could get away before the creatures saw them. While they were running to the bushes to hide, Jane tripped on a rock and fell. Bob stopped, ran back and helped Jane get up because he couldn't let the creatures get his sister! They hid in the bushes for awhile and when it was safe they ran home.

The best way to decide whether to award a score of 2 for initiating event is to first examine the story for the presence of an event that *might* motivate a character into action and then determine if there is a stated action that is clearly linked to that event through causal adverbs (because, so). In the example, the storyteller states, *An airplane landed in Central Park one Saturday morning and some creepy creatures jumped out.* A few sentences later the storyteller says: *They took off running so they could get away before the creatures saw them.* A key factor in determining whether to award a score of 2 rather than a score of 1 for initiating event is the use of a causal term in the description of the stated action that ties it to the potential initiating event. In this example, the children ran *so* they could get away from the creatures. The

use of the term *so* is a clear indication that the action (running) the characters took was motivated by the initiating event (creepy creatures landing, afraid of being eaten). In this case, scores of 2 would be appropriate for the initiating event and the action elements. Had the storyteller stated that the children simply ran without any clear indication as to *why* they were running, both initiating event and action would have been given scores of 1. After looking for a causal relationship between an event and actions motivated by the event, the story element action (or attempt) would be revisited for the presence of a complicating action. In the example above, one of the children tripped over a rock as they were running to hide from the creatures (goal motivated action). This occurrence is scored as a complicating action that interferes with the characters goal motivated action (running to hide in order to avoid being eaten). Rather than giving action a score of 2, under these circumstances, the score would be increased to 3 to indicate that the story was slightly more complex than a simple basic episode.

When scoring internal response the examiner would look for statements indicating that the characters had feelings about the initiating event. For example, in the story above, it was stated, *they were afraid because they thought the creatures might eat them*. The term afraid in this sentence may be causally linked to the initiating event (creatures jumping out of airplane) with the word *because*. If the storyteller had simply said, *They were afraid*, internal response would be given a score of 1 indicating that the use of this element was emerging, but not yet mastered. The story element *plan* is scored by looking for terms that indicate that characters have thoughts and make decisions about what they might do as a result of the initiating event. As with internal response, the use of a causal term that clearly links the thought or decision to the initiating event is necessary to award a score of 2. In the example above there was a clear causal term (because) that linked the thought or plan (decided) to the initiating event (didn't want to be

eaten by creatures), making a score of 2 appropriate. However, had the storyteller stated, *They decided to run*, a score of 1 would have been given for plan to indicate the story element was emerging. If no words related to thoughts (eg., wanted, decided, planned to) were used in the story, a score of 0 would have been given. Finally, the story element *consequence* is scored by examining the story for a statement, or statements indicating the characters had attained or not attained their goal(s). In our example the goal of the characters was *not to be eaten by the creatures*. Since the storyteller stated that the children *ran home safely*, we would award a score of 2 for consequence, noting that there was a clear indication that the characters had attained their goal. Had the story ended with the creatures having a picnic, consequence would have been given a score of 1 because there was no mention of the children having attained or not attained their goal. Scores of 3 are given when the story contains an embedded episode. Consider that the story example ended in this fashion:

They took off running so they could get away before the creatures saw them. While they were running to the bushes to hide, Jane tripped on a rock and fell. Bob stopped, ran back and helped Jane get up because he couldn't let the creatures get his sister! They stumbled together behind some bushes and watched to see what the creatures would do. The creatures looked for a spot to have their picnic. They decided to take out some baskets of food and set it out on the picnic table so they could eat. They ate all of their food and were happy that they found this great picnic spot. After they ate, they packed up, got back in their airplane, and flew into the sky to return home. After they disappeared, the children came out of the bushes and ran home safely.

In this story, the initiating event for the embedded episode was that the aliens wanted to

have a picnic. This event was clearly tied to their actions (looked for spot, ate food), plan (decided to take out baskets), internal response (happy they had found the best spot), and consequence (flew into the sky to return home). Scores of 3 would be awarded for all of the story elements under these circumstances.

Microstructure Scoring Details

Consider the simple story about the creatures above noting the bolded linguistic elements.

An airplane landed in Central Park one Saturday morning and some creepy creatures jumped out. They were looking for a spot to have a picnic. Bob and Jane had been playing when they saw them and they were afraid because they thought the creatures might eat them. They decided to run and hide because they didn't want to be eaten. They took off running so they could get away before the creatures saw them. While they were running to the bushes to hide, Jane tripped on a rock and fell. Bob stopped, ran back and helped Jane get up because he couldn't let the creatures get his sister! They hid in the bushes for awhile and when it was safe they ran home.

In this story, there was one coordinated conjunction used repeatedly (eg., and) earning a score of 1. There were three difference subordinating conjunctions (eg., so, because, when) for a score of 3, two metacognitive verbs (e.g., thought, planned, decided) or metalinguistic verbs (said, yelled, told) or a combination of the two for a score of 2, and the most complex elaborated noun phrase contained two modifiers (eg., one Saturday morning) earning a score of 2.