

■ The Validity of the Joint Story Retell as a Measure of Young Children's Comprehension of Familiar Stories

■ Validité de la *Joint Story Retell* comme mesure de la compréhension d'une histoire connue chez les jeunes enfants

Lynn F. Dempsey
Elizabeth Skarakis-Doyle

Abstract

Increasing emphasis is being placed on early identification of young children with speech and language impairments. Unfortunately, gaps in our assessment batteries for young children limit our ability to achieve this goal. This is particularly evident in the area of language comprehension where many measures focus on vocabulary and syntax but few assess discourse level language. The Joint Story Retell (JSR) is a new measure of oral story comprehension adapted from the cloze procedure. The purpose of this investigation was to evaluate the JSR's validity by examining its relationship with age, traditional comprehension questions, and expressive language skill. Results indicated that the JSR was sensitive to age differences, had a moderately strong, positive relationship with comprehension questions, and limited expressive language demands. These findings suggest that the JSR may be a valid measure of young children's discourse comprehension, and thus may be a meaningful addition to a battery of comprehension measures.

Abrégé

On accorde de plus en plus d'importance au dépistage précoce des troubles de la parole et du langage chez les jeunes enfants. Malheureusement, les lacunes de nos batteries d'évaluations pour ce groupe d'âge limitent notre capacité à poser un diagnostic. C'est notamment le cas pour ce qui est de la compréhension du langage, où nombre des mesures portent sur le vocabulaire et la syntaxe, mais peu sur l'évaluation de la compréhension au niveau de l'histoire. Adaptée de la méthode de closure, la *Joint Story Retell* (répétition conjointe d'une histoire) constitue une nouvelle mesure de la compréhension d'une histoire orale. Le but de la présente enquête consiste à évaluer la validité de cette nouvelle technique en examinant sa relation avec l'âge, les questions de compréhension classiques et les capacités expressives du langage. Les résultats indiquent que la méthode de la répétition conjointe d'une histoire était sensible à la différence d'âge, avait une relation positive moyennement forte avec les questions de compréhension et limitait les exigences d'expression du langage. Ces conclusions suggèrent que la répétition peut constituer une mesure valide de compréhension du discours chez les jeunes enfants et représente ainsi un ajout valable à la panoplie de mesures de la compréhension.

Key words: Joint Story Retell, discourse comprehension, language assessment, measurement validity, early identification, language impairment

The nature of childhood language impairment and its long-term prognosis differ when both expressive and receptive abilities rather than solely expressive abilities are impaired (Craig & Evans, 1989; Thal, 1991; Thal, Tobias, & Morrison, 1991). This makes valid and reliable assessment of comprehension essential for early identification of young children with

Lynn F. Dempsey and
Elizabeth Skarakis-Doyle
University of
Western Ontario
London, Ontario

language impairments. Noting the limitations in our ability to measure comprehension reliably, Thal et al. (1991) maintain that multiple procedures for comprehension assessment are necessary so that a variety of evidence may ensure a comprehensive profile of these abilities. Most assessment procedures for the evaluation of preschool children's language comprehension focus exclusively on vocabulary and syntax. Very few measures of discourse level understanding appropriate for preschool age children have been developed. This is the case even though preschoolers are regularly exposed to discourse in the early years of developing language (Snow, Perlmann, & Nathan, 1987).

Failure to evaluate preschoolers' discourse comprehension could have serious consequences. Numerous studies have demonstrated that difficulties in discourse comprehension may limit the acquisition of literacy skills and social communicative abilities (Dickinson & Smith, 1994; Feagans, 1982; Feagans & Short, 1984). Dickinson and Snow (1997), for example, found a strong relationship between narrative comprehension among kindergarten-aged children and measures of early literacy such as the ability to define words, phonemic awareness, and early print skills. It has been suggested that children at risk for reading and social communication problems may be identified by measuring the strength of their discourse level language skills (McCabe & Rollins, 1994).

One problem associated with the paucity of discourse level comprehension measures suitable for preschool age children is that performance on vocabulary and syntax level measures may be used to draw conclusions about children's discourse level abilities. Regardless of how well a task assesses vocabulary or syntactic comprehension, one cannot necessarily generalize those findings to the broader domain of comprehension. The reason for this is clear when the specific abilities required to comprehend language at each level are considered. At the vocabulary and syntactic levels, typical comprehension tasks are decontextualized and involve identifying the referents of single words and decoding meaning relations within a sentence (Miller & Paul, 1995). At the discourse level, comprehension, in addition to requiring word and sentence understanding, involves making judgments based on social, textual, scriptal, and other forms of prior knowledge to determine what an utterance means in relation to what else has gone on in the discourse (Miller & Paul, 1995; Rees & Shulman, 1978). Since discourse comprehension requires more than simply the ability to understand decontextualized words and sentences, performance on tasks that measure only word and sentence understanding cannot be taken as necessarily indicative of the ability to comprehend longer

segments of language (McCabe, 1996). However, the risk that invalid interpretations about comprehension will be made on the basis of vocabulary and syntactic comprehension measures will remain until an acceptable measure of young children's discourse comprehension is developed. Such a measure, in conjunction with measures of vocabulary and syntactic level understanding, might permit more comprehensive measurement of language comprehension than is currently possible.

As has been discussed, the risk of drawing invalid conclusions about preschoolers' comprehension abilities might be lessened if an appropriate measure of their discourse comprehension was available. For such a measure to contribute meaningfully to comprehension assessment as a whole, it must be valid in its own right. There are several issues associated with the development of a valid assessment task. One critical issue involves ensuring that the task measures only the construct or behavioural domain it is intended to measure (i.e., construct relevance; Messick, 1995). When an assessment procedure contains an enabling factor or task that is extraneous to the behavioural domain being assessed, yet required for its performance, measurement of the target domain may be obscured (McCauley & Swisher, 1984). This, in turn, may make interpretation of task performance more difficult (Messick, 1995).

Two traditional assessment procedures for discourse comprehension contain an enabling factor for task performance that can obscure measurement of preschoolers' comprehension abilities. Story retelling tasks, that require participants to listen to stories and then retell them, and comprehension question tasks, that require participants to respond to questions posed about stories, demand expressive language proficiency. The intrusion of expressive language into tasks intended to measure discourse comprehension might negatively affect performance on the tasks in a manner irrelevant to comprehension ability, since young children's expressive language skills are not as well-developed as those of older children and adults (Applebee, 1978; Carlisle, 1991; Feagans & Farran, 1981; McCabe, 1996; Stein, 1988). That is, children with age-appropriate discourse comprehension ability might perform poorly on traditional discourse comprehension tasks, and thus appear to have comprehension difficulties, when in fact it is the demand for expressive language proficiency, rather than discourse comprehension, that contributes to the poor performance. It must be noted in the case of comprehension questions that not all types are equally demanding of expressive language proficiency. Clearly yes/no type questions and some simple wh-questions (e.g., who-questions) require only a limited verbal response. However, many wh-questions require much more verbal

ability. Most comprehension question tasks employ a mix of yes/no questions which require individuals to judge the truth or falsity of propositions and wh-questions which require individuals to formulate answers from the propositions they hold in memory. Thus, comprehension question tasks may, by virtue of the wh-questions they contain, demand substantial expressive language proficiency. In summary, the expressive language requirements of traditional discourse comprehension tasks such as story retelling and comprehension questions may hinder young children's ability to engage in those tasks and confound interpretations of performance, hence compromising the validity of comprehension assessment (McCabe, 1996).

Skarakis-Doyle and Wootton (1998) developed the Joint Story Retell (JSR) to meet the need for a measure of oral discourse comprehension that would impose limited demands on expressive language abilities. The JSR was adapted from the cloze test, a widely used procedure for assessing reading comprehension in which the participant demonstrates comprehension by providing the word(s) missing from a passage. Both the cloze test and the JSR are based on the principle that individuals who have understood a passage well will be more likely to complete the passage with the deleted word(s) than those who have not. Since children completing the JSR are required to produce only specific and limited elements rather than formulate sentences or entire stories, it is hypothesized that this task places minimal demands on children's expressive language abilities. Thus, the limited response requirements of the JSR may enable young children to demonstrate their story comprehension with limited interference from expressive language. Results of a preliminary investigation (Skarakis-Doyle & Wootton, 1998) supported this hypothesis, demonstrating that although MLU was moderately correlated with the JSR, it did not contribute meaningfully to the prediction of performance on the task. This finding suggests that the JSR does not impose substantial expressive language demands on preschool children, and thus may more directly reflect discourse comprehension abilities. However, more rigorous evidence is needed.

In addition to evidence that the JSR provides a direct assessment of discourse comprehension unobscured by expressive language proficiency, evidence of expected performance differences over time (i.e., developmental sensitivity) is required to establish the task's validity (Anastasi, 1988). This evidence may be obtained through comparison of JSR test scores with chronological age. Since language comprehension is an ability that improves with age, a valid measure of discourse comprehension would be expected to show a similar improvement in performance with age. As yet, the degree to

which JSR performance improves with age has not been firmly established, since, in the preliminary investigation (Skarakis-Doyle & Wootton, 1998), the measure was examined across a limited age-range of children (i.e., 46-58 month olds). Therefore, empirical evidence of the developmental sensitivity of the JSR is still needed.

In addition to evidence of developmental sensitivity, evidence of an expected relationship between the JSR and another test of discourse comprehension would support the JSR's validity. According to Anastasi (1988), correlations between a new test and similar earlier tests of the same domain may be cited as evidence that the new test measures the same general area of behaviour as the other tests, and thus demonstrates concurrent validity. Correlations between the JSR and a currently employed measure of discourse comprehension (i.e., story retelling, comprehension questions) could be examined to determine whether the JSR measures the behaviour, discourse comprehension, it is intended to measure. As yet, the JSR has not been tested against traditionally employed measures of discourse comprehension. Therefore, the concurrent validity of the JSR has yet to be determined.

The purpose of this investigation was to evaluate the validity of the JSR by addressing the aforementioned validity criteria. The following specific questions were posed:

1. Is the JSR a developmentally sensitive measure? It is hypothesized that performance on the JSR will improve with age.
2. What is the relationship of the JSR to a traditional measure of discourse comprehension (i.e., comprehension questions)? It is hypothesized that performance on the JSR will be at least moderately correlated with accuracy on the comprehension questions, indicating that the two procedures measure the same construct, discourse comprehension.
3. What is the relationship of expressive language ability to performance on the JSR? If the JSR imposes only minimal expressive language demands as is hypothesized then it should contribute little to performance on the task.

Method

Participants

Thirty-eight children (24 females and 14 males) between the ages of 30 and 50 months were included in this investigation. All of the children came from homes where English was reported by the parent to be the primary language. None of the participants had any cognitive or other uncorrected sensory (including hearing) or motor impairment, as reported by parents and

as described on the Checklist for Hearing Impairment (CHI; Warr-Leeper et al., 1997). The CHI, a checklist pertaining to a child's developmental and hearing history, consists of yes/no questions and rating scales. All children also demonstrated understanding of the primary experimental task, the JSR, as indicated by at least one self-initiated accurate response during a practice session administered prior to the task.

The children's receptive language abilities were normally developing as indicated by scores within 1.5 standard deviations (*SD*) of the mean for their age on the Auditory Comprehension (AC) subscale of the Preschool Language Scale-3 (PLS-3; Zimmerman, Steiner, & Pond, 1979). All children also possessed normally developing language as indicated by a score above the 10th percentile on the MacArthur Communication Development Inventory III (MCDI-III; Dale, 1996). The MCDI-III, a global measure of language development, is a parent checklist consisting of items that tap expressive vocabulary, sentence structure, and comprehension. A vocabulary pretest constructed by the investigators documented each child's understanding of the words used in the story. The pretest consisted of 16 words from the story including eight nouns, five verbs, two adjectives, and one locative. The test was a word-picture matching task in which the child was required to select one of four pictures based on a word orally presented by the investigator (e.g., "Show me, rubber duckie"). All children obtained scores of 60% or greater on the vocabulary pretest. Children's scores on the vocabulary pretest, as well as their scores on the AC subscale of the PLS-3, are shown in Table 1.

Parent responses to questions on an early literacy questionnaire that was constructed for this investiga-

tion (Dempsey, Perfetti, & Skarakis-Doyle, 1999) characterized the participants' early literacy experience. According to their parents, all of the children participated in listening to stories. Ninety-five percent of the children were read stories more than three times per week.

Materials and Experimental Test Stimuli

The patterned children's storybook, "Splish Splash," created by Skarakis-Doyle and Wootton (1998), was employed in this investigation. A patterned story was employed because it was a familiar and engaging story genre for the age group under investigation. The plot of the story incorporates a familiar bathtime routine. The story is composed of seven episodes that are organized around a central goal. Each episode, in turn, is comprised of a sequence of goal-directed activities. The overall length of the story in total number of words is 398. As in other patterned stories, vocabulary words, sentence patterns, and episodes are repeated and a distinct rhythmic or song-like refrain is incorporated into the text of the story at several intervals. Each page of text is accompanied by a corresponding picture. An audio recording of the story was employed in the story familiarization phase of the investigation to allow for consistency in story presentation across participants.

Two kinds of test stimuli were constructed: a cloze or JSR story version, and two forms of a set of traditional comprehension questions. In the cloze story version (examples in Appendix A), eight elements, (i.e., actions, actors, objects, locations, adjectives) were omitted from the text of the story and substituted with blanks or pauses. Selection of the elements for omission was based on an adaptation of procedures used to construct cloze tests of reading comprehension. Seven of the omitted elements were critical to the progression of the story

Table 1
Mean Ages, and Mean Raw Scores and Standard Deviations
for Participants on Pre-Experiment Tests (N = 38)

	Variables		
	Age ^a	PLS-3 ^b	Vocab. test ^c
M	42.08 (30-50)	35.55	13.11
SD	5.54	5.95	1.74

Note. Age ranges in parentheses

a. Age reported in months

b. PLS-3 maximum raw score = 48

c. Vocabulary pretest maximum raw score = 16

toward its goal (i.e., necessary to achieve the goal); one (i.e., the name of the central character in the story) was supportive. In order that the eight elements might be deleted and the sense of the story maintained, some of the word order was altered from the original version of the story and the length was condensed to 238 words. Three of the eight cloze items were accompanied by pictures that could have revealed the appropriate response. The pictures accompanying the other five items did not provide additional cues to the appropriate responses.

The comprehension questions were designed to tap the same content as the cloze version of the story. As is typical in traditional comprehension question tasks, both yes/no, and wh-questions were included in the set of questions. Thus, the demand for recognition type responses (as in yes/no questions) and recall type responses (as in wh-questions) was balanced. Two forms of the comprehension questions (Form A and Form B; examples in Appendix B) were prepared in order to minimize the probability of responses to the yes/no questions being influenced by guessing.

Procedure

The investigation was separated into two phases. Phase I consisted of preexperiment testing and story familiarization. Phase II involved presentation of the experimental procedures. The two phases occurred over a two-day period for each child. The experimental sessions were recorded on videotape to allow for detailed scoring and analysis. All sessions were conducted either in a university laboratory, in the child's home, or at the child's preschool or daycare centre.

Preexperiment Testing and Story Familiarization

During this phase, children completed the inclusionary testing for participation in the investigation. The children completed a standardized receptive language test, the Auditory Comprehension (AC) subscale of the PLS-3 (Zimmerman et al., 1979), as well as the experimental vocabulary pretest constructed by the investigators. In addition, the parents completed the following questionnaires: the MCDI-III (Dale, 1996); the CHI (Warr-Leeper et al., 1997); and the early literacy questionnaire (Dempsey et al., 1999). The MCDI-III is an upward extension of the MacArthur Communicative Development Inventories for 8- to 30-month-olds that are widely used, reliable, and valid measures of language and communicative development (Fenson et al., 1991). Evidence indicates that the MCDI-III is useful for research calling for an easily obtained overall measure of language development (Dale, Reznick, & Thal, 1998). Like other parent report measures, the MCDI-III has the advantage of providing data that are more rep-

resentative of young children's language than laboratory language samples (Fenson et al., 1991). The MCDI-III was employed in this study for two purposes. First, overall scores on the questionnaire were used to ensure that all participants had normally developing language. Second, performance on two parts of the MCDI-III, the Vocabulary Scale which consists of a 100-item checklist, and the Grammar Scale which consists of 12 sentence pairs from which the parent selects the item most closely resembling the child's current level of language, was used to explore the experimental relationship between expressive language ability and JSR performance.

Each child was presented with the original "Splish Splash" story three times over a twenty-four hour period prior to participating in the experimental tasks. The first story presentation occurred following completion of the standardized testing. The child listened to the audio-recording of the story with the investigator and followed along in the book. The second story presentation occurred between the end of the first session and the beginning of the test phase on the second day. Each parent was provided with a copy of the storybook and the audiotape and was asked to listen to the story with the child one time at home. The parent was instructed to listen to the story with the child during their usual storyreading time and was asked to respond to comments made by the child during the storyreading with neutral acknowledgment (e.g., "Oh," "OK," "Oh," plus repetition of the child's phrase) but not engage in discussion of the story with the child. The third and final story presentation occurred on the second day, immediately prior to the administration of the experimental tasks.

Experimental Test Procedures

Following the third story presentation, children participated in the experimental portion of the investigation where they jointly retold the story with the investigator and answered the comprehension questions. The order of the experimental procedures was counterbalanced across children, such that half of the children participated in the JSR first, while the other half answered the comprehension questions first.

Joint Story Retell Procedure

Each child completed four practice items prior to participating in the JSR. Following completion of the practice items, the investigator read the cloze story to the child as he or she followed along in the book. The child was required to supply the appropriate word or words during pauses made by the investigator. The investigator waited no longer than five seconds for a response. If the child did not respond within five seconds, the inves-

tigator provided the desired response and then continued with the story.

Comprehension Questions Procedure

Administration of the traditional comprehension questions consisted of the investigator asking the child either the Form A or Form B questions. The child was expected to answer the wh-questions verbally; however, either a verbal or a nonverbal (i.e., nodding yes or shaking head no) response was acceptable for the yes/no questions. Each question could be repeated only once. If a child did not respond to a question or gave an incorrect response the investigator responded neutrally and then continued with the next question without providing the correct response. The child was given no longer than five seconds to respond to each question.

Data Analysis

Scoring

Accuracy on the JSR was determined by calculating the total number of items (out of a maximum number of eight) that were given a correct response. Accurate responses were either verbatim from the story or contained minor variations in wording that did not alter the primary meaning of the original utterance (e.g., big bar of soap/ bar of soap/ soap).

Accuracy on the comprehension questions was determined by calculating the number of correct responses to the questions. Verbal responses were expected for wh-questions; both verbal and nonverbal responses to the yes/no questions were accepted. Correct responses to wh-questions received one point each, while correct responses to yes/no questions received half a point each to prevent inflated scores due to the potential for guessing an answer correctly.

Agreement

An independent rater rescored a minimum of 10% of both the JSR and comprehension question protocols in order to determine inter-judge agreement. Agreement for scoring between raters was 100% for both the JSR and comprehension question protocols.

Results and Discussion

Before the primary data analyses were conducted preliminary analyses were undertaken in order to evaluate particular aspects of the test stimuli and their presentation. First, the possibility that pictures accompanying some of the cloze items might have cued a correct response was examined. Second, analyses were conducted to ensure that the two comprehension question forms were equivalent.

The impact of pictures on JSR performance was examined across all children ($N = 38$). The results of a paired t -test indicated that there was no statistically significant difference ($t(37) = 0.02, p > .05$) between performance on the three cloze items where pictures might have cued a correct response ($M = 0.69, SD = 0.27$) and performance on the other five items ($M = 0.69, SD = 0.31$). Therefore, subsequent data analyses did not differentiate between the two types of items; the JSR raw scores were computed from children's responses to all eight items.

The equivalency of the two forms of comprehension questions (Form A and Form B) was also examined in preliminary analyses. An independent t -test was conducted and no statistically significant difference ($t(36) = -1.53, p > .05$) was found between accuracy on Form A ($M = 0.60, SD = 0.14$) and accuracy on Form B ($M = 0.70, SD = 0.24$). Therefore, subsequent data analyses did not differentiate the form of comprehension questions completed by the children. These preliminary analyses hav-

Table 2
Correlation Coefficients for Age, JSR, and Comprehension Questions (N = 38)

Variable	Age	JSR	Questions
Age	--	.61**	.60**
JSR		--	.60**
Questions			--

** $p < .01$.

ing been completed, primary analyses were conducted to evaluate the JSR's developmental sensitivity, its concurrent validity, and its construct relevance (i.e., its ability to measure discourse comprehension directly without the irrelevant skill, expressive language).

First, analyses were conducted to examine whether the JSR met the criterion of developmental sensitivity. Since discourse comprehension abilities are expected to improve with age during childhood, older children should obtain higher test scores on the JSR than younger children, if the measure is valid. Correlational analysis was performed to explore the relationship between age and performance on the JSR for all 38 participants. As shown in Table 2, a statistically significant correlation of moderate magnitude ($r = .61, p < .01$) was found between age and the JSR, indicating that age is one of the multiple factors that play a role in JSR performance and as age increases JSR performance improves. Also shown in Table 2 is the correlation between age and accuracy on the traditional comprehension questions. As would be expected, a statistically significant correlation of moderate strength ($r = .60, p < .01$) was found between age and accuracy on the comprehension questions, indicating that with age comprehension question performance also improves.

Thus, as expected, results of this investigation showed that comprehension questions as measures of discourse comprehension did demonstrate developmental sensitivity. Of greater import was the finding that the JSR is also sensitive to age differences. As would be expected of a task that measures an ability that improves with age, increased age was associated with greater accuracy on the JSR. Analyses were also conducted to evaluate the concurrent validity of the JSR.

Since the JSR is intended to measure discourse comprehension, the same construct that is measured by comprehension questions, the concurrent validity of the JSR was evaluated by comparing children's performance on the JSR with their scores on the traditional comprehension questions. As shown in Table 2, correlational analysis revealed a statistically significant correlation of moderate magnitude ($r = .60, p < .01$) between the JSR and the comprehension questions, indicating that the two measurement tasks share some common features. Also shown in Table 2 is the relationship between each of the comprehension measures and age. Since, in addition to being moderately correlated with each other, both measures were also moderately correlated with age, it was possible that shared variance related to age, rather than to discourse comprehension, might have accounted for the relationship between the JSR and comprehension questions. To determine whether there was any unique contribution of comprehension assessment, the rela-

tionship between the JSR and the comprehension questions needed to be examined with some control placed on age. This was done by analyzing the relationship between the two comprehension measures for a subset of children ($n = 23$) between 42 and 50 months of age. In comparison to the broad age range of the complete group of participants (i.e., 30-50 months), this subgroup was more developmentally stable. That is, the period between 42 and 50 months is one of less rapid and dramatic changes in language ability. The subgroup approach to controlling age was taken instead of statistically partialling age out of the correlation, since age is inextricably linked to comprehension ability. Correlational analysis was performed to examine the relationship between scores on the JSR and scores on the comprehension questions for children within the subgroup. A moderate correlation ($r = .44$) was found between the JSR and the comprehension questions in this age group. Thus, 19% of the variability in scores on the JSR was explained by the relationship between performance on that task and performance on the comprehension questions when age varied only minimally. The magnitude of the correlation between the JSR and the comprehension questions was not as strong within the developmentally stable age group ($r = .44$) as it was within the group as a whole ($r = .60$). Although the magnitude of the correlation between the JSR and comprehension questions decreased when age varied only minimally, the correlation remained in the moderate range, suggesting that the JSR and comprehension questions shared variance that was not solely attributable to age. Given that comprehension questions are a widely employed measure of discourse comprehension, it is plausible that the shared variance indicates that both tasks measure discourse comprehension.

The developmental sensitivity and the concurrent validity of the JSR having been evaluated, the construct relevance of the task was examined. Specifically, the ability of the JSR to provide a direct assessment of discourse comprehension, untainted by the simultaneous measurement of an irrelevant behavioural domain like expressive language skill was evaluated. The JSR was designed to place limited demands on children's expressive language abilities, requiring them to produce specific elements of the text within a supportive context. In order to examine the relationship between expressive language and performance on both the JSR and the comprehension questions, two stepwise regression analyses were conducted. In stepwise regression, the computer selects the variable that has the highest bivariate correlation with the dependent variable and enters it into the equation first. It then enters the predictor with the highest semipartial correlation, continuing until the remaining variables no longer make a significant addi-

tional contribution to the multiple correlation. The first stepwise regression analysis was performed with number of responses correct on the JSR as the dependent variable and age and the expressive language component of the MCDI-III (MCDI-III-production) as the independent variables. The MCDI-III-production score was calculated by summing each participant's score on the Vocabulary and Grammar Scales. Results of the regression analysis revealed that the single best predictor of JSR performance was age, which accounted on its own for 32% of the variability in JSR performance ($r = .57$, $F(1,33) = 15.67$, $p < .01$). MCDI-III-production did not make a unique contribution to the variability, indicating that expressive language ability did not contribute to JSR performance. When a second regression analysis was conducted with comprehension questions as the dependent variable, MCDI-III-production, having had the highest correlation with the comprehension questions, was entered first. It alone accounted for 40% of the variability in performance on the comprehension questions ($r = .63$, $F(1,33) = 21.29$, $p < .01$) indicating that language production was the single best predictor of comprehension question performance. Age, having had the lower correlation with comprehension question performance, was entered next into the equation but only accounted for an additional nine per cent of the variability in performance on the questions ($r = .69$, $F(2,23) = 14.67$, $p < .01$).

Thus, although there were moderate correlations between the MCDI-III-production and both the JSR ($r = .38$, $p < .05$) and the comprehension questions ($r = .63$, $p < .01$), results suggested that global expressive language ability (as measured by the MCDI-III) only contributed uniquely to performance on the comprehension questions. Global expressive language ability did not contribute to performance on the JSR beyond what would be expected due to age alone. The comprehension question task designed for this study required participants to respond to a mix of yes/no and primarily simple wh-questions. The results demonstrated that expressive language was the key factor in predicting performance on the comprehension questions, even though many of the questions had relatively limited expressive language demands. An even stronger contribution of expressive language to comprehension question performance may have been discovered had more complex kinds of wh-questions been utilized.

In addition to the regression analysis, the role of expressive language in JSR performance was also explored with an analysis of the types of errors on the task. This analysis revealed that only 18 of the 96 errors on JSR items across children were attributable to a nonresponse. While incorrect responses may be more clearly attrib-

uted to failures in comprehension, nonresponses might either be attributed to comprehension or to expressive difficulties. The fact that nonresponses accounted for only 19% of the total number of errors means that the number of errors that might possibly be attributed to difficulty with the expressive demands of the task is small.

The finding that expressive language ability is not a key factor in JSR performance suggests that this task may be suitable for assessing the discourse comprehension skills of young children with expressive language impairments, unless their skills are severely delayed. Future research should examine the performance of children with expressive impairments on the JSR to determine more precisely the level of expressive language skill required to complete the task. However, results from this study suggest that the expressive language skill required to complete the JSR is limited with respect to the more traditional approach.

The purpose of this investigation was to examine the validity of the JSR, a new measure of young children's discourse comprehension, by examining its relationship to age, traditional comprehension questions, and expressive language. Results revealed a statistically significant correlation of moderate strength between age and performance on the JSR, suggesting that this measure is sensitive to age differences. In addition, results revealed a moderately strong relationship between the JSR and traditional comprehension questions, indicating that the JSR, like comprehension questions, measures discourse comprehension. Finally, results revealed that expressive language ability did not predict performance on the JSR, indicating that the JSR does not impose substantial demands on young children's expressive language abilities. This finding suggests that the JSR provides a direct assessment of discourse comprehension such that measurement of comprehension is not likely to be obscured by the expressive components of the task. Taken together, all of the findings of this study suggested that the JSR may provide a valid measure of young children's discourse comprehension.

Clinical Implications

In contemporary society increasing emphasis is being placed on early childhood development and timely identification of children at risk for later delays. Speech-language pathologists have a mandate to identify children with speech and language impairments as early as possible. In Ontario, the Ministry of Health has targeted the early identification of and intervention for speech and language impairments as provincial health priorities. Given this emphasis on early identification, it is critical to carefully examine the tools available to achieve

this goal. Do sufficient measures exist to identify children with speech and language impairments at very early ages? At the present time, the answer to this question is no. There are gaps in traditional assessment batteries for very young children, particularly in the area of language comprehension. This hinders early identification and intervention efforts. Thus, it is essential that attention be devoted to the development of assessment tasks that will fill the gaps in batteries for young children and ultimately allow the goal of early identification of speech and language impairments to be met.

One of the notable gaps in traditional assessment batteries is in comprehension measurement. While measures of vocabulary and syntactic comprehension exist, few measures of discourse comprehension appropriate for young children have been developed. Comprehensive assessment of comprehension has been limited by the lack of these measures. As stated previously, the problem associated with the paucity of discourse comprehension measures is that scores on available measures, which are typically only valid for the assessment of decontextualized vocabulary and syntax, might be used to draw conclusions about language comprehension in general. The negative impact of basing conclusions about a child's overall comprehension on vocabulary and syntactic measures alone must not be underestimated. When conclusions about a child's overall comprehension ability are drawn solely from vocabulary and syntactic measures there is a risk that information about the child's ability to construct meaning from longer spans of language may be missed. Such information is an important component of comprehensive assessment profiles and treatment plans for young children. A child who demonstrates age-appropriate vocabulary and syntactic language understanding on standardized measures of language comprehension may nevertheless have difficulty understanding an age-appropriate children's story read to him or her. If appropriate vocabulary and syntactic understanding are taken to mean comprehension skills as a whole are satisfactory, difficulties in using social, textual, scriptal, and other forms of knowledge to determine the meaning of longer units of language may be overlooked. Early identification and treatment of these discourse level difficulties may prevent later problems with the acquisition of literacy and social communication skills. The risk that invalid interpretations about comprehension will be made on the basis of vocabulary and syntactic measures alone will remain until a valid measure of young children's discourse comprehension is developed.

The Joint Story Retell (JSR) is a newly developed measure of preschool children's oral discourse comprehension. Although further investigations of the JSR are

necessary, evidence obtained from this investigation suggests that it has the potential to be a valid measure of oral story comprehension. Thus, it appears that the JSR could be a meaningful addition to the current battery of comprehension measures. Its use, in conjunction with the use of vocabulary and syntactic level measures, should allow more comprehensive assessment of young children's comprehension in the future.

Author Notes

This study was conducted in partial fulfillment of the requirements for the Master of Science degree by the first author under the supervision of the second.

The authors would like to thank Tania Cleveland for assisting with the development of the testing materials and the data collection. The assistance of the staff of the preschools and day care centres and the participation of the children and their parents are also gratefully acknowledged. This research was supported by the Harmonize for Speech Fund, Ontario District.

Portions of this study were presented at the American-Speech-Language-Hearing Association Annual Convention in San Francisco, CA, November, 1999.

Correspondence concerning this article should be addressed to Lynn F. Dempsey, Department of Rehabilitation Science, University of Western Ontario, London, Ontario, N6G 1H1.

References

- Anastasi, A. (1988). Validity: Basic concepts. In A. Anastasi & S. Urbina (Eds.), *Psychological testing* (6th ed., pp. 139-164). New York: Macmillan.
- Applebee, A. N. (1978). *The child's concept of story*. Chicago: University of Chicago Press.
- Carlisle, J. F. (1991). Planning an assessment of listening and reading comprehension. *Topics in Language Disorders*, 12(1), 17-31.
- Craig, H. K., & Evans, J. (1989). Turn exchange characteristics of SLI children's simultaneous and non-simultaneous speech. *Journal of Speech and Hearing Disorders*, 54, 334-347.
- Dale, P. (1996). *MacArthur Communicative Development Inventory- III (MCDI-III)*. Experimental Edition. Seattle, WA: University of Washington.
- Dale, P., Reznick, J. S., & Thal, D. J. (1998, April). A parent report measure of language development in three-year olds. Paper presented at the International Congress for Infant Studies, Atlanta, GA.
- Dempsey, L., Perfetti, J., & Skarakis-Doyle, E. (1999). *Early literacy questionnaire*. Unpublished Manuscript, University of Western Ontario, London, ON.
- Dickinson, D., & Smith, M. (1994). Long term effects of preschool teachers bookreadings on low income children's

vocabulary and story comprehension. *Reading Research Quarterly*, 29, 105-122.

Dickinson, D., & Snow, C. E. (1997). Interrelationships among prereading and oral language skills in kindergardeners from two social classes. *Early Childhood Research Quarterly*, 2, 1-25.

Feagans, L. (1982). The development and importance of narratives for school adaptation. In L. Feagans & D. Farran (Eds.), *The language of children reared in poverty* (pp. 95-116). New York: Academic Press.

Feagans, L., & Farran, D.C. (1981). How demonstrated comprehension can get muddled in production. *Developmental Psychology*, 17(6), 718-727.

Feagans, L., & Short, E. J. (1984). Developmental differences in the comprehension and production of narratives by reading disabled and normally achieving children. *Child Development*, 55, 1727-1736.

Fenson, L., Dale, P., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., Pethick, S., & Reilly, J.S. (1991, November). *Technical manual for the MacArthur Communicative Development Inventories*. San Diego, CA: San Diego State University.

McCabe, A. (1996). Evaluating narrative discourse skills. In K. Cole, P. Dale, & D. Thal (Eds.), *Assessment of communication and language* (pp. 121-142). Baltimore, md: Paul H. Brookes.

McCabe, A., & Rollins, P. R. (1994). Assessment of preschool narrative skills. *American Journal of Speech Language Pathology*, 3, 45-55.

McCauley, R. J., & Swisher, L. (1984). Psychometric review of language and articulation tests for preschool children. *Journal of Speech and Hearing Disorders*, 49, 34-42.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. In A.

E. Kazdan (Ed.), *Methodological issues & strategies in clinical research* (2nd ed., pp. 241-260). Washington, DC: American Psychological Association.

Miller, J. F., & Paul, R. (1995). *The clinical assessment of language comprehension*. Baltimore, MD: Paul H. Brookes

Rees, N. H., & Shulman, M. (1978). I don't understand what you mean by comprehension. *Journal of Speech and Hearing Disorders*, 43, 208-219.

Skarakis-Doyle, E., & Wootton, S. (1998, June). *Measuring preschool children's discourse comprehension: A modified retell procedure*. Paper presented at the 19th annual Symposium on Research in Child Language Disorders, Madison, WI.

Snow, C. E., Perlmann, R., & Nathan, D. (1987). Why routines are different: Toward a multiple-factors model of the relation between input and language acquisition. In K. E. Nelson & A. van Kleeck (Eds.), *Children's Language* (vol. 6, pp. 65-97). Mahwah, NJ: Erlbaum.

Stein, N. (1988). The development of children's storytelling skill. In M. B. Franklin & S. Barten (Eds.), *Child language: A book of readings* (pp. 282-297). New York: Oxford University Press.

Thal, D. (1991). Language and cognition in normal and late-talking toddlers. *Topics in Language Disorders*, 11, 33-42.

Thal, D., Tobias, S., & Morrison, D. (1991). Language and gesture in late talkers: A 1-year follow-up. *Journal of Speech and Hearing Research*, 34, 604-612.

Warr-Leeper, G., Miller, L., Brac, M., Culhane, R., Bernhard, K., & Yearous, J. (1997, October). *Use of questionnaires to identify children at risk of hearing impairment*. Poster session presented at the annual OSLA Conference, Toronto.

Zimmerman, I., Steiner, V., & Pond, R. (1979). *Preschool Language Scale-3*. Columbus, OH: Charles E. Merrill.

Manuscript received: May 28, 2001

Accepted: October 3, 2001



Appendix A

Joint Story Retell Version of "Splish Splash": Examples

One day a little girl named Sarah made twenty very messy mud-pies in the back yard. Sarah's mother took one look at her, and said, (1) "_____." (*Splish splash Sarah needs a bath...*)

So the mother took Sarah upstairs to the bathroom and filled the bathtub with water. Then, the little girl named (2) _____ stuck her big toe into the bathtub and said, "Oh Mommy, the water's too cold. The water must be nice and warm." (*Sarah*)

Appendix B

Comprehension Questions: Forms A and B: Examples

(*Sample answers in parentheses*)

1. What does Sarah's mother say? (*Splish splash Sarah needs a bath...*)
 - a. Is Sarah dirty? (*yes*) or a. Is Sarah clean? (*no*)
 - b. Does mom want Sarah to go outside and play? (*no*) or b. Does mom want Sarah to take a bath? (*yes*)